Interactive comment on "An objective determination of optimal site locations for detecting expected trends in upper-air temperature and total column ozone" by K. Kreher et al.

The reviewers comments are written in italics, the authors' response in bold.

Anonymous Referee #1

General comments:

This paper describes (1) about determination of measurement uncertainty criteria of the upper air sounding observation with considering the uncertainty from the sampling errors using CFSR data and (2) about the method to determine the observation sites for 21st century tropospheric temperature, stratospheric temperature, and ozone. About (1), the new point of this paper is (i) using the latest reanalysis data with the raised model top and (ii) analyzing up to the higher level (1hPa) than the referred paper.

About (2), it is very example to determine the historical observation sites objectively while the selection criteria should be discussed further because of the requirements for the other field. Totally, this paper is very informative for the observation network design.

Thank you very much for your summary and feed-back.

Minor comments:

Introduction)

1. In the line 23 in page 1620, the paper refers GRUAN website for its network. But the web site may become obsolete in the future. So that, the current GRUAN network should be displayed explicitly in this paper. How about showing them in figure 1?

This is a good idea, also because the map of GRUAN sites shown on the web page shows a different set of GRUAN sites to that included in this analysis; GRUAN has expanded somewhat since this analysis was completed and the paper written. The GRUAN sites included in this study are now shown as blue dots in Figure 1 and the relevant text in the manuscript has been changed to reflect this.

In section 2)

2. In the line 11 in page 1621, please show the referred paper used NCEP-NCAR reanalysis data since there are several reanalysis data currently and the characteristics are different. So that it should be shown explicitly.

We have added the reference for this paper in the manuscript as suggested.

3. In the line 16 in page 1622, please show the point selection strategy (as a sample and selected randomly) as shown in the caption of the figure. This information is important and should be stated in the main part of the paper.

We have added the requested material to the manuscript text accordingly.

4. About the figure 3, these sampling scenarios should be summarized in a different table.

We have added a table (Table 1) listing the sampling frequencies because instead of "sampling strategies" we now only use "sampling frequencies" and have dropped the additional change in "time of day". We have also applied this change to Figures 2, 3, 4, 6 and 7.

And, since the referred paper mentioned that "made at least twice daily, at least once every two or three days", why are not there scenarios as "noon and midnight for every 2/4/7 days"?

We could have certainly also explored measurement regimens of noon and midnight measurements every 2, 4 or 7 days. This would have added three more measurement regimens to what was already a rather crowded set of regimens and would not have explored anything different to what Seidel and Free (2006) explored. Our goal was not to replicate what Seidel and Free did but to also explore some novel measurement regimens. Therefore, we felt that adding these three measurement regimens would not add significantly to the conclusions drawn and, in the interests of expediency and clarity, omitted them.

5. In the line 11-12 in page 1623, although it mentioned the criteria (0.2K) is not different for the other points / levels, it is better showing one (or two) other point sample (figure) in the other (lower) latitudes. It is hardly understood that one high latitude point situation can represent the globe.

We agree and have added another figure showing the uncertainties on the monthly mean temperatures for 35°S, 45°E. We have changed the text in the manuscript accordingly.

In section 3)

6. Around the line 5 in 1628, what AMSU data is used?

As detailed in the manuscript, the AMSU data set described in Mears and Wentz (2008) was used.

Just TB? How do you deal with the cloud contamination?

We used the version 3.2 merged MSU and AMSU data set exactly as described in Mears and Wentz (2008). No additional processing was applied.

7. In the line 28 in page 1628, it mentioned about GUAN network but no information was provided. Please mention it and it is better to show GUAN sites.

We agree with the reviewer that we need to add more information but we think that it would unnecessarily bloat the paper to include an additional figure showing the location of all 171 GUAN sites when this is available from other sources. Rather we have cited the seminal GCOS document that described GUAN in detail and have included the link to a web page that provides a map of all GUAN sites).

In section 4)

8. Why table 3 and 4 is so different? Please note any idea about it

We have added a paragraph explaining Table 3 and 4 in more detail in the manuscript text. And since we have added another table, Table 3 and 4 are now Table 4 and 5.

N.R.P. Harris (Referee #2)

This manuscript contains a description of a statistically based approach to identify measurement sites to detect future trends in (a) free tropospheric and stratospheric temperatures and (b) total ozone. The putative future trends are based on CCMVal-2 model runs, and the variability and auto-correlation derived from historic observations are assumed to remain the same in the future. Future measurements are assumed to be stable with all the certainty being included in the instrumental precision. No allowance is made for drifts or offsets. Based on these assumptions, the time to detect a trend is quantified for globally diverse locations (including current measurement sites) and its dependence on measurement frequency *and timing is investigated. This study contains a lot of good work and several interesting results. The temperature part is especially convincing and should provide valuable information for network development. The work should be published. However, I think the presentation study could, and should, be significantly improved before publication.*

Thank you very much for your summary and feed-back.

General comments

There are two main aspects to this: (i) more context and discussion of the results; and (ii) increased clarity of presentation. Neither of these need much (if any) new calculation, just a re-thinking of the presentation. These aspects are discussed in turn

The authors currently include very little discussion of the results. The rationale for this (p. 1629, lines 20-24) is that "The purpose of the exercise is to show that generating fields....... provide(s) one objective method of selecting optimal sites....". That is fair enough at one level, but *I think any interested reader of this study would want to know more about the other factors that could be considered.*

It would also help to have a clearer discussion of the assumptions that have been made. For example, what if the model trends are wrong? How well do they reproduce past trends?

Text to address this has been added in the "Discussion and Summary" section.

What is the effect of specific geophysical features (e.g. broadening of the tropics)? Should some sites be chose with this in mind? What additional information is relevant for knowing an atmospheric change is happening? E.g. change in tropopause height or in vertical/latitudinal shape of trends – or, for ozone, trends in the vertical profile. These are factors that should be considered when designing a network of stations as they could provide additional constraints to the results presented here.

The goal of our paper is not to design a network of stations for detecting changes in atmospheric composition. Much of what the reviewer has requested is therefore well beyond the scope of our paper. The goal of our paper is to present one objective method, which in a real-world application would be used together with the multiple additional considerations that the reviewer alludes to, to select the location for measurements of atmospheric temperature and total column ozone.

Some discussion of possible drifts (how well the network stability is known) and other potential instrumental uncertainties should also be included.

Because our paper is not dealing with a specific network, we cannot make any comment on network stability. This study considers an <u>ideal network</u> rather than an existing network. The issue of measurement drift and other potential instrumental uncertainties is therefore tangential to the scope of this paper. Were we to include analyses of the effects of instrumental drift and/or time dependent systematic biases in the measurements (e.g. those resulting from changes in instrumentation), this would add a whole new dimension to the paper that is currently not considered at all. This would be a topic for a follow-up paper. We therefore cannot see how to accommodate this request by the reviewer.

Precision (which, as the authors point out, has a simple relation with sampling frequency) is an important factor but it is not the only one and may not be the most important.

Our paper does not discuss precision at all and this word appears nowhere in the manuscript. We therefore do not understand this point raised by the reviewer.

I would like to emphasise that I am looking for is a fuller discussion of the assumptions, their implications and the broader debate to which this study contributes.

Specific concerns raised by the reviewer are addressed below.

The clarity of the manuscript should also be improved. Most importantly, the logic of the work needs to be made clearer. Some of the critical steps are described very briefly and were easy to skip over.

We made some changes to the manuscript to improve the clarity and comprehensibility.

Some improvements could be made to the figures so that their meaning is more easily understood. Having said that, I think that Figs 8-10 are very clear and make the main points very well. The build-up to them could be improved though. Some specific comments are given below.

These specific comments have been addressed (details below).

My last general comment concerns the section on ozone trend detection. It is weaker than the section on temperature trends, but that is probably inevitable since only the total column is considered and sampling frequency is not discussed. I am ambivalent on whether it stays in. If it does, then it needs strengthening. There was some early work on measurement strategy *(station location, sampling frequency) by Tiao et al. (1990) which should be mentioned as a background and complement to the work presented here. (Note that further study by them (not sure if it was published – it is referred to at the end of Section 4) showed that persistence of weather systems weakened their conclusion that "the precision of trend estimates (is) very insensitive to changes in the temporal sampling rate of daily data", and led to a conclusion that measurements needed to be spaced regularly through the month, e.g. on the earliest possible day of each week.)*

G.C. Tiao, G.C. Reinsel, D. Xu, et al., Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation. J.G.R. 95(D12), 20507-20517 (1990).

This has been addressed in the text and a discussion of the work by Tiao et al. (1990) and Weatherhead et al. (2000) has been included.

At the same time, the final part of this work (1631, 15-24) needs to be described more fully – it feels like an express train at the moment.

We agree with the criticism and have extended the discussion in the manuscript.

Specific comments

Abstract: The current version of the abstract does not exactly entice the casual reader further in to the paper. The problem is in the first paragraph (which reads like an earlyish draft) as the second one is fine. I am loath to make too many comments as it should basically be rewritten, starting with the aims of the paper and then the three stages should be described (probably without explicitly saying there are three stages) along with some of the assumptions. The aim should be to make it shorter. The start of the summary is clear and would be a good place to start.

The abstract has been extensively modified taking the reviewers comments into account.

Section 2.1: The core of the temperature analysis, in some ways, is the second paragraph. I think it need a bit more substance – how good are the NCEP fields?

An assessment of the quality of the NCEP CFSR temperature fields is beyond the scope of this paper. We refer the reviewer to the outcomes of the SPARC Reanalysis Intercomparison Project which is undertaking a detailed assessment of the quality of the

NCEP CFSR temperature fields. For the purposes of our study, the key characteristics that the temperature fields need to capture are:

1) **The amplitude of day-to-day variability,**
2) **The amplitude of the diurnal cycle in temperature,**
3) **The structure of the seasonal cycle in temperature,**

at each of the sites considered. We have not conducted an assessment of the quality of the reanalyses in regard to these three characteristics.

; state explictly that it is temperature only

We have inserted the word 'temperature' in front of NCEP CFSR to make it clear that it is the temperature data from the NCEP CFSR database that is being used to assess the effects of sampling frequency on the random uncertainty on monthly mean temperatures.

and what spatial resolution the measurements are?

