Using Matching Pursuit to Assess Atmospheric Circulation Changes over the North Pacific

Donald B. Percival[1,3], James E. Overland[2] and Harold O. Mofjeld[2]

[1]Applied Physics Laboratory
Seattle, WA 98195–5640

[2]Pacific Marine Environmental Laboratory/NOAA
Seattle, WA 98115–6349

[3]corresponding author
Applied Physics Laboratory
Box 355640
University of Washington
Seattle, WA 98195–5640
dbp@apl.washington.edu
206.543.1368

For submission to *Journal of Climate*

May 30, 2002

Contribution No. 2476 from NOAA/Pacific Marine Environmental Laboratory

Abstract

Matching pursuit was originally formulated as a technique for identifying the time/frequency content of a time series whose spectral properties evolve over time. The basic idea was to construct a large 'dictionary' of explanatory vectors that are localized both in time and in frequency and then to analyze a time series by projecting it against the vectors in the dictionary. Matching pursuit can be adapted to explore other properties of a time series besides its time/frequency content. In this paper the technique is described in detail and then used to investigate the notion that North Pacific climate time series exhibit sudden changes in levels (regime-like shifts). In particular matching pursuit is used to explore a hypothesis by Minobe (1999) that there are penta- and bi-decadal oscillations in a North Pacific (NP) index of sea level pressures, with sharp transitions between high and low states that cannot be easily modeled by sinusoidal oscillations. When the NP index is analyzed using a dictionary containing square wave oscillations, sinusoidal oscillations and descriptors of isolated events (Haar wavelet and scaling vectors), matching pursuit selects square wave oscillations with periods and phases that agree closely with Minobe's analysis and penta-decadal hypothesis; however, the results call into question the existence of an independent bi-decadal oscillation. Matching pursuit was adjusted to handle a time series of winter air temperature measurements from Sitka, Alaska, which have been recorded since 1829, but with 22 missing values scattered throughout the 1800s. For this series, matching pursuit again picks out a square wave oscillation that is quite similar to the one identified for the NP index, with zero crossings during 1840–1, 1867–8, 1894–5, 1921–2, 1948–9 and 1975–6.

1 Introduction

Climatology data are often collected in the form of a time series, i.e., a sequence of observations collected over time. The goal of time series analysis is to extract information about features or patterns in the series that might further our understanding of the mechanisms generating the series. Several widely used methods for extracting pertinent information are similar in that they are special examples of subjecting the time series to an orthonormal transform. If we let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ denote a column vector containing N samples of a real-valued time series, then an orthonormal transform uses $\mathbf{X}$ to create a new vector $\mathbf{O}$ called the transform coefficients, also of length N. Mathematically we can write $\mathbf{O} = \mathcal{O}\mathbf{X}$, where $\mathcal{O}$ is a real-valued $N \times N$ matrix such that its transpose is its inverse; i.e., $\mathcal{O}^T\mathcal{O}$ is equal to the $N \times N$ identity matrix ($\mathcal{O}$ can also be complex-valued, in which case its Hermitian transpose is its inverse). Examples of orthonormal transforms include the orthonormal discrete Fourier transform (ODFT) and orthonormal discrete wavelet transforms (ODWTs). Using the ODFT, we can detect sinusoidal patterns in a time series. An ODWT yields coefficients that reflect changes in averages of the series over various scales, thus giving us the ability to detect important variations in a time series at different scales.

Orthonormal transforms essentially give us a new representation for a time series that is fully equivalent to the original series. Two important implications of this equivalence are the following. First, any orthonormal transform preserves the 'energy' in the time series; i.e.,

$$\|\mathbf{X}\|^2 \equiv \sum_{t=0}^{N-1} |X_t|^2 = \|\mathbf{O}\|^2.$$

Thus, while the square of the original series indicates how the energy is distributed across time, the squared modulus of the transform coefficients indicates how the energy is distributed in the transform domain. For example, the ODFT tells us how the energy is distributed across frequencies, while the ODWT gives us a scale-based distribution of energy. Second, we can reconstruct our original time series from $\mathbf{O}$ using a linear combination of column vectors whose elements are the complex conjugates of the rows of $\mathcal{O}$. If we let $\mathcal{O}_{j,\bullet}$ represent the jth such vector and if O_j is the

jth element of $\mathbf{O}$, then we have

$$\mathbf{X} = \sum_{j=0}^{N-1} O_j \mathcal{O}_{j,\bullet}.$$

Since $\|\mathcal{O}_{j,\bullet}\|^2 = 1$ is part of the requirement that $\mathcal{O}$ be an orthonormal transform, the most important components in the synthesis (reconstruction) of $\mathbf{X}$ are those having large magnitudes $|O_j|$. Suppose that we pick a small number of vectors $\mathcal{O}_{j,\bullet}$ associated with large $|O_j|$. If we let $\mathcal{J}$ represent the indices of these vectors, then we can write

$$\mathbf{X} = \sum_{j\in\mathcal{J}} O_j \mathcal{O}_{j,\bullet} + \sum_{j\notin\mathcal{J}} O_j \mathcal{O}_{j,\bullet} \equiv \widehat{\mathbf{X}} + \mathbf{R}, \tag{1}$$

where the first sum above yields an approximation $\widehat{\mathbf{X}}$ to $\mathbf{X}$ that captures its important features, while $\mathbf{R}$ is a vector of residuals. Energy is preserved in this decomposition because

$$\|\mathbf{X}\|^2 = \|\widehat{\mathbf{X}}\|^2 + \|\mathbf{R}\|^2 = \sum_{j\in\mathcal{J}} |O_j|^2 + \|\mathbf{R}\|^2. \tag{2}$$

Orthonormal transforms can be very useful, but there is an important limitation to their use. Once a particular transform has been picked, we will only extract features in a time series that can be succinctly represented by that transform. Thus the ODFT is good at picking out sinusoidal variations that persist across the entire time series, but this transform cannot readily pick out important transient events (e.g., a burst of energy); conversely, the ODWT can succinctly capture certain transient events, but is inferior to the ODFT in pulling out sinusoidal variations. In analyzing a time series for which there is no compelling reason to prefer a particular orthonormal transform *a priori*, it is thus undesirable to stick with just a single transform.

In an attempt to circumvent the limitations imposed by using but one transform, Mallat and Zhang (1993) and Davis *et al.* (1994) proposed an approach called matching pursuit. Conceptually matching pursuit allows us to build up a representation for a time series using vectors from a large collection of different orthonormal transforms. By not restricting ourselves to one particular transform, we can build up a representation that is free of the constraints imposed by a single transform. Although matching pursuit in general does not lead to an orthonormal transform, it does have equivalents to the additive and energy decompositions of Equations (1) and (2). We can thus achieve many of the same goals of analysis using matching pursuit as we can using an orthonormal transform.

The goal of this paper is to make matching pursuit more widely known by presenting a self-contained introduction, along with substantive examples of how it can be applied to qualitatively evaluate the nature of climatology time series. After the introduction in Section 2, we apply matching pursuit in Section 3 to the North Pacific (NP) index of average winter sea level pressures for the Aleutian low. Minobe (1999) has postulated the existence of penta- and bi-decadal oscillations in the NP index that 'cannot be attributed to a single sinusoidal-wavelike variability'; i.e., transitions between values above and below the long term mean of the NP index occur much faster than sinusoidal variations can easily account for. To evaluate Minobe's hypothesis, we formulate a matching pursuit consisting of both square wave and sinusoidal oscillations. As a second example, we consider in Section 4 a series of annual air temperatures from Sitka, Alaska. This series has a number of missing observations, but we show that matching pursuit can be adjusted to handle a gappy time series. We make some concluding remarks in Section 5.

