

## An Omnibus Test for Red Noise, with Applications to Climatology Time Series

Don Percival

Applied Physics Laboratory  
Department of Statistics  
University of Washington  
Seattle, Washington, USA

<http://faculty.washington.edu/dbp>

Collaborative effort with You-Gan Wang

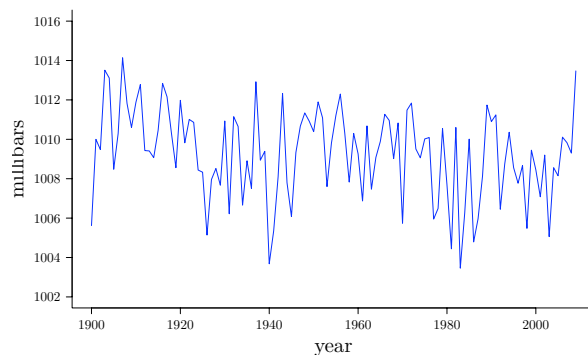
## Overview

- analysis of time series related to climate often rely on ‘red noise’ as a simple model for correlation in the series
- after giving background on
  - red noise,
  - partial autocorrelation sequences (PACFs) and
  - portmanteau tests for white noise,will describe an omnibus test – and a variation thereof – designed to point out when a model other than red noise is needed
- will discuss adaptation of tests to handle time series with missing values (a common occurrence in climatology)
- will demonstrate use of tests on two climatology time series

1

## Example of a Climatology Time Series (NPI): I

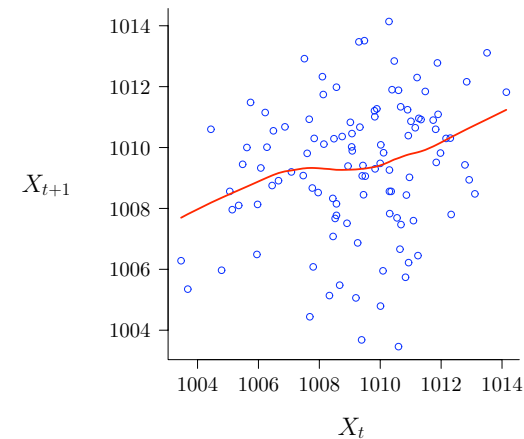
- consider North Pacific Index (NPI): area-weighted sea level pressure over  $30^\circ$  N to  $65^\circ$  N &  $160^\circ$  E to  $140^\circ$  W and over November to March for each year from 1900 to 2009



2

## Example of a Climatology Time Series (NPI): II

- unit lag scatter plot & locally weighted regression fit ( $\hat{\rho}_1 \doteq 0.21$ )



3

## Modelling Correlation in Time Series as Red Noise

- cannot regard NPI and most other climatology time series  $X_t$  as realizations of independent random variables (RVs)
- widely-used simple model for correlated time series is 'red noise'
- red noise is the same as a first-order autoregressive (AR(1)) stationary Gaussian process with a positive correlation at unit lag (see, e.g., von Storch and Zwiers, 1999)
- assuming  $E\{X_t\} = 0$  for convenience, such a process satisfies

$$X_t = \phi X_{t-1} + \epsilon_t,$$

where  $|\phi| < 1$ , and  $\epsilon_t$ 's are IID Gaussian with mean 0 and variance  $\sigma_\epsilon^2$  (i.e., Gaussian white noise)

4

## Properties of AR(1) Processes

- can argue that

$$\sigma_X^2 \equiv \text{var}\{X_t\} = \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad \text{and} \quad \text{cov}\{X_{t+k}, X_t\} = \phi^{|k|} \sigma_X^2,$$

