PERGAMON

Computers & Geosciences 28 (2002) 929–937

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

# Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE ☆

Rich Pawlowicz[a,*], Bob Beardsley[b], Steve Lentz[b]

[a] *Department of Earth and Ocean Sciences, University of British Columbia, 6270 University Blvd., Vancouver, BC, Canada V6T 1Z4*
[b] *Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA*

Received 24 September 2001; received in revised form 13 February 2002; accepted 28 February 2002

## Abstract

A standard part of any oceanic pressure gauge or current meter analysis is the separation of tidal from non-tidal components of the signal. The tidal signal can either be discarded, or its characteristics described in some fashion useful for further analysis. Although tidal signals can be removed by standard high or bandpass filtering techniques, their relatively deterministic character and large amplitude make special techniques more effective. In classical harmonic analysis, the tidal signal is modelled as the sum of a finite set of sinusoids at specific frequencies related to astronomical parameters. A set of programs has been written in MATLAB to (a) perform classical harmonic analysis for periods of about 1 year or shorter, (b) account for (some) unresolved constituents using nodal corrections, and (c) compute confidence intervals for the analyzed components. 

*Keywords:* Tides; Confidence interval; Oceanography

## 1. Introduction

As the earth rotates on its axis, spatially varying gravitational forces from the moon and the sun act on the ocean, generating a forced elevation and current response primarily (but not solely) at diurnal and semi-diurnal frequencies. Body forces act directly on deep oceanic waters. Tidal effects in coastal regions are not directly forced by these astronomical forces. Instead they arise as a side-effect of deep oceanic variability, propagating through shallower coastal waters as a wave or a combination of waves. In a typical oceanic time series, tidal variability is often the largest signal. Power spectra for such time series are often characterized by a broad hump with a low-frequency maximum and a decline at higher frequencies. Superimposed are a number of sharp tidal peaks near diurnal and semi-diurnal frequencies, and sometimes a broader peak associated with Coriolis or inertial effects. Dynamical analysis requires the separation of the tidal signal from sub or super-tidal variations, or in some cases separation of tidal effects from inertial effects at a nearby frequency. The tidal information is either discarded or kept for further analysis.

Standard high/low/bandpass filtering techniques (e.g., Jackson 1986) can be used but in general these are inefficient because fairly narrow filters with a great deal of rejection are needed. Also, although these are useful in analyzing non-tidal variability, they provide no compression of the tidal information. Specialized techniques have been devised to take advantage of the "deterministic" nature of tidal processes. In classical harmonic analysis, the tidal forcing is modelled as a set of spectral lines, i.e., the sum of a finite set of sinusoids at specific frequencies. These frequencies are specified by various combinations of sums and differences of integer multiples of 6 fundamental frequencies arising from planetary motions (Godin, 1972). These fundamental

☆ Code available from server at http://www.iamg.org/CG-Editor/index.htm or http://www.ocgy.ubc.ca/~rich.

*Corresponding author. Tel.: +1-604-822-1356; fax: +1-604-822-6091.

*E-mail address:* rich@ocgy.ubc.ca (R. Pawlowicz).

0098-3004/02/$ - see front matter 
PII: S0098-3004(02)00013-4

parameters represent effects of rotation of the earth (lunar day of 24.8 h), the orbit of the moon around the earth (lunar month of 27 days) and the earth around the sun (tropical year), and periodicities in the location of lunar perigee (8.85 years), lunar orbital tilt (18.6 years), and the location of perihelion ($\approx$21 000 years). The set of 6 signed integers required to describe a particular frequency are called the *Doodson numbers*. Many of the more important frequencies have names such as "$M_2$", "$K_1$", etc. From astronomical considerations alone, an "equilibrium" response can be predicted; this is the phase and amplitude that would be observed if the response of the earth was fast enough that the surface deformation was effectively in equilibrium with the forcing at all times. The real ocean is definitely not in equilibrium with the tidal forcing. However, as tidal amplitudes are small compared with the total ocean depth, the dynamics are very nearly linear, implying that the forced response contains only those frequencies present in the forcing. A least-squares fit can be used to determine the relative phase and amplitude of each frequency in the response. This phase/amplitude data thus provides a compression of the data in the complete time series, which can then be compared with similar data at other locations to understand the characteristics of tidal dynamics, or can be used to synthesize time series of tidal effects at other times for predictive purposes.

