Square Waves, Sinusoids and Gaussian White Noise: A Matching Pursuit Conundrum?

Don Percival

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

`http://faculty.washington.edu/dbp`

Sanity Checkers: Peter Craigmile and Barry Quinn

Introduction

- 'matching pursuit' approximates a vector of time series values
$$\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$$
using a linear combination of vectors picked from a (typically quite large) set of vectors $\mathcal{D}$
- each vector in $\mathcal{D}$ has some interpretation, allowing us to extract features of potential interest from $\mathbf{X}$
- introduced into engineering literature by Mallat & Zhang (1993)
- talk will focus on an unexpected finding (the 'conundrum'!) that appeared when applying matching pursuit to a climatology time series

Overline of Remainder of Talk

- discuss basic ideas behind matching pursuit (MP)
- discuss application of MP to climatology time series that led to conundrum
- discuss tentative – but unsatisfying – explanation of conundrum
- lots of open questions, including what (if anything!) to do next

Matching Pursuit: I

- given a time series $\mathbf{X}$ of dimension N and a vector $\mathbf{d}$ of similar dimension satisfying
$$\|\mathbf{d}\|^2 = \langle \mathbf{d}, \mathbf{d} \rangle = \sum_{t=0}^{N-1} d_t^2 = 1,$$
consider approximating $\mathbf{X}$ using $\mathbf{d}$ in a linear model:
$$\mathbf{X} = \beta \mathbf{d} + \mathbf{e},$$
where β is unknown, and $\mathbf{e}$ is the error in the approximation
- can minimize $\|\mathbf{e}\|^2$ by setting β equal to $\langle \mathbf{X}, \mathbf{d} \rangle = \sum_{t=0}^{N-1} X_t d_t$
- approximation is $\mathbf{A} = \langle \mathbf{X}, \mathbf{d} \rangle \mathbf{d}$ & residuals are $\mathbf{R} = \mathbf{X} - \mathbf{A}$

Matching Pursuit: II

- in addition to additive decomposition $\mathbf{X} = \mathbf{A} + \mathbf{R}$, also have decomposition of sum of squares:

$$\|\mathbf{X}\|^2 = \|\mathbf{A}\|^2 + \|\mathbf{R}\|^2 = |\langle \mathbf{X}, \mathbf{d} \rangle|^2 + \|\mathbf{R}\|^2$$

- now consider a set of vectors $\mathcal{D}$, each $\mathbf{d}_k \in \mathcal{D}$ leading to

$$\mathbf{X} = \mathbf{A}_k + \mathbf{R}_k \;\text{ and }\; \|\mathbf{X}\|^2 = |\langle \mathbf{X}, \mathbf{d}_k \rangle|^2 + \|\mathbf{R}_k\|^2$$

- declare best approximation to be the one for which $\|\mathbf{R}_k\|^2$ is smallest, i.e., for which $|\langle \mathbf{X}, \mathbf{d}_k \rangle|$ is largest – call this approximation $\mathbf{A}^{(1)} = \langle \mathbf{X}, \mathbf{d}^{(1)} \rangle \mathbf{d}^{(1)}$, and let $\mathbf{R}^{(1)}$ be the corresponding vector of residuals so that

$$\mathbf{X} = \mathbf{A}^{(1)} + \mathbf{R}^{(1)} \;\text{ and }\; \|\mathbf{X}\|^2 = |\langle \mathbf{X}, \mathbf{d}^{(1)} \rangle|^2 + \|\mathbf{R}^{(1)}\|^2$$

Matching Pursuit: III

- first stage of MP leads to

$$\mathbf{X} = \mathbf{A}^{(1)} + \mathbf{R}^{(1)} \;\text{ and }\; \|\mathbf{X}\|^2 = |\langle \mathbf{X}, \mathbf{d}^{(1)} \rangle|^2 + \|\mathbf{R}^{(1)}\|^2$$

- second stage treats $\mathbf{R}^{(1)}$ as $\mathbf{X}$ was treated, leading to

$$\mathbf{R}^{(1)} = \mathbf{A}^{(2)} + \mathbf{R}^{(2)} \;\text{ and }\; \|\mathbf{R}^{(1)}\|^2 = |\langle \mathbf{R}^{(1)}, \mathbf{d}^{(2)} \rangle|^2 + \|\mathbf{R}^{(2)}\|^2$$