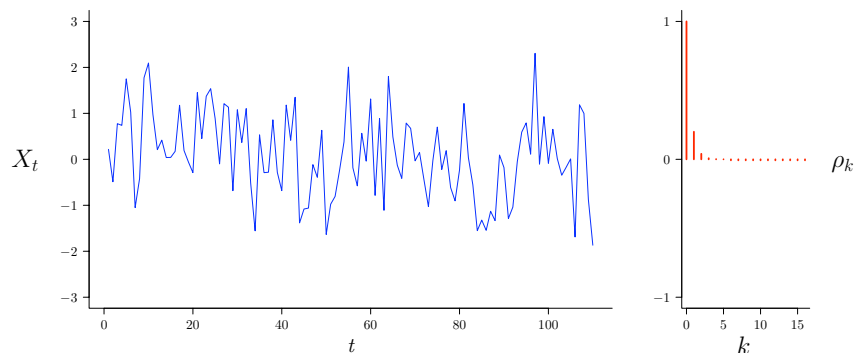
so  $\rho_k \equiv \text{corr}\{X_{t+k}, X_t\} = \phi^{|k|}$  – the autocorrelation sequence (ACS) – dies down exponentially as  $k \rightarrow \infty$  ('short-range' dependence)

- when  $\phi = 0$ , AR(1) process reduced to white noise
- AR(1) process is related to a first-order stochastic differential equation with 'correlation time' dictated by  $\phi$
- AR(1) model is simple enough to offer analytic tractability for calculations (e.g., getting an expression for  $\text{var}\{\bar{X}\}$ , where  $\bar{X}$  is the mean of  $X_0, X_1, \dots, X_{N-1}$ )

5

## Examples of Realizations of AR(1) Processes

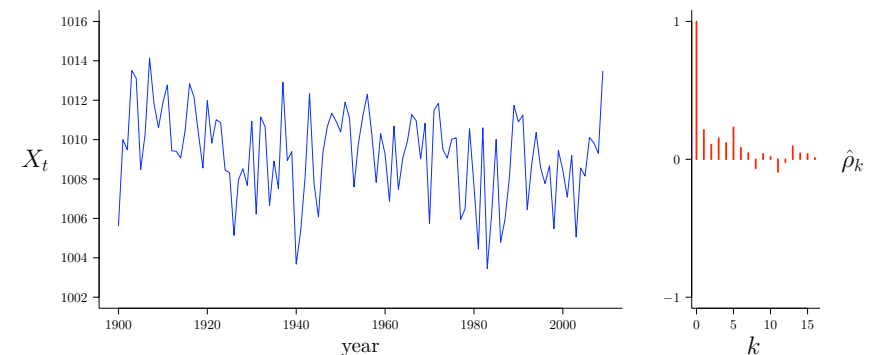
- easy to generate realizations of Gaussian AR(1) processes (see Kay, 1981, for details)
- with  $\sigma_X^2 = 1$ , here is a realization when  $\phi = 0.2$



6

## Examples of Realizations of AR(1) Processes

- easy to generate realizations of Gaussian AR(1) processes (see Kay, 1981, for details)
- for comparison, here is NPI with  $\phi$  estimated by  $\hat{\rho}_1 \doteq 0.21$



7

## Quote from Review of Recent Paper (2010)

“... the paper ... can serve as a cautious reminder to those (probably more than 90% of climate researchers) who blindly use the AR(1) model ... for climate data”

- Q: has red noise been ‘oversold’ within climate community?
- points out need for statistical tests that flag time series for which red noise might be too simplistic of a model
- will now review a concept that, for Gaussian processes, provides a clear distinction between AR(1) processes and all other types of correlated stationary processes – to do so, we first need to consider ‘predicting’  $X_t$  using other RVs in the process

8

## Forward/Backward Prediction of $X_t$ : I

- given a realization of a portion  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$  of a zero mean Gaussian stationary process, suppose we want to predict  $X_t$  based upon some function, say  $g(X_{t-1}, X_{t-2}, \dots, X_{t-k})$ , of these  $k$  RVs
- the ‘best’ predictor of  $X_t$  takes the form

$$g(X_{t-1}, X_{t-2}, \dots, X_{t-k}) = \sum_{j=1}^k \phi_{k,j} X_{t-j} \equiv \vec{X}_t(k),$$

where here ‘best’ means that

$$E\{[X_t - g(X_{t-1}, X_{t-2}, \dots, X_{t-k})]^2\}$$

is minimized over all possible functions of  $X_{t-1}, X_{t-2}, \dots, X_{t-k}$  (note: the  $\phi_{k,j}$  depend just on the ACS for the process)

