

On the statistical significance of surface air temperature trends in the Eurasian Arctic region

C. Franzke¹

Received 16 October 2012; accepted 6 November 2012; published 11 December 2012.

[1] This study investigates the statistical significance of the trends of station temperature time series from the European Climate Assessment & Data archive poleward of 60°N. The trends are identified by different methods and their significance is assessed by three different null models of climate noise. All stations show a warming trend but only 17 out of the 109 considered stations have trends which cannot be explained as arising from intrinsic climate fluctuations when tested against any of the three null models. Out of those 17, only one station exhibits a warming trend which is significant against all three null models. The stations with significant warming trends are located mainly in Scandinavia and Iceland. **Citation:** Franzke, C. (2012), On the statistical significance of surface air temperature trends in the Eurasian Arctic region, *Geophys. Res. Lett.*, 39, L23705, doi:10.1029/2012GL054244.

1. Introduction

[2] The Arctic has experienced some of the most dramatic environmental changes over the last few decades which includes the decline of land and sea ice, and the thawing of permafrost soil. These effects are thought to be caused by global warming and have potentially global implications. For instance, the thawing of permafrost soil represents a potential tipping point in the Earth system and could lead to the sudden release of methane which would accelerate the release of greenhouse gas emissions and thus global warming.

[3] Whilst the changes in the Arctic must be a concern, it is important to place them in context because the Arctic exhibits large natural climate variability on many time scales [Polyakov *et al.*, 2003] which can potentially be misinterpreted as apparent climate trends. For instance, natural fluctuations on a daily time scale associated with weather systems can cause fluctuations on much longer time scales [Feldstein, 2000; Czaja *et al.*, 2003; Franzke, 2009]. This effect is called climate noise. Even very simple stationary stochastic processes can create apparent trends over rather long periods of time; so-called stochastic trends [Cryer and Chan, 2008; Cowpertwait and Metcalfe, 2009; Barbosa, 2011; Fatichi *et al.*, 2009; Franzke, 2010, 2012]. On the other hand, a so-called deterministic trend arises from external factors like greenhouse gas emissions.

[4] Specifically, here I will ask whether the observed temperature trends in the Eurasian Arctic region are outside

of the expected range of stochastic trends generated with three different null models of the natural climate background variability. Choosing the appropriate null model is crucial for the statistical testing of trends in order not to wrongly accept a trend as deterministic when it is actually a stochastic trend [Franzke, 2010, 2012].

[5] There are two paradigmatic null models for representing climate variability: short-range dependent (SRD) and long-range dependent (LRD) models [Robinson, 2003; Franzke, 2010, 2012; Franzke *et al.*, 2012]. In short, SRD models are the most used models in climate research and represent the initial decay of the autocorrelation function very well. For instance, a first order autoregressive process (AR(1)) has an exponential decay of the autocorrelation function. LRD models represent the low-frequency spectrum very well, have a pole at zero frequency and a hyperbolic decay of the autocorrelation function. One definition of a LRD process is that the integral over its autocorrelation function is infinite while a SRD process has always an integrable autocorrelation function [Robinson, 2003; Franzke *et al.*, 2012]. In general, both stochastic processes can generate stochastic trends but stochastic trends of LRD models can last for much longer than stochastic trends of SRD models. This shows that the rate of decay of the autocorrelation function has a strong impact on the length of stochastic trends. In addition to these two paradigmatic models we will also use a non-parametric method to generate surrogates which exactly conserve the autocorrelation function of the observed time series. Figure 1 displays the autocorrelation function for one of the used stations and the corresponding autocorrelation functions of the above three models. It has to be noted that there are a myriad of nonlinear stochastic models which can potentially be used to represent the background climate variability and the significance estimates will depend on the used null model. However, I have chosen the three above models because two of them represent paradigmatic models for representing the correlation structure and one conserves exactly the empirical correlation structure.

2. Data and Methods

[6] I use the daily mean temperatures from 109 stations from the European Climate Assessment and Data archive compiled by Klein Tank *et al.* [2002] poleward of 60°N. Only stations with an almost continuous data coverage are used, these have at most a few days missing at a time, and missing data are interpolated as in Franzke [2010, 2012]. The data are de-seasonalised by subtracting the average temperature of each day. The stations cover time periods starting between 1881 and 1980 and ending between 1994 and 2011. The locations of the stations used is depicted in Figure 2.

¹British Antarctic Survey, Natural Environment Research Council, Cambridge, UK.