2 The Matching Pursuit Algorithm

As formulated in the seminal papers by Mallat and Zhang (1993) and Davis *et al.* (1994), the idea behind matching pursuit is to approximate a function using a linear combination of a small number of more primitive functions. In their work, each of the primitive functions was associated with a particular time and a particular frequency. This setup allows an investigator to obtain an adaptive time/frequency decomposition. The end result is an approximation to the function consisting of a sum of primitive functions whose time/frequency characteristics match those contained in the function. The primitive functions are drawn from a large and redundant collection called a dictionary. Matching pursuit can be easily generalized for purposes other than time/frequency analysis. Here we formulate this approach in the quite general context of approximating a time series using a large collection of vectors (see also Percival and Walden 2000, Sections 6.8 and 6.9).

As before, let $\mathbf{X}$ be an N dimensional column vector $[X_0, X_1, \ldots, X_{N-1}]^T$ containing a real-valued time series. We wish to expand $\mathbf{X}$ into linear combinations of vectors from a dictionary $\mathcal{D}$ containing K vectors; i.e., letting $\mathcal{K}$ denote the set of integers $0, 1, \ldots, K-1$, we have $\mathcal{D} \equiv \{\mathbf{D}_k : k \in \mathcal{K}\}$, where $\mathbf{D}_k = [D_{k,0}, D_{k,1}, \ldots, D_{k,N-1}]^T$ is one of K dictionary vectors (each of dimension N). We allow the elements of $\mathbf{D}_k$ to be either real- or complex-valued, but, if $\mathbf{D}_k$ has complex-valued elements, we assume that its complex conjugate $\mathbf{D}_k^* \equiv [D_{k,0}^*, D_{k,1}^*, \ldots, D_{k,N-1}^*]^T$ is such that the inner product $\langle \mathbf{D}_k, \mathbf{D}_k^* \rangle$ between $\mathbf{D}_k$ and $\mathbf{D}_k^*$ is zero; here

$$\langle \mathbf{U}, \mathbf{V} \rangle \equiv \sum_{t=0}^{N-1} U_t V_t^*$$

for any two complex- or real-valued vectors $\mathbf{U} \equiv [U_0, U_1, \ldots, U_{N-1}]^T$ and $\mathbf{V} \equiv [V_0, V_1, \ldots, V_{N-1}]^T$ (the restriction $\langle \mathbf{D}_k, \mathbf{D}_k^* \rangle = 0$ is satisfied by, e.g., the complex-valued vectors in an ODFT). Each dictionary vector is assumed to have a norm of unity; i.e, $\|\mathbf{D}_k\|^2 = 1$. In order to prove certain convergence properties of matching pursuit (not of interest to us in this paper), it is usually assumed that a subset of the vectors in $\mathcal{D}$ forms a basis for the space $\mathbb{R}^N$ of all possible N dimensional real-valued vectors, of which $\mathbf{X}$ is an element; i.e., $K \geq N$, and there is a set of vectors in $\mathcal{D}$ such that we can express any $\mathbf{X}$ as a linear combination of these vectors. With this assumption, the smallest possible dictionary $\mathcal{D}$ is thus a set of N vectors forming a basis for $\mathbb{R}^N$, but the idea behind matching pursuit is that K should be much larger than N and that we use the redundancy in $\mathcal{D}$ to help extract information of potential interest to us about a time series.

Matching pursuit is an algorithm that successively approximates $\mathbf{X}$ with orthogonal projections onto vectors in $\mathcal{D}$. For any $\mathbf{D}_k \in \mathcal{D}$, we project $\mathbf{X}$ onto this vector to form a real-valued approximation to $\mathbf{X}$. If the vector $\mathbf{D}_k$ is real-valued, the approximation is given by $\mathbf{A}_k \equiv C_k \mathbf{D}_k$, where $C_k \equiv \langle \mathbf{X}, \mathbf{D}_k \rangle$; on the other hand, if $\mathbf{D}_k$ has complex-valued elements, we let $\mathbf{A}_k \equiv C_k \mathbf{D}_k + C_k^* \mathbf{D}_k^*$. We can handle both cases by defining the approximation to be $\mathbf{A}_k \equiv C_k \mathbf{D}_k + C_k^* \mathbf{D}_k^* \mathcal{I}_k$, where $\mathcal{I}_k$ is zero if $\mathbf{D}_k$ is real-valued and is unity otherwise. We then construct the real-valued residual vector $\mathbf{R}_k \equiv \mathbf{X} - \mathbf{A}_k$ so that $\mathbf{X} = \mathbf{A}_k + \mathbf{R}_k$. If we recall our assumption that $\langle \mathbf{D}_k, \mathbf{D}_k^* \rangle = 0$ when $\mathbf{D}_k$ is complex-valued, then we have, for any k,

$$\begin{aligned}
\langle \mathbf{A}_k, \mathbf{R}_k \rangle &= \langle C_k \mathbf{D}_k + C_k^* \mathbf{D}_k^* \mathcal{I}_k, \mathbf{X} - C_k \mathbf{D}_k - C_k^* \mathbf{D}_k^* \mathcal{I}_k \rangle \\
&= \langle C_k \mathbf{D}_k, \mathbf{X} - C_k \mathbf{D}_k - C_k^* \mathbf{D}_k^* \mathcal{I}_k \rangle + \langle C_k^* \mathbf{D}_k^* \mathcal{I}_k, \mathbf{X} - C_k \mathbf{D}_k - C_k^* \mathbf{D}_k^* \mathcal{I}_k \rangle \\
&= C_k \langle \mathbf{D}_k, \mathbf{X} - C_k \mathbf{D}_k \rangle + C_k^* \mathcal{I}_k \langle \mathbf{D}_k^*, \mathbf{X} - C_k^* \mathbf{D}_k^* \rangle \\
&= C_k \Big(\langle \mathbf{D}_k, \mathbf{X} \rangle - \langle \mathbf{D}_k, C_k \mathbf{D}_k \rangle \Big) + C_k^* \mathcal{I}_k \Big(\langle \mathbf{D}_k^*, \mathbf{X} \rangle - \langle \mathbf{D}_k^*, C_k^* \mathbf{D}_k^* \rangle \Big) \\
&= C_k \Big(C_k^* - C_k^* \langle \mathbf{D}_k, \mathbf{D}_k \rangle \Big) + C_k^* \mathcal{I}_k \Big(C_k - C_k \langle \mathbf{D}_k^*, \mathbf{D}_k^* \rangle \Big) = 0,
\end{aligned}$$

and hence

$$\|\mathbf{X}\|^2 = \|\mathbf{A}_k\|^2 + \|\mathbf{R}_k\|^2 = (1 + \mathcal{I}_k)\, |C_k|^2 + \|\mathbf{R}_k\|^2. \tag{3}$$

In order to make the energy in the residuals as small as possible, we choose $k^{(1)} \in \mathcal{K}$ such that $(1 + \mathcal{I}_{k^{(1)}})\, |C_{k^{(1)}}|^2 = \max_{k \in \mathcal{K}} (1 + \mathcal{I}_k)\, |C_k|^2$. The corresponding approximation is given by

$$\mathbf{A}^{(1)} \equiv C_{k^{(1)}} \mathbf{D}_{k^{(1)}} + C_{k^{(1)}}^* \mathbf{D}_{k^{(1)}}^* \mathcal{I}_{k^{(1)}}, \tag{4}$$

and the residual vector is $\mathbf{R}^{(1)} \equiv \mathbf{X} - \mathbf{A}^{(1)}$ so that

$$\mathbf{X} = \mathbf{A}^{(1)} + \mathbf{R}^{(1)}. \tag{5}$$

Equation (3) now becomes

$$\|\mathbf{X}\|^2 = \|\mathbf{A}^{(1)}\|^2 + \|\mathbf{R}^{(1)}\|^2 = (1 + \mathcal{I}_{k^{(1)}})\, |C_{k^{(1)}}|^2 + \|\mathbf{R}^{(1)}\|^2. \tag{6}$$