9

## Forward/Backward Prediction of $X_t$ : II

- given a realization of a portion  $X_{t+1}, X_{t+2}, \dots, X_{t+k}$  of RVs coming *after*  $X_t$ , the best ‘predictor’ of  $X_t$  is given by

$$\overleftarrow{X}_t(k) \equiv \sum_{j=1}^k \phi_{k,j} X_{t+j},$$

i.e., coefficients  $\phi_{k,j}$  in  $\overleftarrow{X}_t(k)$  are the same as those in  $\vec{X}_t(k)$

- given  $\vec{X}_t(k)$  and  $\overleftarrow{X}_t(k)$ , form the corresponding forward and backward prediction errors:

$$\vec{\epsilon}_t(k) \equiv X_t - \vec{X}_t(k) \quad \text{and} \quad \overleftarrow{\epsilon}_t(k) \equiv X_t - \overleftarrow{X}_t(k);$$

note: define  $\vec{X}_t(0) = \overleftarrow{X}_t(0) = 0$  so that  $\vec{\epsilon}_t(0) = \overleftarrow{\epsilon}_t(0) = X_t$

10

## Partial Autocorrelation Sequence (PACS): I

- focusing now on  $X_{t-k}, X_{t-k+1}, \dots, X_{t-1}, X_t$ , we can interpret  $\phi_{k,k}$  in the following interesting manner:

$$\phi_{k,k} = \frac{\text{cov}\{\vec{\epsilon}_t(k-1), \overleftarrow{\epsilon}_{t-k}(k-1)\}}{(\text{var}\{\vec{\epsilon}_t(k-1)\} \text{var}\{\overleftarrow{\epsilon}_{t-k}(k-1)\})^{1/2}};$$

because

$$\vec{\epsilon}_t(k-1) = X_t - \vec{X}_t(k-1)$$

$$\overleftarrow{\epsilon}_{t-k}(k-1) = X_{t-k} - \overleftarrow{X}_{t-k}(k-1),$$

and because both  $\vec{X}_t(k-1)$  and  $\overleftarrow{X}_t(k-1)$  depend on just

$$X_{t-k+1}, \dots, X_{t-1},$$

can regard  $\phi_{k,k}$  as correlation between  $X_{t-k}$  and  $X_t$  after ‘adjustment’ by the intervening  $k-1$  RVs  $X_{t-k+1}, \dots, X_{t-1}$

11

## Partial Autocorrelation Sequence (PACS): II

- $\phi_{k,k}$ ,  $k = 1, 2, \dots$ , is known as the partial ACS (PACS)
- Ramsey (1974): under a Gaussian assumption, a stationary process is an AR(1) process if and only if its PACS is identically zero for all  $k \geq 2$
- can thus use estimators of  $\hat{\phi}_{2,2}, \hat{\phi}_{3,3}, \dots$  to test null hypothesis

$H_0$ : time series is a realization of an AR(1) process

versus nonspecific alternative

$H_1$ : time series is a realization of another stationary process

12

## Partial Autocorrelation Sequence (PACS): III

- given time series that is a realization of  $X_0, X_1, \dots, X_{N-1}$ , can estimate  $\phi_{k,k}$  by fitting  $k$ th order AR process, i.e.,

$$X_t = \sum_{j=1}^k \phi_{k,j} X_{t-j} + \epsilon_t,$$

using a variety of methods (Yule–Walker, Burg, forward least squares (LS), forward/backward LS, maximum likelihood, ...)

- AR( $k$ ) coefficients  $\{\phi_{k,j} : j = 1, \dots, k\}$  and PACS  $\{\phi_{j,j} : j = 1, \dots, k\}$  are equivalent to one another
- large-sample theory says PACS estimators  $\hat{\phi}_{2,2}, \hat{\phi}_{3,3}, \dots, \hat{\phi}_{K,K}$  are approximately IID normal with mean zero and variance  $1/N$  for an AR(1) process and fixed  $K$  (Kay & Makhoul, 1983)

13

## Portmanteau Tests for White Noise: I

- can formulate tests for red noise analogous to portmanteau tests for white noise (Box and Pierce, 1970; Ljung and Box, 1978)
- large-sample statistical theory says that, for a time series coming from a white noise process and for fixed  $K$ , standard ACS estimators  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_K$  are approximately IID normal with mean zero and variance  $1/N$
- Box–Pierce and Ljung–Box–Pierce portmanteau test statistics for white noise versus nonspecific alternative are given by

$$Q_K = N \sum_{k=1}^K \hat{\rho}_k^2 \quad \text{and} \quad \tilde{Q}_K = N(N+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{N-k}$$