There are several drawbacks to classical harmonic analysis. The first is that, ignoring the modulation of perihelion which is effectively constant over historical time, an $\approx$18.6 year time series is required to resolve all of the listed frequencies (that is, the number of wavelengths of each constituent in the record is at least 1 different from all other constituents). In practice, record lengths are often 1 year or shorter. In order to handle this issue an assumption is made that the phase/amplitudes of response sinusoids with similar frequencies (i.e., those whose first three Doodson numbers are identical) are in the same proportion as those of the equilibrium response under the reasonable premise that the ocean response should be similar at similar frequencies. In such a cluster large equilibrium peaks are surrounded by small subsidiary peaks in frequency-space which provide "nodal modulations" (or more correctly, "satellite modulations") to the main peak. The appearance of the total signal will be a sinusoid whose phase and amplitude varies slowly with time. These changes are slow enough to be considered effectively constant for record lengths of up to 1 year. At much shorter record lengths another problem arises. The frequency resolution further degrades until even dissimilar constituents are unresolvable. The best solution is to apply inference. This technique for finding the absolute phase/amplitude requires that the relative differences in phase/amplitude between the two unresolved constituents is known from other nearby data. If this is not the case, it is thought best to either discard the smaller constituents and fit only to the largest in a given frequency interval, or to use the equilibrium response to establish the desired differences.

Another drawback of classical analysis is that it provides no easy way to determine whether the resulting phase/amplitude of a given sinusoid is meaningful in a deterministic way (i.e., it is truly a tidal line), or whether it results from fitting to a component of the non-tidal broad-spectrum variability. In general a fit is likely to include elements of both and some kind of confidence interval for the deterministic part is useful. To address this issue, the "response" method was invented (Munk and Cartwright, 1966). Although this provides better results than classical harmonic analysis, it has not found widespread use.

Further problems with classical harmonic analysis arise in coastal regions where the tidal response is in the form of a wave propagating onshore. In large estuaries, the seasonal change in salinity and flow may change the dynamic response but as these changes can vary from year to year the tidal process is not really stationary. Instead spectral peaks are broadened so that they are no longer pure lines, but, depending on the situation, such variations may be treated as lines in the analysis. Within smaller estuaries, tidal height variations may be significant compared to water column depth and a variety of non-linear effects can occur. For example, flood periods shorten and intensify and ebbs lengthen. As long as these effects are reasonably deterministic they may be handled by adding extra "shallow water" constituents which occur at sum/difference frequencies of the major constituents. More problematic in these regions are the effects of internal variability. Tidal interactions with varying topography can produce large internal waves and bores whose characteristics are highly sensitive to ambient stratification. In such cases the assumption of "line" frequencies becomes questionable and other techniques such as wavelet analysis have been suggested (Jay and Flinchem, 1999). More comprehensive descriptions of analysis techniques, their use, and their limitations is given in, e.g., Foreman et al. (1995) and Godin (1991).

Here, we describe T_TIDE, a package of routines that can be used to perform classical harmonic analysis with nodal corrections, inference, and a variety of user-specified options. Predictions can also be made using the analyzed constituents. There are several novel features. First, although the harmonic analysis algorithm with nodal corrections, etc., itself is not original (other than conversion to complex algebra), it is implemented in MATLAB, an analysis package widely used by oceanographers. This allows for easy use within the framework of a complete analysis involving plotting of raw data, scatter plots, and so forth. Second, the code is written directly in matrix terms and thus relatively easy to understand and modify if required. Finally, in order to differentiate between true deterministic (line) frequencies and broad-spectrum variability, confidence intervals for the estimated tidal parameters are computed using one of several user-selectable algorithms. The package is made up of a number of files each of which contain one or more functions. User-callable functions generally have a "*t_*" prefix to prevent namespace collisions.

The paper is composed of four parts. In the first the form of the equilibrium potential is described. In the second part the mathematical basis of the technique for making phase/amplitude estimates is described. In the third part the generation of confidence intervals is outlined. Finally an illustrative example is discussed.

## 2. Tidal potential

The effect of gravitational force vectors from the sun and moon, $\mathbf{F}$, can be written as the gradient of a scalar potential $V$, $\mathbf{F} = -\nabla V$. The magnitude of this potential at the earth's surface at any time obviously is dependent on the relative positions of the earth, moon, and sun. In Doodson's development (Doodson, 1954, described in Godin, 1972) the potential is written as a function of lunar time $\tau$ (defined to begin at "lunar midnight") and other astronomical variables (which are also functions of time):

$s$ is the mean longitude of moon, $h$ the mean longitude of sun, $p$ the longitude of perigee, $N'$ the negative of the longitude of the ascending node, and $p'$ is the longitude of perihelion, where all terms are in units of cycles. These variables can be evaluated for a given Julian date using the function *t_astron* which implements formulas in Seidelmann (1992). Their effects are combined with the Doodson number set for a particular constituent $\{i', j', k', l', m', n'\}$ into the astronomical argument $V_a = i'\tau + j's + k'h + l'p + m'N' + n'p'$. Sets with a common $i'$ are called a species (thus the slow, diurnal, and semidiurnal species for $i' = 0, 1$, and 2, respectively), and sets with common $i'j'k'$ are called a subgroup. The constituent frequency $\sigma$ is defined as $\sigma = 2\pi\,\mathrm{d}V_a/\mathrm{d}t$. The tidal potential is then written in the form

$$V = \sum_{i'=0}^{3}\Bigg[G_{i'}(\theta)\sum_{j'k'l'm'n'} A'_{i'j'k'l'm'n'}\cos(2\pi V_a) + G'_{i'}(\theta)\sum_{j'k'l'm'n'} B'_{i'j'k'l'm'n'}\sin(2\pi V_a)\Bigg]. \quad (1)$$