The spatial resolution of the temperature fields is 0.5°×0.5° and this is now stated in the manuscript.

How well do they correspond to an in situ radiosonde profile? I assume that many of the radiosondes going in are 12 hourly – if so what is the main influence on the intervening 6 hour temperatures and is the diurnal cycle reproduced well? I guess I worry about the use of the word 'true'.

The reviewer is correct in stating that the reanalyses should not be considered as 'the truth'. The extent to which the 4-dimensional variational assimilation captures the:

1) **The amplitude of day-to-day variability,**
2) **The amplitude of the diurnal cycle in temperature,**
3) **The structure of the seasonal cycle in temperature,**

is what matters for the purpose of our study. However, as stated above, we have not conducted an assessment of the quality of the reanalyses in regard to these three characteristics.

Also, a bit more detail on the Monte-Carlo method would be useful

More detail has been added and the section has been rewritten to clarify the method further.

– should the material at the beginning of Section 2.2 be moved earlier?

We believe it is easier to follow the discussion if the material at the beginning of Section 2.2 stays where it currently is and is not be moved earlier.

Figs 2 and 3: Figure 2 does not show that much which could not be stated in the text

We agree. We have removed Figure 2 and stated the information in the text.

, while Figure 3 has 12 panels in which the most obvious difference is between sampling frequencies rather than sampling time. And not much is said about the differences resulting from changing sampling times in the text. I wonder whether it would make more sense to cut the number of panels in Fig 3 to, say, 6 which show different sampling frequencies and to state in the text that changing the time of day does not have a significant influence on the uncertainty on the monthly mean. The same format could plausibly be used in Figure 2

We have implemented the suggestions and cut the panels in the new Figure 2 and in Figure 3 back to 8 panels only using the sampling frequencies and not the change in time of day. To be consistent, we have also applied this to Figures 4, 5 and 7.

Figure 5 would also be easier to interpret if the x-axis represented different sampling frequencies, i.e. the different sampling strategies were removed.

We agree and have done this (see comment above).

Figs 6 and 7 and accompanying text. It would help the reader if the connection between Figs 6 & 7 could be made clearer. For example, the model trend values for the two cases shown in Fig 7 should be mentioned in the text and/or the Figure caption. This would help give some meaning to equation 1.

The values for the projected trends at the 2 sites have been added to the manuscript.

Fig 7 is presumably meant to build up to Figs 8-10, but I personally find these latter figures much easier to grasp the meaning of (though the assumed sampling frequency should be mentioned in the caption). The main use of the figure it to show the effect of sampling frequency, so again the number of scenarios shown could be reduced.

We agree and have simplified the figure by only using the sampling frequency.

An objective determination of optimal site locations for detecting expected trends in upper-air temperature and total column ozone

K. Kreher[1], G.E. Bodeker[1], and M. Sigmond[2]

[1]{Bodeker Scientific, Alexandra, New Zealand}

[2]{Canadian Centre for Modelling and Analysis, Victoria, BC, Canada}

Correspondence to: K. Kreher (karin@bodekerscientific.com)

Abstract

Detection of climate change requires a network of stable ground-based long-term measurements. Building upon earlier work, we first explore requirements of such measurements (such as maximum random uncertainty and sampling frequency) to ensure a minimum random uncertainty in monthly mean temperatures and to ensure effective detection of trends. In agreement with previous work we find that only for individual measurement random uncertainties >0.2 K does the measurement random uncertainty start to contribute significantly to the random uncertainty of the monthly mean. For trend analysis, we find that only when the measurement random uncertainty exceeds 2 K, and measurements are made just once or twice a month, is the quality of the trend determination compromised.

~~In the first study reported on here, requirements on random uncertainty of instantaneous temperature measurements, sampling frequency, season and pressure, required to ensure a minimum random uncertainty of monthly mean temperatures, have been explored. These results then inform analyses conducted in a second study which seeks to identify the optimal location of sites for detecting projected trends in upper-air temperatures in the shortest possible time. The third part of the paper presents a similar analysis for the optimal locations of sites to detect projected trends in total column ozone. Results from the first study show that only for individual measurement random uncertainties >0.2 K does the measurement random uncertainty start to contribute significantly to the random uncertainty of the monthly mean. Analysis of the effects of the individual measurement random uncertainty and sampling strategy on the ability to detect upper-air temperature trends shows that only when the measurement random uncertainty exceeds 2 K, and measurements are made just once or twice a month, is the quality of the trend determination compromised.~~ In the second part of the study we provide guidance on how to most effectively design a measurement network. To this end we developed a method to objectively identify the optimal location of sites for detecting projected trends in upper-air temperatures and total column ozone in the shortest possible time. This is done by first estimating the spatial distribution of the minimum time measurements required to detect projected trends in temperature and ozone. This quantity is calculated from the unforced variance in the signal and the degree of auto-correlation, both estimated from historical data sets and assumed not to change in the future, and the projected trends as estimated from chemistry climate models. The optimal site locations are then selected by an iterative procedure

Formatted: Normal

based on the minimum time required to detect a trend and a minimal distance between different measurement sites.

~~The time to detect a trend in some upper-air climate variable is a function of the unforced variance in the signal, the degree of autocorrelation, and the expected magnitude of the trend. For middle tropospheric and lower stratospheric temperatures, the first two quantities were derived from Microwave Sounding Unit (MSU) and Advanced Microwave Sounding Unit (AMSU) measurements while projected trends were obtained by averaging 21st century trends from simulations made by 11 chemistry-climate models (CCMs). For total column ozone, variance and autocorrelation were derived from the Bodeker Scientific total column ozone database with projected trends obtained from median values from 21 CCM simulations of total column ozone changes over the 21st century.~~ While the optimal sites identified ~~in this analysis for detecting temperature and total column ozone trends in the shortest time possible~~ here result from our use of only one of a wide range of objective strategies, these results provide additional incentives for initiating measurement programmes at these sites or, if already in operation, to continue to be supported.

1 Introduction

Stratospheric temperatures represent the first order connection between natural and anthropogenically driven changes in radiative forcing and changes in other climate variables at the Earth’s surface. There is, therefore, a strong interest in detecting upper-air temperature trends as efficiently and reliably as possible. The vertical structure of temperature trends also provides important information for climate change attribution since increases in atmospheric long-lived greenhouse gas (GHG) concentrations warm the troposphere, but cool the stratosphere. Ozone also acts as a GHG and absorbs UV radiation in the stratosphere such that changes in ozone concentration also change the temperature structure of the atmosphere. Thus, dependable long-term measurements of temperature and ozone are essential for climate change detection and attribution studies.

Historical observations present challenges for estimating trends since measurement uncertainties can be large. A number of papers (e.g. Free et al., 2002; Wang et al., 2012 and references therein) point to the inherent, and partly irreparable, problems that arise from complicated merging of data sets of changing or unknown quality and different measurement approaches. Homogenisation of merged data sets cannot eliminate the respective uncertainty in derived trends.

Although satellite instruments are capable of measuring the vertical distribution of temperature and ozone globally, the resultant measurement series are often deficient for trend detection since: 1) The calibration of satellites, once in orbit, is a challenging task and slight differences in instrumental design, satellite and satellite-operation design, and retrieval algorithms impose severe difficulties on constructing homogeneous time series (Thompson et al., 2012). 2) Individual satellites often measure over periods that are too short to detect trends, and may stop operating unexpectedly, preventing appropriate continuity or overlap in observations. 3) The vertical and horizontal resolution of satellite measurements may be too coarse to allow for appropriate interpretation and attribution of observed changes.

Stable ground-based long-term temperature and ozone measurements at selected sites, adhering to stringent measurement standards and traceability protocols (e.g. Immler et al., 2010), facilitate the calibration of individual satellite instruments (e.g. Tobin et al., 2006; Balis et al., 2007; Adams et al., 2013) and support the merging of data sets from different satellites with the goal of creating reliable long-term climate data records (e.g. Tummon et al., and references therein). Such high-quality temperature and ozone time series can also support bridging any gaps that may emerge in satellite data records. With unexpected termination of satellite operations, as well as the ongoing change in satellite technology, bridging gaps becomes critical to creating a continuous monitoring system for the global atmosphere.

More importantly though, these data sets could allow trend analysis in their own right. A first requirement would be that measurements are performed with sufficiently low random uncertainty and at a sufficiently high sampling rate. The minimum requirements have been previously explored by Seidel and Free (2006). They found that observations with an uncertainty of ≤0.5 K, made at least twice daily, at least once every two or three days, were sufficient to ensure accurate monthly climate statistics (specifically, monthly mean temperature and standard variation), i.e. only ~5% of monthly statistics will be significantly different from those based on four observations per day.

A second requirement for the measurement network to provide a global picture of the trends is that the ~~If~~ observing sites are strategically placed~~, they might well~~ and sample a sufficiently diverse range of regimes~~ to provide a global picture of trends~~. Most of the current measurement sites, however, are located close to populated areas for ease of access and for historical reason.

With approximately 90% of the global population living in the Northern Hemisphere, measurement sites favour the Northern Hemisphere. As a result, such a distribution of sites is unlikely to be representative of the global climate. An example for this is the distribution of the then 15 ~~initial~~ GRUAN (GCOS Reference Upper-Air Network) sites ~~which were~~ considered at the time of this analysis which are predominantly located at Northern Hemisphere mid-latitudes (see blue dots in Figure 1~~map at www.gruan.org~~). Hence, the need exists to provide an objective approach to determine the optimal location of sites. To address this need, we describe one objective approach for locating sites for early temperature and ozone trend detection.

~~To address this need, we describe one objective approach for locating sites for early temperature and ozone trend detection.~~ In the first part of this study (Section 2), we expand the analysis of Seidel and Free (2006) and examine the effects of the individual measurement random uncertainty (hereafter simply referred to as the ‘measurement uncertainty’ to distinguish from systematic biases) and sampling strategy on the robustness of upper-air temperature trend detection. In the second part, we address the need for an objective approach for a site selection process with t~~T~~he description and outcomes of the site selection process being~~are~~ described in Section 3 for temperature and in Section 4 for ozone.

2 Sampling and trend detection using temperature profiles

The two key questions addressed in this section are: What individual measurement uncertainty and measurement frequency is needed to achieve a certain uncertainty in monthly mean temperature and what are the effects of the individual measurement uncertainty and sampling strategy on the ability to detect upper-air temperature trends?

