Atmos. Meas. Tech. Discuss., 7, 11957–11989, 2014
www.atmos-meas-tech-discuss.net/7/11957/2014/
doi:10.5194/amtd-7-11957-2014

Atmospheric Measurement Techniques Discussions

Techniques for analyses of trends in GRUAN data

G. E. Bodeker and S. Kremser

Bodeker Scientific, Alexandra, New Zealand

Received: 18 August 2014 – Accepted: 7 November 2014 – Published: 2 December 2014

Correspondence to: G. Bodeker (greg@bodekerscientific.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) provides reference quality RS92 radiosonde measurements of temperature, pressure and humidity. A key attribute of reference quality measurements, and hence GRUAN data, is that each datum has a well characterised and traceable estimate of the measurement uncertainty. The long-term homogeneity of the measurement records, and their well characterised uncertainties, make these data suitable for reliably detecting changes in global and regional climate on decadal time scales. Considerable effort is invested in GRUAN operations to (i) describe and analyse all sources of measurement uncertainty to the extent possible, (ii) quantify and synthesize the contribution of each source of uncertainty to the total measurement uncertainty, and (iii) verify that the evaluated net uncertainty is within the required target uncertainty. However, if the climate science community is not sufficiently well informed on how to capitalize on this added value, the significant investment in estimating meaningful measurement uncertainties is largely wasted. This paper presents and discusses the techniques that will need to be employed to reliably quantify long-term trends in GRUAN data records. A pedagogical approach is taken whereby numerical recipes for key parts of the trend analysis process are explored. The paper discusses the construction of linear least squares regression models for trend analysis, boot-strapping approaches to determine uncertainties in trends, dealing with the combined effects of autocorrelation in the data and measurement uncertainties in calculating the uncertainty on trends, best practice for determining seasonality in trends, how to deal with co-linear basis functions, and interpreting derived trends. Synthetic data sets are used to demonstrate these concepts which are then applied to a first analysis of temperature trends in RS92 radiosonde upper air soundings at the GRUAN site at Lindenberg, Germany (52.21° N, 14.12° E).

AMTD
7, 11957–11989, 2014

1 Introduction

Long-term climate data records of essential climate variables such as temperature, water vapour and ozone are a prerequisite for climate change detection and attribution studies. Many decades of measurements are typically required to detect a trend at e.g. the 95 % confidence level. Not only do climate data records need to be temporally homogeneous over many decades, it is also imperative that the uncertainty on the trend can be estimated robustly so that we know what level of confidence to place on the derived trend. Having well characterised measurement uncertainties on the data being analysed, including traceability of uncertainty estimates to internationally recognized measurement standards, is essential in establishing the uncertainty on the resultant trend.

One of the core goals of the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN; www.gruan.org) is to provide vertical profiles of reference measurements suitable for reliably detecting changes in global and regional climate on decadal scales. Reference measurements require that, at a minimum, the uncertainty of the measurement (including uncertainties arising from corrections applied) has been determined, the entire measurement procedure and set of processing algorithms are properly documented and accessible, and that every effort has been made to tie the observations to an internationally accepted traceable standard (Seidel et al., 2009). For vertical profile measurements within GRUAN, uncertainties are also required to be vertically resolved such that each datum in a profile is treated as a single measurement result requiring both the measurement and its uncertainty. Metadata describing how the measurements were made, what corrections were applied, what changes were made to the instruments over the lifetime of the measurements, and the data reduction algorithms used during the observation and post-observation periods, are also imperatives for reference quality observations.

AMTD
7, 11957–11989, 2014

Immler et al. (2010) provided the theoretical basis for developing reference quality upper-air measurements in the form of GRUAN data products. These fundamental guidelines for establishing reference quality atmospheric observations are based on central concepts of metrology and, in particular, traceability. They demonstrate that the detailed analysis of the uncertainty budget of a measurement technique is the critical step for establishing reference quality observations. Detailed knowledge of the calibration procedures and data processing algorithms is required as part of determining the uncertainty budget. Finally, as highlighted by Immler et al. (2010), uncertainties introduced by correction schemes adjusting for systematic biases in the measurement system are an important component of the uncertainty budget. Dirksen et al. (2014) demonstrated the application of that theoretical basis for creating GRUAN data products based on Vaisala RS92 radiosonde measurements of temperature, pressure and humidity. The dominant source of uncertainty for radiosonde temperature measurements is solar radiation heating and Dirksen et al. (2014) report on the laboratory experiments performed to investigate and model the effects of solar radiative heating on the RS92's temperature and humidity measurements. GRUAN daytime humidity profiles show up to 15 % enhancement in humidity over Vaisala processed profiles, of which two-thirds is due to the radiation dry bias correction, and one-third is due to an additional calibration correction. Philipona et al. (2012) reported simultaneous solar shortwave radiation, thermal longwave radiation, and air temperature measurements with radiosondes from the Earth's surface to 35 km altitude and then later demonstrated the use of these measurements during day-time and night-time, under sun-shaded and unshaded conditions, to determine the radiation-induced error on radiosonde air temperature measurements (Philipona, 2013). They showed that, in general, solar radiation produces a radiative heating of about +0.2 K near the surface which linearly increases to about +1 K at 32 km (~ 10 hPa). It is clear from these studies that significant time and effort has been invested in establishing reference quality measurements within GRUAN.

AMTD
7, 11957–11989, 2014

The goal of detecting a trend in upper-air temperature, and the uncertainty on that trend, from point source measurements from a radiosonde and their associated measurement uncertainties, can be challenging. Previous papers (Tiao et al., 1990; Weatherhead et al., 1998) and books (Kutner et al., 2005) have provided detailed statistical descriptions of how to detect trends, and their uncertainties in geophysical time series. However, some of these studies require a higher understanding of statistics than might be available to scientists who want to quickly and reliably determine the trend and its uncertainty in a time series. The purpose of this paper is to take a pedagogical approach, leading the reader through the various perils and pitfalls that need to be surmounted to calculate a trend and its uncertainty. In this way we hope that the considerable effort that has been invested in obtaining reference quality measurements within GRUAN is exploited to the utmost.

2 Construction of a regression model

2.1 Why not just fit a straight line?

The first question someone may ask when analyzing a time series for trends is "Why can I not just fit a straight line to the data and look at its slope?". If you simply want to know whether the data are trending up or down, with no interest in what is driving that change, or whether or not that change is statistically significant, this may be a legitimate approach. However, consider the synthetic data set shown in Fig. 1.

The 10 year time series shown in the upper trace was generated by repeating the same annual cycle and adding some random noise. The data are purely periodic and contain no trend. And yet, if we fit a straight line to these data (dashed line in Fig. 1) we detect a negative trend. This is the case for almost all such time series generated with such random noise. Interestingly, if we derive the trend separately in each month over a large ensemble of such time series, and average those trends, the result is a zero trend. The reason for this is that a sinusoid (an "odd" function) was used to

AMTD
7, 11957–11989, 2014

define the annual cycle, starting with a positive phase in the cycle and ending with a negative phase, thus causing a negative trend. Doubling the length of the record from 10 to 20 years reduces the magnitude of the negative trend, as expected. If we simply want to know whether the data are trending up or trending down, then this is a valid approach and the analyses indeed show that the data are trending down. But is it correct to conclude that there is a geophysical trend in these data that are constructed purely as a (noisy) repeating annual cycle? Almost certainly not. So we acknowledge that a first step in detecting a geophysical trend in a time series is to remove the mean annual cycle. This is shown by the line with filled dots in Fig. 1 with the straight line fit (dashed line).