For a given Doodson number set either $A'$ or $B'$ is non-zero, but not both. These constants are tabulated and stored in data structures that can be loaded using *t_getconsts*. The geodetic functions $G_{i'}$ and $G'_{i'}$ vary with species $i'$ and latitude $\theta$, and also depend on such constants as the radius of the earth and the masses and separations of the earth, moon, and sun. The equilibrium amplitude for a constituent is defined as $GA'/g$ or $G'B'/g$, where $g$ is the gravitational acceleration, and can be generated for a particular latitude using *t_equilib*.

## 3. Phase/amplitude estimates

The algorithm used here for making phase and amplitude estimates is based on algorithms and FORTRAN code described by Godin (1972), Foreman (1977), and Foreman (1978). However, unlike those authors we use complex algebra directly rather than deal with sine and cosine fits separately. This has the advantage of unifying the treatment for scalar (e.g., pressure) and vector (e.g., horizontal currents) time series which are represented as complex numbers $u + iv$. Note that the complex form for currents is based on a physical model of a rotating current vector, and is only valid for linear or nearly linear tidal waves. In some cases it may be better to treat, e.g., along and across-channel currents as two separate scalar time series.

Consider a time series of observations $y(t), t = t_1, t_2, \ldots, t_M$ arranged in a vector, where the observation times are regularly spaced at an interval $\Delta t$ (default 1 h) and $M$ is an odd number (an endpoint is discarded if required). The time axis is defined such that the origin (or central time) is at $t_{(M+1)/2}$. Some missing observations can be handled by using a "missing data" marker in the input vector (by MATLAB convention this is NaN, the IEEE arithmetic representation for Not-a-Number). This regular interval restriction does not arise from the least-squares fit itself but rather from the automated constituent-selection algorithm and is also a requirement when spectra are estimated in one of the confidence interval algorithms. This time series may be composed of either real or complex numbers. The time series is passed to the analysis program *t_tide* along with a variety of (mostly optional) parameters.

The tidal response is modelled as

$$x(t) = b_0 + b_1 t + \sum_{k=1,\ldots,N} a_k e^{i\sigma_k t} + a_{-k} e^{-i\sigma_k t}, \quad (2)$$

where $N$ constituents (each with unique Doodson number sets) are used. Each constituent has a frequency $\sigma_k$ which is known from the development of the potential, and a complex amplitude $a_k$ which is not known, although if $y(t)$ is a real time series $a_k$ and $a_{-k}$ are complex conjugates. A possible offset and (optional) linear drift is handled by the first two terms. The traditional approach uses real sinusoids:

$$x(t) = b_0 + b_1 t + \sum_{k=1,\ldots,N} A_k \cos(\sigma_k t) + B_k \sin(\sigma_k t) \quad (3)$$

and is related to Eq. (2) by $A_k = a_k + a_{-k}$ and $B_k = \mathrm{i}(a_k - a_{-k})$. The real representation is more convenient for the linear error analysis described later.

Constituents can be chosen from a list of 45 astronomical and 101 shallow-water constituents. Data structures containing names and other information about these constituents are loaded using *t_getconsts*. There are several alternatives for selecting constituents. For general use, an automated selection algorithm (following Foreman, 1977) is in place, which works as follows. A basis of all astronomical and 24 of the most important shallow-water constituents are gathered together. All constituents are listed in order of pre-defined importance based on equilibrium amplitudes. Less important constituents whose frequencies are less than a Rayleigh resolution limit $\alpha(N\Delta t)^{-1}$ (with default $\alpha = 1$) apart from more important constituents in frequency are discarded. Additional shallow-water constituents can be specified if required. If the relative phase/amplitude of two constituents that are otherwise unresolvable is known from other sources, then an inference procedure can be carried out. Alternatively, constituent lists can be explicitly specified.

The least-squares fit are the coefficients minimizing

$$E = \sum_m |x(t_m) - y(t_m)|^2 = ||Ta - y||^2, \tag{4}$$

where $y = [y(t_1), y(t_2), \ldots, y(t_M)]'$, $a = [b_0, b_1, a_1, a_{-1}, a_2, a_{-2}, \ldots, a_{-N}]'$, and $T$ is an $M \times 2N + 2$ matrix of linear and sinusoidal basis functions evaluated at observation times. The solution is found using the Matlab "$\backslash$" matrix division operator.