The second step of matching pursuit is to decompose $\mathbf{R}^{(1)}$ by projecting it onto the vector in $\mathcal{D}$ that best matches it; i.e., in analogy to Equations (4) and (5), we have

$$\mathbf{A}^{(2)} \equiv C_{k^{(2)}} \mathbf{D}_{k^{(2)}} + C_{k^{(2)}}^* \mathbf{D}_{k^{(2)}}^* \mathcal{I}_{k^{(2)}} \ \text{ with } C_{k^{(2)}} \equiv \langle \mathbf{R}^{(1)}, \mathbf{D}_{k^{(2)}} \rangle$$

and

$$\mathbf{R}^{(1)} = \mathbf{A}^{(2)} + \mathbf{R}^{(2)}, \tag{7}$$

where $k^{(2)}$ is chosen such that $(1+\mathcal{I}_{k^{(2)}})|C_{k^{(2)}}|^2 = \max_{k\in\mathcal{K}}(1+\mathcal{I}_k)\left|\langle \mathbf{R}^{(1)}, \mathbf{D}_k\rangle\right|^2$. If we combine Equations (5) and (7), we obtain

$$\mathbf{X} = \sum_{n=1}^{2} \mathbf{A}^{(n)} + \mathbf{R}^{(2)}.$$

The analog of Equation (6) is

$$\|\mathbf{R}^{(1)}\|^2 = \|\mathbf{A}^{(2)}\|^2 + \|\mathbf{R}^{(2)}\|^2 = (1+\mathcal{I}_{k^{(2)}})\,|C_{k^{(2)}}|^2 + \|\mathbf{R}^{(2)}\|^2,$$

which, upon combining with (6), yields

$$\|\mathbf{X}\|^2 = \sum_{n=1}^{2} \|\mathbf{A}^{(n)}\|^2 + \|\mathbf{R}^{(2)}\|^2.$$

We can obviously continue to iterate on this procedure. After m steps, we will have

$$\mathbf{X} = \sum_{n=1}^{m} \mathbf{A}^{(n)} + \mathbf{R}^{(m)} \equiv \widehat{\mathbf{X}}^{(m)} + \mathbf{R}^{(m)},$$

where $\widehat{\mathbf{X}}^{(m)}$ is the mth order approximation to $\mathbf{X}$. The above is the matching pursuit analog of Equation (1) for orthonormal transforms. Energy is conserved because

$$\|\mathbf{X}\|^2 = \sum_{n=1}^{m} \|\mathbf{A}^{(n)}\|^2 + \|\mathbf{R}^{(m)}\|^2 = \sum_{n=1}^{m} (1+\mathcal{I}_{k^{(n)}})\,|C_{k^{(n)}}|^2 + \|\mathbf{R}^{(m)}\|^2$$

which is analogous to Equation (2). As we increase m, the residual energy $\|\mathbf{R}^{(m)}\|^2$ cannot increase. Mallat and Zhang (1993) and Mallat (1999) discuss conditions on $\mathcal{D}$ under which $\|\mathbf{R}^{(m)}\|^2$ decreases to zero as m increases.

3 Analysis of the NP Index Using Matching Pursuit

As a first example of the use of matching pursuit to study climatology time series, let us consider the North Pacific (NP) index, which we here define to be the winter (November to March) averaged sea level pressure time series for the Aleutian low (Trenberth and Paolino 1980, Overland *et al.* 1999). This series consists of one value for each year from 1900 to 1999 ($t = 0$ to 99, respectively) and hence has a sample size of $N = 100$. The series is shown in each of the right-hand plots in Figure 2 (and subsequent similar figures) as the thin jagged curve. Recently Minobe (1999) found evidence for climate regime shifts in this index, i.e., transitions from one climate state (represented by, e.g., high values of the NP index) to another (e.g., low values) occurring in a relatively short time interval (this result corroborated several previous studies – see Minobe (1999) for references to this previous work). The following quote from his article emphasizes that the transitions are evidently sharp.

> 'Although the regime shifts in the present century are characterized by a penta-decadal timescale, rapid transitions from one regime to another cannot be attributed to a single sinusoidal-wavelike variability. The rapid-transition nature of 20th century regime shifts suggests that the penta-decadal variability is characterized by a non-sinusoidal variation such as a rectangular wave . . . .'

In addition to a dominant penta-decadal oscillation (characterized as having a period of 30–80 years), Minobe also identified a secondary bi-decadal oscillation (10–30 years) that seemed to have transitions in phase with the penta-decadal transitions. The penta- and bi-decadal climate oscillations to which Minobe alludes in his paper are not taken to be strictly periodic with periods of fifty and twenty years, respectively. For example, Minobe found that there were three transitions during the twentieth century associated with the penta-decadal oscillation. These were centered at 1924–5, 1947–8 and 1976–7, which are not spaced exactly twenty five years apart as strict periodicity would dictate. In addition, at each of these transitions, there was also a transition of the same polarity in the bi-decadal oscillation, an arrangement that could not happen if these oscillations were strictly periodic.

Let us now demonstrate how matching pursuit can be used to assess Minobe's hypotheses about the NP index. To simplify our presentation, we analyze the NP index after it has been recentered by subtracting off its sample mean; i.e., we apply matching pursuit to $X_t \equiv \widetilde{X}_t - \overline{X}$ rather than to the original series $\widetilde{X}_t$, where $\overline{X} \equiv \sum_t \widetilde{X}_t / N$. Without this recentering, the first vector that is picked out by all the matching pursuits considered below is one whose elements are all equal to $1/\sqrt{N}$. When this happens, the approximation $\mathbf{A}^{(1)}$ is a vector whose elements are all $\overline{X}$, and the residual vector $\mathbf{R}^{(1)}$ is the recentered time series. Recentering the series essentially bypasses a step that is common to all the matching pursuits for the NP index and allows us to concentrate on the more interesting parts of the analysis.

After recentering the series, we first consider a matching pursuit using a dictionary consisting of just the $K = N$ vectors from an ODFT. The tth element of the kth such vector is given by $\exp(-i2\pi tk/N)/\sqrt{N}$, where $0 \leq k, t \leq N - 1$. The top row of Figure 1 shows this complex-valued vector for the case $k = 5$ and $N = 100$, for which the real and imaginary components are a cosine and a negative sine, both with a period of twenty. With this dictionary, we approximate the NP index using sinusoids proportional to $\cos(2\pi kt/N + \phi_k)$, $k = 1, \ldots, N/2$, i.e., with periods given by $100/k$ and with corresponding phases ϕ_k set to provide the best possible approximations.