Corresponding author: C. Franzke, British Antarctic Survey, Natural Environment Research Council, High Cross Madingley Road, Cambridge CB3 0ET, UK. (chan1@bas.ac.uk)

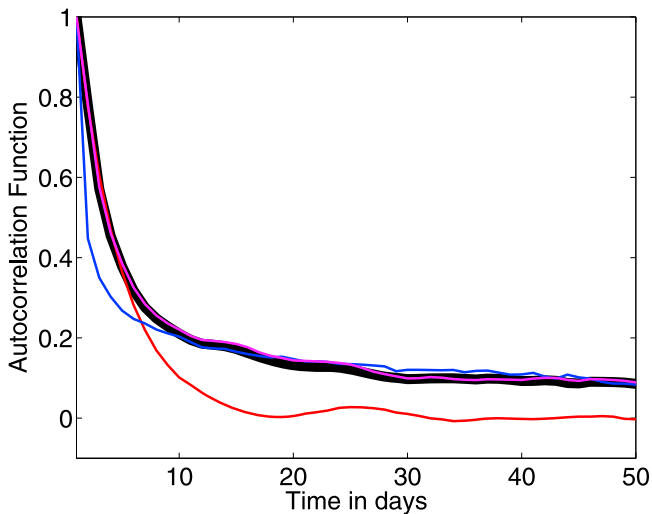


Figure 1. Autocorrelation functions for Falun, Sweden, station (black line), corresponding AR(1) process (red line), ARFIMA process (blue line) and phase scrambling surrogate (magenta line).

[7] The definition of a trend is subjective and depends on the method used [Wu *et al.*, 2007]. In order to evaluate the robustness of trends I use different trend identification methods: (i) ordinary least-squares regression, (ii) robust regression, (iii) generalised linear model regression, (iv) wavelets and (v) ensemble empirical mode decomposition. In the following I will only briefly describe these methods; a more complete exposition is given in Franzke [2012].

[8] (i) Ordinary least-squares (OLS) regression fits polynomial functions of arbitrary order to data. OLS decomposes a time series into a signal and residuals and assumes that the residuals come from a distribution which has a finite variance and are serially uncorrelated. In this study I use linear, quadratic and cubic polynomials as the signal which will then be interpreted as the trend [e.g., Franzke, 2012].

[9] (ii) Robust regression [Draper and Smith, 1998] is a method which can deal with outliers. It is a form of weighted least-squares regression and is done iteratively. At each iteration step a new set of weights are computed by bisquare weighting based on the residuals, with larger residuals having smaller weights (see Franzke [2012] for more details). The weights depend on the residuals and consequently large deviations and outliers are down weighted and have less influence on the regression fit.

[10] (iii) Generalized Linear Model (GLM) regression generalises OLS regression by allowing the dependent variable to stem from a distribution from the exponential family [Draper and Smith, 1998]. GLM can be useful when the residuals are non-Gaussian distributed.

[11] (iv) Wavelets are wave-like functions representing brief oscillations [Mallat, 1999]. Here I use the wavelet approach of Andreas and Trevino [1997] to detect linear and quadratic trends. See Andreas and Trevino [1997] for more details.

[12] (v) The Ensemble Empirical Mode Decomposition method (EEMD) [Wu and Huang, 2009; Huang *et al.*, 1998; Huang and Wu, 2008] decomposes a time series into a finite number of intrinsic mode functions and an instantaneous mean. When this instantaneous mean is not constant I refer

to it as a trend, which is possibly nonlinear [Wu *et al.*, 2007; Franzke, 2009, 2010, 2012] over the time series length.

[13] The above methods will be applied to monthly mean time series which have been computed from the daily data. This has been done to reduce the noise and to concentrate on the longer time scales. A comparison of the trends reveal that a cubic regression fit gives the smallest root mean square error (not shown). This is consistent with the results in Franzke [2012]. Furthermore, a visual inspection gives the impression that the cubic regression fit and the non-parametric EEMD trend are very similar. Also cubic OLS, robust regression and cubic GLM regression give very similar results. This provides further evidence that temperature trends are non-linear [Franzke, 2010, 2012].

[14] Thus, for the significance tests, I will focus on the cubic regression trends. The magnitude of a trend is defined as the range between the minimum and maximum value of the trend line which in most cases corresponds to the start and end of the time series. This is a robust definition because it is a very smooth function and variability on interannual and decadal time scales has thus been removed. The cubic regression is very similar to the EEMD trend and EEMD has been shown to be able to extract climate variability on interannual and decadal time scales [Wu *et al.*, 2007; Franzke, 2009; Franzke and Woollings, 2011] and meaningful trends. Furthermore, defining the magnitude of the trend as the range between the start and end point gives similar results.