Once the fit has been performed, various corrections are applied. These are generated in *t_vuf*. First, the phase of the constituent response is usually reported as "Greenwich phase" $g_k$, that is, phase referenced to the phase of the equilibrium response at 0° longitude (the Greenwich meridian). This can be interpreted as reporting the phase of the response at the time when the equilibrium forcing is at its largest positive value at 0° longitude. It is simplest to find the fitted phase at the central time of the record ($t = 0$); the equilibrium phase $v_k$ is then just $V_a$ for the given constituent computed at the Julian date corresponding to this central time, with possible adjustments of $\frac{1}{4}$, $\frac{1}{2}$, or $\frac{3}{4}$ cycle depending on whether $A$ or $B$ is non-zero, and their signs.

Second, if a latitude is specified, then nodal or satellite corrections are computed as follows. Consider a main peak of index $k$ with satellites with indices $kj$. The effect of the different satellites will be to slowly modulate the phase/amplitude of the main peak over various periods, usually more than 8 years. Our fitted response $\hat{a}_k$ over some period can then be written as a modification of the "true" response of the main constituent $a_k$, in which the amplitude is changed by a factor $f_k$ and the phase by an angle $u_k$ due to the presence of the satellites. $f_k$ and $u_k$ are called the nodal correction amplitude and phase, respectively. That is,

$$\hat{a}_k e^{i\sigma_k t} = f_k a_k e^{i\sigma_k t + iu_k} = a_k e^{i\sigma_k t} + \sum_j a_{kj} e^{i\sigma_{kj} t}. \tag{5}$$

Cancelling common terms, we have

$$f_k e^{iu_k} = 1 + \sum_j \frac{a_{kj}}{a_k} e^{i(\sigma_{kj} - \sigma_k)t} \approx 1 + \sum_j \frac{a_{kj}}{a_k}. \tag{6}$$

The final approximation will hold as long as $(\sigma_{kj} - \sigma_k)t$ remains "small" (i.e., $N\delta t \ll 8$ years). In general the true phases and amplitudes of the satellites are not known. However, since their frequencies are very similar to that of the main peak it is standard to assume the ratio of true amplitudes is the same as the ratio of amplitudes in the equilibrium response, and the difference in true phases will be equal to the difference in equilibrium phases. The nodal corrections are thus computed from the equilibrium response Eq. (1). A latitudinal dependence arises from the geodetic functions. $G_1'$ is zero at the equator and a crude limiting is used to prevent some diurnal corrections from getting overly large. The validity of using the latitude-dependent equilibrium response to predict an aspect of the dynamic behavior in one part of an ocean basin in such a simple way is not clear. If the record length is longer than 1 year the comparison of successive 1 year analyses with and without nodal correction can be used to test the validity of this process. Note that if the time series to be analyzed is longer than 18.6 years in length then the "true" satellite amplitude/phase terms can be estimated directly (Foreman and Neufeld, 1991) but this is not currently possible in T_TIDE.

The products of the analysis above are a pair of complex values $\{a_k, a_{-k}\}$, possibly corrected for nodal modulation, for each constituent $k$. These are converted into standard parameters:

$$L_k = |a_k| + |a_{-k}|, \tag{7}$$

$$l_k = |a_k| - |a_{-k}|, \tag{8}$$

$$\theta_k = \frac{ang(a_k) + ang(a_{-k})}{2} \bmod 180, \tag{9}$$

$$g_k = v_k - ang(a_k) + \theta_k. \tag{10}$$

For horizontal currents, these parameters describe the features of an ellipse traced out by the tip of the velocity vector: the length of the semi-major and semi-minor axis of the ellipse ($L_k, l_k$, respectively), the inclination of the northern semi-major axis counter-clockwise from due east $\theta_k$, and the Greenwich phase $g_k$. If $l_k > 0$ ($<0$) then the ellipse is traced in a counter-clockwise (clockwise) direction. For scalar time series, the parameter $L_k$ is the amplitude, and $l_k, \theta_k \equiv 0$ (the ellipse degenerates to a line along the positive axis). Note that the restriction of the definition of inclination to the northern axis (via the modulo operator) means that analyses of constituents whose ellipses are aligned in an east/west direction may have inclinations that fluctuate between close to $0^\circ$ and nearly $180^\circ$ due to noise. These apparently large jumps are an artifact of the restriction but do not represent similarly large changes in the physical behavior.

Once ellipse parameters are found, these can be used for further analysis. They can also be used to generate predictions at other times using *t_predic*. Nodal corrections in *t_predic* are computed at the time series midpoint so that it is an exact inverse of *t_tide*.

## 4. Confidence intervals

One drawback of classical harmonic analysis is that the degree to which a given constituent represents true tidal energy as opposed to the energy of a broad-band non-tidal process is not determined. This is useful information for two reasons: first, it allows one to make better estimates of the tidal behavior, and second, it can allow one to quantitatively compare different analyses. There are two steps to producing confidence intervals. First, we must form an estimate of the characteristics of non-tidal or residual "noise" affecting the $a_k$ (or $A_k, B_k$). Second, we must convert these estimates into confidence intervals for the standard parameters through a non-linear mapping. We discuss the situation of real time series first.