Figure 2 shows results from the first five steps of matching pursuit with an ODFT dictionary. The series that are plotted in the left-hand column are, from top to bottom, proportional to the real-valued vector $\mathbf{A}^{(m)}$ that best approximates the residual vector $\mathbf{R}^{(m-1)}$ at stages $m = 1, \ldots, 5$, where we let $\mathbf{R}^{(0)} \equiv \mathbf{X}$. The thick curves in the the right-hand column are the corresponding mth order approximations $\widehat{\mathbf{X}}^{(m)}$ to $\mathbf{X}$, while the thin curves show the recentered NP index itself. The percentage of the variance that is explained by $\widehat{\mathbf{X}}^{(m)}$ is also listed (by definition, this is $(\|\mathbf{X}\|^2 - \|\mathbf{R}^{(m)}\|^2)/\|\mathbf{X}\|^2 \times 100\%$). The first approximating vector $\mathbf{A}^{(1)}$ (top row) is a sinusoid with a period of 50 years and explains just 15.6% of the total variance. The second approximating vector has a period of 20 years, and together the first two vectors explain 21.7% of the variance. These vectors evidently correspond to the penta- and bi-decadal oscillations picked out in Minobe's analysis. The next three selected vectors have periods of 4.3, 14.3 and 3.2 years, which are not easily related to his analysis.

Let us now augment the matching pursuit dictionary by including vectors corresponding to strictly periodic square wave oscillations with half periods p ranging from one up to $N/2$ years in combination with all possible shifts s (i.e., placements of the first transition from a negative to a positive value). Mathematically, we can describe these vectors by first defining, for any positive integer p and any integer t,

$$D^{(\mathrm{sw})}_{p,0,t} \equiv \begin{cases} 1/\sqrt{N} & \text{when } t = 0, \ldots, p-1; \\ -1/\sqrt{N} & \text{when } t = p, \ldots, 2p-1; \text{ and} \\ D^{(\mathrm{sw})}_{p,0,t \bmod 2p} & \text{otherwise.} \end{cases}$$

In the above, we interpret '$t \mod 2p$' to be equal to $t+2pn$, where n is the unique integer such that $0 \le t+2pn \le 2p-1$. The tth element $D^{(\mathrm{sw})}_{p,s,t}$, $t = 0, \ldots, N-1$, of a vector containing a square wave with half period p and shift s is defined to be $D^{(\mathrm{sw})}_{p,0,t-s}$. Due to symmetries in the square waves, we can restrict the shifts we need to consider to $s = 0, \ldots, p-1$. The second row of Figure 1 shows a square wave vector for the case $2p = 20$, $s = 0$ and $N = 100$. A matching pursuit with a dictionary consisting of both sinusoid and square wave vectors gives us a chance to see if periodic square waves are to be preferred over sinusoids in approximating the NP index.

Figure 3 shows the first five steps of matching pursuit with the combined ODFT and square wave dictionary (the layout of this and subsequent figures parallels that of Figure 2). Interestingly enough, the first three vectors that matching pursuit picks out are now square waves (the fourth and fifth are from the ODFT). The first two have periods of, respectively, fifty and twenty years, evidently corresponding to Minobe's penta- and bi-decadal oscillations (the upper right-hand plot in Figure 3 closely resembles the penta-decadal oscillation shown in Figure 1 of Minobe's paper). The first three zero crossings of the penta-decadal oscillation occur at 1924–5, 1949–50 and 1974–5, as compared to Minobe's assessment of, respectively, 1924–5, 1947–8 and 1976–7. While matching pursuit did pick out a bi-decadal oscillation as its second choice for an explanatory vector, this oscillation differs qualitatively from Minobe's in that none of its zero crossing align well with those of the penta-decadal oscillation (due undoubtedly to the strictly periodic nature of the square waves used here). Nonetheless, the fact that matching pursuit picked out square waves preferentially over sinusoids offers some support for the hypothesis that transitions from high to low values (and vice versa) in the NP index occur more abruptly than can be accounted for by sinusoids. In addition, the percentage of explained variance is somewhat higher at each stage using this combined dictionary rather than just the ODFT.

Let us now augment our matching pursuit dictionary even further by including functions describing single cycles of a square wave oscillation. The rationale for doing this is to explore if the NP index might be better described in terms of single cycles of a square wave that are not necessarily coupled together tightly as in the case of the strictly periodic square wave oscillations. Vectors containing a single cycle square wave are identical to those used in a discretized version of the Haar wavelet transform. To define these mathematically, for all positive integers p and all integers t, let

$$D^{(\mathrm{Hw})}_{p,0,t} \equiv \begin{cases} 1/\sqrt{(2p)} & \text{when } t = 0, \ldots, p-1; \\ -1/\sqrt{(2p)} & \text{when } t = p, \ldots, 2p-1; \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

The tth element of a vector with half-period p and shift s is given by

$$D^{(\mathrm{Hw})}_{p,s,t} \equiv \begin{cases} D^{(\mathrm{Hw})}_{p,0,t-s}\sqrt{[2p/\min\{N, 2p-|s|\}]} & \text{when } s = -(p-1), \ldots, -1; \\ D^{(\mathrm{Hw})}_{p,0,t-s} & \text{when } s = 0, \ldots, N-2p; \text{ and} \\ D^{(\mathrm{Hw})}_{p,0,t-s}\sqrt{[2p/\min\{N, N-s\}]} & \text{when } s = N-2p+1, \ldots, N-p-1. \end{cases}$$

The dictionary includes vectors with $p = 1, \ldots, N-1$ and corresponding shifts $s = -p+1, \ldots, N-p-1$. The third row of Figure 1 shows the case $2p = 20$, $s = 40$ and $N = 100$.

Figure 4 shows the first five steps of matching pursuit with the combined ODFT, square wave and Haar wavelet dictionary. The first vector picked out is the same as before (a square wave oscillation with a fifty year period), but the second is no longer related to a bi-decadal square wave oscillation. It is now a Haar wavelet vector covering the $2p = 12$ years from 1983 to 1994, with a positive going transition in 1988–9. The third and fourth selected vectors are also Haar wavelet vectors. The third one covers the $2p = 6$ years from 1937 to 1942, with a negative going transition in 1939–40. The fourth formally has a period of 18 years, but is actually capturing a contrast between the years 1900–2 and 1903–11. The fifth vector is from the ODFT with a period of 3.2 years (reflecting high frequency variations). Only the fourth vector is somewhat related to a bi-decadal oscillation.

Finally, let us also allow for the possibility of sudden shifts in the NP index that are not then associated with a symmetric shift in the opposite direction. To do so, we need to include vectors that constitute just the first half of a single cycle of a square wave oscillation. A vector of this form is identical to a discretized version of a Haar scaling function. To define the Haar scaling vectors mathematically, let

$$D^{(\mathrm{Hs})}_{p,0,t} \equiv \begin{cases} 1/\sqrt{p} & \text{when } t = 0, \ldots, p-1; \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

where p is a positive integer, while t can be any integer. The tth element of a Haar scaling vector with extent (or scale) p and shift s is given by $D^{(\mathrm{Hs})}_{p,0,t-s}$. The dictionary includes vectors with $p = 1, \ldots, N-1$ and corresponding shifts $s = 0, \ldots, N-p$. The final row of Figure 1 shows a Haar scaling vector for the case $p = 10$, $s = 40$ and $N = 100$.

When we use a dictionary consisting of ODFT, square wave oscillation, Haar wavelet and Haar scaling vectors, the first five vectors that matching pursuit picked out are the same as displayed in Figure 4. The fact that none of the first five selected vectors changed upon inclusion of the Haar scaling vectors in the dictionary suggests that there are no important shifts that have not already been captured by either the square wave oscillations or the Haar wavelets.

To summarize our findings, the fact that matching pursuit consistently picks out a penta-decadal square wave oscillation as the single best descriptor of the NP index when presented with many other choices lends some credence to the main conclusion in Minobe (1999). The secondary conclusion of a bi-decadal square wave oscillation is supported when we use a dictionary consisting of just ODFT and square wave oscillations, but becomes questionable when we augment the dictionary to include Haar wavelets. In Percival *et al.* (2002), we explore a statistical model for the NP index consisting of the penta-decadal square wave oscillation observed in the presence of Gaussian white noise and find this simple model to be competitive with purely stochastic short- and long-memory models considered previously (Percival *et al.*, 2001). In particular, an examination of the residuals from this simple model does not indicate the need to expand the model to include any other terms, which further suggests that a bi-decadal oscillation is questionable.