- stages $j = 3, 4 \ldots$ give us

$$\mathbf{R}^{(j-1)} = \mathbf{A}^{(j)} + \mathbf{R}^{(j)} \;\text{ and }\; \|\mathbf{R}^{(j-1)}\|^2 = |\langle \mathbf{R}^{(j-1)}, \mathbf{d}^{(j)} \rangle|^2 + \|\mathbf{R}^{(j)}\|^2$$

- defining $\mathbf{R}^{(0)} = \mathbf{X}$, after J such steps, have

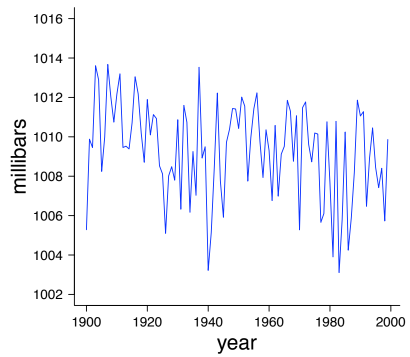
$$\mathbf{X} = \sum_{j=1}^{J} \mathbf{A}^{(j)} + \mathbf{R}^{(J)} \;\text{ and }\; \|\mathbf{X}\|^2 = \sum_{j=1}^{J} |\langle \mathbf{R}^{(j-1)}, \mathbf{d}^{(j)} \rangle|^2 + \|\mathbf{R}^{(J)}\|^2$$

Matching Pursuit: IV

- MP is ‘greedy’ in that, at each stage j, approximating vector is the one maximizing $|\langle \mathbf{R}^{(j-1)}, \mathbf{d}_k \rangle|$ amongst all $\mathbf{d}_k \in \mathcal{D}$
- under certain conditions on contents of $\mathcal{D}$, $\|\mathbf{R}^{(j)}\|^2$ must decrease and reach zero as j increases
- choice of vectors to place in $\mathcal{D}$ is obviously critical to quality of resulting approximation and is application dependent

North Pacific Index (NPI): I

- area-weighted sea level pressure over 30° N to 65° N & 160° E to 140° W & over November to March for each year from 1900 to 1999 (Trenberth & Paolino, 1980; Trenberth & Hurrell, 1994)



North Pacific Index (NPI): II

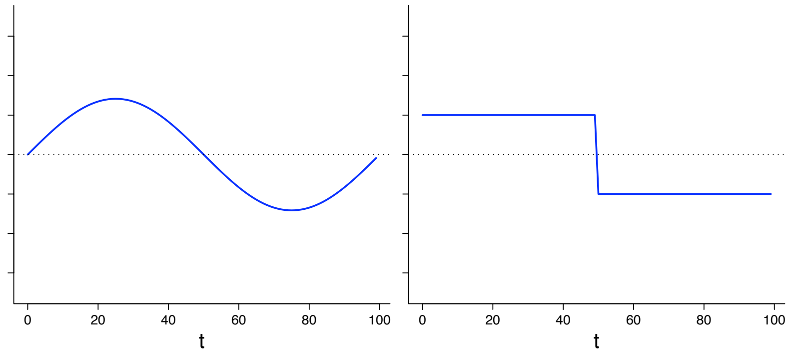
- Minobe (1999) postulated existence of penta- and bi-decadal oscillations in NPI that

  "...cannot be attributed to a single sinusoidal-wavelike variability ...";

  i.e., transitions between values above and below the long term mean of NPI occur much faster than sinusoidal variations can easily account for
- can (informally) evaluate Minobe's hypothesis by subjecting NPI to MP ($\mathbf{X}$ thus contains all $N = 100$ values of NPI, but after centering by subtracting off the sample mean)
- $\mathcal{D}$ consists of both sinusoidal and square wave oscillations, with frequencies dictated by Fourier frequencies $j/100$, $j = 1, 2, \ldots,$ 50 (periods are $100/j$ years), along with all possible phase shifts