### 4.1. Residual noise (real)

After the harmonic analysis for an N-point real time series $y(t)$ is performed, we examine the structure of the residual series. In the simplest situation, the residuals are statistically Gaussian and uncorrelated in time. If this is the case then the total residual power $P_T = \sigma_x^2 = P/\Delta t$, where $P$ is the two-sided spectral density. The amplitude of the fit to sine and cosine terms ($A$ and $B$, respectively) will be contaminated by errors arising from unresolved noise components within a frequency interval of $\Delta f = (N\Delta t)^{-1}$ around the line. Thus $\sigma_A^2 = \sigma_B^2 = P\Delta f = \sigma_x^2/N$. It is unlikely that a geophysical series will be spectrally flat, and a more sophisticated approach used in *t_tide* is to find a local value of $P$ suitable for constituents in that neighborhood by making a spectral estimate from the residual time series (i.e., after the removal of all fitted constituents) and averaging the power over frequency bins in a window around the frequency of any constituent, neglecting bins in which fitted constituents reside. Here we chose a sequence of windows of width 0.4 cpd centered on $1, 2, 3, \ldots$ cpd (actually on multiples of the $M_2$ frequency, see the code for details). The value of $P$ appropriate to, say, semi-diurnal constituents would be estimated from the second of these bins.

### 4.2. Conversion to standard parameters (real)

A conversion from errors in the cos/sine amplitudes to errors in standard parameters (amplitude and phase) can be done using a linearized analysis. Consider a constituent $k$. Let $\xi = F(A_k, B_k)$ be a non-linear function of these parameters, either the amplitude or the Greenwich phase. Then if $\{A_k, B_k\}$ are independent random variables, we can find a linearized estimate of the standard error of $\xi$ in terms of the standard errors of the sinusoid amplitudes:

$$\sigma_\xi^2 = \left(\frac{\partial F}{\partial A_k}\right)^2 \sigma_A^2 + \left(\frac{\partial F}{\partial B_k}\right)^2 \sigma_B^2, \tag{11}$$

where the partial derivatives can be derived exactly (but tediously) from Eqs. (7)–(10).

Alternatively the non-linear mapping can be handled directly using a "parametric bootstrap" (Efron and Tibshirani, 1993). In this situation the residual variance estimates are used by the code to simulate a number of realizations or replications of the analysis by taking the estimates of the sinusoid amplitudes and adding Gaussian noise with the appropriate variance to them. All of these realizations are then converted non-linearly to standard parameters using Eqs. (11)–(10) and an estimate of the standard error computed from this replicate data set directly.

Once a standard error is determined, 95% confidence intervals can be estimated using standard techniques. Alternatively, a signal-to-noise power ratio (SNR) can be computed based on the square of the ratio of amplitude to amplitude error. Simulations performed in *t_synth* (and described in the text file *t_errors*) in which the variability of analyses carried out on a fixed data set with different noise realization are compared with estimated confidence intervals show that the linear procedure appears to be adequate for real time series (e.g. tidal height), as long as the SNR $> 10$, and is probably not bad for SNR as low as 2 or 3. The non-linear procedure gives similar results to the linearized procedure at high SNR, and is more accurate at low SNR.

### 4.3. Residual noise (complex)

A complex residual time series $u + iv$ can be modelled as bivariate white noise and variances $\sigma_u^2$, $\sigma_v^2$ and covariance $\sigma_{uv}$ computed. If we assume further that the noise in both components is not correlated ($\sigma_{uv} \approx 0$), then a coloured bivariate noise model can be used and variances assigned to real and imaginary parts of constituent amplitude separately on the basis of local spectral densities as described above. If it is suspected that $\sigma_{uv} \neq 0$ then it is recommended that the time series be rotated into a coordinate system in which this is true