But why stop there? There may be other cyclical signals in the time series, such as the solar cycle or El Niño Southern Oscillation (ENSO), that should be removed to avoid their interference with the underlying geophysical trend. Therefore, a typical statistical trend model may be of the form:

$$V_t = \alpha + (\beta \times t) + (\gamma \times \mathrm{QBO}_t) + (\delta \times \mathrm{ENSO}_t) + (\epsilon \times \mathrm{SOLAR}_t) + R \quad (1)$$

where V_t is the data value at time step t (typically year and month), QBO_t is the quasi-biennial oscillation (Reed et al., 1961; Ribera et al., 2003) with a prescribed value at time step t, ENSO_t is the normalised Tahiti minus Darwin sea level pressure (southern oscillation index) at time step t, and SOLAR_t is the solar cycle with some prescribed value at time step t. The coefficients α, β, γ, δ and ϵ are fit coefficients, typically calculated using a linear least squares regression approach (Moore and McCabe, 2003). The fit coefficient β represents the trend in the time series. R is the residual, i.e. that part of the signal that cannot be tracked by the statistical model and is usually derived by subtracting the model fit (after it has been calculated) from the original data.

It is possible, maybe even likely, that the trend will be different during different months of the year. For that matter, perhaps the QBO has a stronger effect in some months than others. The next section considers the options of how best to deal with possible seasonal dependence in the fit coefficients.

AMTD
7, 11957–11989, 2014

2.2 Dealing with seasonality in the fit coefficients

One approach is to fit a regression model in the form of e.g. Eq. (1) completely independently for each month, i.e. first fit the model to only the January data, then fit the model to only the February data etc. This results in 12 trend regression coefficients, one for each month, that capture the seasonal dependence of the data on each basis function. The disadvantage of this approach is that the number of fit coefficients increases by a factor of 12. This significantly increases the uncertainty on the fit coefficients. The approach also assumes complete independence from one month to the next, e.g. the trend in month M is completely independent of the trend in month $M-1$ and in month $M+1$. This is unlikely to be the case. Let us assume, as an example, that the fit coefficients are a function of season. So the trend coefficient β in Eq. (1) can be expanded as: $\beta = f(t)$. Clearly the value of β on 31 December is going to be essentially identical to the value of β on 1 January i.e. the dependence should be cyclic in season. It should also be smooth, and it should be a linear superposition of functions that, when added together, can describe any seasonal periodic structure. These requirements are met by expanding the coefficient in Fourier series, e.g. the trend coefficient β is expanded as

$$\beta = \beta_0 + \beta_1 \times \sin\left(2\pi\frac{M}{12}\right) + \beta_2 \times \cos\left(2\pi\frac{M}{12}\right) + \beta_3 \times \sin\left(4\pi\frac{M}{12}\right) + \beta_4 \times \cos\left(4\pi\frac{M}{12}\right)$$

where M is the month number. When inserted back into Eq. (1), this becomes:

$$\beta \times t = \beta_0 \times t + \beta_1 \times \sin\left(2\pi\frac{M}{12}\right) \times t + \beta_2 \times \cos\left(2\pi\frac{M}{12}\right) \times t + \beta_3 \times \sin\left(4\pi\frac{M}{12}\right) \times t + \beta_4 \times \cos\left(4\pi\frac{M}{12}\right) \times t$$

Now, the number of fit coefficients increases not from one to twelve but from one to five and the value of the coefficient in month M is mathematically constrained to be similar to the value in month $M-1$ and $M+1$.

AMTD
7, 11957–11989, 2014

A synthetic time series consisting of an annual cycle, a seasonally dependent trend, and some Gaussian noise added for realism is shown in the inset in Fig. 2. The trends extracted from this synthetic time series, first by applying a truncated version of Eq. (1) to include only the α and β terms, month by month (black dots in Fig. 2), and then by applying the truncated Eq. (1) with the α coefficient expanded in four Fourier pairs and the β coefficient expanded in two Fourier pairs (black line in Fig. 2), is shown in the main plot in Fig. 2. How do we select how many Fourier pairs to use in each expansion? This is, in part, a judgment call. Expanding the α coefficient in Fourier pairs becomes the mean annual cycle because it is the seasonally dependent offset across the whole time series. The calculation of the uncertainties on the derived trend coefficients is detailed in Sect. 4. It is clear from Fig. 2 that expanding the trend coefficient in Fourier pairs results in a more coherent picture of the annual structure of the trend than if linear fits are used to derive trends individually for each month. The uncertainty on the trends (grey shaded area in Fig. 2) is also smaller when the seasonality is accounted for using Fourier series compared to fitting individually to each month.

2.3 Pre-treating basis functions

Consider some geophysical quantity measured over the period 2002 to 2009 that has no trend and consists of nothing but an annual cycle and some unforced variability. Four synthetic time series of this nature have been created, differing only by the randomly generated unforced variability added to the sinusoidal annual cycle. The results from the fits of Eq. (1) to the four synthetic time series are shown in Fig. 3. In this case the ENSO basis function was excluded for clarity. Even though the time series was constructed to have no trend, one of the four trend results is found to be statistically significantly different from zero at the 1σ level. This occurs because the combination of the linear trend term, and the solar cycle term, over the period 2002 to 2009 period, may, purely fortuitously, track some of the unforced variability (generated randomly in our synthetic time series). As can be seen in Fig. 3, when the regression model assigns variance between the solar cycle and linear trend, solar cycle fits that induce

AMTD
7, 11957–11989, 2014

a negative trend in the time series are always matched with a positive linear trend. This problem results, in large part, from the fact that the linear trend and the solar cycle over this 2002 to 2009 are close to degenerate, i.e. they are very much like linear scalings of each other. This makes the combination of the linear trend and solar cycle basis function very susceptible to fitting the unforced variability on the signal. A non-zero QBO fit coefficient occurs for the same reason. One way to avoid this from happening is to take the view that any trend in the signal must be assigned exclusively to the trend basis function. The only way to ensure that this happens is to detrend each basis function time series. To show the effects of detrending the solar cycle and QBO basis functions before fitting Eq. (1), 20 000 trend results equivalent to those listed in Fig. 3b have been generated from 20 000 time series including random noise as in Fig. 3b. First the original solar cycle and QBO basis functions were used and then solar cycle and QBO basis functions were detrended before their use in the regression. Histograms of the 20 000 trend results from these two tests are shown in Fig. 4. It is clear from Fig. 4 that when the QBO and solar cycle basis functions are detrended, the likelihood of obtaining a trend that is statistically significantly different from zero is very much reduced i.e. the uncertainty on the trend halves.

