GEOPHYSICAL RESEARCH LETTERS, VOL. 36, L19806, doi:10.1029/2009GL040044, 2009

What are the demands on an observational program to detect trends in upper tropospheric water vapor anticipated in the 21st century?

R. Boers[1] and E. van Meijgaard[1]

Received 19 July 2009; revised 20 August 2009; accepted 24 August 2009; published 10 October 2009.

[1] An ensemble of runs made by a regional climate model was used to estimate the expected change in humidity in the atmosphere between 1950 and 2100. Output from the model runs at a midlatitude European site is used as a proxy for a perfect climate record of specific humidity at the 300 hPa level from which trends were derived. The trend analysis demonstrates that it takes 30 years for a trend to show up in the perfect climate record. The model output was degraded to simulate observations made with the best radiosonde equipment for recording water vapor. Results indicate that it would be necessary to conduct observations for at least five decades until a trend expected by current models could be confirmed by observations to within 20%.

Citation: Boers, R., and E. van Meijgaard (2009), What are the demands on an observational program to detect trends in upper tropospheric water vapor anticipated in the 21st century?, *Geophys. Res. Lett.*, *36*, L19806, doi:10.1029/2009GL040044.

1. Introduction

[2] As atmospheric temperatures increase due to the continued anthropogenic injection of carbon dioxide, it is expected that water vapor concentrations will increase as well, inducing an enhancement of the greenhouse effect. All climate models predict this enhancement, but diversity in sensitivity to changes in radiative perturbations is present amongst the models [*Held and Soden*, 2000].

[3] Upper tropospheric water vapor, i.e. water vapor between the levels of 250–400 hPa plays a special role in our climate. Even though the amount of water vapor at those levels is small when compared to the boundary layer, the sensitivity of the outgoing radiation to water vapor is dominated by upper-tropospheric water vapor [*Soden and Bretherton*, 1993]. Consequently, it is of great importance to accurately record the upper tropospheric water vapor.

[4] Traditional radiosonde equipment to record the thermodynamic state is not suited to the task of constructing climate records. Atmospheric water vapor at high altitudes can only be detected by specialized systems, and this problem has spawned a research effort into designing, improving and optimizing such systems [*Vömel et al.*, 2007]. Even so, it is not clear whether the equipment currently available meets the expectations of the climate community because these expectations are ill-defined. In the design for the Global Climate Observing Systems (GCOS) Upper Atmosphere Network (GRUAN) uncertainty ranges for the Essential Climate Variables (ECV's) are specified [*Global Climate Observing System* (*GCOS*), 2007]. But the uncertainty requirements for ECV's are uncertain themselves because they depend on the variability of the climate record to be studied. An ECV with a large trend will accept a detection system with a larger uncertainty than an ECV with a small trend. Therefore, only the ECV to be measured in the future will tell us whether the detection system will have met its purpose.

[5] Here the demands on an observational program for measuring upper tropospheric water vapor are assessed in light of its expected future variability. Although the answer to this question might have been informed by time series of upper tropospheric water vapor in the past, upper air water vapor data of climate quality from the last 50 years are not available. Thus, there is no alternative but to resort to a modeling study using the output from climate simulations.

[6] The parameter to be studied is the water vapor specific humidity at the 300 hPa level. Our procedure is to replace a perfect climate record (i.e., an unbiased time series with perfect accuracy) by output from a transient climate simulation using a regional climate model embedded in a state-of-the-art global climate model. The model output is thus a proxy for a climate record at this level and by means of statistical methods the existence and value of a trend can be established. This trend is our best estimate of what to expect in the future.

[7] A monitoring program of upper tropospheric water vapor requires a regular probing of these levels using water vapor detection systems for which the uncertainty in observable can be ascertained. Such an observation program can be simulated by subsampling the 'perfect' climate record and by imposing an uncertainty on the 'observations'. Our questions then are a) Given a realistic uncertainty in the output from our detection system: how long and with what frequency will we need to repeat these observations in order to see a trend? And b) How does the observed trend compare to the true trend (i.e., the trend calculated from the perfect climate record)?

[8] These questions are not new: *Seidel and Free* [2006] focused on temperature, which is a parameter that can be measured with higher accuracy than water vapor. They used re-analysis data to check variability, trends and subsampling issues rather than using projections of future climate. Because of our approach questions such as whether climate projections represent a realistic climate signal remain unanswered.