4 Analysis of the Sitka Air Temperature Using Matching Pursuit

As a second example of the use of matching pursuit, let us consider a time series of winter air temperatures from Sitka, Alaska. This series consists of $N' = 146$ data values collected over an $N = 168$ year period (1829–1996), so there are 22 years for which there were no observed values. The series is shown in each of the right-hand plots in Figure 5 as the thin jagged broken curves. This example illustrates that matching pursuit can be easily modified to handle time series with missing observations, a situation that occurs quite commonly in climatology data.

Let $\widetilde{X}_{t_n}$, $n = 0, \ldots, N'-1$, represent the N' measured air temperatures, where $t_n + 1829$ is the year in which the nth measurement was taken (thus $t_0 = 0$ and $t_{145} = 167$). As before, we form the recentered series $X_{t_n} \equiv \widetilde{X}_{t_n} - \overline{X}$, where $\overline{X} \equiv \sum_n \widetilde{X}_{t_n}/N'$. We define our vector of observations to be $\mathbf{X} = [X_{t_0}, X_{t_1}, \ldots, X_{t_{N'-1}}]^T$.

We construct a matching pursuit dictionary that is analogous to the last one that we considered in our discussion of the NP index, i.e., one containing vectors related to the ODFT, square wave, Haar wavelet and Haar scaling vectors defined previously. Let us first consider how we can adjust a vector describing a square wave oscillation to handle a time series with missing observations. For a given half-period p and shift s, we start by forming a vector appropriate for a time series of length N, namely,

$$[D^{(\mathrm{sw})}_{p,s,0}, D^{(\mathrm{sw})}_{p,s,1}, \ldots, D^{(\mathrm{sw})}_{p,s,N-1}]^T.$$

We then shorten this vector to length N' by retaining just those elements corresponding to the years for which we have actual measurements, after which we multiply each retained element by $\sqrt{(N/N')}$ so that the resulting vector has a norm of unity; i.e., we form

$$\frac{\sqrt{N}}{\sqrt{N'}}[D^{(\mathrm{sw})}_{p,s,t_0}, D^{(\mathrm{sw})}_{p,s,t_1}, \ldots, D^{(\mathrm{sw})}_{p,s,t_{N'-1}}]^T,$$

which will then become one of the vectors in our dictionary. We can adjust the Haar wavelet and Haar scaling vectors in a similar manner, but in general we will need to use a factor different from $\sqrt{(N/N')}$ to achieve a unit norm (we also have to be aware that, after shortening, some vectors will contain nothing but zeros and hence cannot be included in the dictionary).

Adjustment of the complex-valued ODFT vectors to handle missing observations is somewhat more involved. In the case of equally spaced observations, at the mth stage of matching pursuit, we use the two ODFT vectors associated with frequency $0 < f_k \equiv k/N < 1/2$ to form an approximation to $\mathbf{R}^{(m)}$ that takes the form

$$A_{k,t} \equiv c_k \cos(2\pi f_k t + \phi_k) = a_k \cos(2\pi f_k t) + b_k \sin(2\pi f_k t), \quad t = 0, \ldots, N-1.$$

The values for a_k and b_k are determined by the complex-valued ODFT coefficient $C_k \equiv \langle \mathbf{X}, \mathbf{D}_k \rangle$, but an equivalent interpretation for them is that they are the values we would obtain from a least squares fit of $\mathbf{R}^{(m)}$ to the model $A_{k,t} + \epsilon_t$, where $\{\epsilon_t\}$ is a sequence of uncorrelated errors. We can make use of this least squares interpretation to construct ODFT-like dictionary vectors for a time series with missing observations in the following manner. Let $R^{(m)}_{t_n}$ be the nth element of the vector $\mathbf{R}^{(m)}$. For each frequency f_k, we fit the model

$$R^{(m)}_{t_n} = a_k \cos(2\pi f_k t_n) + b_k \sin(2\pi f_k t_n) + \epsilon_n, \quad n = 0, \ldots, N' - 1,$$

via least squares to obtain the estimated values

$$\begin{bmatrix} \hat{a}_k \\ \hat{b}_k \end{bmatrix} \equiv \begin{bmatrix} \sum_n \cos^2(2\pi f_k t_n) & \sum_n \cos(2\pi f_k t_n) \sin(2\pi f_k t_n) \\ \sum_n \cos(2\pi f_k t_n) \sin(2\pi f_k t_n) & \sum_n \sin^2(2\pi f_k t_n) \end{bmatrix}^{-1} \begin{bmatrix} \sum_n \cos(2\pi f_k t_n) R^{(m)}_{t_n} \\ \sum_n \sin(2\pi f_k t_n) R^{(m)}_{t_n} \end{bmatrix}.$$

The vector that we place into the matching pursuit dictionary has elements that are proportional to $\hat{a}_k \cos(2\pi f_k t_n) + \hat{b}_k \sin(2\pi f_k t_n)$, $n = 0, \ldots, N' - 1$, where the constant of proportionality is used to force the vector to have unit norm.

Figure 5 shows the first five steps of matching pursuit using a dictionary containing ODFT, square wave, Haar wavelet and Haar scaling vectors adjusted to accommodate the gappy Sitka temperature measurements. The first vector that matching pursuit picks out is a square wave oscillation with a period of fifty four years, which is quite close to the penta-decadal oscillation found for the NP index. In the twentieth century the zero crossings for this oscillation occur at 1921–2, 1948–9 and 1975–6, which agree reasonably well with the zero crossings for the NP index

(1924–5, 1949–50 and 1974–5). This connection is not surprising as the Sitka temperature regime is related to advection of ocean surface temperatures (Royer 1993). There are additional zero crossings in the nineteenth century at 1840–1, 1867–8 and 1894–5. The second picked vector is a Haar scaling vector with of unit period (extent), which takes care of an anomalous cold temperature in 1876. The third and fourth vectors are both Haar wavelet vectors having periods of, respectively, eighty and sixteen years. The third vector is evidently fine-tuning the overall approximation to account for a shift in 1887–8, while the fourth is capturing a contrast between the years 1949–56 and 1957–64. The fifth vector is ODFT-like and is capturing some high frequency fluctuations with a period of 2.2 years, i.e., close to the Nyquist period of two years.

The most interesting result here obviously is the fact that matching pursuit has selected as the best single approximation a square wave oscillation that matches the NP index both in terms of its period and alignment. The fact that the measurement for 1876 merits a separate treatment by matching pursuit at stage $m = 2$ might make us concerned about how much influence it had on the choice of the square wave oscillation during the first stage; however, if we repeat our analysis with this year removed, the first vector chosen remains the same, and the vectors that are selected at stages $m = 2, 3$ and 4 are the same as shown in Figure 5 for stages $m = 3, 4$ and 5. Evidence for an extended bi-decadal oscillation is missing, although the period of sixteen for the time-limited Haar wavelet vector chosen at the fourth stage agrees somewhat with such an oscillation.

5 Conclusions

Matching pursuit was originally introduced with the idea of extracting time/frequency content from a time series. By using dictionaries based upon the discrete wavelet transform, Walden and Contreras Cristan (1998) demonstrated that this technique can also extract time/scale structure. Here we demonstrate through two examples that matching pursuit can help us informally assess the nature of atmospheric circulation changes over the North Pacific. We present these examples not only because they address a scientific hypothesis of recent interest, but also to motivate readers to consider matching pursuit for analogous investigations of other climatology time series.