But why stop there? By detrending all of the basis functions we ensure that the trend basis function is orthogonal to all other basis functions. But those remaining basis functions may not be orthogonal to each other. The basis functions can all be made orthogonal to each other through a Gram–Schmidt orthogonalization procedure (Nering, 1963). Consider two basis functions BF1 and BF2, where $BF_{\perp 1} = BF_1$ and the $\perp$ subscript denotes the orthogonalized version of the basis function. To find a second vector $BF_{\perp 2}$ that is orthogonal to $BF_{\perp 1}$, we need to find a number X such that:

$$BF_{\perp 2} = X \times BF_{\perp 1} + BF_2 \quad (2)$$

where $BF_{\perp 1}$ is orthogonal to $BF_{\perp 2}$. Two vectors are orthogonal to each other if their inner product (denoted here as $\langle BF_{\perp 1}, BF_{\perp 2} \rangle$) is equal to zero. Taking the inner product

AMTD
7, 11957–11989, 2014

of Eq. (2) with $BF_{\perp 1}$ and solving for X leads to:

$$X = \frac{\langle BF_{\perp 1}, BF_{\perp 2}\rangle}{\langle BF_{\perp 1}, BF_{\perp 1}\rangle}$$

substituting into Eq. (2) results in:

$$BF_{\perp 2} = BF_2 - \frac{\langle BF_{\perp 1}, BF_{\perp 2}\rangle}{\langle BF_{\perp 1}, BF_{\perp 1}\rangle} BF_{\perp 1}$$

If we have a third basis function, we derive $BF_{\perp 1}$ and $BF_{\perp 2}$ as described above and we find $BF_{\perp 3}$ in the form of:

$$BF_{\perp 3} = X_1 \times BF_{\perp 1} + X_2 \times BF_{\perp 2} + BF_3$$

where $BF_{\perp 1}$ is orthogonal to $BF_{\perp 2}$ and $BF_{\perp 2}$ is orthogonal to $BF_{\perp 3}$. We take the inner product with $BF_{\perp 1}$ to find X_1 and the inner product with $BF_{\perp 2}$ to find X_2. This will then result in:

$$BF_{\perp 3} = BF_3 - \frac{\langle BF_{\perp 1}, BF_{\perp 3}\rangle}{\langle BF_{\perp 1}, BF_{\perp 1}\rangle} BF_{\perp 1} - \frac{\langle BF_{\perp 2}, BF_{\perp 3}\rangle}{\langle BF_{\perp 2}, BF_{\perp 2}\rangle} BF_{\perp 2}$$

The same can be done for the nth basis function to ensure that all basis functions used within the regression are orthogonal to each other. The generic form of the Gram–Schmidt orthogonalization is:

$$BF_{\perp n} = BF_n - \sum_{i=1}^{n-1} \frac{\langle BF_{\perp i}, BF_{\perp n}\rangle}{\langle BF_{\perp i}, BF_{\perp i}\rangle} BF_{\perp i}$$

Orthogonalization of the basis functions ensures that each additional basis function only describes the variance not already explained by the existing basis functions. However, using orthogonal basis functions in a regression model obfuscates the physical

AMTD
7, 11957–11989, 2014

interpretation of the contribution of each basis function to the signal, which is represented by the derived fit coefficients. The contribution of each basis function to the signal then depends on the order in which the basis functions were orthogonalized and might lead to different conclusions if the basis functions are orthogonalized in a different order. If orthogonal basis functions have been used within the regression, care needs to be taken when interpreting the derived fit coefficients.

3 Dealing with lags between cause and effect

It is conceivable that there may be a delay between the change in a basis function and its effect on the geophysical quantity being modelled. For example, the effect of the QBO on ozone can often be shifted in phase; the QBO signal in tropical ozone is out of phase with the QBO signal in extra-tropical ozone (see e.g. Fig. 4 of Bodeker et al., 2013).

Consider first the more abstract case of fitting a sinusoidal function, with an arbitrary phase shift, to some geophysical variable, i.e.:

$$V(t) = \alpha \times \sin(t + \beta) \tag{3}$$

where α is the amplitude of the signal and β is the phase shift. Noting that:

$$\alpha \times \sin(t + \beta) = \alpha \times \sin(t)\cos(\beta) + \alpha \times \cos(t)\sin(\beta)$$

and substituting $A = \alpha \times \cos(\beta)$ and $B = \alpha \times \sin(\beta)$, Eq. (3) becomes:

$$V(t) = A \times \sin(t) + B \times \cos(t)$$

with A and B becoming the new fit coefficients. Once A and B have been determined by fitting the two terms to the data, the original α and β values can be derived from:

$$\alpha = \sqrt{A^2 + B^2} \text{ and } \beta = \arctan\left(\frac{B}{A}\right)$$

AMTD
7, 11957–11989, 2014

The key to this solution is that the cosine signal is orthogonal to the sine signal such that by mixing these signals with different amplitudes, a sinusoidal signal with arbitrary phase can be generated. For some studies, e.g., Steinbrecht et al., 2003; Sioris et al., 2014, two different QBO basis functions are selected at two different pressure levels such that the signals are approximately a quarter of a cycle out of phase and therefore close to orthogonal to each other. This allows for an arbitrary phase shift in the QBO to be fitted. Austin et al. (2008) artificially constructed an orthogonal QBO time series to the base QBO time series to fit for a phase shift in the base time series, while other studies, e.g., Randel and Wu, 1996 simply use multiple QBO time series to accommodate phase shifts.

But how can we accommodate lags in basis functions which are not cyclical e.g. basis functions accounting for the effects of volcanic eruptions? Generally this is best done by running the regression model multiple times with different lags applied and determining which lag leads to a minimum in the sum of the squares of the regression model residuals.

4 Determining the uncertainty on the trend

One approach to determining the uncertainty on regression model fit coefficients is the bootstrapping method (Efron and Tibshirani, 1986). For example, to obtain the uncertainty on the trend coefficient in a regression model:

1. fit the regression model (e.g. Eq. 1) to the data.
2. Subtract the regression model fit from the data to obtain the residuals.
3. Take the signal produced by the regression model fit and, for each data point in that signal, randomly select one residual value and add it to the data point. Do this for the whole signal to generate a new signal which, while having the same underlying structure as the original signal, now has different random noise. Fit the regression model and record the trend value.

AMTD
7, 11957–11989, 2014

4. Repeat step 3 many times (e.g. 10 000) to generate 10 000 estimates of the trend. Calculate the standard deviation of those 10 000 values to obtain the estimated uncertainty on the trend.

Bootstrapping works because a very large ensemble of signals, each with the same underlying structure as the original signal, but with different noise characteristics, is created. Each of these signals, while statistically indistinguishable, has a slightly different trend resulting from the noise characteristics of the original data set. It is therefore natural that the standard deviation of the trends, derived from the ensemble of signals, is the uncertainty on the trend. There are some complications with the application of bootstrapping that need to be overcome in some circumstances, e.g. when the time series is autocorrelated, as discussed in Sect. 5. In this section we use bootstrapping to test some methodological choices in the calculation of uncertainties on trend coefficients.

One would expect that the uncertainty on the regression model fit coefficients, such as the trend, should depend on, inter alia:

1. the noise (unexplained variance) on the data signal – the greater the noise the less certain the regression model fit coefficients should be.

2. The measurement uncertainty on each datum comprising the data signal – the greater the measurement uncertainties the less certain the regression model fit coefficients should be.

Press et al. (1989) allude to the diagonal elements of the inverse matrix from the regression model being the variances (squared uncertainties) of the fitted parameters. The inverse matrix is given by $(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}$ where $\mathbf{A}$ is the so-called design matrix whose $N \times M$ components are constructed from the M basis functions evaluated at the N times for which data are available, and from the N measurement uncertainties (σ_i), i.e.:

$$A_{i,j} = \frac{X_j(x_i)}{\sigma_i} \text{ with } i = 1,2,\ldots,N \text{ and } j = 1,2,\ldots,M \quad (4)$$

AMTD
7, 11957–11989, 2014

where $X_j(x_i)$ is the jth basis function evaluated at the ith time for which data are available. While this inverse matrix $(\mathbf{A}^T\mathbf{A})^{-1}$, and hence the uncertainties on the regression model fit parameters, is sensitive to the measurement uncertainties (satisfying expectation 2 above), it is insensitive to the noise on the signal since it is a function of the basis functions only and therefore does not take into account in any way the unexplained variance in the signal being fitted. Press et al. (1989) suggest that to account for the variance, one can use the bootstrapping method to estimate quantitative confidence limits on the fitted parameters. Just using the square root of the diagonal elements of the inverse matrix would not account for all uncertainties and would most likely underestimate the uncertainties on the derived trend.

