Continental Shelf Research ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Continental Shelf Research

journal homepage: www.elsevier.com/locate/csr

Enhancing tidal harmonic analysis: Robust (hybrid L^1/L^2) solutions

Keith E. Leffler *, David A. Jay

Department of Civil and Environmental Engineering, Portland State University, PO Box 751, Portland, OR 97207-0751, USA

ARTICLE INFO

Article history:
Received 1 February 2007
Received in revised form
24 February 2008
Accepted 28 April 2008

Keywords:
Tides
Tidal analysis
Harmonic analysis
Robust statistics

ABSTRACT

Traditional harmonic analysis of tides is highly sensitive to omnipresent environmental noise. Robust fitting is an extension of the ordinary least squares calculation of harmonic analysis that is more resistant to broad spectrum noise. Since the variance of the amplitude and phase is calculated from the power spectrum of the residual, a calculation that filters broad spectrum noise and reduces the residual variance will increase the confidence of the computed parameters, and also allows more low-amplitude constituents to be resolved from the background noise. Improvement in confidence and resolution of more constituents has obvious benefits to the resolution of both seasonal and long-term variation of amplitude and phase of tidal constituents. Using a 6 month calculation window, confidence intervals were systematically reduced by 30–85% over results calculated with standard methods, with an increase in resolved constituents of 20–75%. The analysis was carried out with Matlab, using the t-tide package [Pawlowicz, R., Beardsley, B., Lentz, S., 2002. Classical tidal harmonic analysis with errors in matlab using t-tide. Computers and Geosciences 28, 929–937], with modifications to accommodate Matlab's implementation of the Iteratively Reweighted Least Squares algorithm.

1. Introduction

There is an increasing need to understand the dynamic behavior of coastal and estuarine waters and the response of these coastal systems to anthropogenic and climate changes. The predevelopment state of many estuaries is of great interest for management purposes, and understanding estuarine evolution poses interesting dynamical challenges. Tides provide the longest instrumental records available to address such questions. To fully utilize historical tidal records, however, it is necessary to develop better analysis strategies, because these records are often short, sparse and/or noisy. Non-tidal noise is introduced into the tidal signal by recording and transcription errors and climatic events. Traditional ordinary least squares (OLS) minimization is highly sensitive to these non-tidal components in the observed signal. Essentially, OLS over-fits the non-tidal components in an attempt to minimize the total residual error. This is not ideal, because it causes the results to be quite sensitive to omnipresent environmental noise. Noise affects not only the estimated tidal constituent amplitude and phase, but, equally importantly, the estimated variance of these quantities (Munk and Hasselmann, 1964; Pawlowicz et al., 2002).

It is tempting to exclude 'bad' data, and include only 'good' data in the calculations. This approach is impractical, for several reasons. First, it is labor intensive and inherently subjective. Without a consistent, reliable rejection method, only the most egregious outliers may be conclusively identified as 'bad'. At the other extreme, removing all data contaminated by storm events may leave insufficient data for analysis. For many analyses, it is necessary to resolve K_1, P_1 and O_1 from the diurnal band, and K_2, M_2, N_2 and S_2 from the semi-diurnal band. Some of these constituents (e.g. K_1 and P_1, K_2 and S_2) differ by ~2 cycles/yr, and resolution of these constituents requires approximately 190 days of nearly continuous hourly data (Foreman, 1977, revised 1996). Estuarine tidal records may show variation in tidal parameters due to riverine flow variations and other causes, and it is desirable to minimize the calculation window length, while still resolving an adequate set of constituents.