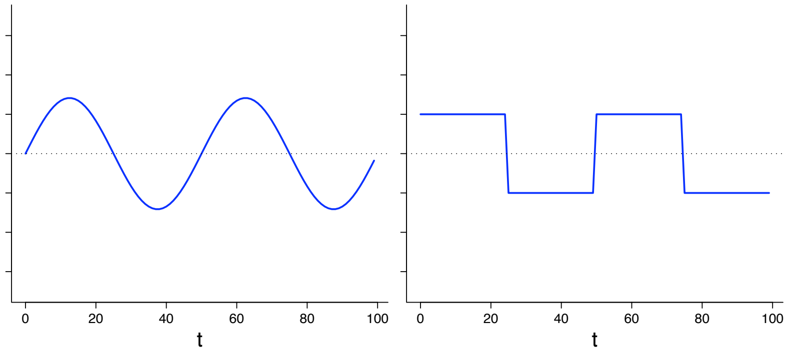
Examples of Vectors in $\mathcal{D}$

- period of 100 years, and one of 50 possible phase shifts



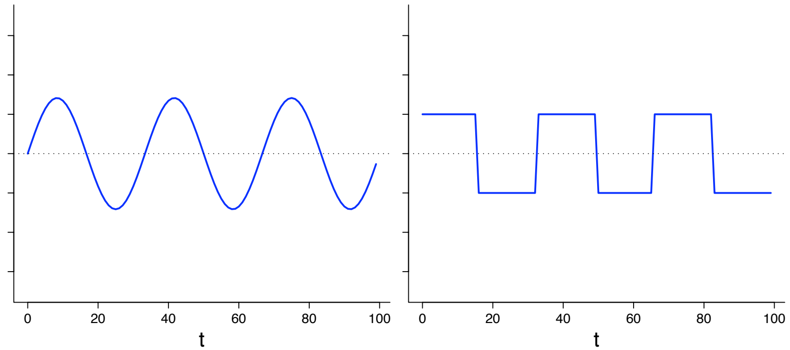
Examples of Vectors in $\mathcal{D}$

- period of 50 years, and one of 25 possible phase shifts



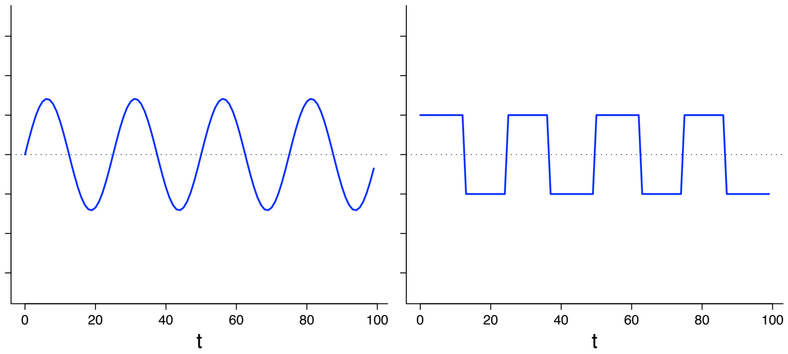
Examples of Vectors in $\mathcal{D}$

- period of 100/3 years (other phase shifts not shown)



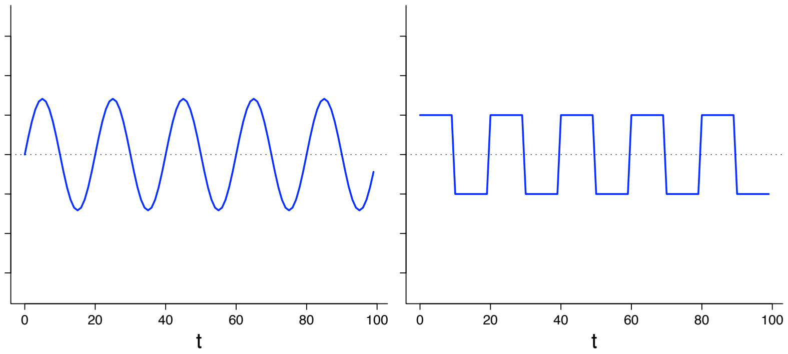
Examples of Vectors in $\mathcal{D}$

- period of 25 years (other phase shifts not shown)