The temperature profile data used within this study are 6 hourly data from the Climate Forecast System Reanalysis (CFSR; Saha et al., 2010) produced by the National Centers for Environmental Prediction (NCEP). Seidel and Free (2006), using the reanalysis of the climate of the past half century (Kistler et al., 2001) as a model of temperature variations over the next half century, tested various data collection protocols to develop recommendations for observing system requirements to monitor upper-air (here we define ‘upper-air’ as the free troposphere and above) temperature trends. The analysis of Seidel and Free (2006) focussed on estimating monthly average temperature and its standard deviation, as well as multi-decadal trends in monthly temperatures at specific locations, from the surface to 30 hPa. The analysis presented here repeats, in part, that of Seidel and Free (2006), but extends above 30 hPa (the highest level analysed by Seidel and Free) using NCEP CFSR temperature data to extend the results to 1 hPa and to add to some of their conclusions especially with the goal in mind to provide site location recommendations.

~~Seidel and Free (2006) also assessed the effects on monthly mean temperatures of increasing the uncertainty of temperature measurements, incomplete sampling of the diurnal cycle, incomplete sampling of the days in the month, imperfect long-term stability of the observations, and changes in observation schedule. They found that observations with an uncertainty of ≤0.5 K, made at least twice daily, at least once every two or three days, were sufficient to ensure accurate monthly climate statistics (specifically, monthly mean temperature and standard variation), i.e. only ~5% of monthly statistics will be significantly different from those based on four observations per day.~~

2.1 Effects of sampling on monthly mean uncertainty

To corroborate the findings of Seidel and Free (2006), and to extend the analysis into the upper stratosphere, a similar approach has been followed here, where for a number of selected locations, the uncertainty on monthly mean temperatures is determined as a function of the uncertainty on each contributing instantaneous measurement, sampling frequency, season, and pressure. For this study, we have used 72 locations in 90° longitude zones and 10° latitude zones and added the 15 initial GRUAN sites, which results in a total of 87 locations as shown in Figure 1.

The analysis is based on sampling of NCEP CFSR temperature fields with a spatial resolution of 0.5°×0.5°, assuming that sampling at the highest possible frequency (6 hourly) produces the ‘true’ monthly mean. Then, by simulating different sampling strategies, with different simulated uncertainties on each measurement, and doing this in a Monte Carlo framework, the standard deviation of the differences between the calculated monthly means and the true monthly means can be determined.

Formatted: Font: Not Bold

~~Figure 2 shows the uncertainty on the monthly mean temperature as a function of season and the uncertainty on each individual measurement for 6-hourly sampling throughout the month. There is no contribution to the uncertainty on the monthly mean from sampling because the same 6-hourly sampling is used to derive the ‘true’ monthly mean. Therefore, the uncertainty on the monthly mean is about an order of magnitude smaller than the uncertainty on each instantaneous measurement, which is to be expected when averaging ~120 measurements~~

Formatted: Space After: 0 pt

~~through the month i.e. 1/√120≈0.1. As can be seen clearly in Figure 1, the seasonal influence is minimal.~~

Figure 2~~3~~ shows the uncertainty on the monthly mean temperature~~s~~ at 50 hPa and at 85°N, 135°W as a function of season and as a function of the uncertainty on each individual measurement for a range of different sampling frequencies listed in Table 1. The location has been selected randomly as an example and Figure 3 shows the same information for a second randomly selected location (35°S, 45°E). The first top panel of Figures 2 and 3 shows the uncertainty on the monthly mean for 6-hourly sampling throughout the month. There is no contribution to the uncertainty on the monthly mean from sampling because the same 6-hourly sampling is used to derive the 'true' monthly mean. Therefore, the uncertainty on the monthly mean is about an order of magnitude smaller than the uncertainty on each instantaneous measurement, which is to be expected when averaging ~120 measurements through the month i.e. 1/√120≈0.1. As can be seen clearly~~ in Figure 1~~, the seasonal influence is minimal.

~~for the same location and pressure level as in Figure 2 but now for a range of different sampling strategies, including the 6-hourly sampling already displayed in Figure 2.~~ Note that this is the uncertainty on the monthly means, neglecting any systematic errors (offsets) as these are less important for trend analysis – so while sampling every 24 hours at noon would produce monthly mean temperatures very different to what would be achieved when sampling every 24 hours at midnight, the standard deviation of the differences between the calculated monthly mean and the true monthly mean (rather than the absolute value) is what is assessed. The uncertainty on the monthly mean now shows a clear seasonal cycle for 12-hourly sampling, or coarser, since the temperatures show a higher degree of variability in the winter months at this location and level. At this pressure level (50 hPa), reductions in measurement uncertainty below 0.2 K have little effect on the uncertainty on the monthly mean because it is the uncertainty resulting from incomplete sampling that dominates. It is only for a measurement uncertainty greater than 0.2 K that the measurement uncertainty begins to make an appreciable contribution to the uncertainty on the monthly mean. This 0.2 K threshold does not only apply to the two~~one~~ locations displayed in Figures 2 and 3 but is also valid for the other locations at 50 hPa as well as most other sites at 500 hPa (not shown here). The 0.2 K threshold also supports the GRUAN target of less than 0.2 K uncertainty for instantaneous stratospheric temperature measurements

(Immler et al., 2010). The permissible measurement uncertainty varies with pressure and season.

The permissible uncertainty of individual temperature measurements required to avoid increasing the uncertainty on the monthly means by more than 10% above what would be achieved when sampling with 0.01 K uncertainty is shown in Figure 4. Results from all the 87 locations selected for this analysis and for all months were averaged to produce this figure, with the individual curves showing the permissible uncertainty for each of the 7~~11~~ sampling frequencies~~schemes~~.

When sampling every 12 hours, at noon/midnight~~ or at 6am/6pm~~ (solid blue ~~and cyan ~~curve~~s~~), in the upper stratosphere, measuring with 0.5 K uncertainty is sufficient to avoid affecting the uncertainty of the monthly means by more than 10%; this reduces to 0.25 K at ~20 hPa and to 0.15 K in the free troposphere. If the frequency of sampling decreases, the sampling uncertainty comes to dominate, resulting in less stringent requirements on the uncertainty on each individual measurement. For example, for operational radiosonde sites making twice daily temperature profile measurements, there is something to be gained by reducing the uncertainty on each measurement to 0.2 K or less since this minimizes the uncertainty on the resultant monthly means, thereby allowing for more robust estimates of upper-air temperature trends. For sites sampling only once per week, or less frequently (red~~blue~~, cyan~~green~~ and dark green~~red dashed ~~curves in Figure 4), a measurement uncertainty of 0.5 K is sufficient to ensure that there is no additional increase in the random uncertainty of the resultant monthly means. Of course with such infrequent sampling the monthly means will have greater uncertainties than with more frequent sampling.

2.2 Sampling strategies, measurement uncertainty, and trend detection

The effects of individual measurement uncertainty and sampling strategy on the ability to detect upper air temperature trends has also been investigated using the NCEP CFSR reanalyses temperature profiles. Temperature trends were calculated at each of the 37 pressure levels, for each of the 87 locations using a state-of-the-art regression model (Bodeker et al., 1998).

This method was applied to each of the monthly mean time series, as generated above, based on different assumptions about the uncertainty of each of the individual temperature measurements, and the 8 different sampling frequencies (see Table 1). A Monte Carlo bootstrap approach was used to estimate the uncertainty on the derived trends. In each case, 1000 statistically identical time series were generated by randomly sampling the initial regression model residuals and adding these residuals to the sum of the regression model basis function contributions, i.e. the forced part of the signal attributable to the different basis functions included in the regression model. In this case ‘statistically identical’ refers to the 1000 time series having the same underlying trend and forced variability but different structure of unforced variability. These 1000 time series are then also passed through the regression model to obtain 1000 trend values which are used

~~Residuals from the regression model fit were then used in a Monte Carlo bootstrap resampling to create 1000 statistically identical time series, each of which was passed through the regression model~~ to create a histogram of trends. ~~Blocks of residuals are selected so as to preserve the autocorrelation structure in the original time series. This method was applied to each of the monthly mean time series, as generated above, based on different assumptions about the uncertainty of each of the individual temperature measurements, and the 12 different sampling strategies (see Figure 3 and caption of Figure 4).~~

Two examples of the effects of (1) uncertainty on individual measurements and (2) sampling frequency~~strategy~~ on the quantification of temperature trends are displayed in Figure 5. The graph shows that at this location and pressure, only sampling less frequently than once weekly, and with measurement uncertainty ≥2 K is the quality of trend detection significantly degraded. At 50 hPa and 39.95°N, 105.2°W (lower panel of Figure 5), temperature trends of ~-0.032 K/decade are statistically highly significant in that none of the 1000 Monte Carlo simulations produced positive trends, and are robust against almost all combinations of measurement uncertainty and sampling ~~strategy~~frequency. As in the previous example, it is only when the measurement uncertainty exceeds 2 K, and measurements are made only once or twice per month, is the robustness of the trend determination compromised.

Formatted: Space After: 6 pt

3 Site selection for temperature trend detection

In this section, we address the question: Which of the existing sites engaged in upper-air temperature measurements are best located to detect expected future trends in upper-air temperatures within the shortest time possible? To do so, we explore and discuss one objective method (without claiming that it is the best or only method) for selecting the optimal locations for detecting projected 21st century temperature trends at approximately 5 km and 17.5 km altitude in the shortest time possible.

To provide specific guidance based on the material presented in Section 2, we investigated the number of years it would take to detect projected trends in upper-air temperatures for specified sampling regimens (both in terms of frequency and measurement uncertainty). Figure 6 shows expected 21st century trends in upper-air temperatures obtained by averaging trends from REF-B2 simulations made by 11 chemistry-climate models as part of the SPARC CCMVal-2 activity (e.g. Young et al., 2013). REF-B2 is the so-called reference simulation and is a self consistent transient simulation from 1960 to 2100 (Eyring et al., 2010). In this simulation the surface time series of halocarbons are based on the adjusted A1 scenario from WMO (2007). The adjusted A1 halogen scenario includes the earlier phase out of hydrochlorofluorocarbons (HCFCs) that was agreed to by the Parties to the Montreal Protocol in 2007 (Eyring et al, 2010). The long-lived GHG surface concentrations are taken from the SRES (Special Report on Emission Scenarios) GHG scenario A1B (IPCC, 2000).