14

## Portmanteau Tests for White Noise: II

- for either  $Q_K$  or  $\tilde{Q}_K$ , reject null hypothesis of white noise at significance level  $\alpha$  when statistic exceeds  $(1 - \alpha) \times 100\%$  percentage point for chi-square distribution with  $K$  degrees of freedom
- literature recommends setting  $K = \max\{2, \min\{20, N/10\}\}$
- Baragona and Battaglia (2000) and Kwan (2003) consider analogous tests for white noise based on PACS estimators rather than ACS estimators (for a white noise process and for fixed  $K$ , PACS estimators  $\hat{\phi}_{1,1}, \hat{\phi}_{2,2}, \dots, \hat{\phi}_{K,K}$  are asymptotically IID normal with mean zero and variance  $1/N$ , i.e., the same result as for  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_K$ )

15

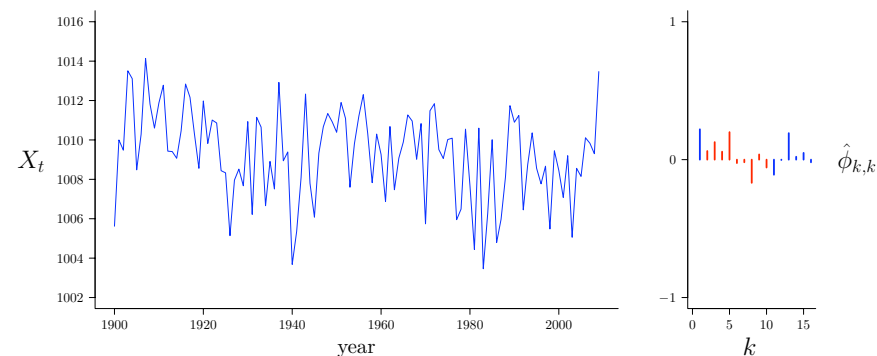
## Omnibus Portmanteau-based Tests for Red Noise

- since large sample properties of
  - $\hat{\rho}_k$  for  $k \geq 1$  under white noise hypothesis and
  - $\hat{\phi}_{k,k}$  for  $k \geq 2$  under red noise hypothesis
 are identical, can test hypothesis of red noise versus nonspecific alternative using the following analogs of  $Q_K$  or  $\tilde{Q}_K$ :
 
$$T_K = N \sum_{k=2}^K \hat{\phi}_{k,k}^2 \quad \text{and} \quad \tilde{T}_K = N(N+2) \sum_{k=2}^K \frac{\hat{\phi}_{k,k}^2}{N-k}$$
- for either test statistic, reject null hypothesis of red noise at significance level  $\alpha$  when statistic exceeds  $(1 - \alpha) \times 100\%$  percentage point for chi-square distribution with  $K - 1$  degrees of freedom

16

## Use of Tests on NPI Time Series

- for NPI series with  $K = 10$ , get  $T_{10} \doteq 10.69$  and  $\tilde{T}_{10} \doteq 11.47$ , yielding  $\hat{\alpha} = 0.30$  and  $0.24$ , so cannot reject red noise hypothesis at any reasonable significance level



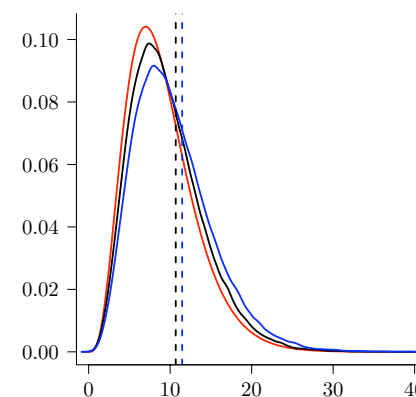
17

## Assessing $\chi^2$ Approximation

- generated 100,000 Gaussian AR(1) time series of length  $N = 110$  using  $\phi$  estimated from NPI data and computed  $T_{10}$  and  $\tilde{T}_{10}$  for each series
- used these as input to **density** function in **R** to estimate probability density functions (PDFs) for comparison with  $\chi^2_9$  PDF
- repeated above, but with  $\phi = 0.9$ , to see if results depend on  $\phi$