A major objective of this work is to test the 'Iteratively Reweighted Least Squares' (IRLS) algorithm, which consistently and automatically reduces the influence of non-tidal variation and minimizes data rejection, using only the tidal record itself. This is especially important for historic records, which often consist of partially complete, manually transcribed, high–low data (four observations/tidal day). IRLS is a form of robust statistical fitting that reduces the contribution of high-leverage data points, in order to improve the overall fit. It has been successfully applied to other geophysical problems (e.g. Bube and Langan, 1997). IRLS is a common implementation of hybrid L^1/L^2 minimization, a subset of the general L^p minimization problem. For low-noise data, the algorithm retains the high frequency resolution of an L^2 algorithm (Jay and Flinchem, 1999), while the handling of high-amplitude outliers approaches that of an L^1 minimization, reducing their

* Corresponding author. Tel.: +1 503 725 2961; fax: +1 503 725 5950.
E-mail addresses: leffler@cecs.pdx.edu (K.E. Leffler), djay@cecs.pdx.edu (D.A. Jay).

0278-4343/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.csr.2008.04.011

influence on the overall solution (Darche, 1989). By reducing the influence of outliers, IRLS fitting reduces the estimated variance, yielding increased confidence in the parameter estimates. We present the mathematical details of the algorithm in Section 2, followed, in Section 3, by numerical examples that show how IRLS more accurately recovers a known signal contaminated with outliers, Gaussian noise, and a simulated storm surge. A comparison of the performance of OLS and IRLS methods applied to observed tidal height data follows in Section 4. Implementations for analysis of coastal records are discussed in Section 5.

The analyses presented here were performed using a modified version of the t-tide Matlab package (Pawlowicz et al., 2002). To simplify the development, we have restricted our investigations to uniformly time-sampled tidal height, and have investigated neither two-dimensional currents nor irregularly sampled data.

2. Mathematical background

Harmonic tidal analysis is usually conducted using OLS, which gives each data point equal weight in the solution. It is well known that tidal records include non-tidal events such as storm surges. IRLS introduces an algorithmically calculated weighting function to reduce the contribution of large residuals on the overall solution. For tidal analysis, the weighting function reduces the influence of non-tidal events. It is helpful to review the mathematical background of OLS and IRLS fitting and the harmonic tidal model as a prelude to tests of IRLS with artificial and real data.

2.1. Ordinary and iteratively reweighted least squares

Common general regression techniques, including ordinary and weighted least squares, are based on maximum likelihood estimates, termed M-estimators (Fox, 2002). Assume a linear model $h = Ax$, with h the observed values, A the basis function and x the set of unknown coefficients. For tidal analysis, the system is overdetermined, since the number of observations will inevitably exceed the number of constituents (Munk and Cartwright, 1966). The normal course is to seek a solution for x that minimizes an objective function ρ of the residual r, with the residual defined as $r = h - x^{\mathrm{T}}A$. The general M-estimator minimizes

$$\sum_{i=1}^{n} \rho(r_i) = \sum_{i=1}^{n} \rho(h_i - x_i^{\mathrm{T}} A) \tag{1}$$

Letting $\psi = \partial\rho/\partial x$, taking $\partial/\partial x$ (1), and setting the result equal to zero produces a set of equations (2) for the coefficients x:

$$\sum_{i=1}^{n} \psi(h_i - x_i^{\mathrm{T}} A) x_i^{\mathrm{T}} = 0 \tag{2}$$

By defining a weight function ω as

$$\omega(r) = \frac{\psi(r)}{r} \tag{3}$$

the estimating functions may be written

$$\sum_{i=1}^{n} \omega_i (h_i - x_i^{\mathrm{T}} A) x_i^{\mathrm{T}} = 0 \tag{4}$$

which minimizes the sum of weighted residuals

$$\sum_{i=1}^{n} \omega_i^2 r_i^2$$

Eq. (4) may be solved as (Moler, 2004)

$$x = (A^{\mathrm{T}} \omega A)^{-1} A^{\mathrm{T}} \omega h \tag{5}$$

If ω is set to the diagonal of the identity matrix I, the equation reduces to the OLS solution (Moler, 2004):

$$x = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} h \tag{6}$$

For the weighted case, ω is implicitly defined, and Eq. (5) is solved iteratively for x and w, while the OLS case (6) is directly solvable for x. The weighting function w is defined recursively, by relating the residual at iteration i to the weight at iteration $i+1$. The usual course is to set $\omega = diag(I)$ for the first iteration.