How well matching pursuit works in particular applications will depend on what is put into the dictionary. Here we were motivated by Minobe's hypothesis to initially consider a dictionary with just sinusoidal oscillations to interpret the NP index. The first two sinusoidal oscillations that were picked out by matching pursuit had penta- and bi-decadal periods, in agreement with the periods – but not the shapes – found by Minobe. To address the shape question, we then augmented the dictionary with square waves to serve as competitors to the sinusoidal oscillations. The first two vectors selected by matching pursuit were now penta- and bi-decadal square waves, thus lending support to Minobe's contention that transitions between regimes occur more rapidly than what can be readily expressed by single sinusoids. We then further augmented the dictionary with Haar wavelet and scaling vectors, which are in essence snippets of square wave oscillations. This augmentation gives us the ability to assess the persistency of the penta- and bi-decadal oscillations over time. While the penta-decadal square wave oscillation emerges still as the best single descriptor for the NP index, the bi-decadal oscillation disappears in favor of Haar wavelet vectors describing isolated transitions. By successively augmenting the dictionary with vectors that offer support both for and against Minobe's hypothesis, we were able to informally assess various aspects of the hypothesis in a manner that lets the data 'speak for themselves' through the matching pursuit algorithm. For other investigations, different basis vectors and different augmentation schemes will undoubtedly be called for to study other hypothesized patterns in time series.

As we have stressed, the assessment of a hypothesis by matching pursuit is qualitative since this technique is not a formal statistical test. To what degree statistical significance can be attached to the results of a matching pursuit is a question for future research.

With regard to the two examples that we have concentrated on in Sections 3 and 4, our analysis lends support to the idea of penta-decadal oscillations in both time series, but support for bi-decadal oscillations is weaker. In the case of the NP index, this conclusion is further supported in Percival *et al.* (2002), where we find that a simple 'signal plus noise' model consisting of the penta-decadal oscillation and white noise fits the data well; i.e., diagnostic statistics do not indicate the need for any additional terms such as a bi-decadal oscillation. Our results add additional insight into the interpretation of the Pacific Decadal Oscillation (Mantua *et al.* 1997). Both the northern region of the North Pacific Ocean and the overlying atmosphere show tendency for simultaneous shifts on the rough time scale of twenty five years.

Finally we note that alternatives to the version of matching pursuit presented here are discussed in Mallat and Zhang (1993), Davis *et al.* (1994) and Mallat (1999). These alternatives are computationally more complex and are aimed at getting around the sequential nature of matching pursuit by simultaneously fitting collections of identified vectors to a time series. While these alternatives are of interest, the simplicity of the approach we have presented here has much to offer.

Acknowledgments. This contribution was supported in part by the Steller Sea Lion Program of the National Oceanic and Atmospheric Administration. We thank Dr. Thomas Royer for use of the Sitka data record.

REFERENCES

Davis, G., S. G. Mallat and Z. Zhang, 1994: Adaptive time-frequency approximations with matching pursuits. In *Wavelets: Theory, Algorithms, and Applications*, edited by C. K. Chui, L. Montefusco and L. Puccio, Academic Press, 271–293.

Mallat, S. G., 1999: *A Wavelet Tour of Signal Processing* (Second Edition). Academic Press, 637 pp.

Mallat, S. G., and Z. Zhang, 1993: Matching pursuits with time-frequency dictionaries. *IEEE Trans. Sig. Proc.*, **41**, 3397–3415.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.*, **78**, 1069–1079.

Minobe, S., 1999: Resonance in bidecadal and pentadecadal climate oscillations over the North Pacific: Role in climate regime shifts. *Geophys. Res. Lett.*, **26**, 855–858.

Overland, J. E., J. M. Adams and N. A. Bond, 1999: Decadal variability of the Aleutian low and its relation to high-latitude circulation. *J. Climate*, **12**, 1542–1548.

Percival, D. B., J. E. Overland and H. O. Mofjeld, 2001: Interpretation of North Pacific variability as a short- and long-memory process. *J. Climate*, **14**, 4545–4559.

Percival, D. B., J. E. Overland and H. O. Mofjeld, 2002: Modeling North Pacific climate time series. Under review.

Percival, D. B., and A. T. Walden, 2000: *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 594 pp.

Royer, T. C., 1993: High-latitude oceanic variability associated with the 18.6-year nodal tide. *J. Geophys. Res.*, **98**, 4639–4644.

Trenberth, K. E., and D. A. Paolino, 1980: The Northern Hemisphere sea level pressure data set: Trends, errors and discontinuities. *Mon. Weather Rev.*, **108**, 855–872.

Walden, A. T., and A. Contreras Cristan, 1998: Matching pursuit by undecimated discrete wavelet transform for non-stationary time series of arbitrary length. *Statistics and Computing*, **8**, 205–219.

FIGURE CAPTIONS

Figure 1. Examples of dictionary vectors $\mathbf{D}_k$ used in various matching pursuits of the NP index. The elements $D_{k,t}$, $t = 0, \ldots, 99$, for each $\mathbf{D}_k$ are plotted versus $t + 0.5$. The vector in (a) is a complex-valued vector from an orthonormal discrete Fourier transform (the real and imaginary parts are indicated by, respectively, solid dots and open circles). The period associated with this vector is twenty. In (b), the vector contains a square wave oscillation, also with a period of twenty. In (c), the vector is created from a discretized Haar wavelet function associated with changes on a scale of ten, while (d) shows one from a corresponding Haar scaling function.

Figure 2. Matching pursuit of NP index using dictionary consisting of just sinusoids. The thin jagged curve in each right-hand plot shows the NP index $\mathbf{X}$. The thick curves in the left-hand plots depict the sinusoid that was selected in steps $m = 1, \ldots, 5$ (top to bottom, respectively). The thick curves in the right-hand plots show the corresponding approximation $\widehat{\mathbf{X}}^{(m)}$. The period associated with each sinusoid is stated in the left-hand margin, while the right-hand margin lists the percentage of the variance that is explained by $\widehat{\mathbf{X}}^{(m)}$ (by definition, this is $(\|\mathbf{X}\|^2 - \|\mathbf{R}^{(m)}\|^2)/\|\mathbf{X}\|^2 \times 100\%$, where $\mathbf{R}^{(m)} = \mathbf{X} - \widehat{\mathbf{X}}^{(m)}$).

Figure 3. As in Figure 2, but now using a dictionary consisting of sinusoids and square wave oscillations.

Figure 4. As in Figure 2, but now using a dictionary consisting of sinusoids, square wave oscillations and Haar wavelet vectors. Identical results were obtained when the dictionary was further augmented to include Haar scaling vectors.

Figure 5. As in Figure 4, but now using the Sitka air temperatures.