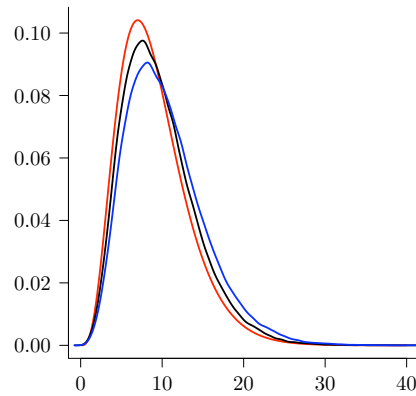
18

## PDFs for $T_{10}$ , $\tilde{T}_{10}$ and $\chi^2_9$ with $\phi \doteq 0.21$



19

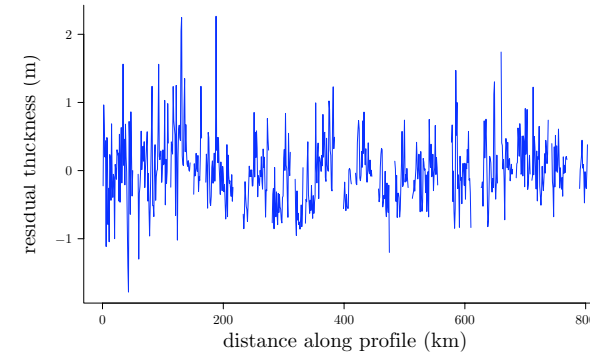
### PDFs for $T_{10}$ , $\tilde{T}_{10}$ and $\chi_9^2$ with $\phi = 0.9$



20

### Extension to Handle ‘Gappy’ Time Series: I

- climatology time series often have missing values, as is true for a ‘time’ series of Arctic sea-ice thickness measurements (172 out of 803 observations missing – 21% of data)



21

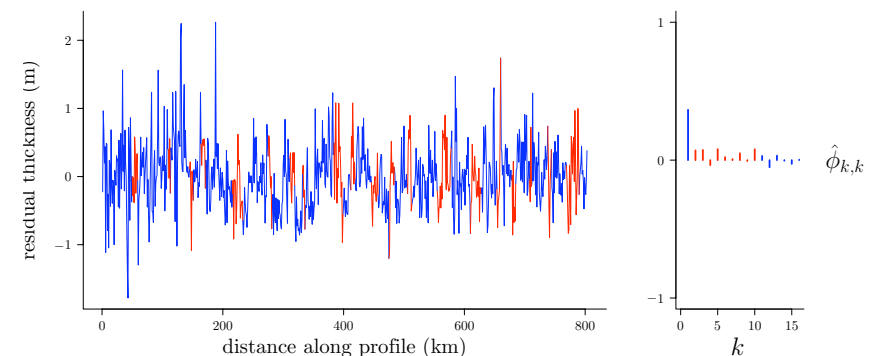
### Extension to Handle ‘Gappy’ Time Series: II

- assuming null hypothesis of red noise to be true, can estimate model parameters for gappy time series using maximum likelihood (Jones, 1980), yielding  $\hat{\phi} \doteq 0.36 (\pm 0.04)$
- let  $\mathbf{X}_O$  and  $\mathbf{X}_M$  be vectors of RVs containing observed and missing parts of time series
- using  $E\{\mathbf{X}_M | \mathbf{X}_O\}$  and  $\text{var}\{\mathbf{X}_M | \mathbf{X}_O\} = \Sigma_{M|O}$  formed under null hypothesis and conditioned on AR(1) parameter estimates, can generate realizations of missing data
  - formally requires Cholesky factorization of  $\Sigma_{M|O}$
  - due to special properties of AR(1) model, can reduce to factorization of a set of much smaller matrices
- can compute  $T_K$  and  $\tilde{T}_K$  for many such realizations