[15] After identifying the trends I have to assess their statistical significance. This has been done by examining how often they are outside the trend ranges of the ensembles of surrogate time series generated by the three null models representing the background climate variability of the respective stations. To create ensembles of surrogate time series I use a first order autoregressive model (AR(1) [Franzke, 2010, 2012]) as a SRD model, and an autoregressive fractionally integrated moving average model (ARFIMA(0,d,0)) [Robinson, 2003; Franzke, 2010, 2012] as a LRD model, where d denotes the LRD parameter. As a non-parametric way of computing surrogate data with exactly the same autocorrelation function I use the phase scrambling method by Theiler *et al.* [1992]. This method computes the power spectrum of a time series and then randomises the

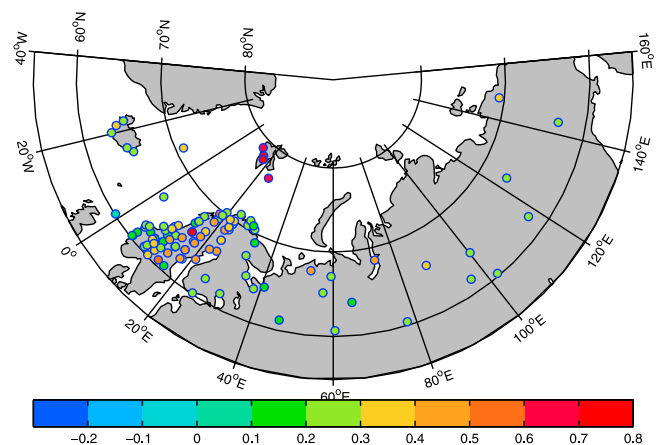


Figure 2. Map of stations: Magnitude of the observed trend in °C per decade.

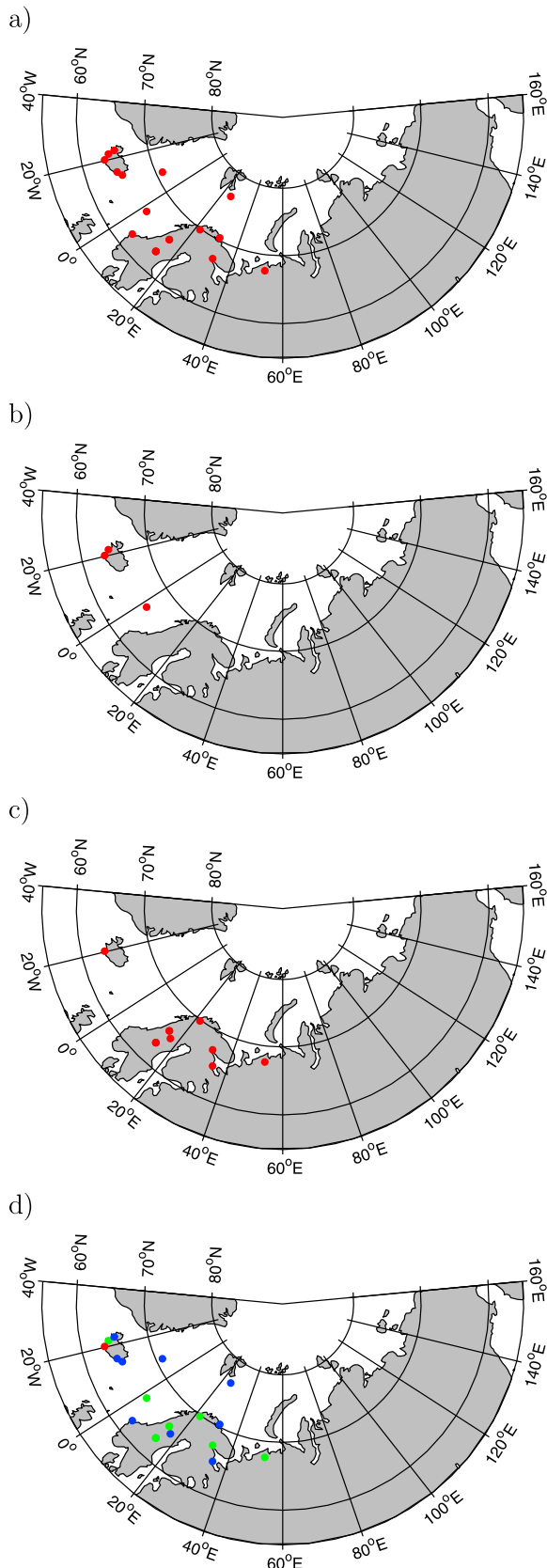


Figure 3. Stations with a statistically significant trend against (a) AR(1), (b) ARFIMA, (c) phase scrambling null model and (d) stations with a significant trend: blue: weak evidence, green: moderate evidence and red: strong evidence.