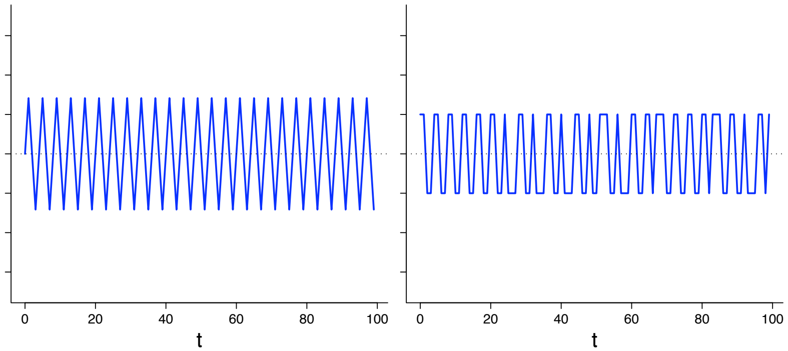
Examples of Vectors in $\mathcal{D}$

- period of 20 years (other phase shifts not shown)



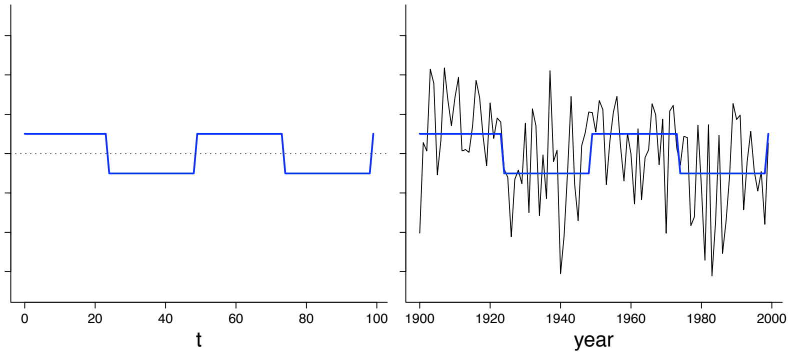
Examples of Vectors in $\mathcal{D}$

- period of 4 years (other phase shifts not shown)