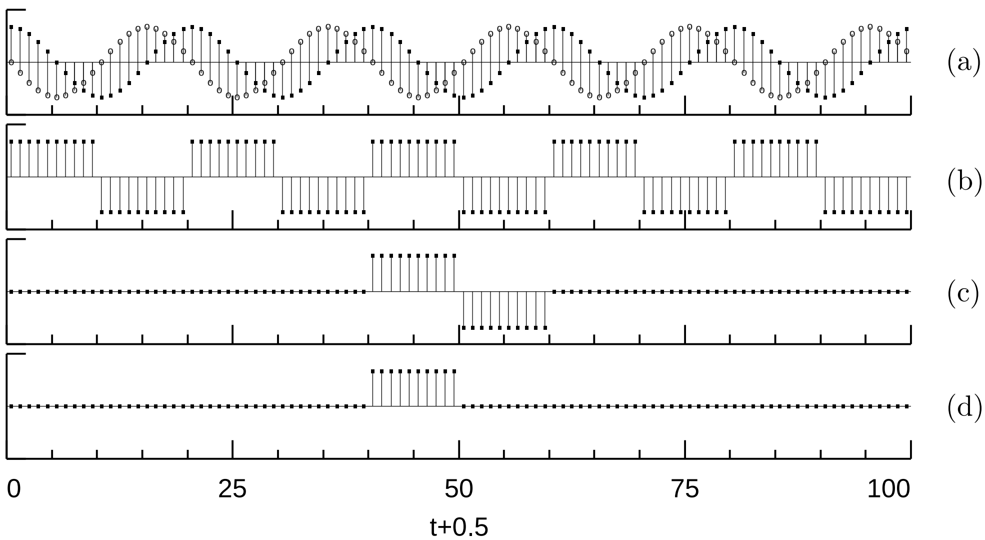

Figure 1: Examples of dictionary vectors $\mathbf{D}_k$ used in various matching pursuits of the NP index. The elements $D_{k,t}$, $t = 0, \ldots, 99$, for each $\mathbf{D}_k$ are plotted versus $t + 0.5$. The vector in (a) is a complex-valued vector from an orthonormal discrete Fourier transform (the real and imaginary parts are indicated by, respectively, solid dots and open circles). The period associated with this vector is twenty. In (b), the vector contains a square wave oscillation, also with a period of twenty. In (c), the vector is created from a discretized Haar wavelet function associated with changes on a scale of ten, while (d) shows one from a corresponding Haar scaling function.

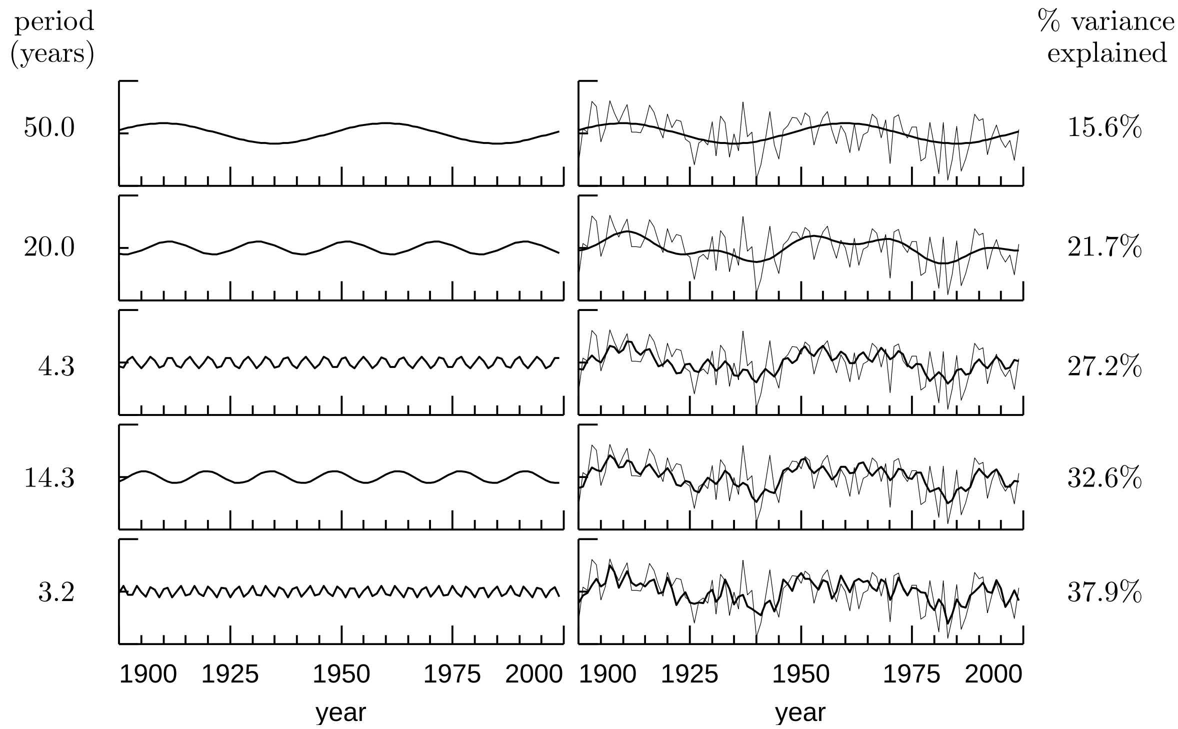

Figure 2: Matching pursuit of NP index using dictionary consisting of just sinusoids. The thin jagged curve in each right-hand plot shows the NP index $\mathbf{X}$. The thick curves in the left-hand plots depict the sinusoid that was selected in steps $m = 1, \ldots, 5$ (top to bottom, respectively). The thick curves in the right-hand plots show the corresponding approximation $\widehat{\mathbf{X}}^{(m)}$. The period associated with each sinusoid is stated in the left-hand margin, while the right-hand margin lists the percentage of the variance that is explained by $\widehat{\mathbf{X}}^{(m)}$ (by definition, this is $(\|\mathbf{X}\|^2 - \|\mathbf{R}^{(m)}\|^2)/\|\mathbf{X}\|^2 \times 100\%$, where $\mathbf{R}^{(m)} = \mathbf{X} - \widehat{\mathbf{X}}^{(m)}$).

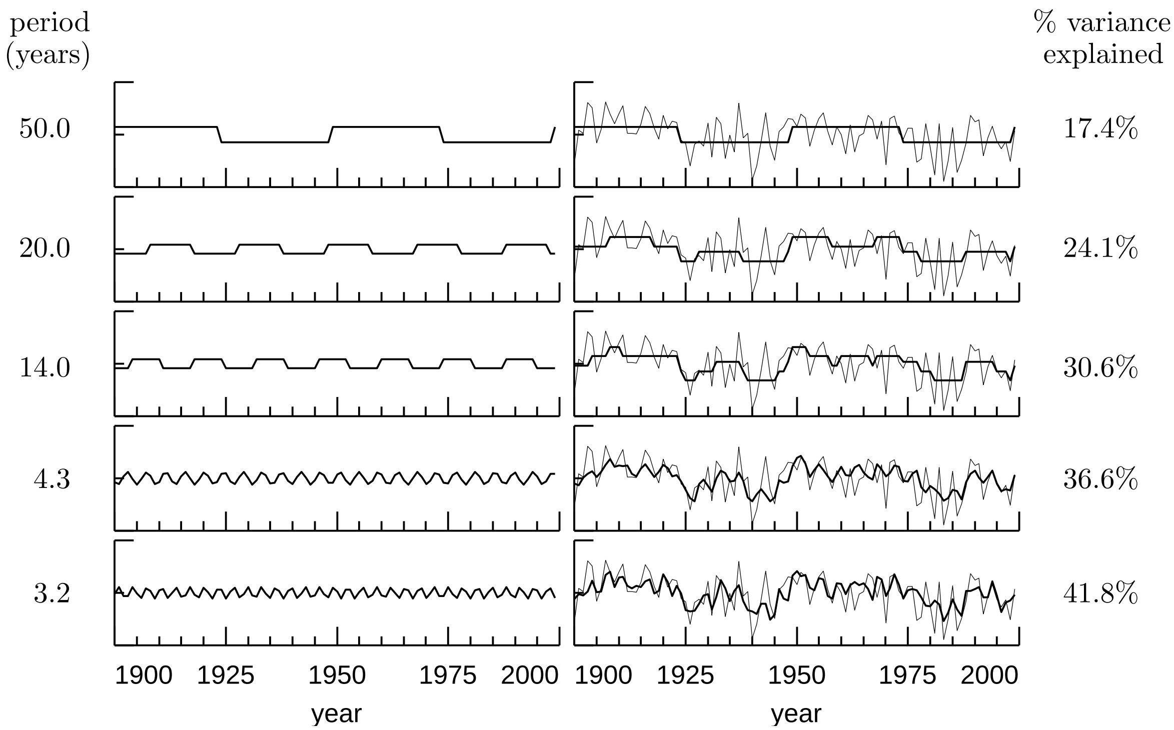

Figure 3: As in Figure 2, but now using a dictionary consisting of sinusoids and square wave oscillations.