2. Climate Simulations and Trend Detection

2.1. Simulation of Observations and Radiosonde Launches

[9] Model output is taken from 150-year (1950–2100) transient climate model simulations made with the KNMI

[1]Royal Netherlands Meteorological Institute, De Bilt, Netherlands.

0094-8276/09/2009GL040044$05.00

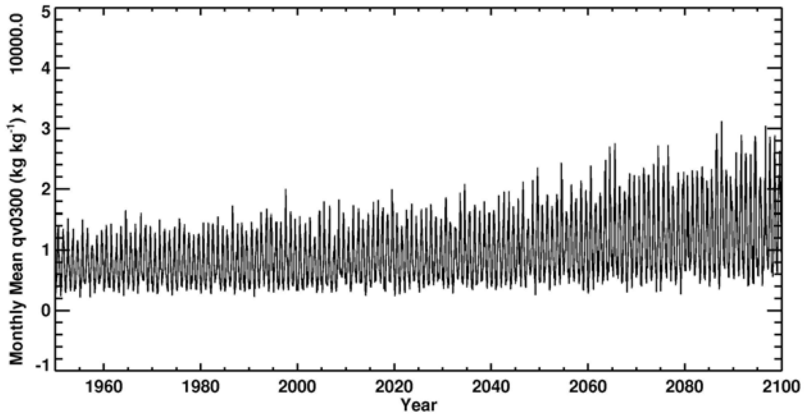

Figure 1. Time series of specific humidity at 300 hPa based on the r3 regional climate simulation.

Regional Climate Model RACMO2.1 [*van Meijgaard et al.*, 2008], partly in contribution to the 6th framework EU Integrated Project ENSEMBLES. RACMO's lateral and sea surface boundary conditions were provided by the Special Report on Emission Scenario's (SRES) -A1B emission scenario forced climate change simulations with the global climate models European Centre Hamburg Model (ECHAM5) (realizations 1, 2, and 3, hereafter referred to as r1, r2 and r3) and the Model for Interdisciplinary Research on Climate version 3.2 hires (MIROC).

[10] From here onwards the terms 'observations' and 'climate record' mean the output and processed output from the transient climate model simulations. The regional climate model simulations are gridded at 50 km resolution, so the climate record at individual stations can be simulated by means of a geographic interpolation. The station of Cabauw, Netherlands (51°58′N, 4°55′E) was chosen to perform this procedure. Many other stations could have been used. However, Cabauw is one of the initial sites for GRUAN [*GCOS*, 2007] and currently no reference radiosondes are launched there. Thus, the study provides input for a decision to start a long-term observation program involving a large expense, and thus it has a financial/economic as well as a scientific rationale. We chose the level of 300 hPa to simulate the climate observations. This level falls within the region of interest (upper troposphere, 9–9.5 km altitude). Four time series of 24 hr resolved specific humidity of 150 years duration were derived from a set of four RACMO integrations, each forced by a different climate change simulation. These four simulations represent a limited spread in expected atmospheric behavior in the 21st century.

[11] Output from the simulations consists of climate model data points recorded at midnight of each day. To obtain the climate record, these data were averaged over a month, and the monthly mean values over the entire 150 year period were subtracted from the data. To simulate a radiosonde observation it needs to be recognized that such an observation is obtained somewhere in the 24 hr period between the two consecutive data points in the original climate record, and that there needs to be a measurement error superimposed. Thus, a data point at 300 hPa was simulated as the climate record data point at midnight incremented by a random Gaussian perturbation with a standard deviation equivalent to the average difference between three consecutive 24 hr climate data points (i.e., the average of the difference between T and T −24 hrs, and T and T+24 hrs). To some (but unknown) extent this uncertainty also simulates the displacement of a radiosonde observation from the point location as the radiosonde travels upward into the atmosphere and therefore no longer represents the point location. Second, a standard Gaussian error of 10% was added to simulate the error in a single radiosonde observation of specific humidity [*Vömel et al.*, 2007]. The value of 10% is subject to debate and reflects a compromise in an unresolved discussion in the literature on the relative merits of the highly specialized systems to detect upper tropospheric water vapor [*Miloshevich et al.*, 2009].

[12] To obtain an equivalent climate record with trend the simulated radiosonde observations were averaged in the same manner as the original climate record. This degraded or subsampled climate record forms the basis for our calculations in the next sections to be compared against the original record.

