Matching Pursuit – Basics

- idea: approximate $\mathbf{X}$ using a few # of 'time/frequency' vectors from large set of such vectors (cf. best basis)
- form 'dictionary' of vectors $\mathcal{D} \equiv \{\mathbf{d}_\gamma : \gamma \in \Gamma\}$
  - $\mathbf{d}_\gamma = \left[d_{\gamma,0}, d_{\gamma,1}, \ldots, d_{\gamma,N-1}\right]^T$
  - each vector has unit norm: $\|\mathbf{d}_\gamma\|^2 = \sum_{l=0}^{N-1} d_{\gamma,l}^2 = 1$
  - γ is vector of parameters connecting $\mathbf{d}_\gamma$ to time/frequency; e.g., $\gamma = [j, n, t]^T$ for WP table dictionary
  - $\Gamma =$ finite set of possible values for γ
  - $\mathcal{D}$ contains basis for $\mathcal{R}^N$, but can be highly redundant (helps identify time/frequency content in $\mathbf{X}$)
- matching pursuit successively approximates $\mathbf{X}$ with orthogonal projections onto elements of $\mathcal{D}$

Background Material

- recall that we can reconstruct a time series $\mathbf{X}$ from its DWT coefficients $\mathbf{W}$ via $\mathbf{X} = \mathcal{W}^T\mathbf{W}$, where $\mathbf{W} \equiv \mathcal{W}\mathbf{X}$
- jth coefficient in $\mathbf{W}$ is $\langle \mathbf{X}, \mathcal{W}_{j\bullet} \rangle$, i.e., the inner product of $\mathbf{X}$ & a column vector $\mathcal{W}_{j\bullet}$ whose elements are the jth row of $\mathcal{W}$
- hence we can write

$$\mathbf{X} = \mathcal{W}^T\mathbf{W} = [\mathcal{W}_{0\bullet}, \mathcal{W}_{1\bullet}, \ldots, \mathcal{W}_{N-1\bullet}] \begin{bmatrix} \langle \mathbf{X}, \mathcal{W}_{0\bullet} \rangle \\ \langle \mathbf{X}, \mathcal{W}_{1\bullet} \rangle \\ \vdots \\ \langle \mathbf{X}, \mathcal{W}_{N-1\bullet} \rangle \end{bmatrix}$$

$$= \sum_{j=0}^{N-1} \langle \mathbf{X}, \mathcal{W}_{j\bullet} \rangle \mathcal{W}_{j\bullet}$$

- regard $\langle \mathbf{X}, \mathcal{W}_{j\bullet} \rangle \mathcal{W}_{j\bullet}$ as approximation to $\mathbf{X}$ based on just $\mathcal{W}_{j\bullet}$

Matching Pursuit Algorithm: I

- for $\mathbf{d}_{\gamma_0} \in \mathcal{D}$, form $\langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \mathbf{d}_{\gamma_0}$, and define residual vector: $\mathbf{R}^{(1)} \equiv \mathbf{X} - \langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \mathbf{d}_{\gamma_0}$ so that $\mathbf{X} = \langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \mathbf{d}_{\gamma_0} + \mathbf{R}^{(1)}$
- note that $\mathbf{d}_{\gamma_0}$ and $\mathbf{R}^{(1)}$ are orthogonal (this is Exer. [240]):

$$\begin{aligned} \langle \mathbf{d}_{\gamma_0}, \mathbf{R}^{(1)} \rangle &= \left\langle \mathbf{d}_{\gamma_0}, \mathbf{X} - \langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \mathbf{d}_{\gamma_0} \right\rangle \\ &= \langle \mathbf{d}_{\gamma_0}, \mathbf{X} \rangle - \left\langle \mathbf{d}_{\gamma_0}, \left\langle \mathbf{X}, \mathbf{d}_{\gamma_0} \right\rangle \mathbf{d}_{\gamma_0} \right\rangle \\ &= \langle \mathbf{d}_{\gamma_0}, \mathbf{X} \rangle - \langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle = 0 \end{aligned}$$