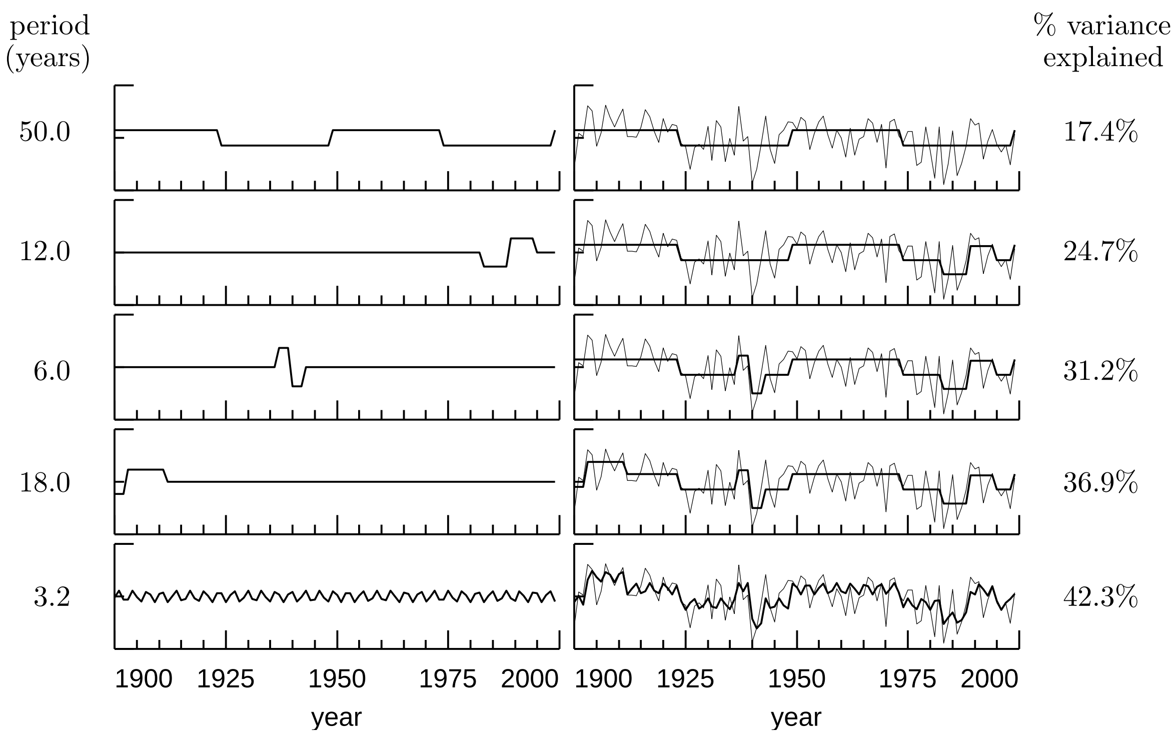

Figure 4: As in Figure 2, but now using a dictionary consisting of sinusoids, square wave oscillations and Haar wavelet vectors. Identical results were obtained when the dictionary was further augmented to include Haar scaling vectors.

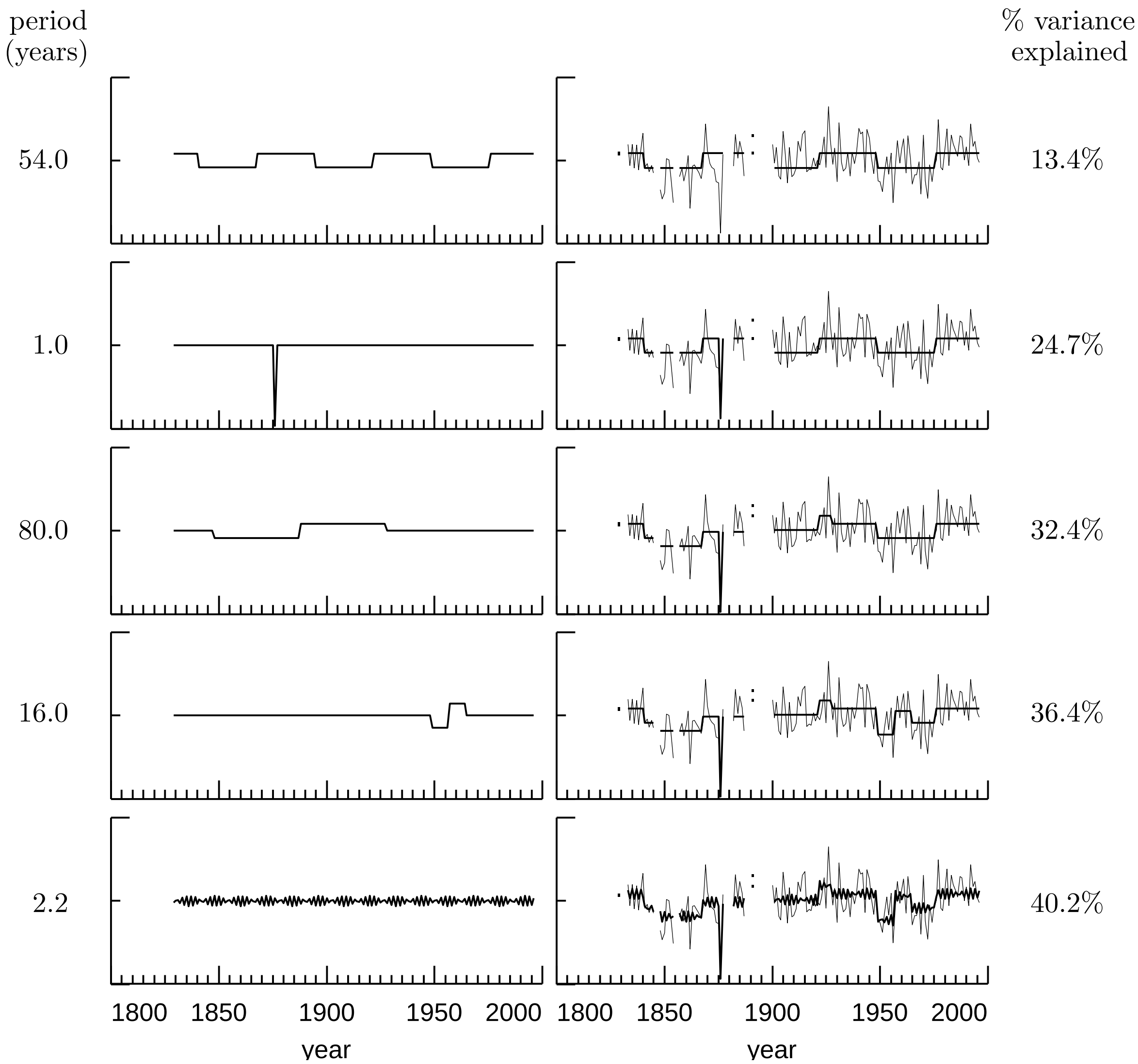

Figure 5: As in Figure 4, but now using the Sitka air temperatures.