The number of years of measurements required to detect a trend at the 95% confidence level with a probability of 0.9 can be approximated by (Whiteman et al., 2011):

$$n^{*} = \left[\frac{3.3\sigma_N}{|\omega_0|} \sqrt{\frac{1+\phi_N}{1-\phi_N}} \right]^{2/3} \qquad (1)$$

where σ_N is the standard deviation of the unforced variability in the time series, i.e. the standard deviation of the residuals after the application of the regression model (described in Section 2.2) to remove all known sources of variability, ω_0 is the trend magnitude in K/year (see Figure 6), and ϕ_N is the auto-correlation in the residuals (Tiao et al., 1990).

This equation implies that after the calculated number of years, there is a 90% probability that a trend of the correct sign will have been detected, if we assume that detecting a trend means identifying a trend at the 95% confidence level. σ_N and ϕ_N values for the 87 analysis locations at the 37 pressure levels were calculated from the NCEP CFSR time series for ~~the~~ 120 different sampling regimens (i.e. 12 sampling strategies for 10 different measurements uncertainties ranging from 0.01 K – 10 K).

When used in equation (1) together with the projected 21st century temperature trends shown in Figure 6, examples of results for two sites are shown in Figure 7. The projected trend at 30 hPa for 85°N (top plot of Figure 7) is -0.01612 K/year and for 25°N -0.03627 K/year (bottom plot). Calculations were made for 3219 cases (87 locations and 37 pressure levels). Typically, it is only when the uncertainty on each measurement exceeds 2 K, is the ability to detect trends significantly compromised, consistent with the findings presented in Section 2.2.

When comparing the results for the two sites displayed in Figure 7, one in the tropics and one at high latitudes, it is clear that the uncertainty on each temperature measurement, has little impact on the time required to detect the projected trend. Similarly, it is only for sampling regimens of every 4 days, or less often~~frequently~~, that the sampling frequency affects the number of years required to detect the projected trend (see also Seidel and Free, 2006). The biggest effect on the time required to detect the projected trend stems from the natural variability (the noise) in the time series, the auto-correlation in the data and the magnitude of the expected trend. While for the site at 25°N the projected trend is expected to be detected within 30 years or less, for the site at 85°N, the projected trend will likely not be detected within 100 years.

To further synthesize the results, three pressure levels, viz. 50 hPa, 10 hPa, and 1 hPa were selected to investigate which measurement regimens, if any, allow for the detection of a temperature trend within 30 years, assuming an uncertainty on each measurement of 1 K. It is apparent from the analysis (not shown here) that in the upper stratosphere (1 hPa), it is possible to detect temperature trends in the tropics (30°S to 30°N) with almost any measurement programme - even one measurement per month would be sufficient to detect the trend within 30 years. Over the Arctic however, no measurement regimen, no matter how frequently the measurements are made, and even if the measurements are made with very small uncertainty (at 0.01 K), would detect the annual temperature trend within 30 years. In contrast to this, in the Antarctic, most measurement regimens (at 1 K uncertainty) would detect the trend within 30 years. Over the southern mid-latitudes, the trends would be detected at only one location within 30 years, whereas over northern mid-latitudes, trends may be detected at several locations.

At 10 hPa, the situation is similar to the 1 hPa level, with tropical trends being detected more easily than extra-tropical ones, but the robustness now also extends to northern mid-latitudes. The trend detection over the Antarctic is less robust. At 50 hPa, trends may be detectable at up to half of the locations within the tropics, whereas in the extra-tropical regions, no measurement regimen would lead to the detection of the expected temperature trends within 30 years.

This analysis was performed using annually averaged trends and it might well be that trends in some seasons are more likely to be detectable than in the annual mean, either because the trend is steeper in that season, or because the variability in that season is smaller, or both. For the purposes of trend detection, analyses such as those summarized in Figure 7 should be conducted for any proposed measurement site to define the required random uncertainty on the measurements, the measurement regimen, and the time it is likely to take to detect the expected trend in temperature. Sites should then be selected based on the magnitude of the expected trend, the natural variability, and the auto-correlation in the data as detailed in equation (1).

To identify such preferable sites where temperature trends could be identified sooner than elsewhere, an analysis based on Microwave Sounding Unit (MSU) and Advanced Microwave Sounding Unit (AMSU) temperature measurements, available from remote sensing systems (Mears and Wentz, 2008), was carried out. Figure 8 shows the results of this analysis for the merged MSU channel 2 and AMSU channel 5 temperatures. These are indicative of the middle troposphere with the weighting function peaking at ~5 km altitude. The standard deviation of the residuals from the application of the regression model to monthly mean temperatures (panel (a) of Figure 8) and the first order auto-correlation coefficient (panel (b) of Figure 8) are two of the quantities needed to calculate the number of years required to detect a prescribed temperature trend as detailed in equation (1).

The month-to-month variability in the data minimizes in the tropics and maximizes over high latitudes, particularly over the Canadian Arctic. This would suggest that the tropics would be ideally suited to long-term temperature trend detection in the middle troposphere. However, as shown in panel (b) of Figure 8, the auto-correlation in the temperature time series also maximizes in the tropics. When the standard deviation on the monthly means and the calculated first order auto-correlation are used together with a prescribed trend of 0.5 K/decade in equation (1), the results shown in panel (c) of Figure 8 are obtained. Large regions of the tropics and sub-tropics have temperature time series that would be amenable to detection of mid-troposphere temperature trends of 0.5 K/decade within ~10 years. However, as seen in Figure 6, temperature trends at 5 km are not everywhere 0.5 K/decade. If we use the expected temperature trends at 5 km from Figure 6 in equation (1) then the results displayed in panel (d) of Figure 8 are obtained. This is the optimal figure to use for deciding where to locate measurement sites for detecting trends in mid-tropospheric temperatures.

One objective strategy (but certainly not the only strategy) is to select an existing site from the relevant global observation networks closest to the minimum value shown in panel (d) of Figure 8. For the purposes of this study, only sites from ~~the~~ GRUAN and GUAN (GCOS Upper-Air

Network; GCOS-73, 2002, http://www.wmo.int/pages/prog/gcos/documents/GUAN_map_2014.pdf) ~~networks~~ were considered for this selection. The site closest to the minimum value was found to be the GUAN site at Guam. The next site with the next shortest time to detect expected mid-tropospheric temperature trends, which is at least 6000 km from Guam (since it is not necessary to have sites very close together), is the GUAN station on Tromelin Island. We then continue to look through the list of existing measurement sites, ordered by the number of years required to detect trends, selecting sites that are at least 6000 km away from the already selected sites. The resultant distribution of sites is shown in panel (d) of Figure 8 and also listed in Table 2~~1~~. Such a selection of sites would provide good global coverage with a preference for sites in regions where the time to detect expected trends in mid-troposphere temperatures is minimal.

Figure 9 shows the results of a similar analysis, but now using merged MSU channel 4 and AMSU channel 9 temperatures indicative of the lower stratosphere (weighting functions peaking at ~17.5 km). The approach described above is used for selecting the optimal measurement sites, now resulting in different sites including one site in the Arctic (Barrow), as well as one Antarctic site (Amundsen-Scott, South Pole), with less emphasis on tropical sites. The sites shown in Figure 9 are also listed in Table 3~~2~~.

Note that this is just one possible strategy for selecting sites for detecting expected long-term trends in mid-troposphere and lower stratosphere temperatures. Clearly, different strategies would result in a different list of ideal sites and strategies need to be tailored to accommodate other factors such as cost, accessibility, measurement capability etc.. The purpose of this exercise is to show that generating fields, such as those shown in panel (d) of Figure 8 and panel (d) of Figure 9 provide one objective method of selecting the optimal location of sites for detecting long-term temperature trends in different regions of the atmosphere within the shortest possible time.

4 Site selection criteria for the detection of ozone trends

As it was done for upper-air temperature trends, we demonstrate a similar technique for objectively selecting optimal locations for detecting expected future trends in total column ozone. Expected ozone trends for different periods (see below) were obtained from 21 CCM simulations of total column ozone changes over the 21st century under the CCMVal2 REF-B2 scenario. Except for one model (CMAM), sea-surface temperatures and sea-ice concentrations are prescribed from coupled ocean model simulations, either from simulations with the ocean coupled to the underlying general circulation model, or from coupled ocean-atmosphere models used in the IPCC 4th assessment report under the same GHG scenario. At each latitude and longitude, the median ozone trend value from the 21 CCM simulations available was extracted and used as the indicative total column ozone trend.

Trends in total column ozone, unlike those in temperature, are not expected to be linear over the coming century over many regions of the globe. It is therefore less relevant to consider the time to detect expected 21st century trends in total column ozone as an indicator of where total column ozone observing sites should be located. For example, if in some region of the globe, such as the tropics, where ozone is expected to increase until the middle of the 21st century and then to decrease thereafter, the time to detect the expected trend until 2100 may be significantly longer than the time to detect the trend until 2050. The approach taken is therefore to first conduct an analysis, similar to that for temperature, but considering expected trends in ozone from 2010 to 2020 and identifying which set of locations would be best suited for detecting those expected trends. The trend period is then extended by one year to consider trends from 2010 to 2021, and a second set of sites is identified. This is repeated until 2010-2050, thereby creating 31 sets of optimal sites for detecting ozone trends. An example of the outcomes of this analysis for the 2010-2050 analysis is shown in panel (a) of Figure 10.