The 'traditional' weighting functions are Huber's function and Tukey's bisquare, defined in terms of a normalized residual R. Huber's weight function is defined as

$$\omega_{huber} = \frac{1}{\max(1, |R|)}$$

while Tukey's bisquare is defined as

$$\omega_{bisquare} = \begin{cases} (1 - R^2)^2, & |R| < 1 \\ 0, & |R| \geqslant 1 \end{cases}$$

The normalized residual R is computed as function of generic signal deviation statistic $\hat{s}$ and tuning constant τ. Standard deviation is a non-robust statistic, because a single sample of arbitrary deviation will affect the standard deviation by an arbitrary amount. The median absolute deviation of the residual $MAD(r)$ is a more robust statistic, as one half of the samples may be affected by noise without affecting the statistic (Fox, 2002). A common estimate of deviation is

$$\hat{s} = \frac{MAD(r)}{0.6745} \tag{7}$$

The constant 0.6745 makes the estimated parameters unbiased for a normally distributed error (Mathworks, 2006). The tuning constant τ essentially controls the width of the weighting window relative to the distribution of the residuals. A lower value of tuning constant imposes a greater penalty on outlying data.

Details of the IRLS algorithm and alternate weighting functions are provided in Appendix A.

2.2. Harmonic tidal model

The traditional representation of the tidal height h_i, at time t_i, with known tidal constituent frequencies $\theta_{1\ldots n}$ and unknown amplitudes a, b and c_0 is

$$h_i = c_0 + \sum_{k=1}^{n} (a_k \cos(2\pi\theta_k t_i) + b_k \sin(2\pi\theta_k t_i)) \tag{8}$$

or $h = c_0 + Ax$ with basis function

$$A = \begin{bmatrix} \cos(2\pi\theta_1 t_1) & \cdots & \cos(2\pi\theta_n t_1) & \sin(2\pi\theta_1 t_1) & \cdots & \sin(2\pi\theta_n t_1) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \cos(2\pi\theta_1 t_m) & \cdots & \cos(2\pi\theta_n t_m) & \sin(2\pi\theta_1 t_m) & \cdots & \sin(2\pi\theta_n t_m) \end{bmatrix} \tag{9}$$

and unknown parameters

$$x = [a_1 \;\; \ldots \;\; a_n \;\; b_1 \;\; \ldots \;\; b_n]^{\mathrm{T}} \tag{10}$$

The standard OLS solution is obtained from Eq. (6), while the IRLS solution for x is obtained by iterative application of Eq. (5), using a weight function as defined in Appendix A. The constituent frequencies θ are selected according to the length of record (LOR) (Foreman, 1977, revised 1996). LOR and noise inevitably limits the frequency resolution for any harmonic tidal analysis. Given a record of n samples spaced evenly spaced by Δt in time, the Rayleigh criterion states that the minimum resolvable frequency difference is $(n\Delta t)^{-1}$. The Foreman procedure selects constituents based on expected relative importance and frequency separation. Because many constituents are separated by 1 cycles/yr or less, only a fraction of the >500 astronomical constituents can be resolved for any reasonable length record. The Foreman procedure identifies 45 astronomical and 24 shallow water candidate constituents for records up to 1.3 years in length, though some constituents may be excluded from the final analysis based on a signal-to-noise significance test, described below. A record greater than 18.6 years of hourly data will resolve a sizable fraction of the 500 astronomical constituents. However, the necessary assumption that all constituents are constant for a period of 18.6 years may not be justified for many coastal and estuarine records.

Once the coefficients x are estimated, the amplitude α of constituent j is found as $\alpha_j = \sqrt{a_j^2 + b_j^2}$, and the phase angle β is found as $\beta_j = Imag(\log(a_j + ib_j))$ (Mathworks, 2006). Nodal, or satellite, corrections are applied to correct for low-frequency variations in constituent amplitudes, and phase angles are corrected to Greenwich phase, as described by Pawlowicz et al. (2002).