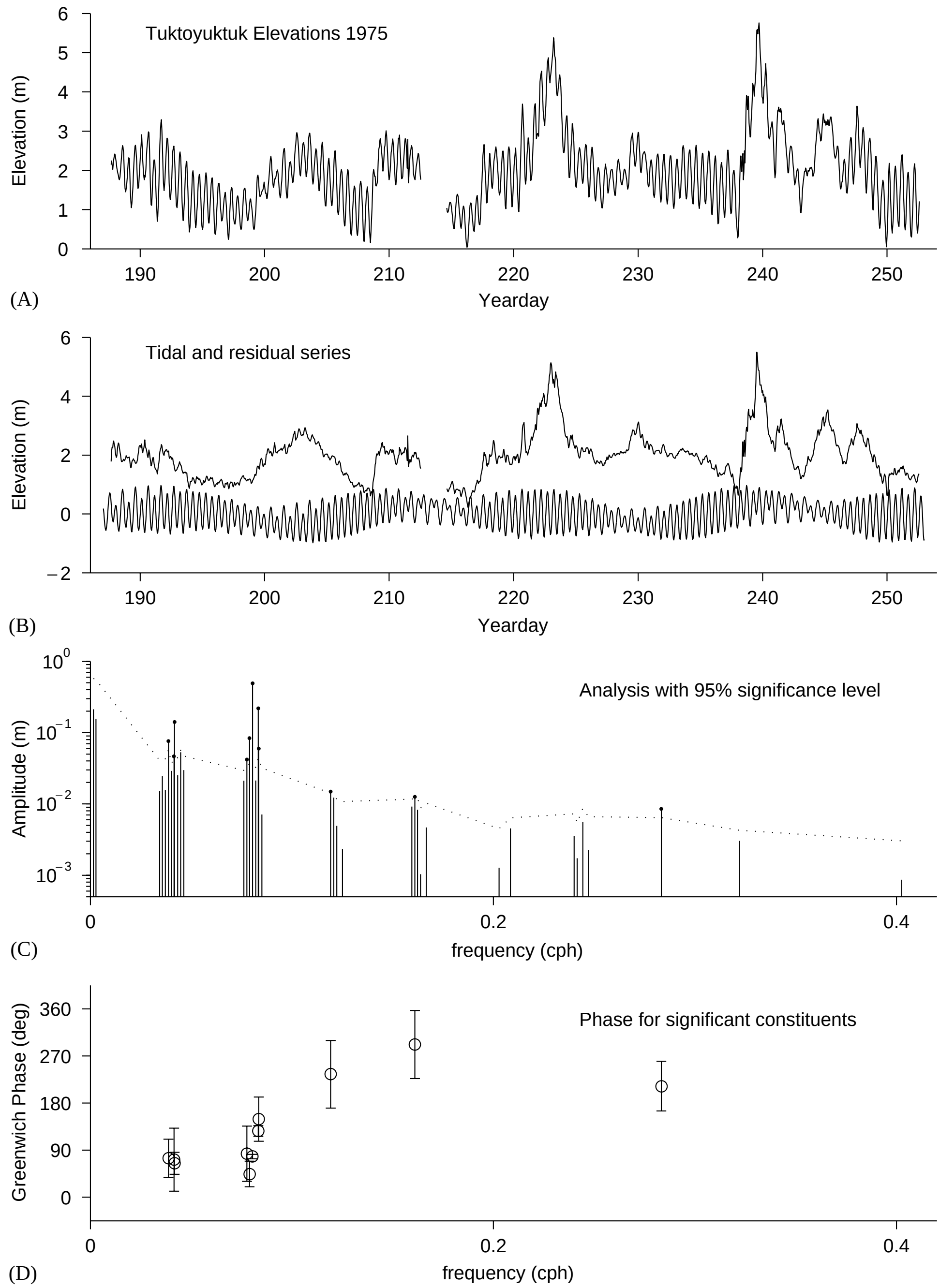


Fig. 1. Tuktoyuktuk analysis example. (A) Raw time series. (B) Upper curve is residual time series after removal of tidal signal. Lower curve is synthesized tidal series using significant constituents. (C) Amplitude of all analyzed components with 95% significance level. Note frequency dependence. Significant constituents are marked with solid circle. (D) Phase of significant constituents with 95% confidence interval.

Table 1
Analysis results for Tuktoyuktuk data set

File name: PAPEROUT.txt
Date: 17-Aug-2001
Nobs = 1584, ngood = 1510, record length (days)=66.00
Start time: 06-Jul-1975 01:00:00
Rayleigh criterion = 1.0
Greenwich phase computed with nodal corrections applied to amplitude and phase relative to center time
x0 = 1.98, x trend = 0
Var(x) = 0.82196 var(xp)= 0.21224 var(xres) = 0.60972
Percent var predicted = 25.8%
Tidal amplitude and phase with 95% CI estimates