Bodeker et al. (1998), hereafter referred to as B98, implemented a revision of the measurement uncertainties according to Tiao et al. (1990) (hereafter referred to as T90). Their approach is to run the regression model according to Press et al. (1989) as a first step but then to account for the unexplained variance in the signal by substituting the measurement uncertainties in Eq. (4) with σ_{new} given by:

$$\sigma_{\text{new}} = \sqrt{\sigma_\epsilon^2 + \sigma_{\text{orig}}^2} \tag{5}$$

where σ_{orig} is the original measurement uncertainty and σ_ϵ^2 is the variance of the residuals from the first fit derived individually for each month (T90).The regression model fit is then repeated using σ_{new} as the revised measurement uncertainties. As a result, the derived uncertainties on the regression model fit coefficients (obtained from the diagonal elements of the inverse matrix) become sensitive to both the original measurement uncertainties and the unexplained variance.

Weatherhead et al. (1998), hereafter referred to as W98, derived an approximation for the uncertainty (ignoring autocorrelation – see Sect. 5) on the regression model estimate of the trend as:

$$\sigma_{\hat{\omega}} = \frac{\sigma_N}{(N/12)^{3/2}} \tag{6}$$

AMTD
7, 11957–11989, 2014

where σ_N is the standard deviation of the residuals of the time series to which the regression model is being fitted and N is the number of data points in the time series. However, this approximation for the uncertainty on the trend is insensitive to the measurement uncertainties and no indication is given in Weatherhead et al. (1998) on how to accommodate that. To include the W98 approach in our methodological testing, Eq. (5) is applied to σ_N of Eq. (6) similar to how it was done in B98. In this way, the W98 approach now accounts for both measurement uncertainties and unexplained variance in determining the uncertainty on the trend.

The B98 and W98 approaches to estimate the trend uncertainty are now compared to each other and to the traditional bootstrapping approach. Monthly mean synthetic time series, containing only an annual cycle and no trend and no autocorrelation, extending from 1990 to 2010, with varying levels of measurement uncertainty and unexplained variance, were generated. Different combinations of measurement uncertainties together with different degrees of unexplained variance were created to investigate what effect the measurement uncertainties and the variance have on the derived uncertainty on the trend (which is zero). Clearly, if the measurements have a larger random uncertainty, the derived trend should become less certain than for a set of data where the measurements have a small uncertainty. The same is true for the unexplained variance on the signal. If we have noisy data (high variance) the estimated trend will have a larger uncertainty than for a signal with lower variance since the true trend might be buried in the noise.

To derive uncertainties that are statistically representative of the derived trend in each generated time series with random noise and random measurement uncertainties, and to be not effected by the chosen randomness of the noise, 500 synthetic time series for each combination of unexplained variance and measurement uncertainties (see Table 1) were generated. Three sets of measurement uncertainties, viz.:

- uncertainties set to 1.0 for all measurements – hereafter referred to as uncert 1.0

AMTD
7, 11957–11989, 2014

- Uncertainties selected randomly within a Gaussian distribution of mean 1.5 and standard deviation of 0.5 – hereafter referred to as uncert 1.5σ0.5
- Uncertainties selected randomly within a Gaussian distribution of mean 4.5 and standard deviation of 1.5 – hereafter referred to as uncert 4.5σ1.5

and three sets of unexplained variance on the signal, viz.:

- no additional variance
- Medium additional variance – random values are chosen from a Gaussian distribution with a mean 0.0 and a standard deviation of 3.0
- High additional variance – random values are chosen from a Gaussian distribution with a mean 0.0 and a standard deviation of 8.0

Together these are used to create 9×500 synthetic time series. For every synthetic time series, the uncertainty on the trend was derived following the B98 and W98 approach described above. As a reference, the bootstrapping method was also employed to derive the uncertainties on the trends. An ensemble of 2000 Monte Carlo simulations was performed for this purpose. The mean uncertainty of the 500 trend values derived for each method, was calculated. The results are summarized in Table 1.

It is clear from Table 1 that for all cases both the B98 and W98 approaches come close to matching the results from bootstrapping. Larger uncertainties on the trends are derived for all three methods when the unexplained variance on the signal and/or the measurement uncertainties increase. By construction all signals were generated without a trend, with the result that the derived trends are not statistically significantly different from zero within the derived 1σ uncertainties. The results presented in Table 1 show that both B98 and W98 are valuable approaches to estimate uncertainties on the derived trend.

While the bootstrapping approach accounts for both the unexplained variance on the signal and the measurement uncertainties, bootstrapping is computationally demanding and it does not account for autocorrelation in the signal. Both B98 and W98

AMTD
7, 11957–11989, 2014

can be used if the signal is autocorrelated (see below) and both methods are less computationally demanding than bootstrapping. W98 is only applicable for estimating uncertainties on the trend. To derive uncertainties on other fit coefficients W98 cannot be used. Therefore, as the B98 approach can be used to derive uncertainties on all fit coefficients, this method is used in all further examples and tests below.

5 Accounting for autocorrelation

Time series of atmospheric climate data records are often autocorrelated e.g. higher than normal temperature on one day is often followed by anomalously high temperature on the next day. This autocorrelation results from natural temporal and spatial scales of many atmospheric phenomena. Such autocorrelations have the effect of reducing the quantity of information that would be available from the same number of independent data points, generally increasing the size of the error estimates (T90). T90 show that the uncertainty on trend estimates depends critically on the magnitude of month to month autocorrelation (serial autocorrelation) in the measurements. For contiguous evenly spaced data, a first order autoregressive model is constructed as:

$$R_t = \phi R_{t-1} + a_t$$

where R_t is the residual at time t (see Eq. 1), R_{t-1} is the residual at the previous time step, ϕ is the autocorrelation coefficient, and the a_t are the final residuals which should be free of any autocorrelation. For unequally spaced measurements, or for measurement series with many gaps, a second order autocorrelation model of the form:

$$R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + a_t$$

is required (Reinsel et al., 1987).

Appendix A of T90 provides a convenient way to account for the autocorrelation in the application of Eq. (1). The data to be regressed are replaced by:

$$V_t^* = V_t - (\phi \times V_{t-1})$$

AMTD
7, 11957–11989, 2014

and all basis function time series are similarly transformed i.e.

$$BF_t^* = BF_t - \phi BF_{t-1}$$

The regression model is rerun and the usual standard error formulae are used to calculate the uncertainty on the regression model fit coefficients.

In the presence of autocorrelation, additional changes are required, e.g. Eq. (5) becomes:

$$\sigma_{\text{new}} = \sqrt{\frac{\sigma_\epsilon^2}{1-\phi^2} + \sigma_{\text{orig}}^2}$$

or, if second order autocorrelation is accounted for:

$$\sigma_{\text{new}} = \sqrt{\left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma_\epsilon^2}{[1-\phi_2]^2 - \phi_1^2} + \sigma_{\text{orig}}^2}$$

and finally, Eq. (6) becomes:

$$\sigma_{\hat{\omega}} = \frac{\sigma_N}{(N/12)^{3/2}}\sqrt{\frac{1+\phi}{1-\phi}} \tag{7}$$

i.e. the presence of autocorrelation inflates the uncertainty on the regression model fit coefficients by $\sqrt{(1+\phi)/(1-\phi)}$.