Monthly mean total column ozone data obtained from the Bodeker Scientific total column ozone database[1] spanning the period November 1978 to August 2012 were then analysed for their standard deviation and first order auto-correlation, two of the quantities needed to calculate the number of years required to detect a prescribed total column ozone trend using equation (1). The model used to derive the residuals was similar to that used in Bodeker et al. (2001), which includes terms accounting for the mean annual cycle, the linear trend, the quasi-biennial oscillation (QBO), the El Niño Southern Oscillation (ENSO), the solar cycle, and the El Chichón and Mt Pinatubo volcanic eruptions. The resultant standard deviation of the monthly means and the first order auto-correlation coefficient are displayed in panels (b) and (c) of Figure 10. Month-to-month variability in the data minimizes in the tropics and maximizes over high latitudes, particularly over Siberia. This would suggest that the tropics would be ideally suited to long-term total column ozone trend detection. However, as shown in panel (c) of

---

[1] http://www.bodekerscientific.com/data/total-column-ozone

Figure 10, the auto-correlation in the total column ozone also maximizes in the tropics. The auto-correlation in ozone and other atmospheric trace gases is the result of the time and spatial scale of weather patterns as well as possible long-term forcing mechanisms (Tiao et al., 1990). Such auto-correlations have the effect of reducing the amount of information that would be available from the same number of independent measurements and generally increases the size of the measurement uncertainty.

When the standard deviation of the regression model residuals and the first order auto-correlation are used together with the projected trends in total column ozone, the results shown in panel (d) of Figure 10 are obtained. As can be seen clearly in Figure 10, the magnitude of the auto-correlation in the total column ozone has a strong impact on the estimate of the number of years of measurements required to detect a trend based on equation (1). This agrees with previous work done by Tiao et al. (1990); they showed that a large positive auto-correlation in the monthly mean data (e.g. total column ozone) will have a severe effect on the uncertainty of trend estimates and hence substantially increase the length of data records required to achieve the same degree of low uncertainty, compared to a situation where the data would be independent over time.

The distribution for the number of years required to detect the expected trends in total column ozone shown in panel (d) of Figure 10 overall agrees with results from an earlier study by Weatherhead et al. (2000). Both studies show e.g. that the areas of high detectability are in the Southern Hemisphere around New Zealand/eastern Australia and southern South America and that locations close to the equator require the longest time for trend detection. The study by Weatherhead et al. (2000) also shows that the detection of expected trends in most parts of the Northern Hemisphere will take longer than in the Southern Hemisphere (their Plate 5) which is also evident in our Figure 10 but not as pronounced. It should be noted that Weatherhead et al. (2000) use a similar technique to calculate the expected number of years for the ozone trend detection but a different model (Goddard Space Flight Centre 2-dimensional chemical model) to predict the trends and a different ozone data set (Nimbus 7 TOMS data).

In analogy to the temperature trends, one objective strategy (but certainly not the only strategy) to use panel (d) of Figure 10 to determine optimal locations for measurement sites is to select an existing site closest to the minimum value shown in panel (d). In this case only sites from WOUDC, SHADOZ and NDACC networks were considered.

The site closest to the minimum value was found to be the historical WOUDC site at Ushuaia (II). We now look for the next site with the shortest time to detect expected total column ozone trends that is at least 6000 km from Ushuaia. This is found to be Hobart. We then continue to look through the list of existing measurement sites, ordered by the number of years required to detect trends, selecting sites that are at least 6000 km away from sites already selected. The resultant distribution of sites is shown in panel (d) of Figure 10 and the 9 sites are listed in Table 4~~3~~. We expect that such a selection of sites would provide sufficient global coverage for trend detection with a preference for sites in regions where the time to detect expected total column ozone trends is as short as possible.

To provide a perspective on how these 9 proposed sites, selected for trend detection in total column ozone from 2010-1050, compare to the other 30 sets of sites, selected for each of the other trend periods (2010 – 2020 to 2010 – 2049), we have collated a list of all the sites selected for the 31 trend periods. Within this analysis, a total of 66 sites were selected with 23 of these sites being located in the Southern Hemisphere and Hobart being the most frequently selected site (23 times). We then ranked the list of sites accordingly to how frequently they were selected and in Table 5, we show the 5 most frequently selected sites for the Northern and Southern Hemisphere each. Three of the sites listed in Table 4~~3~~ (Moscow, Papeete and Hobart) are also on the list of the 10 most often selected sites ~~within the 31 sets of optimal sites discussed above and~~ summarized in Table 5~~4~~.~~ In this analysis, 66 sites were selected in total with 23 of these sites being located in the Southern Hemisphere.~~

5 Discussion and s~~S~~ummary

For a number of globally distributed locations around the globe (87 in total, see Figure 1), the dependence of the uncertainty on monthly mean temperatures on individual measurement uncertainty, sampling frequency, season, and pressure was assessed using NCEP CFSR reanalyses. Our results show that only for individual temperature measurement uncertainties greater than 0.2 K, does the measurement uncertainty start to contribute significantly to the uncertainty on the monthly mean. In practical terms, this means that for operational radiosonde stations which carry out temperature profile measurements twice daily, it is worthwhile to work to reduce the uncertainty on each measurement to ≤0.2 K since this minimizes the uncertainty of the resultant monthly means, which should lead to more robust estimates of upper-air temperature trends. However, there is little to be gained by reducing the measurement uncertainty to much less than 0.2 K. This conclusion~~s~~ supports the recommendations made by GRUAN.

With a reduction in sampling frequency, the sampling uncertainty starts to dominate, such that less rigorous criteria regarding the uncertainty requirements for each individual measurement are acceptable. For example, for sites where sampling is only weekly or less frequently, measurement uncertainties of 0.5 K are sufficient to ensure that there is no additional increase in the random uncertainty on the resultant monthly means by more than 10% above what would be achieved when sampling with 0.01 K uncertainty. This concurs with the findings of Seidel and Free (2006) who found that if the individual measurement uncertainty is at least 0.5 K, monthly means are accurate to within ~0.05 K, and standard deviations are accurate to within 10%.

Seidel and Free (2006) also found that increasing the uncertainty on temperature measurements has minor effects on the accuracy of the monthly means and standard deviations and is not an important factor in determining multi-decadal trends. The latter is consistent with our finding that only when the measurement uncertainty exceeds 2 K, and measurements are made just once or twice a month or less frequently, the quality of the trend determination is compromised. We find that for a wide range of uncertainties and sampling frequencies, these aspects of a monitoring programme have little impact on the number of years required to detect the projected trend which depends more on the natural variability and auto-correlation in the time series. As a result, at some locations such as in the tropics, the projected temperature trend is expected to be detected within 30 years or less, while for locations in the northern high latitudes, the projected trend will likely not be detected even within 100 years.

Given these constraints, we have endeavoured to find an objective selection process for the most suitable measurement sites where temperature trends in the mid-troposphere and lower stratosphere could be identified sooner than elsewhere. Note that this is just one example of an objective site selection strategy and that the resulting maps depend on the criteria used.

A similar technique was applied to find an optimal distribution of measurements sites to detect ozone trends in the shortest time possible. Since trends in total column ozone are not expected to be linear over the coming century over many regions of the globe, it is less pertinent to consider the time to detect expected 21st century trends in total column ozone as an indicator of where total column ozone observing sites should be located. We have therefore investigated different time periods from 2010 – 2020 up to 2010 – 2050 to generate 31 sets of optimal sites for ozone trend detection and the 10 measurement sites appearing most often within these 31 sets are listed in Table 5~~4~~.

The objective method to determine optimal measurement sites presented here is based on an estimation of the geographical distribution of the minimum time to detect the projected trend. To estimate this quantity, we estimated the unforced variance in the signal and the degree of auto-correlation from historical data. The underlying assumption of our analysis is that climate change would not significantly affect these parameters. The minimum time to detect future trends also depends on the magnitude of the projected trends, which was estimated from chemistry climate simulations. These models are the best tools we currently have to estimate future trends and have been shown to reasonably capture past trends in Southern Hemisphere stratospheric temperatures (Young et al., 2013). It should be noted though that there is a large uncertainly in current estimates of past stratospheric temperature trends (Thompson et al., 2012) which limits our ability to validate past temperature trends simulated by these models.

While our proposed method for future site selection only depends on the geographical distribution of the minimum time to detect the projected trend and the geographical distance between measurement sites for two selected pressure levels, other factors may be considered as well. For example to be able to detect projected changes in the width of the tropics, it would be beneficial to select a station close to the boundary of the tropics (e.g. GRUAN-RP-4, 2014). Proximity to the source region of El Nino might be another consideration given that trends in this region will likely have global impacts. Finally, one might be interested in detecting changes in tropopause height, a factor not considered in our study.

Studies such as the one presented here provide a sound scientific basis for decision making with regard to new and existing measurements sites and can help reduce costs and concentrate efforts where they are the most needed and most effective.

Acknowledgements

This work was funded in part through the Support to Science Element ESA (European Space Agency) SPARC (Stratosphere-troposphere Processes And their Role in Climate) Initiative (ESA Contract No. 4000105291/12/I-NB) and we take this opportunity to acknowledge ESA for this support. Funding from the NOAA GCOS office, through the Meteorological Service of

New Zealand Ltd, was also essential for completing this work. We would also like to thank Chi-Fan Shih from the National Center for Atmospheric Research in Boulder, CO, for providing the NCEP CFSR reanalyses. We acknowledge the Chemistry-Climate Model Validation (CCMVal) activity for providing model results from the REF-B2 reference simulations. We also thank the University of Alabama (UAH) and Remote Sensing Systems (RSS) for access to the MSU and AMSU temperature data. The authors wish to thank Anna Mikalsen for helpful comments on the manuscript.

References

Adams, C., Bourassa, A.E., Sofieva, V., Froidevaux, L., McLinden, C.A., Hubert, D., Lambert, J.-C., Sioris, C.E., and Degenstein, D.A.: Assessment of Odin-OSIRIS ozone measurements from 2001 to the present using MLS, GOMOS, and ozonesondes, Atmos. Meas. Tech., 7, 49–64, 2014.

Balis, D., Kroon, M., Koukouli, M.E., Brinksma, E.J., Labow, G., Veefkind, J.P., and McPeters, R.D.: Validation of Ozone Monitoring Instrument total ozone column measurements using Brewer and Dobson spectrophotometer groundbased observations, J. Geophys. Res., 112, D24S46, doi:10.1029/2007JD008796, 2007.

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.

Bodeker, G.E., Scott, J.C., Kreher, K., and McKenzie, R.L.: Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978-1998, J. Geophys. Res., 106(D19), 23029-23042, 2001.

Eyring, V., Cionni, I., Bodeker, G.E., Charlton-Perez, A.J., Kinnison, D.E., Scinocca, J.F., Waugh, D.W., Akiyoshi, H., Bekki, S., Chipperfield, M.P., et al.: Multi-model assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models, Atmos. Chem. Phys., 10, 9451–9472, 2010.