22

### Extension to Handle ‘Gappy’ Time Series: III

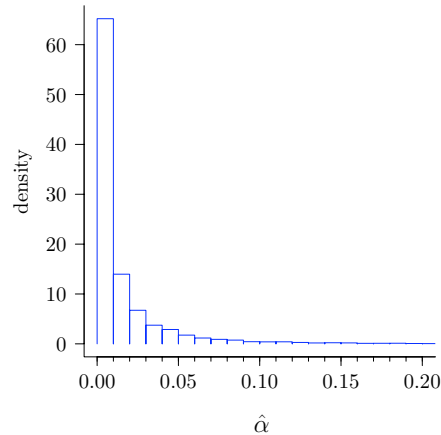
- get  $T_{10} \doteq 21.97$  and  $\tilde{T}_{10} \doteq 22.18$  for this particular stochastic interpolation of Arctic sea-ice series, yielding  $\hat{\alpha} \doteq 0.009$  and  $0.008$



23

## Extension to Handle ‘Gappy’ Time Series: IV

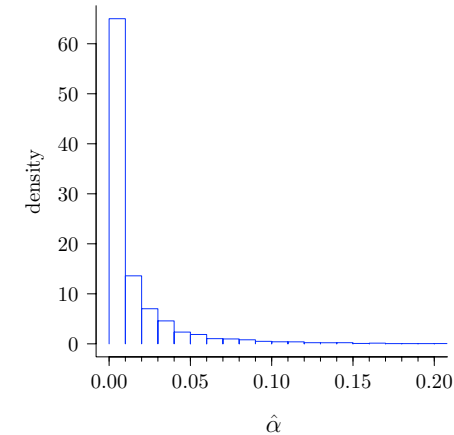
- histogram of 10,000  $\hat{\alpha}$ 's for  $T_{10}$  (one for  $\tilde{T}_{10}$  virtually the same)



24

## Extension to Handle ‘Gappy’ Time Series: V

- variation: account for uncertainty in ML parameter estimates



25

## Future Directions

- critique: tests based on gap-filling biased toward *not* rejecting null hypothesis
- another approach for handling gappy time series (under study):
  - base tests on PACS estimates from fitting  $AR(k)$  models via maximum likelihood (Jones, 1980)
  - method slow (numerical optimization over  $k$  parameters)
  - approach based on constraining ML estimators via Levinson–Durbin recursions leads to sequence of one-dimensional optimization problems and hence potential speed-up
  - small sample properties of simplified approach under study
- lots of other avenues to look into!

26

## Thanks to ...

- conference organizers for opportunity to talk
- numerous folks at CSIRO who made my visit possible

27

## References: I

- R. Baragona and F. Battaglia (2000), 'Partial and Inverse Autocorrelations in Portmanteau-type Tests for Time Series,' *Communications in Statistics-Simulation and Computation*, **29**, pp. 971–86
- G. E. P. Box and D. A. Pierce (1970), 'Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models,' *Journal of the American Statistical Association*, **65**, pp. 1509–26
- R. H. Jones (1980), 'Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations,' *Technometrics*, **22**, pp. 389–95
- S. M. Kay (1981), 'Efficient Generation of Colored Noise,' *Proceedings of the IEEE*, **69**, pp. 480–1
- S. M. Kay and J. Makhoul (1983), 'On the Statistics of the Estimated Reflection Coefficients of an Autoregressive Process,' *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **31**, pp. 1447–55
- A. C. C. Kwan (2003), 'Sample Partial Autocorrelations and Portmanteau Tests for Randomness,' *Applied Economics Letters*, **10**, pp. 605–9
- G. M. Ljung and G. E. P. Box (1978), 'On a Measure of Lack of Fit in Time Series Models,' *Biometrika*, **65**, pp. 297–303

## References: II

- F. L. Ramsey (1974), 'Characterization of the Partial Autocorrelation Function,' *Annals of Statistics*, **2**, pp. 1296–301
- K. E. Trenberth and J. W. Hurrell (1994), 'Decadal Atmosphere-Ocean Variations in the Pacific,' *Climate Dynamics*, **9**, pp. 303–19
- K. E. Trenberth and D. D. Paolino (1980), 'The Northern Hemisphere Sea Level Pressure Data Set: Trends, Errors, and Discontinuities,' *Monthly Weather Review*, **108**, pp. 855–72
- H. von Storch and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge, England: Cambridge University Press