6 GRUAN data

6.1 Application of trend analysis methods to Lindenberg upper air temperature data

At the time of writing this paper, RS92 radiosonde measurements of upper air temperatures currently comprise the only official GRUAN data product. RS92 radiosonde

AMTD
7, 11957–11989, 2014

flights are made at a number of sites typically either two or four times per day (see Fig. 5). The site with the most comprehensive data record to date is the Lindenberg site which is used below in the example trend analyses. Figure 5 is not a reflection of the number of radiosonde flights made at each site, but rather the number of radiosonde flights that passed all QA/QC checks imposed in the centralised data processing; the raw data from these flights are processed through a centralised RS92 radiosonde data processing facility located at the GRUAN Lead Centre in Lindenberg, Germany. A detailed description of the GRUAN RS92 radiosonde product, and how this product conforms to the requirements of a reference measurement, is described in detail in Dirksen et al. (2014). While the currently available RS92 radiosonde data set is too short for a robust long-term trend analysis, application of the methodology outlined above is demonstrated here on the data that are available. The challenges with applying the method to such a short time series are outlined below. The results of this very preliminary trend analysis, presented in Sect. 6.2, should not be over-interpreted as they are certainly not indicative of longer term trends in upper air temperatures. The time series of GRUAN radiosonde upper air temperatures available from Lindenberg is rather short i.e. a five year period extending from mid-2009 to mid-2014. The brevity of this time series makes the fitting of longer-term geophysical drivers of temperature change such as the QBO, ENSO, and solar cycle difficult, if not impossible. And yet the increase in 10.7 cm solar flux from mid-2009 to mid-2014 may well have driven changes in temperature above Lindeberg over this period. If the question to be answered is "What were the temperature trends over Lindenberg from mid-2009 to mid-2014 as a function of season and altitude over and above the possible trends induced by periodic drivers such as QBO, ENSO and solar cycle?", then basis functions for these drivers should be included in addition to a trend basis function. The difficulty then is that if there was a trend in temperatures over this period, but this trend was not induced by the solar cycle, then the trend may well still be ascribed to the solar cycle since the 10.7 cm solar flux showed a steady increase from mid-2009 to mid-2014. The measurement time series is simply not long enough to determine the correlation be-

AMTD
7, 11957–11989, 2014

tween the temperature changes and the solar cycle. The same may be true for ENSO and QBO. If the question to be answered is "What were the temperature trends over Lindenberg from mid-2009 to mid-2014 as a function of season and altitude irrespective of the drivers of those changes?", then a regression model consisting of no more than an offset and trend term in Eq. (1) may be sufficient. Alternatively, a regression model including all potential drivers of temperature changes, but orthogonalizing all basis functions so that all basis functions other than the trend basis function are trend-free, would be appropriate. With these caveats in mind, three different constructs of the regression model have been applied to the data:

1. a regression model including offset, trend, QBO, ENSO and solar cycle basis functions – hereafter referred to as the all regression.
2. A regression model including only offset and trend basis functions – hereafter referred to as the reduced regression.
3. A regression model including offset, trend, QBO, ENSO and solar cycle basis functions, but where all basis functions have been sequentially orthogonalized – hereafter referred to as the all-orthog regression.

The measurement uncertainties, as provided in the GRUAN data files, are provided to the regression model along with the measurements. Individual measurements are used i.e. monthly means are not calculated.

6.2 Upper air temperature trends at Lindenberg

Examples of regression model fits to 00:00 UT temperature measurements above Lindenberg are shown in Fig. 6. The all-orthog fits are identical to the all fits and are therefore not shown; the partitioning of the variance between the basis functions between the all-orthog and all regression model runs of course differs, as will be shown below. Only sporadic measurements are available before late 2009 and during these periods, the uncertainty on the regression model fit is larger. It is also not immediately obvious

AMTD
7, 11957–11989, 2014

that the all fit is superior to the reduced fit though, in some cases (e.g. early 2011 at 20 and 25 km) the all fit does capture the anomalously higher temperatures. Similar fits for 12:00 UT data are not shown. The resultant 00:00 UT trends from the three different applications of the regression model are shown in Fig. 7. There are clear differences in the trends between the three different regression model runs. These differences can aid the interpretation of the causes of the trends. For example, a positive temperature trend, maximizing at 1.96 K year^{-1}, is seen in the reduced results in May at ~ 13 km. A much stronger trend of 2.71 K year^{-1} is derived from the all regression model fit suggesting that some other basis function (in this case the solar cycle basis function) is driving a negative trend in temperature which is then partially offset by the more positive temperature trend produced by the trend basis function to derive a net trend similar to that seen in the reduced regression. Two interpretations of this result are possible:

1. changes in solar activity resulted in cooling in May at ~ 13 km. Other factors, as accounted for by the linear trend term (e.g. changes in greenhouse gas concentrations) caused even larger warming resulting in the small positive trends in temperature seen in the reduced panel of Fig. 7. The trend in temperature, over and above that resulting from the QBO, ENSO and solar cycle basis functions, is that derived from the all regression.

2. Changes in solar activity did not cause cooling in May at ~ 13 km. Rather, shorter-term anti-correlations between temperature and solar activity aliased into a simulated longer-term negative trend in temperature which then needed to be compensated by a more positive linear trend contribution to track the true secular change in temperature as quantified in the reduced regression results.

A regression analysis, on it's own, cannot determine which of these two interpretations is correct. It is a fundamental limitation of regression analyses that no physical understanding of the system is incorporated in the analyses. If, based on physically implausible mechanisms, an anti-correlation between temperature and solar activity at 13 km could be excluded, then interpretation (1) above could be excluded – in which case the

AMTD
7, 11957–11989, 2014

trends quantified in the reduced analyses would best represent reality. We caution that trends derived from application of regression models need to be interpreted.

12:00 UT trends from the three different applications of the regression model to Lindenberg radiosonde temperature measurements are shown in Fig. 8. In general, as with the 00:00 UT trends, the 12:00 UT trends are seldom statistically different from zero at the 2σ level. Similar arguments regarding interpretation of these trends as for Fig. 7 can be made. The winter-time cooling signal at the higher altitudes likely results from an increase in the frequency of polar vortex air coming over Lindenberg in recent years (H. Vömel, personal communication, 2014).

7 Conclusions

Regression analysis is seldom a panacea for extracting trends from geophysical time series. This paper has explored some of the methodological perils and pitfalls of trend determination using regression analysis and has reiterated the caveats by demonstrating the application of the regression model to short time series of upper air temperature measurements at the GRUAN site at Lindenberg. The paper demonstrates how the GRUAN measurement uncertainties, the focus of much of the effort within GRUAN, can be propagated through to uncertainties on derived trends in temperature. The Lindenberg measurements series are too short to derive indicative upper air temperature trends from the data but are useful as a pedagogical example of the application of the regression model.

Acknowledgements. We would like to thank the staff at the GRUAN site at Lindenberg for their tireless efforts to conduct reference quality observations of temperature in the context of GRUAN's stringent requirements. We would also like to thank the NOAA GCOS office, through the Meteorological Service of New Zealand Limited, for supporting this research.