Constituent parameters calculated by harmonic methods are statistical estimates, and hence have associated variances. Confidence intervals for the amplitudes and phases of the tidal constituents, and a signal-to-noise ratio (Pawlowicz et al., 2002) are calculated from the estimated variance. Pawlowicz et al. (2002) discusses methods of estimating parameter variance and confidence intervals, including a linearized analysis and nonlinear bootstrap procedures, assuming either white or colored noise. By default, we assume colored noise, and use the corresponding bootstrap method.

The nonlinear bootstrap method for colored noise bases the error estimate on average spectral density of the residual within bands centered on $0, 1, 2, \ldots, 8$ cycles/day (Pawlowicz et al., 2002). Similarly, Munk and Hasselmann (1964) use the background noise spectrum to estimate variance, though by a different method. A distinction needs to be made here between the weighted residual $R_\omega = \omega(h - Ax)$, used by the algorithm for calculating the estimated parameters and errors, and the unweighted residual $R_{uw} = h - Ax$, the difference between the predicted and observed signals. The weighted least squares calculation finds coefficients that minimize the sum of the weighted residuals. It is therefore appropriate to estimate the variance of the calculated amplitude and phase from the spectrum of the weighted residual $R_w = \omega(y - Ax)$.

For a given constituent with estimated amplitude $\hat{a}$, and estimated variance $\hat{\sigma^2}$, the signal-to-noise ratio (SNR) can be estimated as (Pawlowicz et al., 2002)

$$SNR = \left(\frac{\hat{a}}{\hat{\sigma}}\right)^2 \tag{11}$$

The nonlinear bootstrap method accurately estimates parameter amplitudes down to $SNR \approx 2-3$. Only tidal constituents with an estimated $SNR \geqslant 2$ are included in the results presented here. Parameters with an $SNR < 2$ are excluded from the calculation of predicted tidal height. As an example, the OLS calculation for calendar year 1999 for Astoria rejects 16 of 67 constituents because the SNR for these constituents is <2.

3. Numerical examples

In this section, we describe two examples that demonstrate that IRLS fitting can reduce the influence of non-tidal signal components on the overall solution. In the first, we use a simple linear model, affected by Gaussian noise and extreme outliers, placed to have maximum effect on the results. The OLS solution is significantly different than the known signal, while IRLS fitting removes most of the influence of the outliers and closely matches the known signal. In the second, we simulate a simple tide affected by noise, and then add a simulated storm surge. Though the constituent amplitudes are not statistically different from the known signal using either the OLS and IRLS methods, the confidence intervals of these parameters are significantly smaller for the IRLS analysis.

3.1. *The impact of outliers: a simple example*

The simplest data problem is a high-amplitude spike, a common manifestation of transcription errors and instrument glitches. To show the impact that well placed outliers can have, we use a simple linear model

$$y = at + b + \eta \tag{12}$$

with t is the time, a and b constant and η the Gaussian noise ($\mu = 0, \sigma = 0.5$). Outliers were added at early and late times in the signal record, to maximize their influence, or leverage, on the predicted slope of the line. The outliers, given the same weight as the other samples by the OLS calculation, effectively rotates the predicted line, as shown in Fig. 1a. The calculated fit differs from the pattern described by the majority of the data. In contrast, IRLS fitting down-weights the outliers, and closely matches the majority of the data.

The estimated values of the parameters a and b, and their 95% confidence are shown in Table 1. Notice the poor performance of OLS in the presence of outliers: The correct value of the slope, 0.5, is not within the 95% confidence interval (0.385 ± 0.064 for the OLS/outliers case). Likewise, the correct value of the intercept, 1.0, is not within the 95% confidence interval (6.83 ± 3.73). Fig. 1b shows the distribution of residuals, with a scaled Huber weighting curve overlaid for comparison. With this weighting curve, the influence of the outliers is significantly reduced, but not totally removed. The calculated values of the slope and intercept are much closer to the noise-only cases, and the correct values are within the 95% confidence interval. The confidence intervals of the robust fit case are wider than the noise-only cases since the outliers continue to have some influence, though it is significantly reduced from the OLS case.