Free, M., Durre, I., Aguilar, E., Seidel, D., Peterson, T.C., Eskridge, R.E., Luers, J.K., Parker, D., Gordon, M., Lanzante, J., Klein, S., Christy, J., Schroeder, S., Soden, B., Mcmillan, L.M., and Weatherhead, E.: Creating climate reference datasets; CARDS Workshop on Adjusting

Radiosonde Temperature Data for Climate Monitoring, Bull. Amer. Meteor. Soc., 83, 891-899, 2002.

GCOS-73: Guide to the GCOS surface and upper-air networks: GSN and GUAN (Version 1.1), WMO/TD No. 1106, 2002.

GRUAN-RP-4: Outcomes of the GRUAN Network Expansion Workshop, available from www.gruan.org, 2014.

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, 2010.

IPCC (Intergovernmental Panel on Climate Change): Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change, 599 pp., Cambridge University Press, Cambridge, UK, 2000.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. Bull. Amer. Meteor. Soc., 82, 247–267, 2001.

Mears, C.A. and Wentz, F.J.: Construction of the Remote Sensing Systems V3.2 Atmospheric Temperature Records from the MSU and AMSU Microwave Sounders, J. Atmos. Oceanic Technol., 26, 1040-1056, 2008.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al.: The NCEP Climate Forecast System Reanalysis, Bull. Amer. Meteor. Soc., 91, 1015-1057, 2010.

Seidel, D.J. and Free, M.: Measurement Requirements for Climate Monitoring of Upper-Air Temperature Derived from Reanalysis Data, J. Climate, 19, 854-871, 2006.

Thompson, D.W.J., Seidel, D.J., Randel, W.J., Zou, C., Butler, A.H., Mears, C., Osso, A., Long, C., and Lin, R.: The mystery of recent stratospheric temperature trends, Nature, 491, 692-697, 2012.

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., 95D12, 20507-20517, 1990.

Formatted: Font: (Default) Times New Roman, Not

Formatted: Font: (Default) Times New Roman, Not

Formatted: Font: (Default) Times New Roman

Formatted: Font: (Default) Times New Roman, Not

Formatted: Font: (Default) Times New Roman, Not

Tobin, D.C., Revercomb, H.E., Knuteson, R.O., Lesht, B.M., Strow, L.L., Hannon, S.E., Feltz, W.F., Moy, L.A., Fetzer, E.J., and Cress, T.S.: Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor retrieval validation, J. Geophys. Res., 111, D09S14, doi:10.1029/2005JD006103, 2006.

Tummon, F., Hassler, B., Harris, N.R.P., Staehelin, J., Steinbrecht, W., Anderson, J., Bodeker, G.E., Bourassa, A., Davis, S.M., Degenstein, D., et al.: Intercomparison of vertically resolved merged satellite ozone data sets: interannual variability and long-term trends, Atmos. Chem. Phys. Discuss., 14, 25687–25745, 2014.

Wang, J.S., Seidel, D.J., and Free, M.: How well do we know recent climate trends at the tropical tropopause? J. Geophys. Res., 117, doi:10.1029/2012JD017444, 2012.

Weatherhead, E.C., Reinsel, G.C., Tiao, G.C., Jackman, C.H., Bishop, L., Hollandsworth Frith, S.M., DeLuisi, J., Keller, T., Oltmans, S.J., et al.: Detecting the recovery of total column ozone, J. Geophys. Res., 105D17, 22201-22210, 2000.

Whiteman, D.N., Vermeesch, K.C., Oman, L.D., and Weatherhead, E.C.: The relative importance of random error and observation frequency in detecting trends in upper tropospheric water vapor, J. Geophys. Res., 116, D21118, doi:21110.21029/22011JD016610, 2011.

World Meteorological Organization (WMO)/United Nations Environment Programme (UNEP), Scientific Assessment of Ozone Depletion: 2006, Global Ozone Research and Monitoring Project, Report No. 50, 572 pp., Geneva, Switzerland, 2007.

Young, P.J., Butler, A.H., Calvo, N., Haimberger, L., Kushner, P.J., Marsh, D.R., Randel, W.J., and Rosenlof, K.H.: Agreement in late twentieth century Southern Hemisphere stratospheric temperature trends in observations and CCMVal-2, CMIP3, and CMIP5 models, Journal of Geophysical Research: Atmospheres, 118, 605-613, 2013.

Table 1. List of applied sampling frequencies

Sampling frequency
Every 6 hours
Every 12 hours at noon/midnight
Every 24 hours at noon
Every 2 days at noon
Every 4 days at noon
Every week at noon
Every 2 weeks at noon
Once a month at noon

Table 2~~1~~. Proposed measurement sites for the detection of 21st century temperature trends at the middle troposphere.

Site Name	Latitude	Longitude	Observation network
Annette Island	55.0°N	131.3°E	GUAN
La Coruna	43.3°N	8.5°W	GUAN, WOUDC
Kashi	39.3°N	75.6°E	GUAN
Kingston	17.6°N	76.5°W	GUAN
Guam	13.3°N	144.5°E	GUAN
Tromelin Island	15.5°S	54.3°E	GUAN
St. Helena	15.6°S	5.4°W	GUAN
Rarotonga	21.1°S	159.5°W	GUAN
Puerto Montt	41.3°S	73.1°W	GUAN
Dumont d’Urville	66.7°S	140.0°E	GUAN, NDACC

Table 3~~2~~. Proposed measurement sites for the detection of 21st century temperature trends at the lower stratosphere.

Site Name	Latitude	Longitude	Observation network
Barrow	71.3°N	156.6°W	GRUAN, ARM, GAW
Key West	24.3°N	81.5°W	GUAN
Asswan	23.6°N	32.5°E	GUAN
Chichijima	27.1°N	142.1°E	GUAN
Rapa	27.4°S	144.2°W	GUAN
Perth Airport	31.6°S	115.6°E	GUAN
Cape Town	33.6°S	18.4°E	GUAN
Ezeiza Aero	34.5°S	58.3°W	GUAN
South Pole	90.0°S	0.0	GUAN, NDACC, GAW

Table 4~~3~~. Proposed measurement sites for the detection of ozone trends from 2010-2050.

Site Name	Latitude	Longitude	Observation network
Cold Lake	54.8°N	110.1°W	Historical WOUDC
Moscow	55.7°N	37.5°E	Historical WOUDC
Vladivostok	43.1°N	131.9°E	Historical WOUDC
Barbados	13.1°N	59.5°W	Historical WOUDC
Kodaikanal	10.2°N	77.4°E	Historical WOUDC
Papeete	18.0°S	149.0°W	Historical WOUDC & Historical SHADOZ
Springbok	29.7°S	17.9°E	Historical WOUDC
Hobart	42.9°S	147.5°E	Historical WOUDC
Ushuaia II	54.9°S	68.4°W	Historical WOUDC

Table 5~~4~~. The five most frequently selected Northern Hemisphere sites in the 31 sets of optimal sites followed by the five most frequently selected Southern Hemisphere sites.

Site Name	Latitude	Longitude	Observation network
Kyiv-Goloseyev	50.3°N	30.5°E	Current WOUDC
Sapporo	43.0°N	141.3°E	Current WOUDC, GUAN
Moscow	55.7°N	37.5°E	Historical WOUDC
Edmonton/Stony Pl.	53.5°N	114.1°W	Historical WOUDC
Coolidge Field	17.3°N	61.8°W	Historical WOUDC
Hobart	42.9°S	147.5°E	Historical WOUDC
Papeete	18.0°S	149.0°W	Historical WOUDC & Historical SHADOZ
La Reunion Island	21.1°S	55.5°E	Historical WOUDC
Ushuaia	54.9°S	68.3°W	Current WOUDC, GAW
Ascension Island	8.0°S	14.5°W	Historical WOUDC & Historical SHADOZ

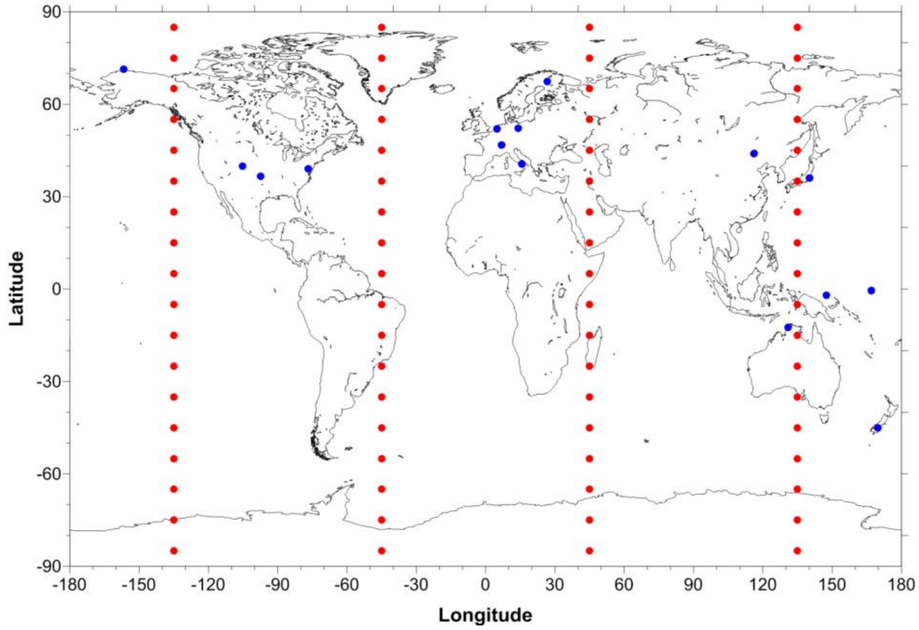

Figure 1. Map of the 87 locations used for the data analysis; 15 of the locations are the initial GRUAN sites (blue dots) and the other 72 of the locations (red dots) are positioned in 90° longitude zones and 10° latitude zones as shown on the map.