phase spectrum. Because the power spectrum is the Fourier transform of the autocorrelation function (Wiener-Khinchin theorem) randomising the phase spectrum does not affect the autocorrelation function (see Figure 1).

[16] The parameters of the AR(1) and ARFIMA models have been estimated from the observed daily data [Franzke, 2009, 2010, 2012]. I use daily data because were the time series really to stem from a AR(1) process then one could relate those parameters to the corresponding AR(1) parameters for averaged data [Kushnir *et al.*, 2006; D. I. Vyushin *et al.*, Modelling and understanding persistence of climate variability, submitted to *Journal of Geophysical Research*, 2012]. Similarly, if the data were indeed from a ARFIMA(0,d,0) process then the parameters from daily data would also be the same as for averaged data (Vyushin *et al.*, submitted manuscript, 2012). However, using monthly or seasonally averaged data would increase the estimation error. Surrogates were then created by the phase scrambling method (see Franzke [2012] for more details). I created 1000 surrogate time series from each of the three models for each station. For the AR(1) and ARFIMA models parameter estimation uncertainty also was taken into account for the Monte Carlo experiments [Franzke, 2010, 2012]. The trends in the surrogate time series were then computed using cubic OLS. This results in a distribution of trends. If now the observed trend is outside the 95% percentile of the distribution of the stochastic trends then I claim that the observed trend is a deterministic trend with respect to the chosen null model of climate variability and likely due to external factors like greenhouse gas emissions.

3. Results

[17] Figure 2 displays the location of all stations and the colour coding indicates the magnitude and sign of the temperature trends. The first thing to note is that all stations experience a warming trend over their respective observational periods. The largest trends (more than 0.4°C per decade) are in central Scandinavia and Svalbard. Most of Siberia experienced warming trends of about $0.2\text{--}0.3^{\circ}\text{C}$ per decade.

[18] After finding evidence for warming trends we have now to assess their statistical significance; do the magnitudes of the observed trends lie already outside of the expected range of natural climate variability? The above three significance tests reveal that 17 of the 109 stations are significant against an AR(1) null model (Figure 3a), 3 stations are significant against a ARFIMA null model (Figure 3b), and 8 stations are significant against a climate noise null hypothesis using phase scrambling surrogates (Figure 3c). All these trends are significant at the 97.5% confidence level. This shows that while the Eurasian Arctic region shows a widespread warming trend, only about 15% of the stations are significant against any of the three significance tests.

[19] Using the three different null models enables us to introduce degrees of significance using the scale introduced in Franzke [2012]. I claim strong evidence of a significant trend if the observed trend is significant against all three null models; I claim moderate evidence of a significant trend if the observed trend is significant against two of the null models; and consequently I claim weak evidence of a