- hence $\langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \mathbf{d}_{\gamma_0}$ & $\mathbf{R}^{(1)}$ are also orthogonal, showing that

$$\|\mathbf{X}\|^2 = \|\langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \mathbf{d}_{\gamma_0}\|^2 + \|\mathbf{R}^{(1)}\|^2 = \left| \langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \right|^2 + \|\mathbf{R}^{(1)}\|^2$$

- minimize energy in residuals by choosing $\gamma_0 \in \Gamma$ such that

$$\left| \langle \mathbf{X}, \mathbf{d}_{\gamma_0} \rangle \right| = \max_{\gamma \in \Gamma} \left| \langle \mathbf{X}, \mathbf{d}_{\gamma} \rangle \right|$$

Matching Pursuit Algorithm: II

- after first step of algorithm, second step is to treat the residuals in the same manner as $\mathbf{X}$ was treated in first step, yielding
$$\mathbf{R}^{(1)} = \langle \mathbf{R}^{(1)}, \mathbf{d}_{\gamma_1} \rangle \mathbf{d}_{\gamma_1} + \mathbf{R}^{(2)},$$
with $\mathbf{d}_{\gamma_1}$ picked such that
$$\left|\langle \mathbf{R}^{(1)}, \mathbf{d}_{\gamma_1} \rangle\right| = \max_{\gamma \in \Gamma} \left|\langle \mathbf{R}^{(1)}, \mathbf{d}_{\gamma} \rangle\right|$$
- letting $\mathbf{R}^{(0)} \equiv \mathbf{X}$, after m such steps, have additive decomposition:
$$\mathbf{X} = \sum_{k=0}^{m-1} \langle \mathbf{R}^{(k)}, \mathbf{d}_{\gamma_k} \rangle \mathbf{d}_{\gamma_k} + \mathbf{R}^{(m)}$$

Matching Pursuit Algorithm: III

- also have an energy decomposition:

$$\|\mathbf{X}\|^2 = \sum_{k=0}^{m-1} \|\langle \mathbf{R}^{(k)}, \mathbf{d}_{\gamma_k} \rangle \mathbf{d}_{\gamma_k}\|^2 + \|\mathbf{R}^{(m)}\|^2$$

$$= \sum_{k=0}^{m-1} |\langle \mathbf{R}^{(k)}, \mathbf{d}_{\gamma_k} \rangle|^2 + \|\mathbf{R}^{(m)}\|^2$$

- note: as m increases, $\|\mathbf{R}^{(m)}\|^2$ must decrease (must reach zero under certain conditions)

Matching Pursuit Dictionaries: I

- key to matching pursuit is dictionary
- simplest dictionary: DWT dictionary
  - $\mathcal{D}$ contains $\mathbf{d}_\gamma \equiv \mathcal{W}_{j\bullet}$, $j = 0, \ldots, N-1$
  - $\gamma = [j]$ associates $\mathcal{W}_{j\bullet}$ with time/scale
  - $\langle \mathbf{X}, \mathbf{d}_\gamma \rangle = W_j$ is jth DWT coefficient
  - 1st step picks W_j with largest magnitude:

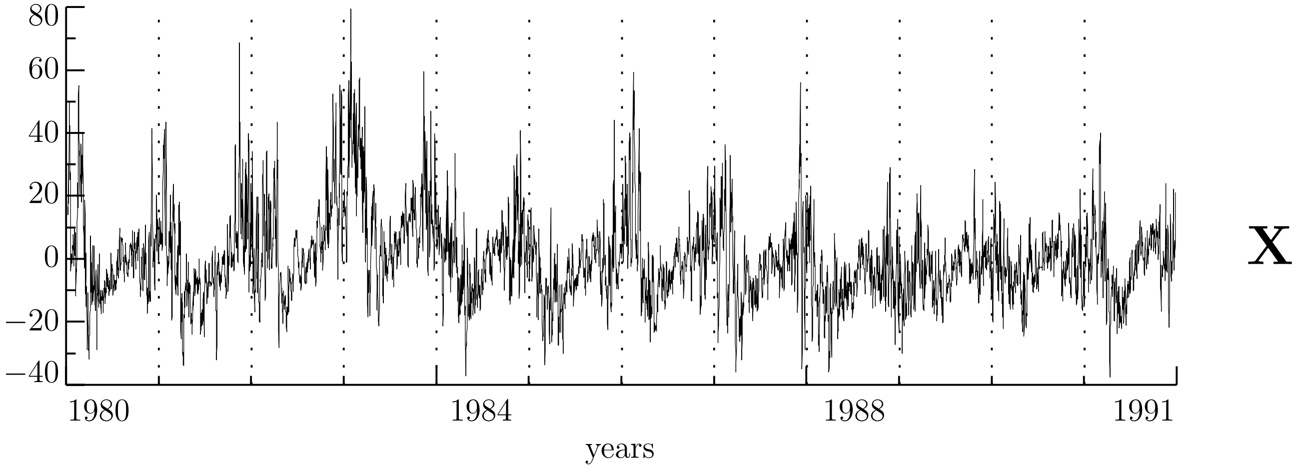
$$\mathbf{X} = W_{(0)}\mathbf{W}_{(0)} + \mathbf{R}^{(1)} \text{ with } \mathbf{R}^{(1)} = \sum_{j \neq (0)} W_j \mathbf{W}_{j\bullet}$$