2.2. Methods of Trend Detection

[13] Within the recent literature on trend detection in climate records there are two prominent approaches. One approach popular in hydrological studies uses the non-parametric Mann-Kendall test which examines the distribution of differences between sets of independent observations within a time series [*Mann*, 1945; *Kendall*, 1975]. Prerequisite is that successive observations are independent, which may not be the case.

[14] Alternatively, *Weatherhead et al.* [1998] proposed a test which encompasses both the natural variability of the signal and the possible existence of serial correlations between observations, and this procedure now enjoys wide following. We decided to choose the latter on account of its explicit inclusion of serial correlation in the testing for trends and accompanying uncertainties.

3. Results

3.1. Trend in Model Data

[15] Figure 1 shows the monthly averaged climate record of specific humidity at 300 hPa for the period 1950–2100 over Cabauw (based on the r3 simulation). The seasonal

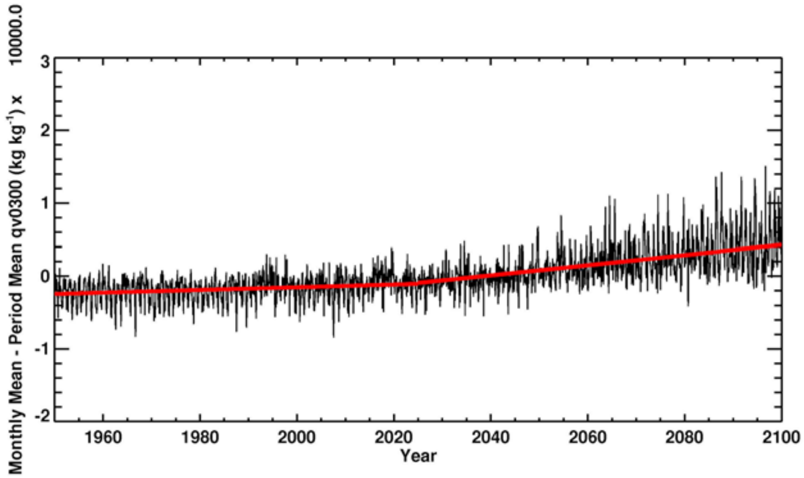

Figure 2. Time series of the specific humidity at 300 hPa with monthly mean removed. The two red trend lines are for the periods 1950–2025, and 2025–2100.

amplitude increases towards the end of the 21st century. Furthermore, the mean value of the specific humidity at 300 hPa doubles between 1950 and 2100. The increased seasonal amplitude is due to increased specific humidity in the summer months. Figure 2 shows the data from Figure 1, but with the seasonal cycle removed from the observations. Two least-squares line fits through the data are superimposed as red lines, one indicating the trend from 1950–2025, the other from 2025–2100. Although we show here only one example of output, all four simulations exhibit similar trends. This is not surprising as they are regional downscaling results from established global IPCC (TAR) – climate model runs.

[16] Using the procedure of *Weatherhead et al.* [1998] it can be shown that both trends are statistically significant at the 5% level, but that the trend almost doubles in the second period. There are isolated large positive spikes present in the record beyond 2060. Inspection of a zoomed – in picture of the years at the end of the 21st century showed that these spikes are due to summertime convective excursions into this level.

[17] A trend analysis was performed on time periods of increasing duration starting at the year 2008 (with increments of five years) to assess how long a perfect climate record would need to be so that a trend shows up in the data. Because each of the simulations (r1, r2, r3, MIROC) represents a realization of 21st century climate, this trend analysis was performed for all four realizations. A thirty year period (2008–2038) was sufficient for a statistically significant trend to show up in all four realizations. However, the spread in calculated trends was more than 100%. The results from this section are the base material against which the calculations of the degraded climate record, i.e., the radiosonde observations are compared.

3.2. Trend in Simulated Radiosonde Observations

[18] To simulate trends in radiosonde observations the trend statistics were calculated on three different radiosonde observation frequencies, namely, every four days, eight days and fifteen days. Time periods of increasing length [with five year increments] were considered with a minimum length of 25 years. The trend analysis was repeated one hundred times for different realisations of the random Gaussian noise superimposed on the radiosonde observation. A set of statistical numbers was thus generated for each of the time periods for which trends were calculated.