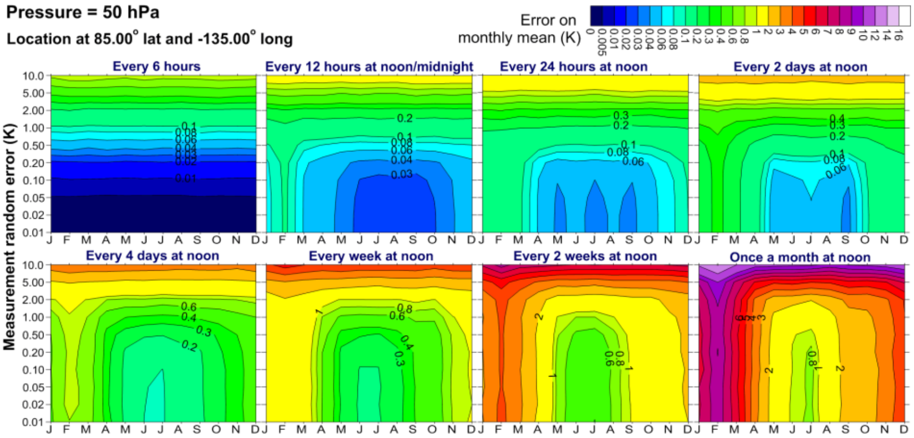

Figure 2. The uncertainty on the monthly mean temperatures at 50 hPa, 85°N, 135°W, for a range of sampling frequencies~~strategies~~, as a function of the random uncertainty of each instantaneous measurement.

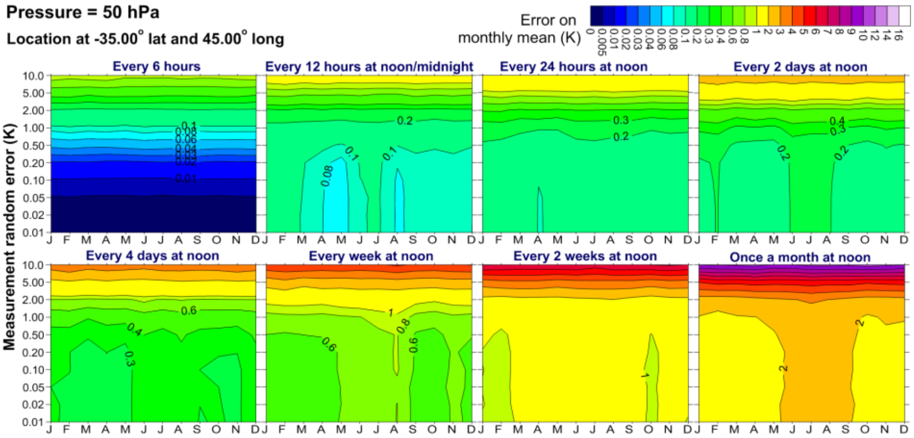

Figure 3. The uncertainty on the monthly mean temperatures at 50 hPa, 35°S, 45°E, for a range of sampling frequencies~~strategies~~, as a function of the random uncertainty of each instantaneous measurement.

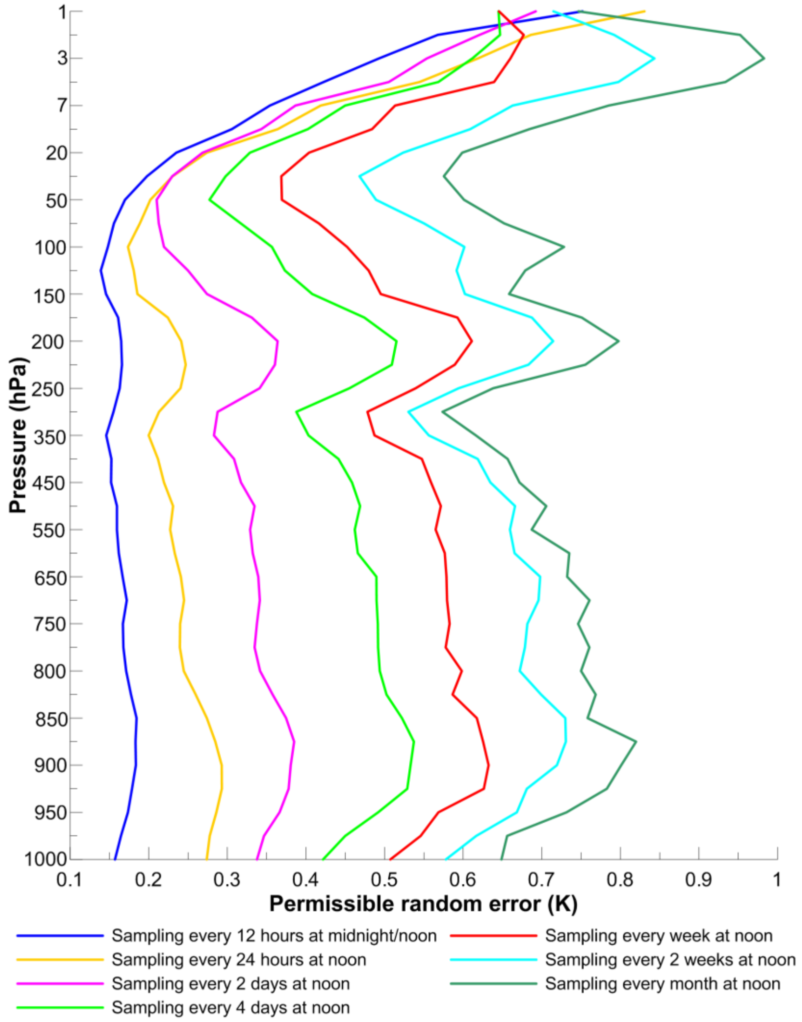

Figure 4. The permissible uncertainty on temperature measurements for~~across~~ a range of sampling frequencies~~strategies~~ required to avoid more than 10% increase in the uncertainty on the monthly means compared to the monthly mean uncertainty that would result from sampling with 0.01 K uncertainty. Results from all 87 sites selected for this analysis and for all months were averaged to produce this figure.

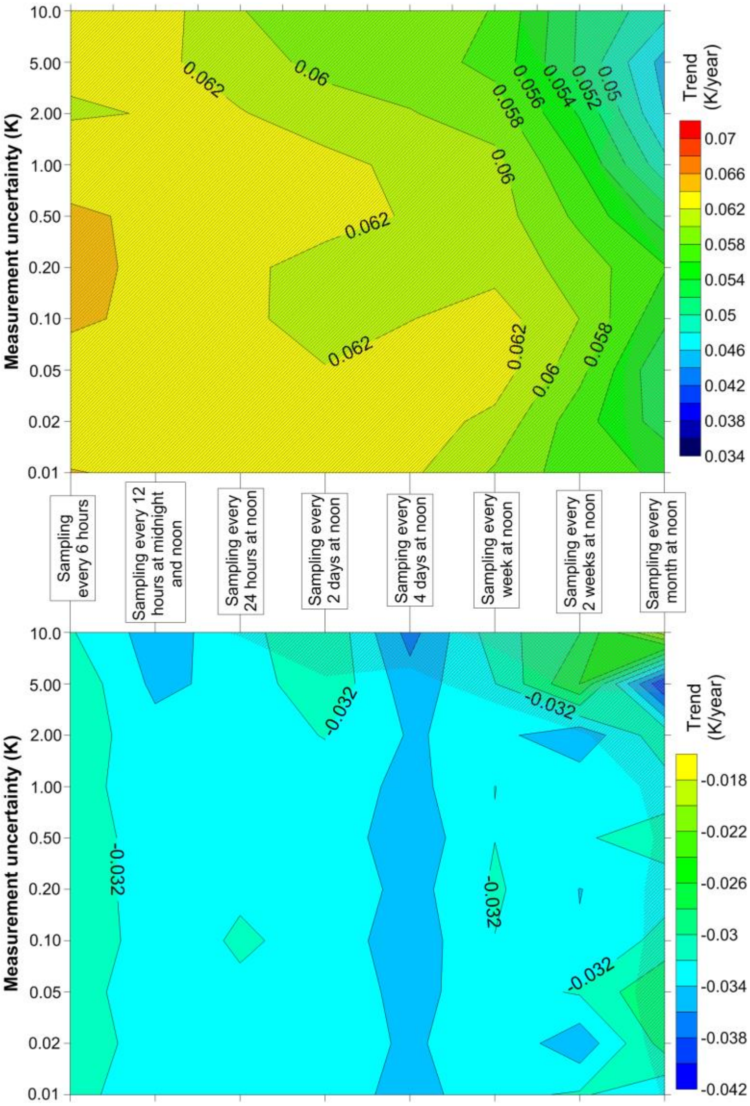

Figure 5. Upper panel: Annual mean trends at 1 hPa, 85°N, 135°W, as a function of individual measurement uncertainty used to calculate the monthly means used as input to the regression analysis, and sampling ~~regimen~~ frequency. Regions with single hatching show where trends are statistically significantly different from zero at between 1σ- and 2σ. Regions with double hatching show where the trend is not statistically significantly different from zero at 1σ. Lower panel: Same analysis as upper panel but for 39.95° N, 105.2° W at 50 hPa,.

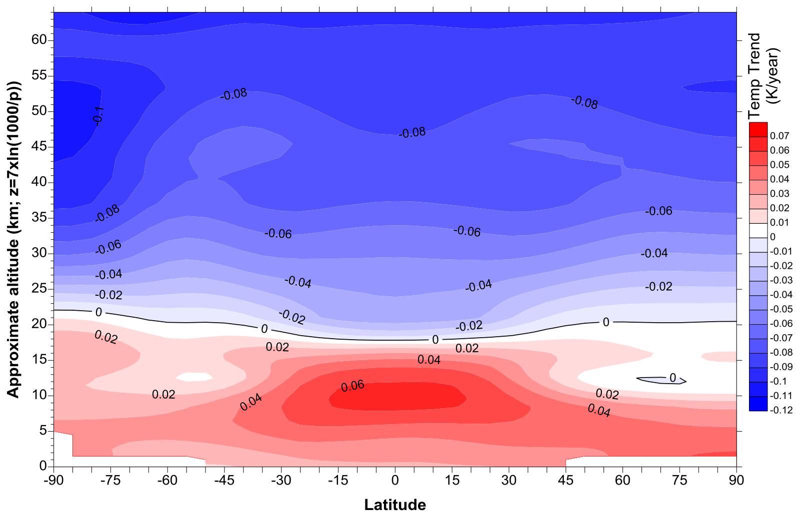

Figure 6. Projected trends in upper-air temperatures for 2000-2099 from 11 chemistry-climate models running the REF-B2 simulation from CCMVal-2.