  - 2nd step picks out W_j with 2nd largest $|W_j|$
  - for any orthonormal $\mathcal{D}$, matching pursuit approximates $\mathbf{X}$ using coefficients with largest magnitudes

Matching Pursuit Dictionaries: II

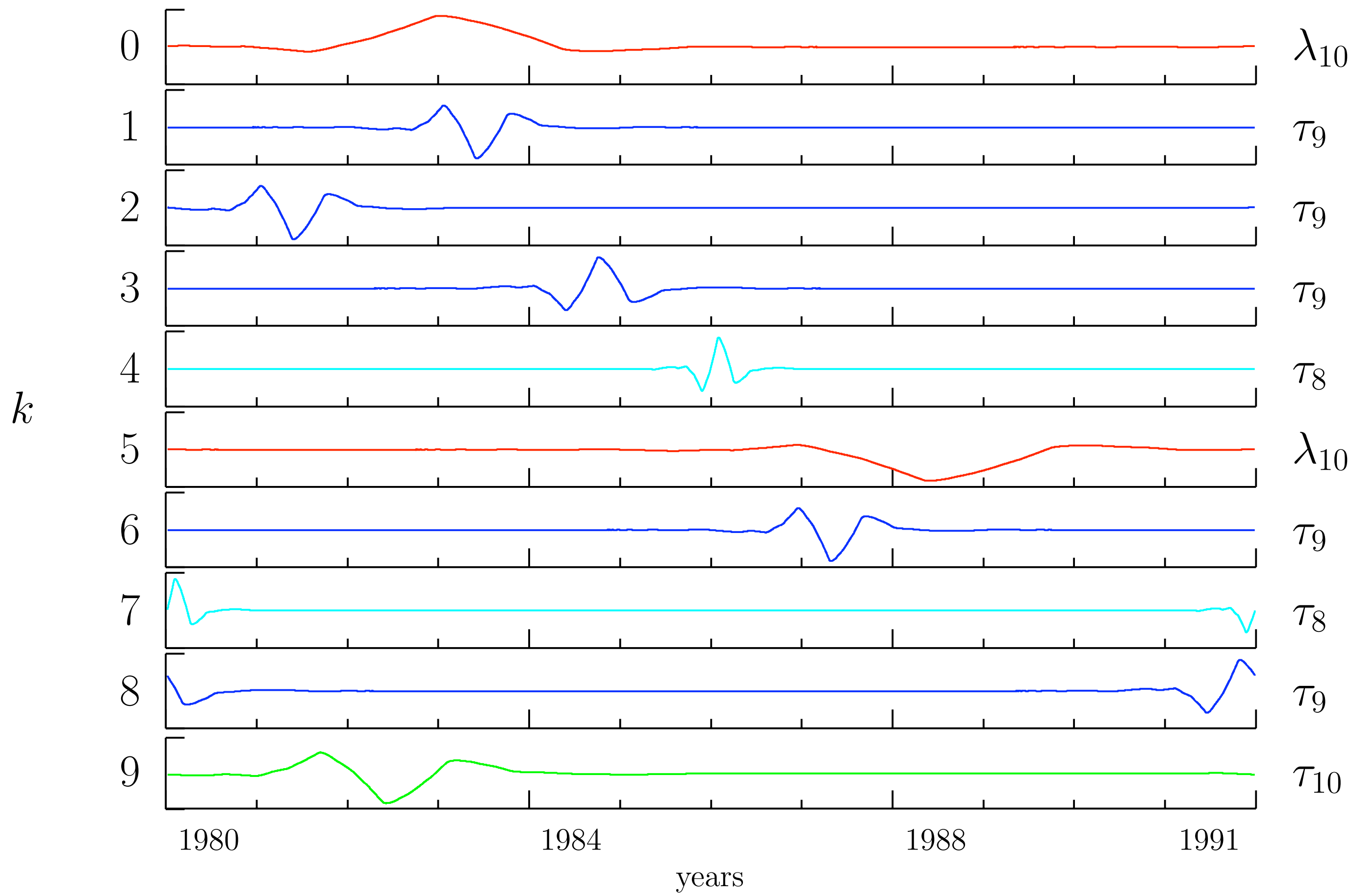
- larger dictionary: wavelet packet table dictionary (more flexible than best basis)
- even larger dictionary: above combined with basis vectors corresponding to a discrete Fourier transform (DFT)
- level J_0 MODWT dictionary
  - works for all N, shift invariant, redundant
  - $\mathcal{D}$ contains vectors whose elements are either
    - normalized rows of $\widetilde{\mathcal{W}}_j$, $j = 1, \ldots, J_0$, or
    - normalized rows of $\widetilde{\mathcal{V}}_{J_0}$