[19] The following statistics were further studied: a) the number of trend values within the sample of 100 trend realisations that demonstrated a statistically significant trend (expressed as a percentage of the entire sample of 100 realisations) designated as statistic A, and b) the maximum difference between the true trend and the collection of detected trends that were deemed statistically significant by the trend analysis, designated as statistic B. The former is significant because it informs us on the chance for success of detecting a trend within the time period specified given a single series of radiosonde launches over the entire period. The latter is significant because it quantifies the quality of the trend detected from the radiosonde observations.

[20] The process outlined here was repeated for all four climate simulations and the four sets of statistics were averaged to produce a final result. The salient features of both statistics are summarised in Figure 3. Figure 3 indicates that to be 100% certain that a statistically significant trend is present in the data, a time series of 45 years duration needs to be generated by sampling at least every 8th day. If a radiosonde is launched every 4 days, this period can be reduced to 40 years. For a (less expensive) twice-monthly sampling rate more than 50 years of observations will be necessary. None of these efforts will mean that the observed trend is close to the real trend, as shown at the bottom right of Figure 3. For example, even though a statistically significant trend is present in the radiosonde observations launched every 8 days for a period of 45 years, the observed trend is still 42% different from the real trend. If climate monitoring requirements specify that the observed trend be within 20% of the real trend then the sampling frequency needs to be doubled to once every four days, and the total time period of observation should be extended to 50 years.

3.3. Implications for an Observation Program

[21] The present study indicates that for a program of upper level specific humidity observations at a midlatitude site of Cabauw, at least 50 years of radiosonde ascents will

Sampling, assuming a noisy measurement	Percentage of time that a statistically significant trend was detected			Maximum difference between true trend and detected trend [in percentage difference from the true trend]		
Sampling / Time period	Every 15 days	Every 8 days	Every 4 days	Every 15 days	Every 8 days	Every 4 days
2008 – 2033	19	37	55	>100	>100	70
2008 – 2038	39	63	85	>100	96	40
2008 – 2043	62	84	94	76	55	37
2008 – 2048	76	95	100	75	50	31
2008 – 2053	92	100	100	59	42	27
2008 – 2058	99	100	100	50	32	20

Figure 3. Summary statistics of trend detection in a water vapor specific humidity signal at 300 hPa over the site of Cabauw, Netherlands. Color coding: 1) Red: The true trend is not statistically significant, but a trend in the observations is sometimes detected nonetheless 2) Pink: The true trend is statistically significant, but it is not 100% certain that a trend will be detected 3) Light green: It is 100% that a trend is detected in the climate record, but the observed trend differs from the true trend by more than 30% 4) Dark green: It is 100% certain that a trend is detected in the climate record and the observed trend is within 30% of the true trend.

be necessary with a sampling rate of every four days using a measurement device with an accuracy of 10% or better to demonstrate that a statistically significant trend is present. This is a daunting requirement for an observation program and arguments in favor of such a program would need to be weighted against diverging scientific or economic considerations. Such considerations include those favoring systems that can sample with much higher frequency but may not always have comparable accuracies (for example a Raman – lidar system).

[22] Although many remote sensing systems may not have the same accuracy as in situ systems in measuring climatically important variables, they do have the advantage of being able to sample atmospheric regions that are inaccessible by other means with unsurpassed sampling rates. A high sampling frequency is a powerful means of reducing the uncertainty of a mean observation. In statistical terms: The uncertainty of the mean is inversely proportional to the square root of the number of observations. So, if one observation system can sample the same parameter 100 times as frequent as another system, then we can accept a measurement uncertainty of the former that is 10 times as large as that of the latter system.

[23] The numbers generated in Figure 3 are averages based on four sets of simulations, each of which is based on one particular realization of a global climate model. If more climate models would have been used, no doubt the results would have been different, although we believe that the numbers shown here are substantial and indicative of the issues faced when a long-term measurement program is conceived.

[24] The study by *Weatherhead et al.* [1998] and others presume that climate monitoring requirements can be obtained from parameter variability in past climate records. Our results demonstrate that this may not always be the case. Figures 1 and 2 demonstrate that there is an increase in trend in upper atmospheric specific humidity starting at about the present time when compared to the last 75 years, combined with an increase in variability towards the end of the 21st century. So, the observed level of natural variability and trend, if any, in the past is no guarantee for what can be expected in the future. This is a reason to rely on climate projections in order to gauge climate monitoring accuracy requirements.