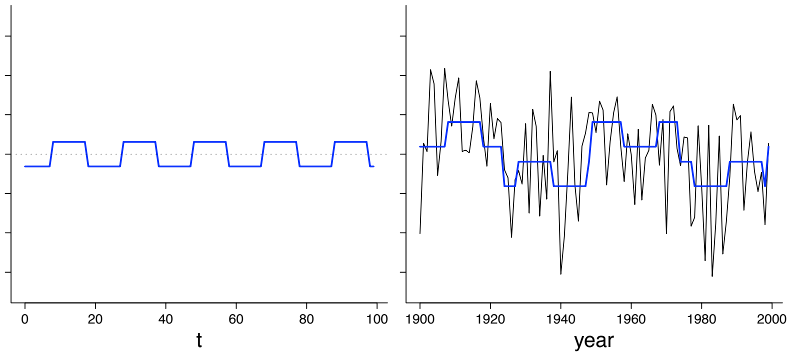
Matching Pursuit of NPI: I

- $j = 1$: square wave, 50 years; 17.4% of variance explained



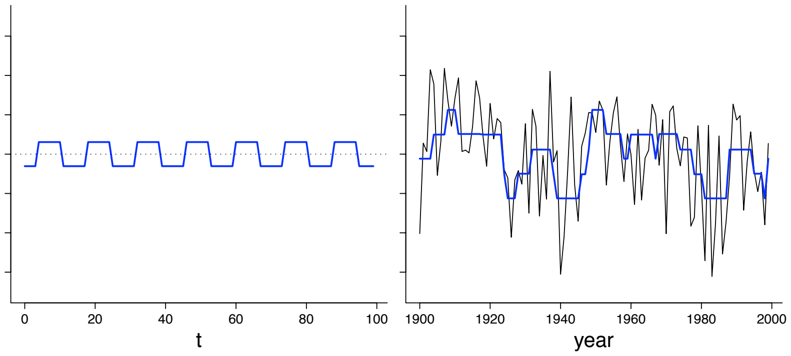
Matching Pursuit of NPI: II

- $j = 2$: square wave, 20 years; 24.1% of variance explained



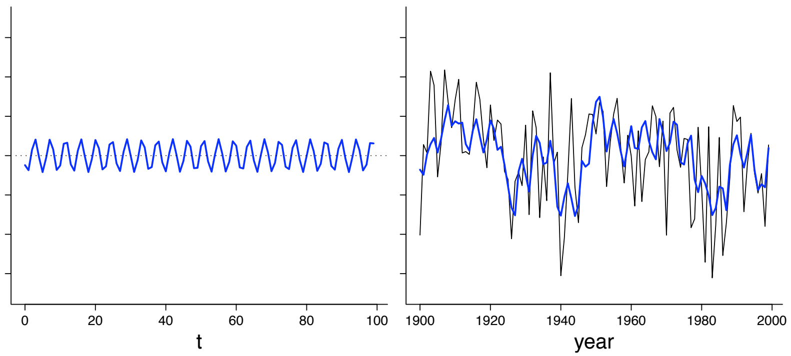
Matching Pursuit of NPI: III

- $j = 3$: square wave, 14 years; 30.6% of variance explained



Matching Pursuit of NPI: IV

- $j = 4$: sinusoid, 4.3 years; 36.4% of variance explained



Matching Pursuit of NPI: V

- MP lends credence to Minobe's hypothesis (penta- and bi-decadal oscillations with faster above/below transitions than sinusoids can explain)
- Q: what (if anything) can we say about statistical significance of patterns picked out by MP?

The Conundrum: I

- to address question of significance, need to consider what MP does under various null hypotheses
- simpliest such hypothesis is that $\mathbf{X}$ is Gaussian white noise (i.e., independent and identically distributed normal random variables) – note that $\mathbf{X}$ should have no discernable structure
- will take $\mathbf{X}$ to have zero mean and covariance/correlation matrix I_N (Nth order identity matrix)
- let K denote number of vectors $\mathbf{d}_k$ in set $\mathcal{D}$, and let $D = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$ so that kth element of $\mathbf{Y} \equiv D^T\mathbf{X}$ is $\langle \mathbf{X}, \mathbf{d}_k \rangle$
- $\mathbf{Y}$ is multivariate Gaussian with zero mean and with $\Sigma \equiv D^T D$ as its covariance/correlation matrix
- note that (j, k)th element of Σ is $\mathbf{d}_j^T \mathbf{d}_k$

The Conundrum: II

- first step of MP picks element of $\mathbf{Y}$ with largest magnitude, so distribution of this pick depends just on multivariate Gaussian correlation matrix Σ
- if $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2\}$, then

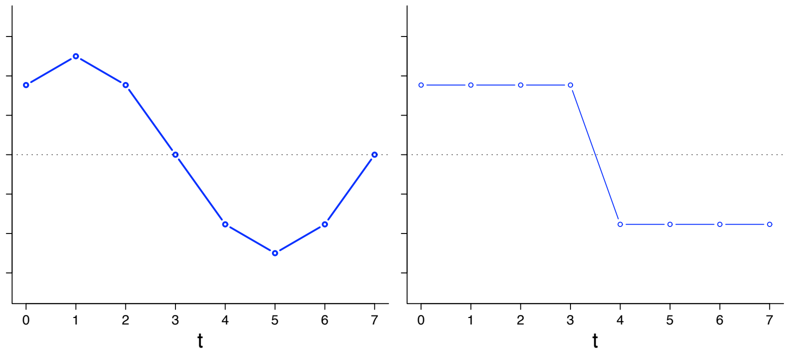
$$\Sigma = \begin{bmatrix} 1 & \mathbf{d}_1^T \mathbf{d}_2 \\ \mathbf{d}_2^T \mathbf{d}_1 & 1 \end{bmatrix},$$

  and, by symmetry, MP will pick $\mathbf{d}_1$ & $\mathbf{d}_2$ each 50% of the time, not matter what they are (e.g., a sinusoid & a square wave)
- if $\mathcal{D}$ has more then two elements, analysis becomes messy, but can resort to Monte Carlo experiments
- using same $\mathcal{D}$ as in NPI analysis (50% of vectors are sinsuoids, and 50% are square waves), MP picks sinusoids 15% of the time and square waves 85% of the time!?!