Example – Subtidal Sea Levels: I



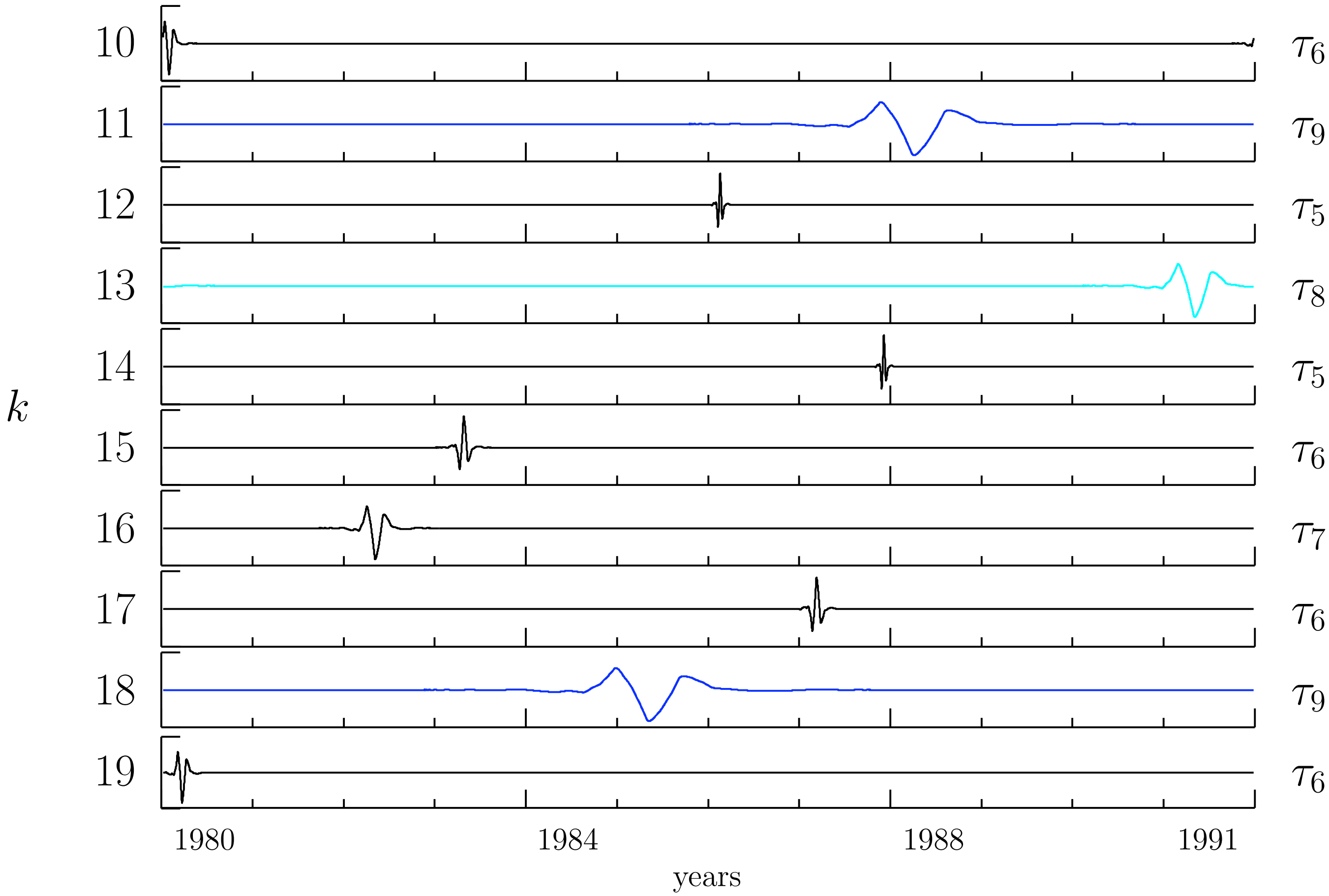
- recall subtidal sea level series **X** for Crescent City, CA

