Title of article: Relative sea levels of Japan from tide-gauge records

Author(s): D. G. Aubrey and K. O. Emery

Contents: 4 pages

Appendix
APPENDIX

Two specific types of information are contained in sea-level records: first, the relative rates of sea-level rise, including acceleration or deceleration of this rise; second, the spatial and temporal structure of relative sea-level rise. Linear and non-linear regression techniques previously used do not provide the second type of information, nor do they determine coherent modes of sea-level change to minimize local aberrations in sea level. Instead, they impose a subjective shape to sea-level curves, and optimally fit a local record to that shape. We use here an objective method to determine uncorrelated modes of sea-level change: eigenanalysis.

Eigenanalysis is a well-known widely-used objective technique for determining dominant modes of variation in data sets. The purpose of eigenanalysis is to separate a data set into orthogonal spatial and temporal modes that most efficiently describe the variability of that data set (no other orthogonal functions can more efficiently represent that same data set). The result is a concise description of the spatial and temporal structure of variability in that data set.

Mathematically, a data set \( \eta(x,t) \) can be decomposed into spatial and temporal functions:

\[
\eta(x,t) = \sum_{k=1}^{N} C_k(t) e_k(x) (\lambda_k n_x n_t)^{1/2},
\]

where \( \eta(x,t) \) is a spatial grid sampled through time with its mean removed, the \( C_k(t) \) represent temporal eigenfunctions, \( e_k(x) \) represent spatial eigenfunctions, \( \lambda_k \) are the eigenvalues, \( n_x \) is the number of spatial points, \( n_t \) is the number of temporal points, and \( N \) is the lesser of \( n_x \) and \( n_t \).
Spatial eigenfunctions are determined from the following matrix operation:

\[(A - \lambda I)e = 0,\]

where

\[A = \frac{1}{n_x n_t} \eta \eta^T.\]  \hspace{1cm} (2)

The superscript T refers to the matrix transpose operator, I is the identity matrix, \(\eta\) is an \((n_x, n_t)\) matrix of data, and \(A\) has size \((n_x, n_x)\). \(A\) is a covariance matrix as shown in (2); if each station variance is set to unity, then \(A\) is a correlation matrix. Temporal eigenfunctions are determined from the matrix equation:

\[(B - \lambda I)c = 0.\]

where

\[B = \frac{1}{n_x n_t} \eta^T \eta.\]  \hspace{1cm} (3)

\(B\) is a \((n_t, n_t)\) matrix. The \(\lambda\)'s are identical for equally ranked spatial or temporal eigenfunctions. Eigenfunctions are ranked according to their eigenvalues; the first eigenfunction has the largest eigenvalue, the second eigenfunction has the next largest eigenvalue, and so on. The eigenfunctions obey a least-squares criterion: the first eigenfunction best describes the data in a least-square sense; the second eigenfunction best describes the residual of the data in a least-squares sense, and so on. All eigenfunctions are orthonormal (or uncorrelated):

\[cc^T = I\]

\[ee^T = I.\]

Eigenfunction equations (2) and (3) are solved numerically (see, for example, Wilkinson and Reinsch, 1971), after matrices \(A\) and \(B\) are calculated. Operationally, once either set of eigenfunctions is determined numerically, the second set can be determined from the inner product of the eigenfunction and the data matrix. Once eigenfunctions are calculated, the original time
series (data set) can be reconstructed completely using equation (1). If the first \( m \) eigenvalues dominate all other eigenvalues (where \( m < N \)), a filtered data set \((\eta'(x,t))\) can be generated using (1), by summing only up to \( k = m \). This new filtered data set \((\eta'(x,t))\) can then be used to represent the "predictable" or information-rich part of the original signal. In the following analysis, we generally obtain a useful \( \eta'(x,t) \) by letting \( m = 2 \). A more complete discussion of statistical methods for selecting information-rich eigenvectors is presented in Preisendorfer et al. (1981).

Since either the \( C_k(t) \) or \( e_k(x) \) eigenfunctions can be viewed as weighting functions, we use the convention of presenting data as normalized \( C_k(t) \) for temporal variability, and physical quantities \( e_k(x)b_k \) with units of meters for spatial variability, where

\[
b_k = \left( \lambda_k n_x n_t \right)^{\frac{1}{2}}.
\]

Modified eigenanalysis permits use of stations with unequal data lengths (gappy data). The matrix \( A \) is formed with elements \( a_{jk} \) summed only over the overlapping segments of station \( j \) and station \( k \). As a result, mean product elements \( a_{jk} \) will be formed over unequal numbers of samples. Although the interpretation of resulting eigenvectors must be done with care, for data with few gaps the method greatly extends the data available for analysis without resorting to interpolation or extrapolation which leads to unacceptable smoothing. Error analysis for this work is discussed by Aubrey and Welch (in prep).

Computationally, the (time) mean is removed from each station for the entire period during which it recorded. Mean product elements are calculated from these series from which means have been removed. Since mean product elements are formed of overlapping segments only, they are in general calculated over dissimilar lengths of time. If the series are stationary during
these various time intervals, the resulting covariance (or correlation) matrix is not affected by gaps. If the individual series are not stationary during these various intervals, errors arise. To reduce the errors to acceptable levels, we used stations with more than 15 years of record (most of which recorded only between 1953 and 1980).

When each station variance is set to unity, resulting spatial functions are not scaled correctly to provide true sea-level rise estimates. In this case, the weighting function, $b_k$, is defined:

$$b_k = (\lambda_k n_x n_t V_k)^{\frac{1}{2}}$$

where $V_k$ is the sample station variance.