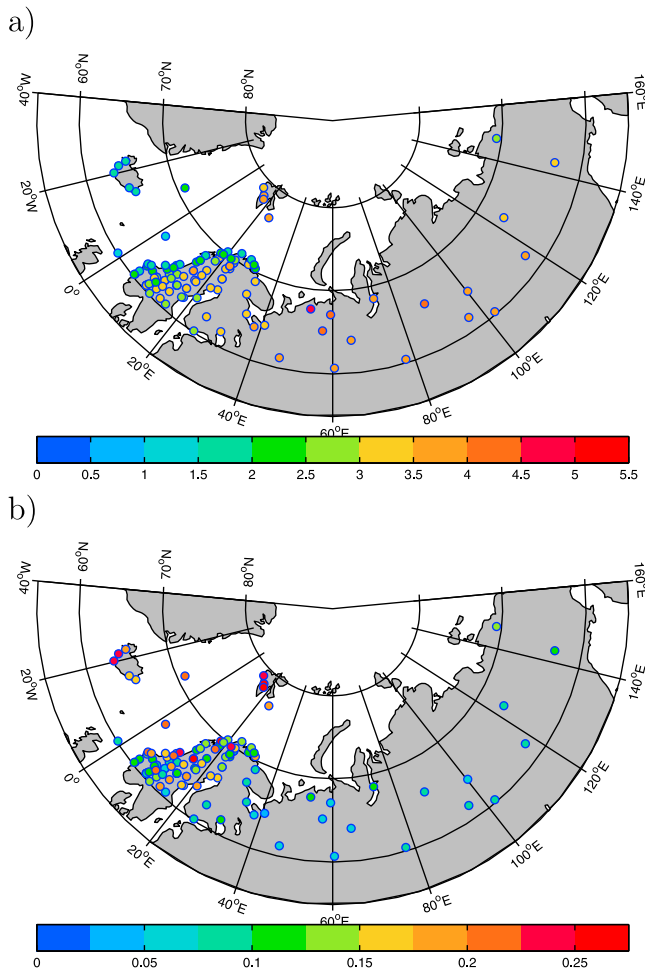


Figure 4. (a) Standard deviations of de-seasonalised temperature at the stations in $^{\circ}\text{C}$ and (b) signal to noise ratio.

significant trend if the observed trend is significant against just one of the null models.

[20] Applying this scale to the results I find just 1 station with strong evidence, 7 stations with moderate evidence and 9 stations with weak evidence for a warming trend which cannot be explained as arising from a stochastic trend (Figure 3d). The stations with significant warming trends are located in Iceland, the North Atlantic, Scandinavia and north-east Russia. The station showing strong evidence is located in Iceland. While Svalbard and Siberia are experiencing large warming trends these trends are not significant; thus they can be explained as arising from natural climate variability. My interpretation of this finding is that this is likely due to the relatively high temperature variance in this region. Figure 4a shows the standard deviation of the de-seasonalised station temperature time series. The largest standard deviations are in Siberia, Svalbard and the interior of Scandinavia. This is confirmed by the signal to noise ratio computed for each station (Figure 4b). This shows that most of Siberia has low signal to noise ratios while Scandinavia has higher ratios.

[21] While the data coverage period of the stations can vary, all stations with significant trends cover periods of at least 55 years and many cover even longer periods. This gives confidence in the robustness of the results and that at

least at some sites there is a long-term warming trend which is already outside of the range of natural climate variability.

4. Summary

[22] In this study I have investigated the statistical significance of air temperature trends of 109 stations in the Eurasian Arctic region. This region experiences strong climate variability on all time scales. This strong natural climate variability can create apparent trends which needs to be captured in the null model for any statistical significance tests. Thus, for this reason we use three different models to increase our confidence in the significance of the results.

[23] I found evidence for significant temperature increases in Scandinavia and the North Atlantic at 17 of the 109 stations. 9 of these 17 trends are only significant against one of the used null models representing the background climate variability. This means that there is only weak evidence for a significant trend at these 9 stations. 7 of the significant trends are significant against two of the null models; thus there is moderate evidence for a temperature trend at these stations. Only for one station, on Iceland, there is strong evidence for a significant trend. This station time series is significant against all 3 null models used.

[24] These results come with the caveat that for relatively short time series we might not be able to identify the ‘true model’ of the background climate variability and the used models might all pass diagnostic tests [Percival *et al.*, 2004] while they might imply different long-term consequences. This can partly be rectified by using longer climate records like ice cores. On the other hand, using longer climate records may also invalidate some of the results based on the currently available data. However, I have provided here evidence for Arctic temperature warming trends based on modern techniques of time series analysis and a high quality temperature data set.

[25] The thawing of the permafrost soil in Siberia is widely seen as a potential tipping point in the climate system. While I do not find evidence for a significant warming trend in Siberia the raw data still indicate a widespread temperature increase in Siberia (Figure 2). Given that the temperature fluctuations in Siberia are large, this portends to the possibility that the warming signal in Siberia has not yet reached its time of emergence when it will be outside of the range of natural temperature variability.

[26] **Acknowledgments.** I thank M. Freeman and two anonymous reviewers for their helpful comments on an earlier version of this manuscript. I thank KNMI for providing me with the ECA&D temperature data. This study is part of the British Antarctic Survey Polar Science for Planet Earth Programme. It was funded by the Natural Environment Research Council.

[27] The Editor thanks the two anonymous reviewers for their assistance in evaluating this paper.