Example – Subtidal Sea Levels: II



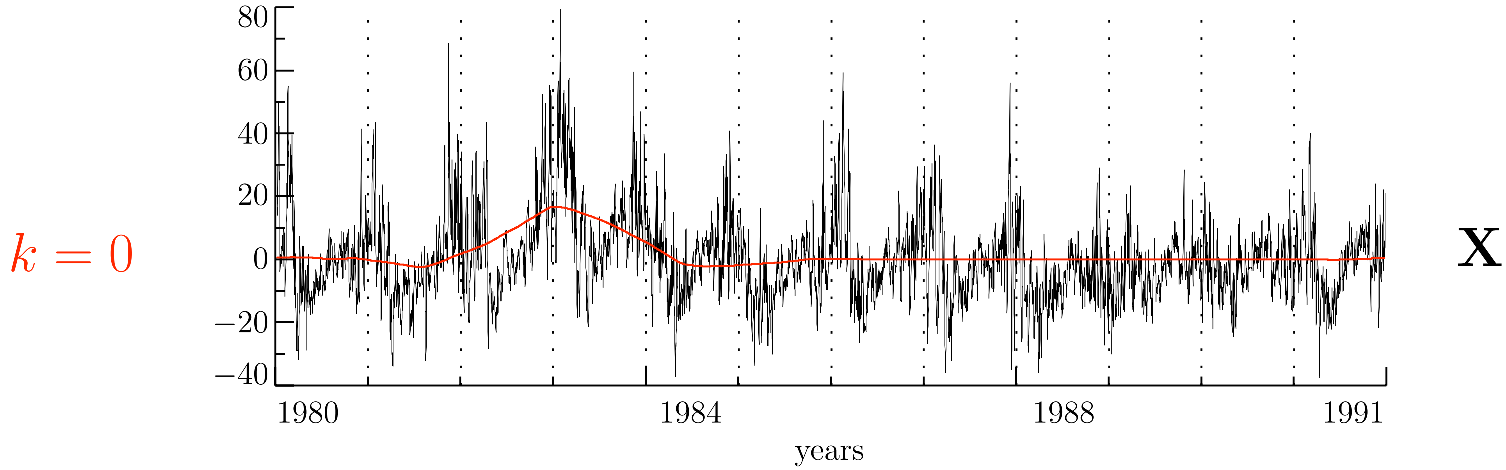
- use $J_0 = 10$ LA(8) MODWT dictionary (96,206 vectors in all)
- above shows first 10 vectors picked by matching pursuit ($\times \pm 1$)