| Tide | Freq | Amp | Amp_err | Pha | Pha_err | Snr |
|---|---|---|---|---|---|---|
| MM | 0.00151 | 0.2121 | 0.503 | 263.34 | 161.41 | 0.18 |
| MSF | 0.00282 | 0.1561 | 0.526 | 133.80 | 188.82 | 0.088 |
| ALP1 | 0.03440 | 0.0152 | 0.044 | 334.95 | 150.82 | 0.12 |
| 2Q1 | 0.03571 | 0.0246 | 0.044 | 82.69 | 106.21 | 0.31 |
| Q1 | 0.03722 | 0.0158 | 0.045 | 65.74 | 160.30 | 0.12 |
| *O1 | 0.03873 | 0.0764 | 0.055 | 74.23 | 43.35 | 1.9 |
| NO1 | 0.04027 | 0.0290 | 0.035 | 238.14 | 74.68 | 0.69 |
| *P1 | 0.04155 | 0.0465 | 0.045 | 71.88 | 70.96 | 1.1 |
| *K1 | 0.04178 | 0.1405 | 0.059 | 64.81 | 23.49 | 5.7 |
| J1 | 0.04329 | 0.0253 | 0.050 | 7.32 | 129.76 | 0.25 |
| OO1 | 0.04483 | 0.0531 | 0.059 | 235.75 | 72.96 | 0.81 |
| UPSI | 0.04634 | 0.0298 | 0.055 | 91.73 | 137.06 | 0.29 |
| EPS2 | 0.07618 | 0.0211 | 0.030 | 184.59 | 104.65 | 0.51 |
| *MU2 | 0.07769 | 0.0419 | 0.034 | 83.23 | 48.82 | 1.5 |
| *N2 | 0.07900 | 0.0838 | 0.035 | 44.52 | 25.54 | 5.9 |
| *M2 | 0.08051 | 0.4904 | 0.035 | 77.70 | 4.51 | 1.9e+02 |
| L2 | 0.08202 | 0.0213 | 0.037 | 35.22 | 113.22 | 0.33 |
| *S2 | 0.08333 | 0.2197 | 0.038 | 126.72 | 9.14 | 34 |
| *K2 | 0.08356 | 0.0598 | 0.043 | 149.12 | 46.60 | 2 |
| ETA2 | 0.08507 | 0.0071 | 0.033 | 246.05 | 207.25 | 0.048 |
| *MO3 | 0.11924 | 0.0148 | 0.014 | 234.97 | 67.38 | 1.1 |
| M3 | 0.12077 | 0.0123 | 0.014 | 261.57 | 62.11 | 0.81 |
| MK3 | 0.12229 | 0.0049 | 0.012 | 331.60 | 144.92 | 0.18 |
| SK3 | 0.12511 | 0.0023 | 0.010 | 237.69 | 219.86 | 0.054 |
| MN4 | 0.15951 | 0.0092 | 0.011 | 256.47 | 69.76 | 0.68 |
| *M4 | 0.16102 | 0.0126 | 0.011 | 291.78 | 65.09 | 1.4 |
| SN4 | 0.16233 | 0.0083 | 0.011 | 270.85 | 91.22 | 0.54 |
| MS4 | 0.16384 | 0.0010 | 0.008 | 339.35 | 248.82 | 0.015 |
| S4 | 0.16667 | 0.0047 | 0.010 | 299.56 | 142.32 | 0.23 |
| 2MK5 | 0.20280 | 0.0013 | 0.005 | 310.10 | 181.03 | 0.067 |
| 2SK5 | 0.20845 | 0.0045 | 0.006 | 104.00 | 99.71 | 0.64 |
| 2MN6 | 0.24002 | 0.0035 | 0.007 | 271.24 | 133.30 | 0.22 |
| M6 | 0.24153 | 0.0017 | 0.006 | 158.88 | 197.43 | 0.093 |
| 2MS6 | 0.24436 | 0.0056 | 0.008 | 306.10 | 90.03 | 0.54 |
| 2SM6 | 0.24718 | 0.0023 | 0.007 | 298.92 | 175.13 | 0.11 |
| *3MK7 | 0.28331 | 0.0086 | 0.006 | 212.25 | 44.21 | 2.1 |
| MS | 0.32205 | 0.0030 | 0.004 | 42.43 | 75.29 | 0.55 |
| M10 | 0.40256 | 0.0009 | 0.003 | 198.23 | 209.99 | 0.089 |

(e.g., into along/across channel axes or into the principal axes).

### 4.4. Conversion to standard parameters (complex)

The linearized analysis now involves functions of four variables, since both $A_k = A_r + iA_i$ and $B_k = B_r + iB_i$ have real and imaginary parts, and analytic expressions for partial derivatives

$$\sigma_\xi^2 = \left(\frac{\partial F}{\partial A_r}\right)^2 \sigma_{A_r}^2 + \left(\frac{\partial F}{\partial A_i}\right)^2 \sigma_{A_i}^2 + \left(\frac{\partial F}{\partial B_r}\right)^2 \sigma_{B_r}^2 + \left(\frac{\partial F}{\partial B_i}\right)^2 \sigma_{B_i}^2 \tag{12}$$

become large. Some analytical simplification is possible by assuming that all four variables are independent.

The bootstrap approach can also be applied to the complex coefficients $a_k$. One minor complication that arises is that unless the noise is circular ($\sigma_u^2 = \sigma_v^2, \sigma_{uv} = 0$), the errors in $a_k$ and $a_{-k}$ are correlated with each other. The bootstrap process requires the generation of correlated noise replicates.