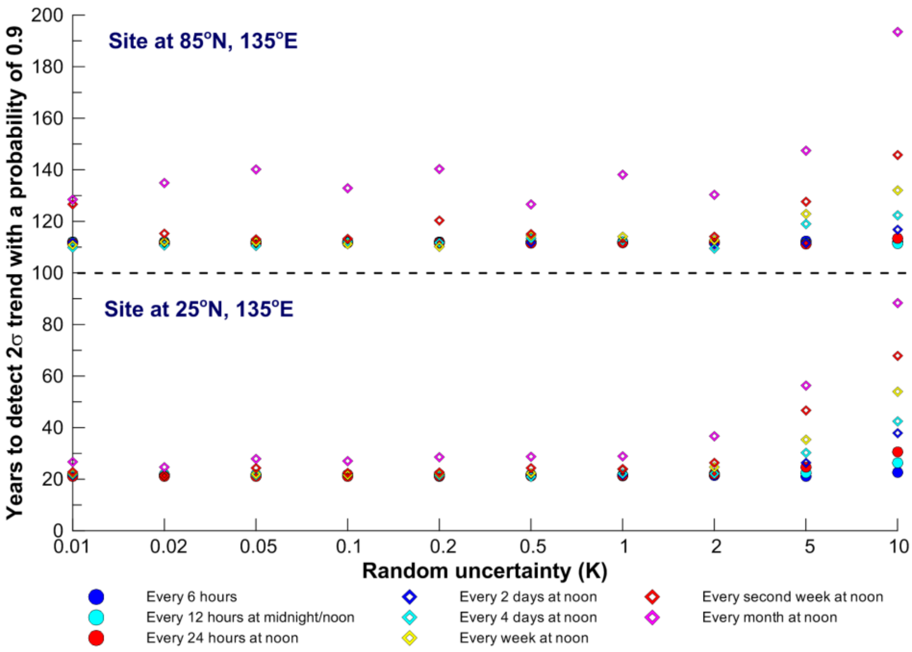

Figure 7. The time to detect projected 21st century temperature trends at 30 hPa at two sites for ~~120~~ different sampling regimens that include a variety of measurement frequencies~~y~~ and measurement uncertainties. Trends were calculated using a standard least squares regression model taking as input monthly means, calculated from individual measurements at the stated frequency and measurement uncertainty (indicated by the coloured circles and diamonds).

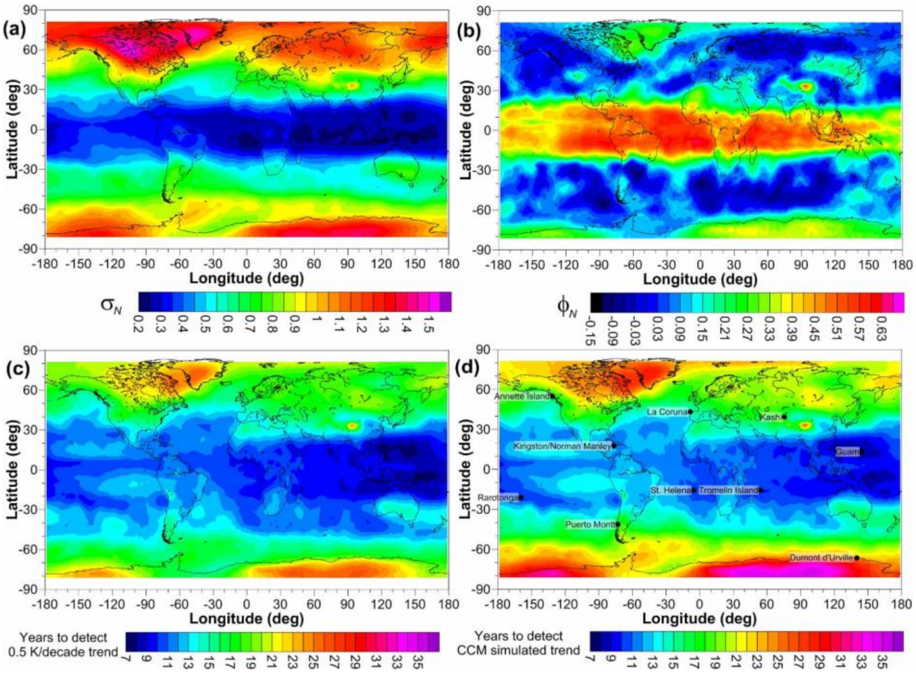

Figure 8. Analyses of merged MSU channel 2 and AMSU channel 5 temperature data – 1978 to 2013. (a) standard deviation of regression model residuals, (b) the first order auto-correlation coefficient of the residuals (c) the number of years required to detect a trend of 0.5 K/decade, and (d) the number of years required to detect the trend at 5 km altitude (close to where the MSU channel 2 and AMSU channel 5 weighting functions peak) as shown in Figure 6. Only existing GUAN and GRUAN stations have been used as basis to select the optimal sites for early trend detection shown here.

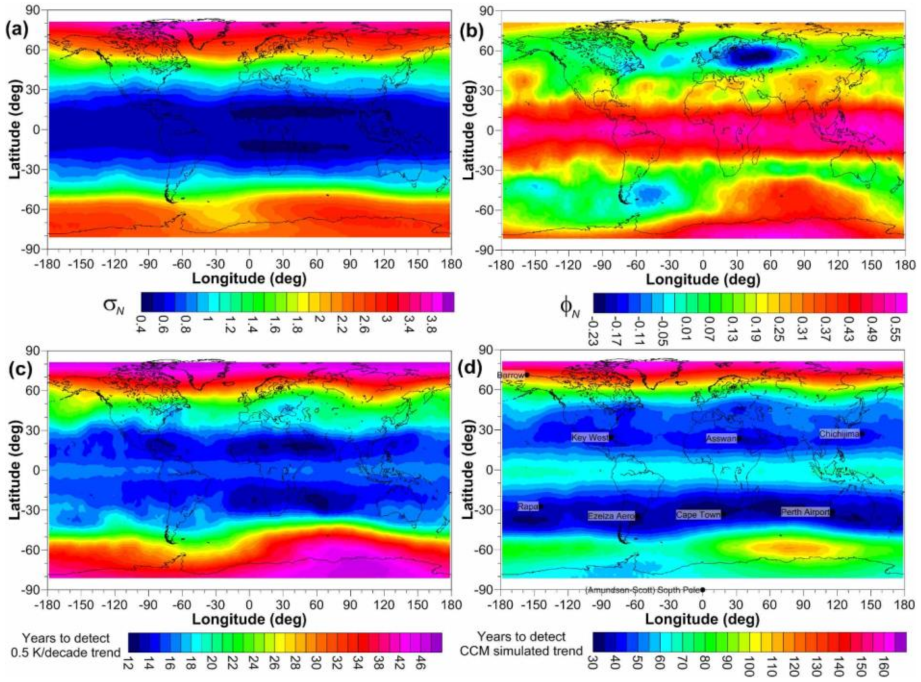

Figure 9. Analyses of merged MSU channel 4 and AMSU channel 9 temperature data - 1978 to 2013. (a) standard deviation of regression model residuals, (b) the first order auto-correlation coefficient of the residuals, (c) the number of years required to detect a trend of 0.5 K/decade, and (d) the number of years required to detect the trend at 17.5 km altitude (close to where the MSU channel 4 and AMSU channel 9 weighting functions peak) as shown in Figure 6. Only existing GUAN and GRUAN stations have been used as basis to select the optimal sites for early trend detection shown here.

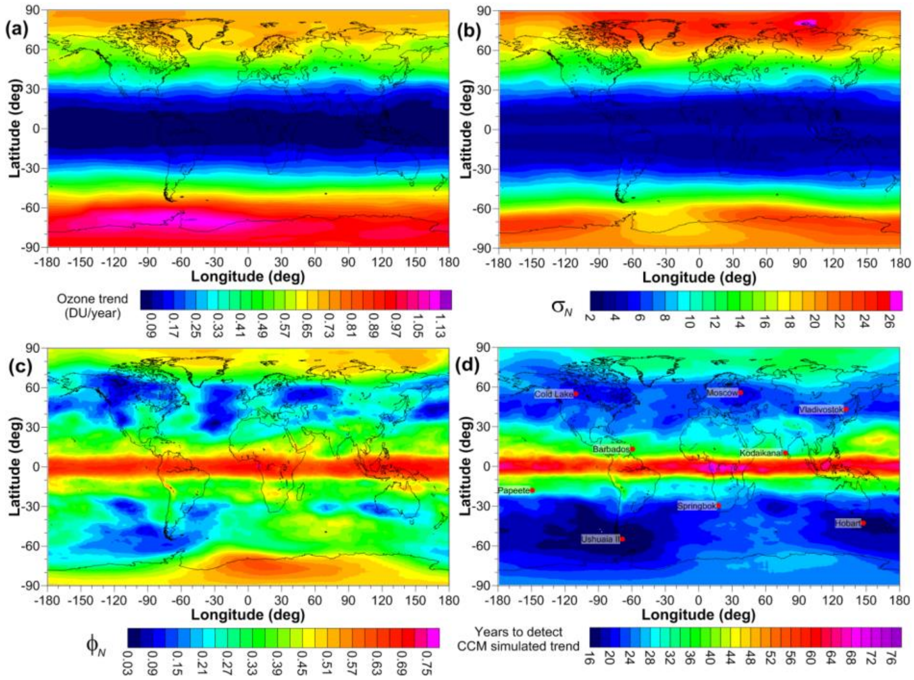

Figure 10. (a) Total column ozone trends in DU/year obtained from median values of trends calculated from 21 CCM projections of ozone over the period 2010 to 2050, (b) the standard deviation in regression model residuals in monthly mean total column ozone calculated from the Bodeker Scientific total column ozone database, (c) the first order auto-correlation coefficient of the residuals, (d) the number of years required to detect the expected total column ozone trends displayed in panel (a). Also shown in panel (d) are selected locations which are at least 6000 km apart but sample regions of short periods to detect expected trends.