Example – Subtidal Sea Levels: III



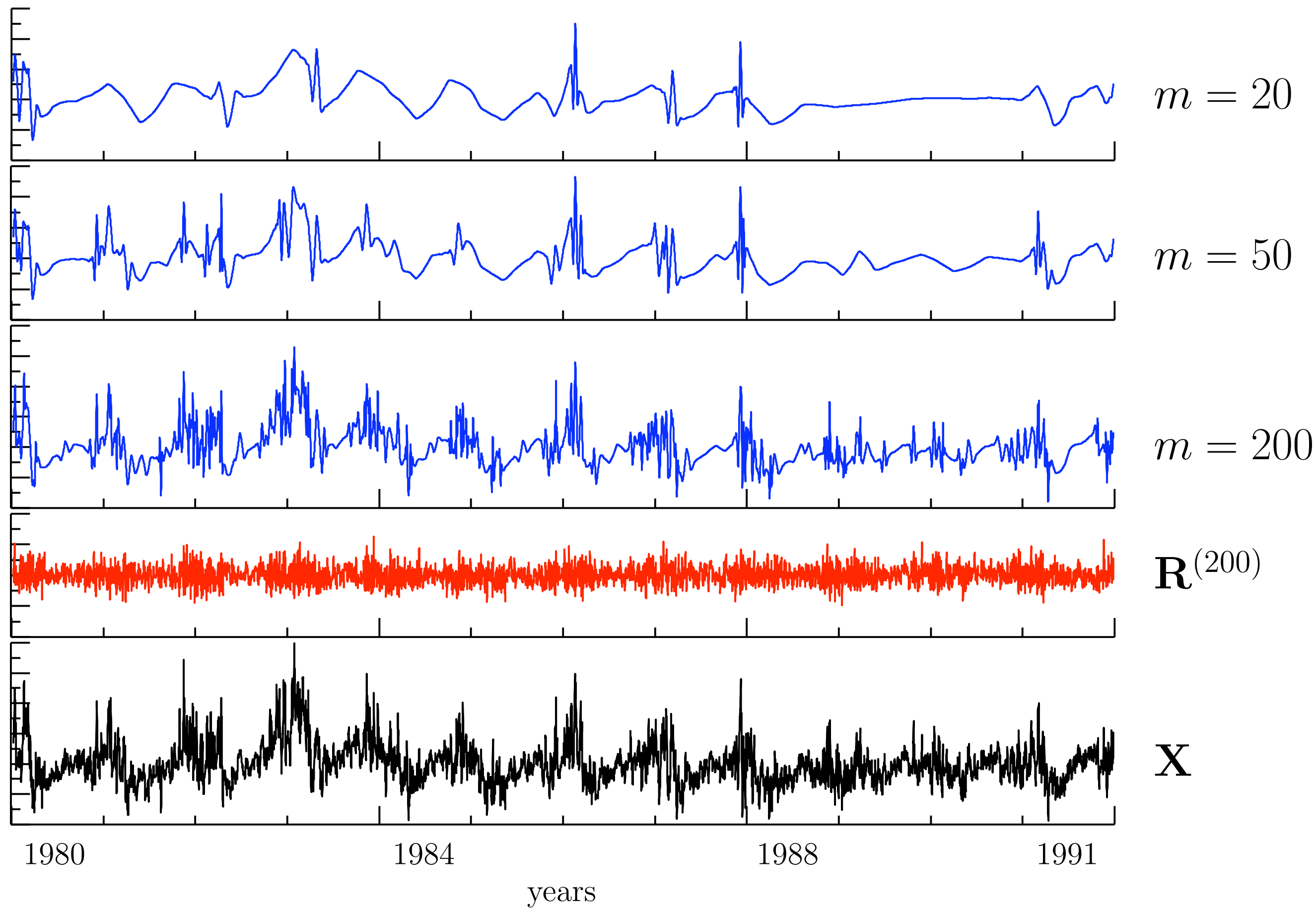
- next 10 vectors picked by matching pursuit ($\times \pm 1$)