## 5. Example

The analysis of an example data set provided in Foreman (1977) is shown in Fig. 1 and Table 1. This example is included in datafile *t_example.mat*. The example data set consists of 66 days of hourly elevations with a 3 day gap. A tidal variation is visible superimposed on subtidal variability. The time series can be loaded and analyzed using the demonstration script *t_demo*. Code within this script illustrates how the programs are called. In this example, the automated constituent selection algorithm is used and it selects 35 constituents. In addition, one shallow water constituent ($M_{10}$) is manually added and two other constituents analyzed via inference. The $P_1$ constituent is inferred from $K_1$, and $K_2$ is inferred from $S_2$. Nodal corrections are performed. A linear trend is not included in the analysis. The coloured bootstrap analysis is used to determine significance and confidence intervals. Table 1 gives the output of the program. In the first column the name of the constituent is given. Significant constituents (those with SNR in the last column > 1) are marked with a "*". The SNR is the squared ratio of amplitude (third column) to the error in amplitude (fourth column). The error factors (and hence SNR) will change slightly in repeated analyses due to the stochastic nature of the bootstrap procedure but amplitudes and phases themselves will be invariant. Frequencies (first column) are listed in cph and Greenwich phase/phase error (fifth and sixth columns) in degrees. Eleven constituents were judged to be significant (only 6 would be significant at an SNR cutoff of 2). Fig. 1B shows the residual series and an elevation series synthesized from the significant constituents (note that it spans the 3-day gap). Results of the analysis are shown in a spectrum in Fig. 1C. Most of the significant constituents are in the diurnal and semidiurnal bands ($\approx$0.04, and $\approx$0.08 cph, respectively) although several higher-frequency constituents also appear to be marginally significant. In spite of the large amount of energy in the fortnightly band ($\approx$0.002 cph) the fitted constituents there are apparently not significant. The analyzed phases of significant constituents are shown in Fig. 1D. The significant constituents generally have reasonably small phase errors.

## 6. Summary

Separation of tidal and non-tidal energy is an important task in any analysis of oceanic time series. Here, we discuss the theoretical foundation and implementation details of a MATLAB package for classical harmonic analysis. The package can also compute confidence intervals for the tidal parameters using one of three different sets of assumptions about the structure of residual noise. An example is provided to show typical results. The code is available at http://www.ocgy.ubc.ca/~rich, or the IAMG Server.

## Acknowledgements

Elements of this MATLAB toolbox for tidal harmonic analysis were developed over the last few years as part of the authors' research into tidal phenomena. We want to first acknowledge Mike Foreman's major contribution in providing the FORTRAN codes and documentation to the community in 1977–78 that served as our starting point. Rich Signell first derived the formulae for the linearized error analysis used here, and Julio Candela and Jim Irish also provided useful input. Support for this effort was provided by Natural Sciences and Engineering Research Council of Canada under grant OGPO194270 (RP) and the National Science Foundation and Office of Naval Research (SL and RB).

## References

Doodson, A.T., 1954. Appendix to circular-letter 4-H. The harmonic development of the tide-generating potential. International Hydrographic Review 31, 37–61.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap, Vol. 57. Monographs on Statistics and Applied Probability. Chapman & Hall, New York, 436pp.

Foreman, M.G.G., 1977. Manual for tidal heights analysis and prediction. Pacific Marine Science Report 77-10, Institute of Ocean Sciences, Patricia Bay, Sidney, BC, 97pp.

Foreman, M.G.G., 1978. Manual for tidal currents analysis and prediction. Pacific Marine Science Report 78-6, Institute of Ocean Sciences, Patricia Bay, Sidney, BC, 57pp.

Foreman, M.G.G., Crawford, W.R., Marsden, R.F., 1995. De-tiding: theory and practise. In: Lynch, D.R., Davies, A.M. (Eds.), Quantitative Skill Assessment for Coastal Ocean Models. Vol. 47. Coastal and Estuarine Studies. American Geophysical Union, pp. 203–239.

Foreman, M.G.G., Neufeld, E.T., 1991. Harmonic tidal analyses of long time series. International Hydrographic Review 68 (1), 85–108.

Godin, G., 1972. The Analysis of Tides. University of Toronto Press, Toronto, 264pp.

Godin, G., 1991. The analysis of tides and currents. In: Parker, B.B (Ed.), Tidal Hydrodynamics. Wiley, New York, pp. 675–709.

Jackson, L.E., 1986. Digital Filters and Signal Processing. Kluwer Academic Publishers, New York, 259pp.

Jay, D.A., Flinchem, E.P., 1999. A comparison of methods for analysis of tidal records containing multi-scale non-tidal background energy. Continental Shelf Research 19, 1695–1732.

Munk, W.H., Cartwright, D.E., 1966. Tidal spectroscopy and predication. Philosophical Transactions of the Royal Society of London, Series A 259, 533–581.

Seidelmann, P.K. (Ed.), 1992. Explanatory Supplement to the Astronomical Almanac. United States Naval Observatory, Mill Valley, Calif., University Science Books. 752pp.