References

- Andreas, E. L., and G. Trevino (1997), Using wavelets to detect trends, *J. Atmos. Oceanic Technol.*, *14*, 554–564.
- Barbosa, S. M. (2011), Testing for deterministic trends in global sea surface temperature, *J. Clim.*, *24*, 2516–2522.
- Cryer, J. D., and K.-S. Chan (2008), *Time Series Analysis*, 2nd ed., 491 pp., Springer, New York.
- Czaja, A., A. W. Robertson, and T. Huck (2003), The role of Atlantic Ocean–atmosphere coupling in affecting North Atlantic Oscillation variability, in *The North Atlantic Oscillation: Climatic Significance and*

- Environmental Impact, Geophys. Monogr. Ser.*, vol. 134, edited by J. W. Hurrell et al., pp. 147–172, AGU, Washington, D. C.
- Cowpertwait, P. S. P., and A. V. Metcalfe (2009), *Introductory Time Series With R*, 256 pp., Springer, New York.
- Draper, N. R., and H. Smith (1998), *Applied Regression Analysis*, 706 pp., Wiley Interscience, New York.
- Fatichi, S., S. M. Barbosa, E. Caporali, and M. E. Silva (2009), Deterministic versus stochastic trends: Detection and challenges, *J. Geophys. Res.*, *114*, D18121, doi:10.1029/2009JD011960.
- Feldstein, S. B. (2000), The timescale, power spectra, and climate noise properties of teleconnection patterns, *J. Clim.*, *13*, 4430–4440.
- Franzke, C. (2009), Multi-scale analysis of teleconnection indices: Climate noise and nonlinear trends, *Nonlinear Processes Geophys.*, *16*, 65–76.
- Franzke, C. (2010), Long-range dependence and climate noise characteristics of Antarctic temperature data, *J. Clim.*, *23*, 6074–6081, doi:10.1175/2010JCLI3654.1.
- Franzke, C. (2012), Nonlinear trends, long-range dependence and climate noise properties of surface air temperature, *J. Clim.*, *25*, 4172–4183.
- Franzke, C., and T. Woollings (2011), On the persistence and predictability properties of North Atlantic climate variability, *J. Clim.*, *24*, 466–472.
- Franzke, C., T. Graves, N. W. Watkins, R. B. Gramacy, and C. Hughes (2012), Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity, *Philos. Trans. R. Soc. A*, *370*, 1250–1267, doi:10.1098/rsta.2011.0349.
- Huang, N. E., and Z. Wu (2008), A review on Hilbert-Huang transform: Method and its applications to geophysical studies, *Rev. Geophys.*, *46*, RG2006, doi:10.1029/2007RG000228.
- Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu (1998), The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London A*, *454*, 903–995.
- Klein Tank, A. M. G., et al. (2002), Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, *Int. J. Climatol.*, *22*, 1441–1453.
- Kushnir, Y., W. A. Robinson, P. Chang, and A. W. Robertson (2006), The physical basis for predicting Atlantic sector seasonal-to-interannual climate variability, *J. Clim.*, *19*, 5949–5970.
- Mallat, S. (1999), *A Wavelet Tour of Signal Processing*, 2nd ed., 637 pp., Academic, San Diego, Calif.
- Percival, D. B., J. E. Overland, and H. O. Mofjeld (2004), Modelling North Pacific climate time series, in *Time Series Analysis and Applications to Geophysical Systems*, edited by D. Brillinger, E. Robinson, and F. Schoenberg, pp. 151–167, Springer, New York.
- Polyakov, I. V., R. V. Bekryaev, G. V. Alekseev, U. S. Bhatt, R. L. Colony, M. A. Johnson, A. P. Maskhtas, and D. Walsh (2003), Variability and trends of air temperature and pressure in the maritime Arctic, 1875–2000, *J. Clim.*, *16*, 2067–2077.
- Robinson, P. M. (2003), Long-memory time series, in *Time Series With Long Memory*, edited by P. M. Robinson, pp. 4–32, Oxford Univ. Press, Oxford, U. K.
- Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer (1992), Testing for nonlinearity in time series: The method of surrogate data, *Physica D*, *58*, 77–94.
- Wu, Z., and N. E. Huang (2009), Ensemble empirical mode decomposition: A noise-assisted data analysis method, *Adv. Adapt. Data Anal.*, *1*, 1–41.
- Wu, Z., Huang, N. E., Long, S. R. and Peng, C. K. (2007), On the trend, detrending, and variability of nonlinear and nonstationary time series, *Proc. Natl. Acad. Sci. U. S. A.*, *104*, 14,889–14,894.