Example – Subtidal Sea Levels: IV



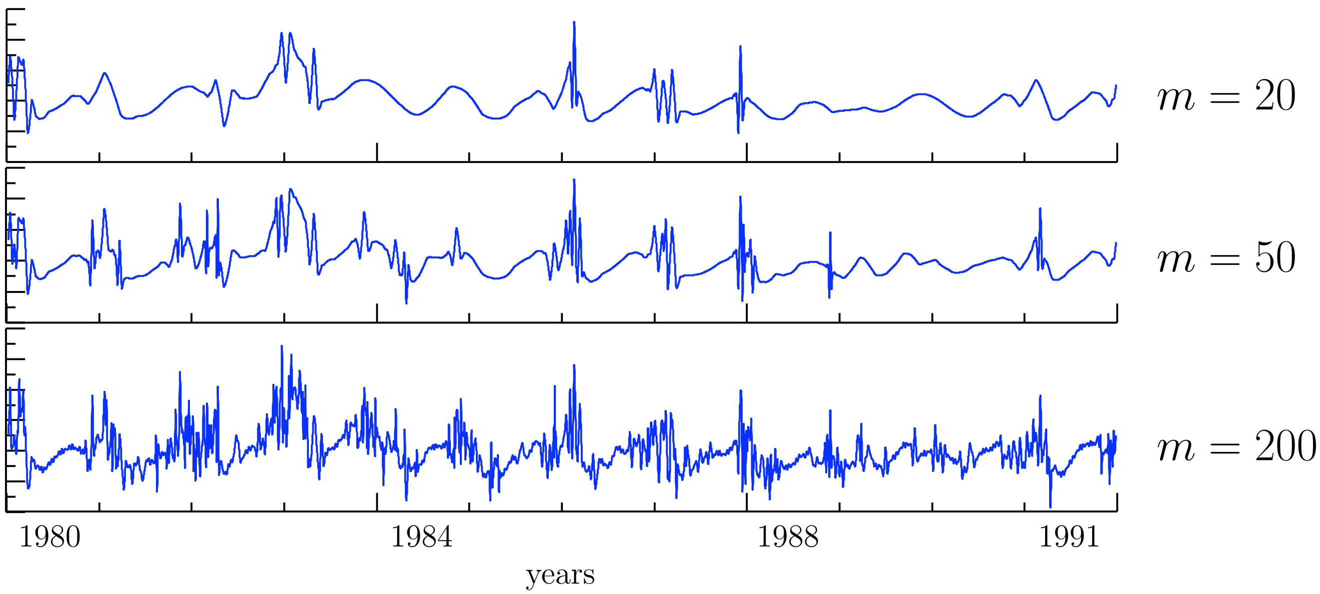
- very first ($k = 0$) associated with overall increase in 1982–3
- first 10 are for $\tau_8 \, \Delta t = 64$ to $\lambda_{10} \, \Delta t = 512$ days
- 7 of first 20 are associated with $\tau_9 \, \Delta t = 128$ days (needed to account for seasonal variabilty)
- $k = 3$ has inverted sign & picks out gradual dip in Spring, 1984 (cf. 1981, 3, 5, 7 & 8); $k = 8$ also inverted, but is a boundary effect

Example – Subtidal Sea Levels: V



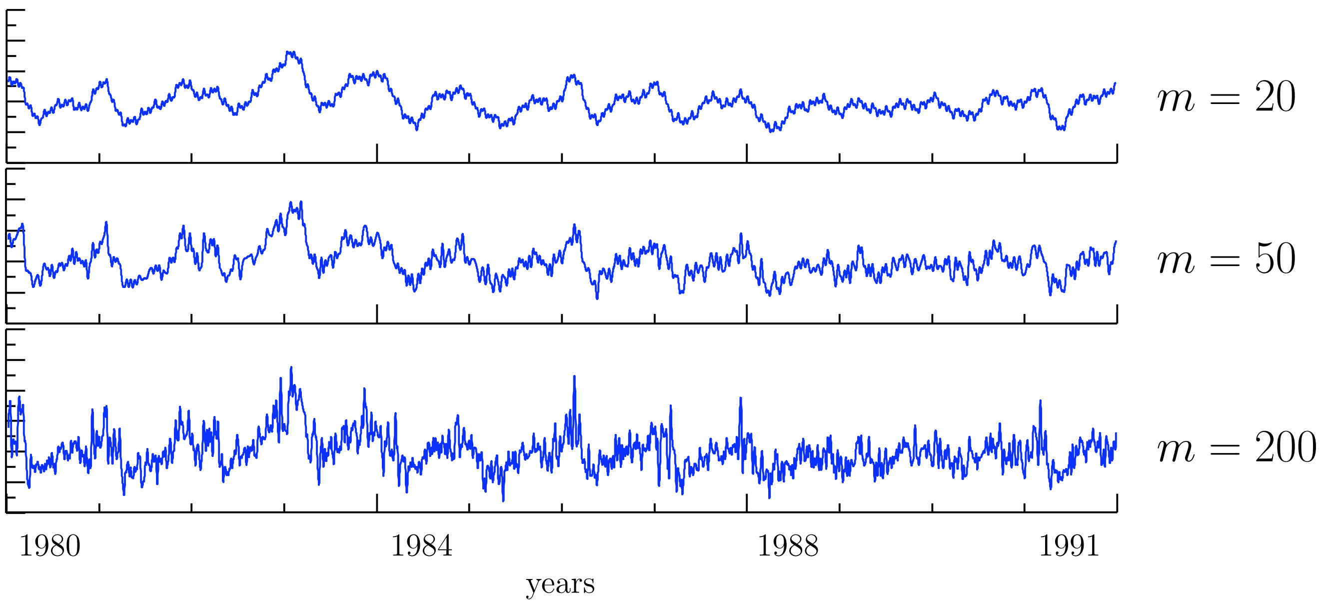
- matching pursuit approximations of orders $m = 20, 50$ and 200, along with residuals for $m = 200$