3.2. *A simple tide model with a synthetic storm surge*

Storm surges are a common source of large amplitude environmental noise in tidal records. We construct a simple tide signal comprised of a mean sea level Z_0, a diurnal component K_1, semi-diurnal component M_2, Gaussian noise η, ($\mu = 0$, $\sigma = 0.1$) and synthetic storm surge s. K_1 and M_2 are each described by an amplitude a, phase ϕ and frequency θ. The sampling interval is given by Δt. At time step i, the height y is

$$\begin{aligned} y_i = Z_0 &+ a_{K1}\cos(2\pi\theta_{K1} i\Delta t - \phi_{K1}) \\ &+ a_{M2}\cos(2\pi\theta_{M2} i\Delta t - \phi_{M2}) + \eta_i + s_i \end{aligned} \tag{13}$$

As a first example, the storm surge is set to zero. As a second example the storm surge s builds exponentially over 10 h, peaks at 1 m, and exponentially decays over the next 48 h. A total of 144 hourly data points were used for the calculation. The noise + surge observations and the surge are shown in Fig. 2a. For comparison, histograms of the first-iteration residuals with a scaled Huber weight function were overlaid for the noise-only and noise + surge cases are shown in Fig. 2b and c, respectively. The noise + surge histogram shows obvious skew introduced by the surge.

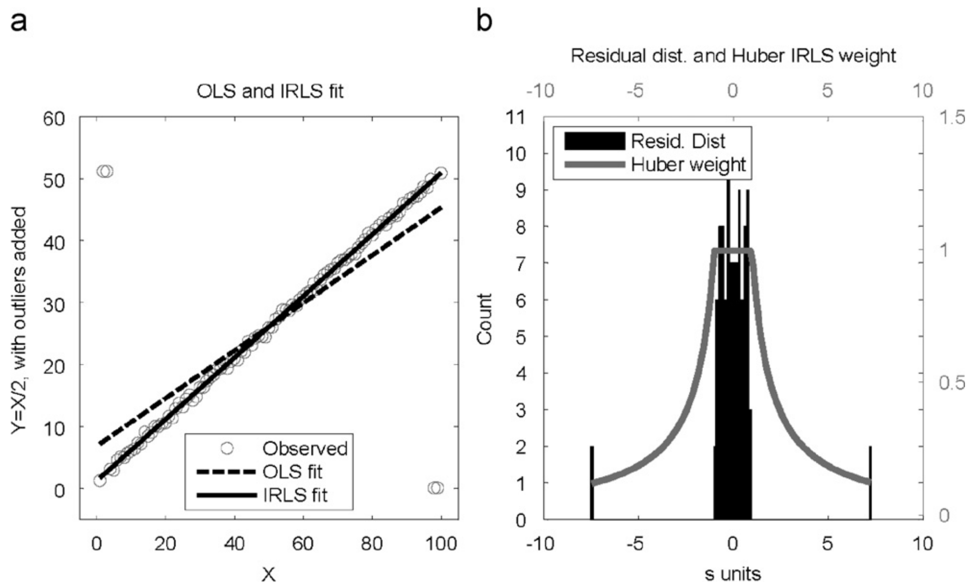

Fig. 1. A simple using artificial data. (a) Shows the OLS and IRLS fit and (b) shows a histogram of the initial (first iteration) residuals, with the scaled Huber weighting function overlaid.