AMTD
7, 11957–11989, 2014

References

Austin, J., Tourpali, K., Rozanov, E., Akiyoshi, H., Bekki, S., Bodeker, G. E., Brühl, C., Butchart, N., Chipperfield, M., Deushi, M., Fomichev, V. I., Giorgetta, M. A., Gray, L., Kodera, K., Lott, F., Manzini, E., Marsh, D., Matthes, K., Nagashima, T., Shibata, K., Stolarski, R. S., Struthers, H., and Tian, W.: Coupled chemistry climate model simulations of the solar cycle in ozone and temperature, J. Geophys. Res., 113, D11306, doi:10.1029/2007JD009391, 2008. 11968

Bodeker, G. E., Boyd, I. S., and Matthews, W. A.: Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986–1996, J. Geophys. Res., 103, 28661–28681, 1998. 11970

Bodeker, G. E., Hassler, B., Young, P. J., and Portmann, R. W.: A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations, Earth Syst. Sci. Data, 5, 31–43, doi:10.5194/essd-5-31-2013, 2013. 11967

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech. Discuss., 7, 3727–3800, doi:10.5194/amtd-7-3727-2014, 2014. 11960, 11975

Efron, B. and Tibshirani, R.: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Stat. Sci., 1, 54–77, 1986. 11968

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217–1231, doi:10.5194/amt-3-1217-2010, 2010. 11960

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li. W.: Applied Linear Statistical Models, 5th edn., McGraw Hill, Singapore, 1396 pp., 2005. 11961

Moore, D. S. and McCabe, G. P.: Introduction to the Practice of Statistics, W. H. Freeman and Company, New York, 828 pp., 2003. 11962

Nering, E. D.: Linear Algebra and Matrix Theory, John Wiley and Sons, New York, 352 pp., 1963. 11965

Philipona, R.: Solar and thermal radiation errors on upper-air radiosonde temperature measurements, J. Atmos. Ocean. Tech., 30, 2382–2393, 2013. 11960

Philipona, R., Kräuchi, A., and Brocard, E.: Solar and thermal radiation profiles and radiative forcing measured through the atmosphere, Geophys. Res. Lett., 39, L13806, doi:10.1029/2012GL052087, 2012. 11960

AMTD
7, 11957–11989, 2014

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: Numerical Recipes in Pascal. The Art of Scientific Computing, New York, Cambridge University Press, 759 pp., 1989. 11969, 11970

Randel, W. J. and Wu, F.: Isolation of the ozone QBO in SAGE II data by singular-value decomposition, J. Atmos. Sci., 53, 2546–2559, 1996. 11968

Reed, R. J., Campbell, W. J., Rasmussen, L. A., and Rogers, D. G.: Evidence of downward-propagating annual wind reversal in the equatorial stratosphere, J. Geophys. Res., 66, 813–818, 1961. 11962

Reinsel, G. C., Tiao, G. C., Miller, A. J., Wuebbles, D. J., Connell, P. S., Mateer, C. L., and DeLuisi, J. J.: Statistical analysis of total ozone and stratospheric Umkehr data for trends and solar cycle relationship, J. Geophys. Res., 92, 2201–2209, 1987. 11973

Ribera, P., Gallego, D., Peña-Ortiz, C., Gimeno, L., Garcia-Herrera, R., Hernandez, E., and Calvo, N.: The stratospheric QBO signal in the NCEP reanalysis 1958–2001, Geophys. Res. Lett., 30, 1691, doi:10.1029/2003GL017131, 2003. 11962

Seidel, D. J., Berger, F. H., Diamond, H. J., Dykema, J., Goodrich, D., Immler, F., Murray, W., Peterson, T., Sisterson, D., Sommer, M., Thorne, P., Vömel, H., and Wang, J.: Reference upper-air observations for climate: rationale, progress, and plans, B. Am. Meteorol. Soc., 3, 361–369, 2009. 11959

Sioris, C. E., McLinden, C. A., Fioletov, V. E., Adams, C., Zawodny, J. M., Bourassa, A. E., Roth, C. Z., and Degenstein, D. A.: Trend and variability in ozone in the tropical lower stratosphere over 2.5 solar cycles observed by SAGE II and OSIRIS, Atmos. Chem. Phys., 14, 3479–3496, doi:10.5194/acp-14-3479-2014, 2014. 11968

Steinbrecht, W., Hassler, B., Claude, H., Winkler, P., and Stolarski, R. S.: Global distribution of total ozone and lower stratospheric temperature variations, Atmos. Chem. Phys., 3, 1421–1438, doi:10.5194/acp-3-1421-2003, 2003. 11968

Tiao, G. C., Reinsel, G. C., Xu, D., Pedrick, J. H., Zhu, X., Miller, A. J., DeLuisi, J. J., Mateer, C. L., and Wuebbles, D. J.: Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation, J. Geophys. Res., 95, 20507–20517, 1990. 11961, 11970

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E.: Factors affecting the detection of trends: statistical considerations and applications to environmental data, J. Geophys. Res., 103, 17149–17161, 1998. 11961, 11970, 11971

AMTD
7, 11957–11989, 2014

Table 1. The mean 1σ uncertainties of the 500 trend values derived for the B98, W98 and bootstrapping approaches. For more details see text.

Degree of unexplained variance on signal	Measurement uncertainties	1σ uncertainty on the trend		
		B98	W98	Bootstrapping
none	uncert 1.0	0.0104	0.0104	0.0104
medium	uncert 1.0	0.0316	0.0323	0.0322
high	uncert 1.0	0.0799	0.0823	0.0823
none	uncert 1.5σ0.5	0.0125	0.0156	0.0165
medium	uncert 1.5σ0.5	0.0340	0.0350	0.0347
high	uncert 1.5σ0.5	0.0810	0.0850	0.0832
none	uncert 4.5σ1.5	0.0374	0.0468	0.0494
medium	uncert 4.5σ1.5	0.0527	0.0563	0.0581
high	uncert 4.5σ1.5	0.0929	0.0957	0.0954

AMTD
7, 11957–11989, 2014

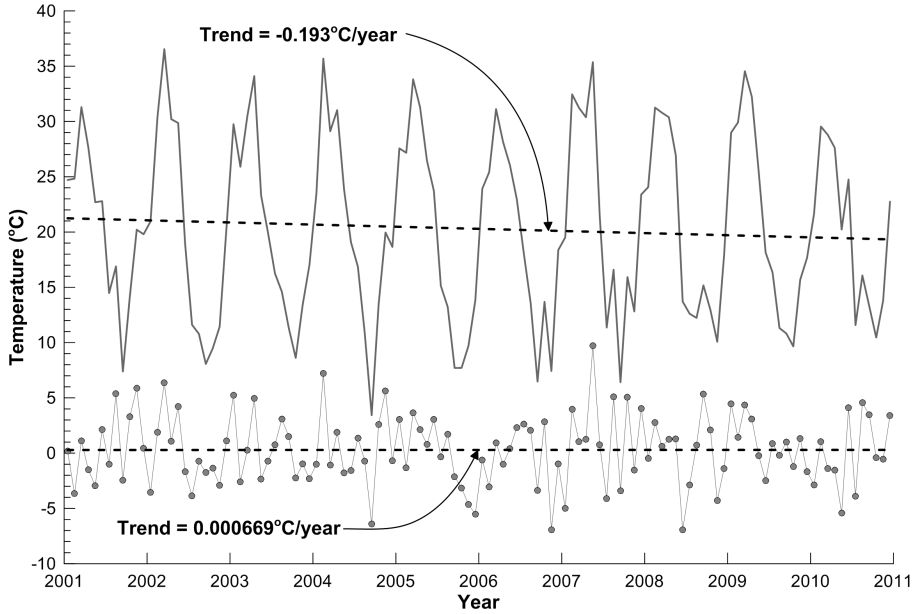

Figure 1. Solid grey line: a synthetic time series created from a purely periodic signal (plus noise) but still exhibiting a non-zero trend (upper dashed straight line fit). Line with filled dots: the deseasonalized signal with a straight line fit showing no statistically significant trend.