Example – Subtidal Sea Levels: VI



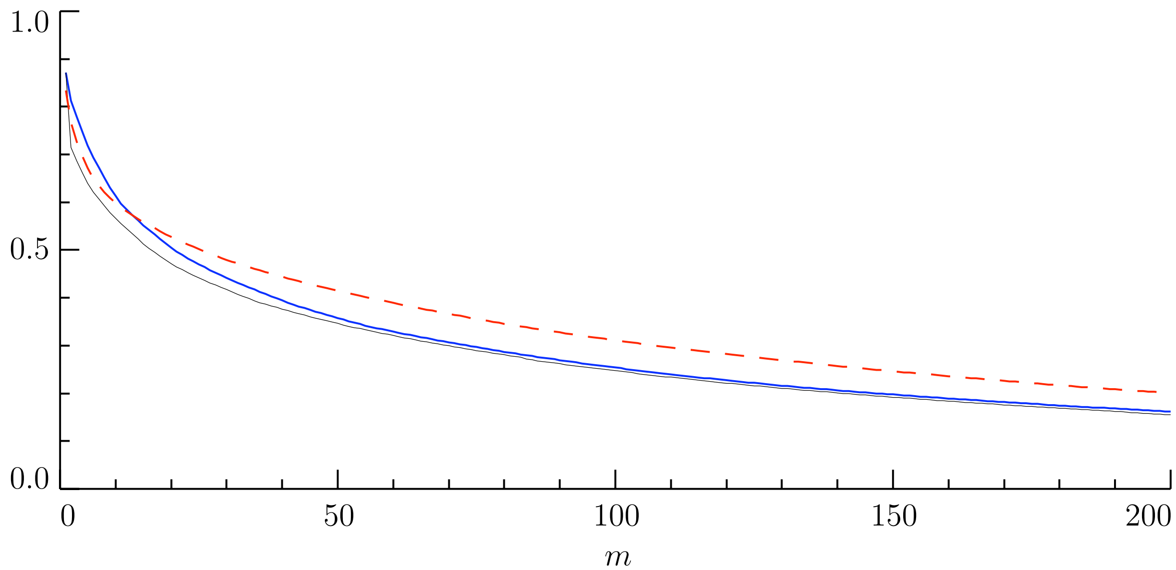
- matching pursuit approximations of orders $m = 20, 50$ and 200, but now using a dictionary augmented to include basis vectors corresponding to the DFT
- $k = 0$ choice same as before, but $k = 1$ choice is DFT vector with period close to one year
- for $2 \le k < 200$, only $k = 65, 84$ and 192 are DFT vectors

Example – Subtidal Sea Levels: VII



- matching pursuit approximations of orders $m = 20, 50$ and 200, but now using a dictionary consisting of just the basis vectors corresponding to the DFT

Example – Subtidal Sea Levels: VIII



- normalized residual sum of squares $\|\mathbf{R^{(m)}}\|^2/\|\mathbf{X}\|^2$ versus number of terms m in matching pursuit approximation using the MODWT dictionary (thick curve), the DFT-based dictionary (dashed) and both dictionaries combined (thin)
- combined dictionary does best for small m, but MODWT dictionary by itself becomes competitive as m increases