Table 1
Linear model: estimated amplitude and 95% confidence intervals

Known and estimated parameters, 95% confidence intervals, by experiment	*a*	95% CI on *a*	*b*	95% CI on *b*
Model values	0.5		1	
OLS: noise only	0.4993	±0.0030	1.0574	±0.1722
IRLS (Huber): noise only	0.4992	±0.0030	1.0767	±0.1764
OLS: noise + outliers	0.3848	±0.0641	6.8251	±3.7278
IRLS (Huber): noise + outliers	0.49	±0.0130	1.1940	±0.7539

The linear model is $y = ax + b + \eta$, with a and b constant, and η Gaussian noise.

The estimated values of the amplitude and phase of the constituents for the test cases are shown in Table 2, with the corresponding 95% confidence intervals shown in Table 3. The amplitudes computed by OLS for the noise + surge case are shifted on the order of centimeters from the model values, especially for the lower-frequency constituents Z_0 and K_1. The IRLS fit values are closer to the noise-only model values, but are still shifted to some degree. A dramatic difference occurs in the confidence interval for K_1. IRLS reduces the estimated confidence interval width by a factor of 8 for the K_1 amplitude, and a factor of 7 for the K_1 phase. There is a small (factor 1.5) improvement in the confidence interval of Z_0, and minor reduction of the confidence intervals for M_2. This example clearly demonstrates that selective rejection of broad-spectrum disturbances by IRLS can improve tidal estimates from a synthetic signal.

4. Application of IRLS to estuarine tidal data

This section compares the performance of OLS and IRLS calculations applied to instrumental data. A variety of analyses were performed on a 6 month window (1 January – 1 July 1999) of hourly tidal height data, from Astoria (Tongue Point), Oregon, USA. Subsequently, 25 station years of Astoria data were analyzed to examine the performance of OLS and robust fitting over a longer record. This station is located at river mile 18 (river km 29) on the Columbia River. The station experiences a significant seasonal river flow variation, ranging from roughly 2500 to 25,000 $m^3 s^{-1}$, and is subject to storm surge from strong winter storms.

Fig. 3a shows the 1999 hourly data, with the residual of the OLS harmonic fit overlaid. Examination of the residual reveals a seasonal variation consistent with Pacific Northwest weather patterns. Stronger and more frequent storms occur during the winter, with relative calm during the summer. The residuals for the winter months have higher amplitudes and greater variability. We can assume that storm-induced height variations affect the OLS calculated values of amplitude, phase and the associated confidence intervals. Of the 67 candidate constituents (Foreman, 1977, revised 1996; Pawlowicz et al., 2002), 51 passed the significance test of $SNR \geqslant 2$. This is an important statement about noise within the signal. The calculation discarded 25% of the possible tidal constituents (though not 25% of the energy), because they were unresolvable from background noise. Fig. 3b shows the amplitudes of the significant constituents, overlaid on the power spectrum of the observed signal. The harmonically fit constituent amplitudes match the FFT power well in the tidal species peaks, though there is energy in the spectrum at points other than spectral lines.

Fig. 3c shows the power spectrum of the OLS residual. This calculation again shows energy at non-tidal frequencies, and there is still some energy above background levels in the tidal bands, especially the semi-diurnal and higher frequencies. This is an expected result. One year of hourly data does not provide sufficient frequency resolution to resolve all constituents within each band. In this case, only 24 of nearly 500 astronomical constituents were used, along with 43 of approximately 145 shallow water constituents. The tidal record is affected by both river flow variations and storms. Nonlinearities within the estuarine system act to broaden the frequency spectrum, an effect which is more pronounced at the higher frequencies. Longer-term changes in the estuary, such as channel dredging, diking and changes in sediment supply also occur. These factors constitute one reason for implementing robust fitting methods. Seasonal and interannual variations in tidal properties dictate that the shortest possible record be used that is consistent with the desired frequency resolution.