AMTD
7, 11957–11989, 2014

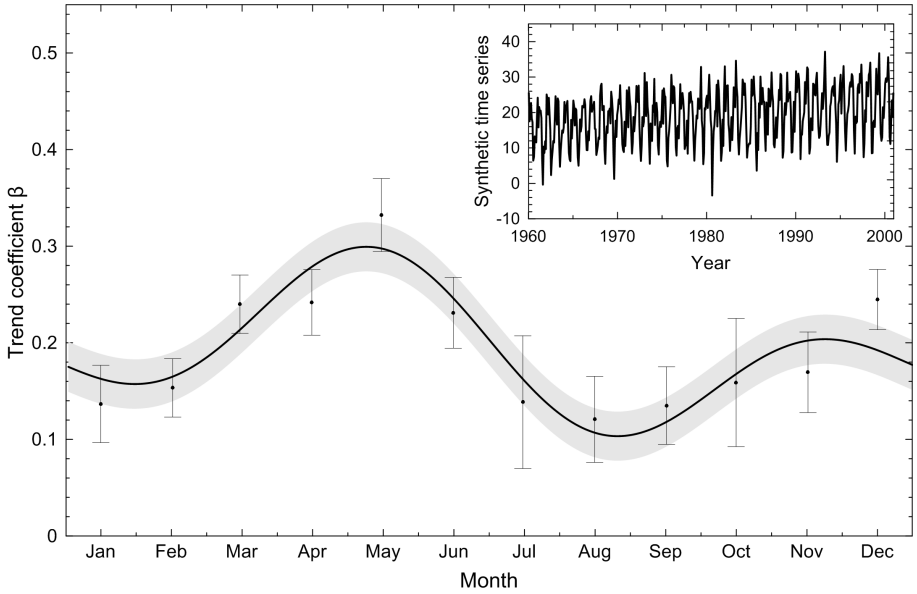

Figure 2. The seasonally resolved trend coefficient derived from fitting Eq. (1) (including only the α and β terms) to the synthetic time series shown in the inset together with its 1σ uncertainty (solid black line with grey shaded area). The trends obtained from linear fits to the data from each month separately, together with their 1σ uncertainties, are shown with black dots and vertical error bars respectively.

AMTD
7, 11957–11989, 2014

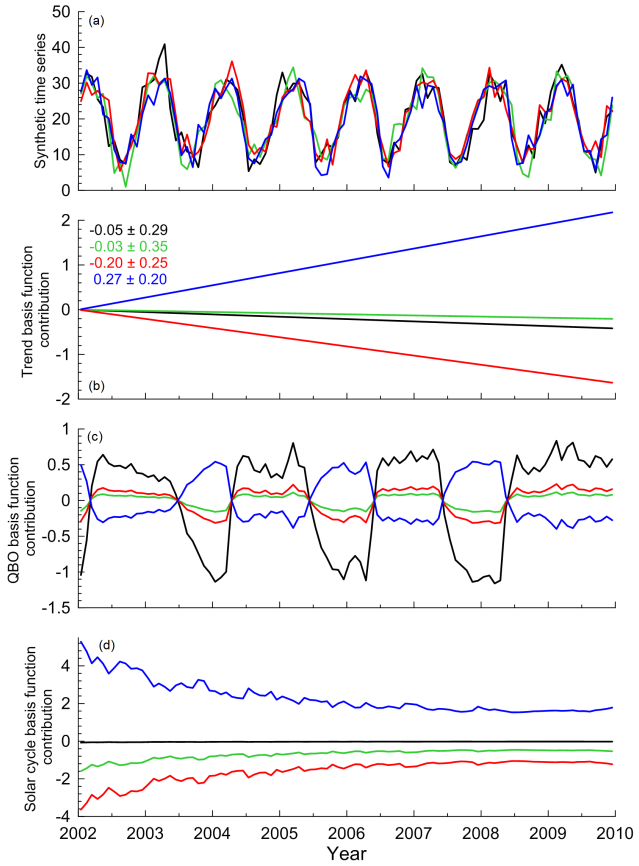

Figure 3. Four synthetic monthly mean time series calculated by adding Gaussian distributed random noise to a repeating mean annual cycle **(a)** together with the contributions from the trend **(b)**, QBO **(c)** and solar cycle **(d)** basis function derived by fitting Eq. (1) to the time series in **(a)**. The quantity plotted is artificial and therefore unitless. The trend values listed in **(b)** include their 1σ uncertainties.

AMTD
7, 11957–11989, 2014

Techniques for analyses of trends in GRUAN data

G. Bodeker and S. Kremser

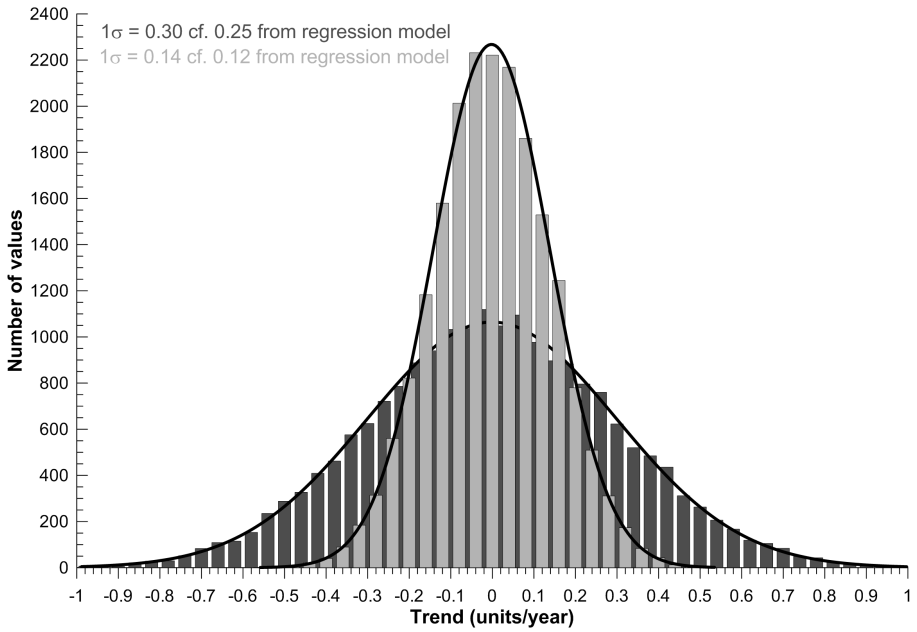

Figure 4. Histograms generated from 20 000 simulations of trend results derived from application of Eq. (1) to synthetic time series (as shown in Fig. 3a) first with the original solar cycle and QBO basis functions (dark grey) and then with detrended QBO and solar cycle basis functions (light grey). The 1σ standard deviation of Gaussian fits to the histograms are listed together with the mean trend uncertainties derived from the regression model. The mean uncertainties were obtained by averaging the uncertainties from the 20 000 trend values.