Slouching Towards an Explanation: I

- why does Gaussian white noise match up better with square waves than sinusoids?
- consider case $N = 8$ with $\mathcal{D}$ containing four sinusoids ($\mathbf{d}_1$, $\mathbf{d}_2$, $\mathbf{d}_3$ and $\mathbf{d}_4$) and four square waves ($\mathbf{d}_5$, $\mathbf{d}_6$, $\mathbf{d}_7$ and $\mathbf{d}_8$), all with a period of 8

Two of Eight Vectors in $\mathcal{D}$



Slouching Towards an Explanation: II

- Monte Carlo experiments indicate that MP picks a sinusoid 29% of the time and a square wave 71% of the time
- correlation matrix Σ in this case looks like the following:

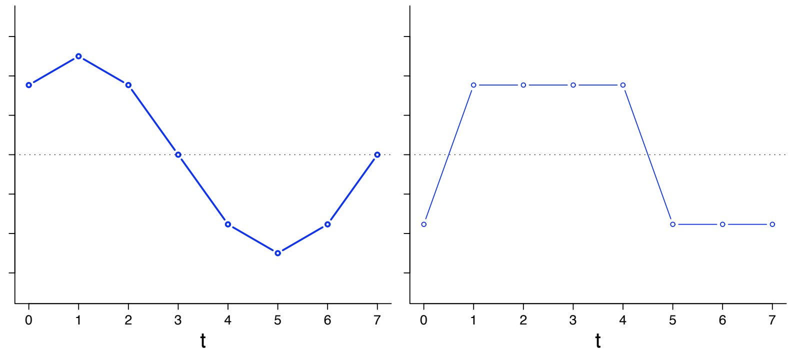
	$\mathbf{d}_1$	$\mathbf{d}_2$	$\mathbf{d}_3$	$\mathbf{d}_4$	$\mathbf{d}_5$	$\mathbf{d}_6$	$\mathbf{d}_7$	$\mathbf{d}_8$
$\mathbf{d}_1$	1.0							
$\mathbf{d}_2$	0.7	1.0						
$\mathbf{d}_3$	0.0	0.7	1.0					
$\mathbf{d}_4$	−0.7	0.0	0.7	1.0				
$\mathbf{d}_5$	0.9	0.9	0.4	−0.4	1.0			
$\mathbf{d}_6$	0.4	0.9	0.9	0.4	0.5	1.0		
$\mathbf{d}_7$	−0.4	0.4	0.9	0.9	0.0	0.5	1.0	
$\mathbf{d}_8$	−0.9	−0.4	0.4	0.9	−0.5	0.0	0.5	1.0

- sinusoids have more extreme cross-correlations than do square waves – is this part of the explanation?

Slouching Towards an Explanation: III

- consider another $\mathcal{D}$, this time with two sinusoids ($\mathbf{d}_1$ and $\mathbf{d}_2$) and two square waves ($\mathbf{d}_3$ and $\mathbf{d}_4$), all again with a period of 8

Two of Four Vectors in $\mathcal{D}$



Slouching Towards an Explanation: IV

- Monte Carlo experiments indicate that MP picks a sinusoid 48.5% of the time and a square wave 51.5% of the time
- correlation matrix Σ in this case looks like the following:

	$\mathbf{d}_1$	$\mathbf{d}_2$	$\mathbf{d}_3$	$\mathbf{d}_4$
$\mathbf{d}_1$	1.00			
$\mathbf{d}_2$	0.00	1.00		
$\mathbf{d}_3$	0.35	0.85	1.00	
$\mathbf{d}_4$	−0.35	0.85	0.50	1.00

- sinusoids now have zero cross-correlation, whereas square waves have a positive cross-correlation, yet square waves are still preferred (but just slightly so)
- cannot explain conundrum in terms of just cross-correlations

Hmmm . . .

References

- S. G. Mallat and Z. Zhang (1993), 'Matching Pursuits with Time-Frequency Dictionaries,' *IEEE Transactions on Signal Processing*, **41**, pp. 3397–3415
- S. Minobe (1999) 'Resonance in Bidecadal and Pentadecadal Climate Oscillations over the North Pacific: Role in Climate Regime Shifts,' *Geophysical Research Letters*, **26**, pp. 855–858
- D. B. Percival, J. E. Overland and H. O. Mofjeld (2002), 'Using Matching Pursuit to Assess Atmospheric Circulation Changes over the North Pacific,' unpublished manuscript available at

  `http://faculty.washington.edu/dbp/research.html`
- K. E. Trenberth and J. W. Hurrell (1994), 'Decadal Atmosphere–Ocean Variations in the Pacifc,' *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