4. Conclusions

[25] In this study observation requirements for a climate monitoring program are derived from transient climate simulations. We focused on upper tropospheric specific humidity because this parameter is climatologically relevant, but at the same time difficult to measure. The climate simulations were used as a proxy for a perfect climate record. The radiosonde observations used to probe this perfect record were derived by sampling at regular time intervals, superimposing a 10% random error. At a midlatitude site, a record of at least fifty years of radiosonde observations is found necessary before a trend can be established that is statistically significant and at least within 20% of the true trend. Climate simulations on which these results are based encompass a limited spectrum of possible solutions. Yet, all regional model output is based on accepted climate scenarios with well-prescribed increases in greenhouse gases and there is consensus that an increase in humidity throughout the troposphere will occur.

[26] The decision to start a monitoring program of radiosonde observations of upper tropospheric humidity is a scientific as well as an economic one. In case the expense of starting such a program is prohibitive alternative options present themselves. The use of a Raman lidar yields similar results because the sampling rates are higher than a single radiosonde ascent, while the accuracy in some Raman-lidar observations now approaches that of an in–situ radiosonde [*Whiteman et al.*, 2006]. Raman observations taken at a frequency of 10 Hz are possible at any time if there are no clouds in the atmospheric column above the site. Also, in a given climatic regime it may be useful to designate a single anchor site to perform the expensive radiosonde observations and periodically calibrate the cheaper remote sensing systems at alternate sites against these in–situ observations. If these other systems, such as lidar systems, are capable of maintaining a stable calibration in the intermediate period between the periodic reference calibrations this would limit the expense of investment into in-situ systems, the performance of which in the future remains uncertain.

[27] Ultimately, this study shows that more effort is necessary to measure and understand upper tropospheric humidity for the purpose of obtaining a calibrated long-term climate signal. The water vapor at only a single station was studied, and work needs to be done in order to understand the water vapor behavior at all altitudes and in all climatic regimes.

[28] **Acknowledgments.** Regional climate model simulations were in part funded by the EU-FP6 ENSEMBLES project through contract GOCE-CT-2003-505539.

References

Global Climate Observing System (GCOS) (2007), GCOS Reference Upper Atmosphere Network (GRUAN): Justification, requirements, siting and instrumentation options, *GCOS 112/World Meteorol. Organ. Tech. Doc. 1379*, World Meteorol. Organ., Geneva, Switzerland.

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

Kendall, M. G. (1975), *Rank Correlation Methods*, Charles Griffin, London.

Mann, H. B. (1945), Non-parametric tests against trend, *Econometrica*, *13*, 245–256, doi:10.2307/1907187.

Miloshevich, L. M., H. Vömel, D. M. Whiteman, and T. LeBlanc (2009), Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements, *J. Geophys. Res.*, *114*, D11305, doi:10.1029/2008JD011565.

Seidel, D., and M. Free (2006), Measurement requirements for climate monitoring of upper-air temperature derived from reanalysis data, *J. Clim.*, *19*, 854–871, doi:10.1175/JCLI3666.1.

Soden, B. J., and F. P. Bretherton (1993), Upper tropospheric relative humidity from the GOES 6.7 μm channel: Method and climatology for July 1987, *J. Geophys. Res.*, *98*, 16,669–16,688, doi:10.1029/93JD01283.

van Meijgaard, E., L. H. van Ulft, W. J. van de Berg, F. C. Bosveld, B. J. J. M. van den Hurk, G. Lenderink, and A. P. Siebesma (2008), The KNMI regional atmospheric climate model RACMO, version 2.1, *KNMI Tech. Rep. 302*, 43 pp., R. Neth. Meteorol. Inst., De Bilt, Netherlands.

Vömel, H., D. E. David, and K. Smith (2007), Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: Instrumental details and observations, *J. Geophys. Res.*, *112*, D08305, doi:10.1029/2006JD007224.

Weatherhead, E. C., et al. (1998), Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.*, *103*, 17,149–17,161, doi:10.1029/98JD00995.

Whiteman, D. N., et al. (2006), Analysis of Raman lidar and radiosonde measurements from the AWEX-G field campaign and its relation to Aqua validation, *J. Geophys. Res.*, *111*, D09S09, doi:10.1029/2005JD006429.

R. Boers and E. van Meijgaard, Royal Netherlands Meteorological Institute, PO Box 201, NL-3730 AE De Bilt, Netherlands. (reinout.boers@knmi.nl)