AMTD
7, 11957–11989, 2014

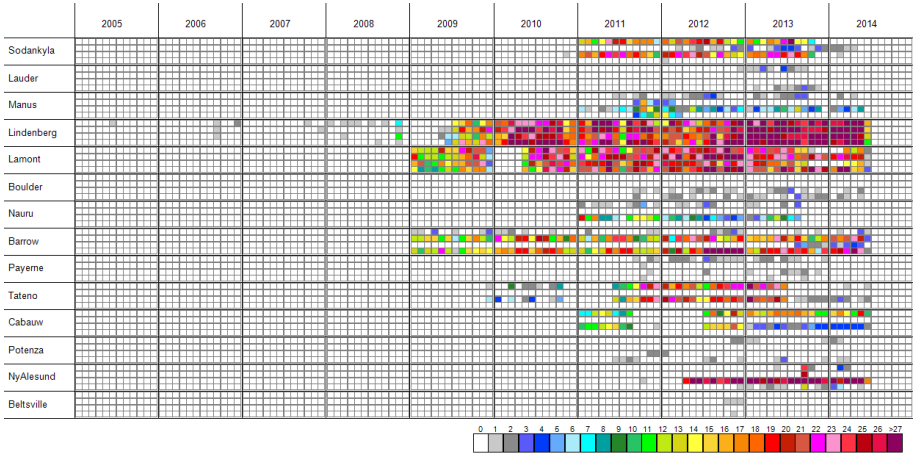

Figure 5. The availability of RS92 radiosonde data at a sub-set of the GRUAN sites (not all GRUAN sites currently provide RS92 radiosonde measurements) at the time of the writing of this manuscript. Each site shows data in four rows indicative of the four 6 h periods through the day. Vertical columns in the grid indicate individual months with the years are listed across the top. The number of RS92 radiosonde flights available in each month in each 6 h period is shown using the colour scale in the bottom right corner of the figure.

AMTD
7, 11957–11989, 2014

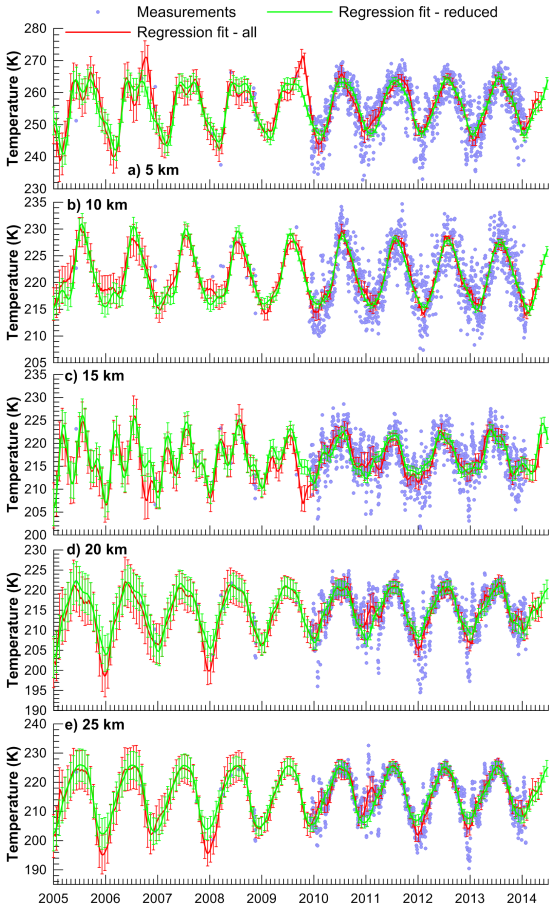

Figure 6. Examples of regression model fits to the 00:00 UT measurements at five selected altitude levels spaced 5 km apart. 1σ uncertainties are shown as error bars on the regression model fits. Because the measurement random uncertainties are small i.e. ~ 0.1 K, they are not shown.

AMTD
7, 11957–11989, 2014

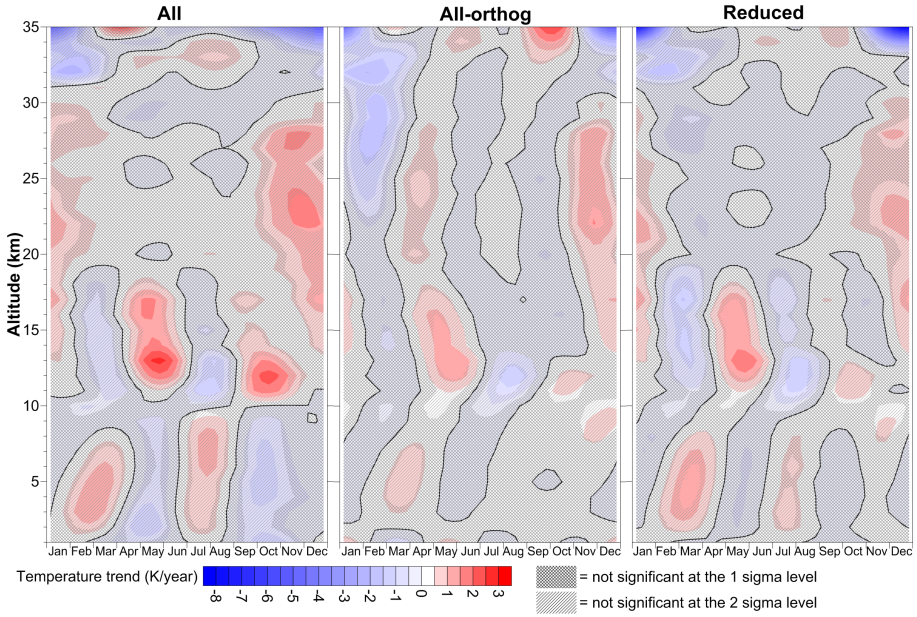

Figure 7. 00:00 UT Lindenberg temperature trends as a function of altitude and season for the three different constructs of the regression model. The trends, in K year^{-1}, are shown using the colour scale at the bottom left. Trends that are not statistically significantly different from zero at the 1σ level are covered in cross-hatching while trends that are not statistically significantly different from zero at the 2σ level are covered in single-hatching.

AMTD
7, 11957–11989, 2014

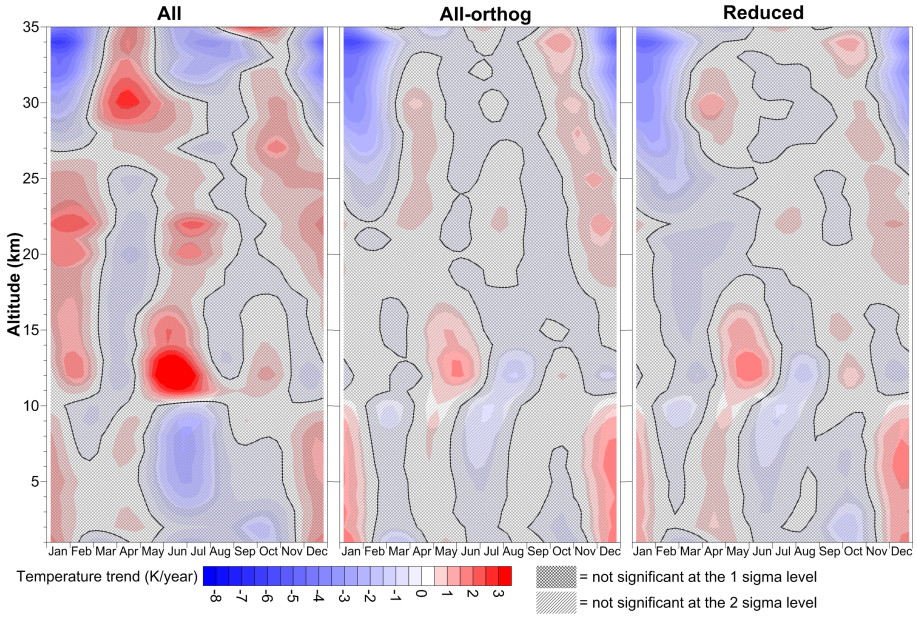

Figure 8. As for Fig. 7 but for 12:00 UT.

AMTD
7, 11957–11989, 2014

Techniques for analyses of trends in GRUAN data

G. Bodeker and S. Kremser