Use of IRLS brings about a decrease in the estimated parameter error. Because the parameter error estimates are based on the

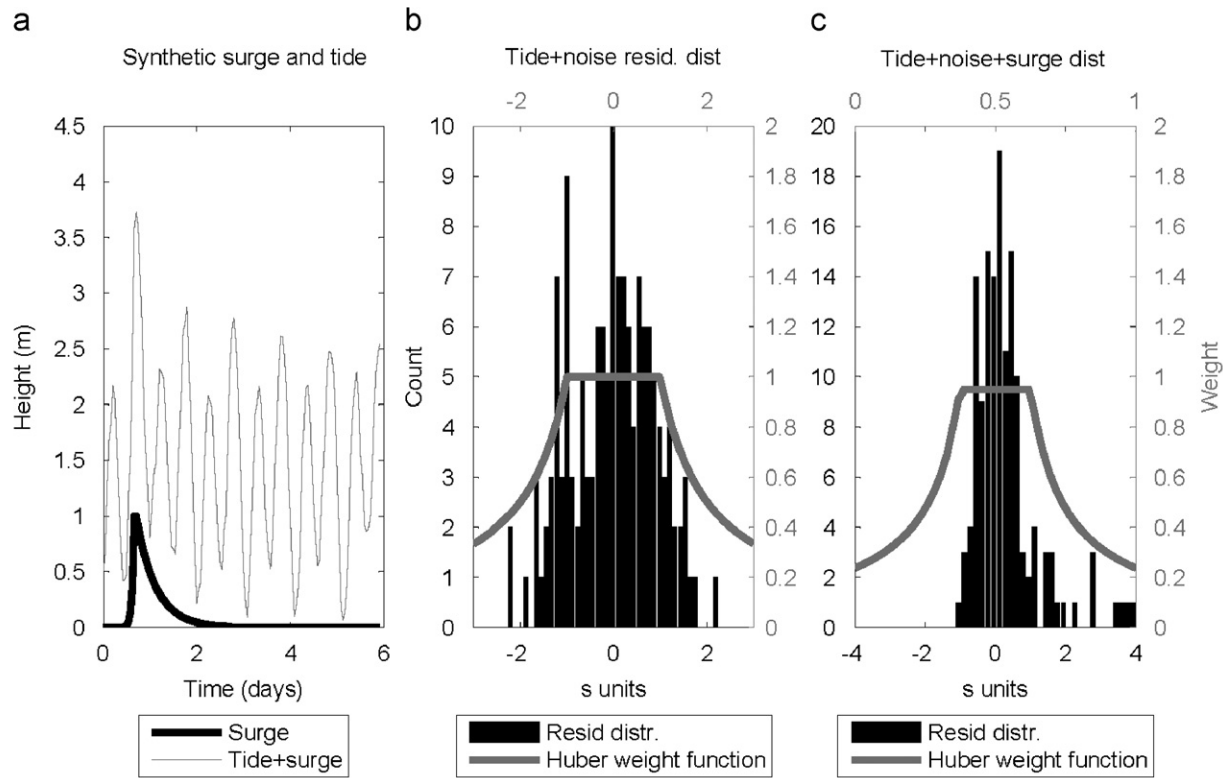

Fig. 2. An example with a synthetic tide, noise and a synthetic surge. (a) Shows the 'surge' and the 'observed' signal. (b) Shows the distribution of the initial residual with the Huber function overlaid. (c) Shows the distribution of the initial residual for the tide + surge + noise case.

Table 2
Simple tide model: estimated amplitude and phase

Estimated parameters, by experiment	Z_0 (m)	K_1 amp (m)	K_1 phase (deg)	M_2 amp (m)	M_2 phase (deg)
Model values	1.4	0.35	265	1.0	145
OLS: noise only	1.396	0.352	262.004	0.986	145.575
IRLS (Huber): noise only	1.396	0.351	262.005	0.985	145.637
OLS: noise + surge	1.482	0.396	268.696	0.967	147.066
IRLS (Huber): noise + surge	1.435	0.358	263.190	0.983	146.405

Table 3
Simple tide model, computed 95% confidence intervals of amplitude and phase

Parameter 95% confidence intervals, by experiment	Z_0 (m)	K_1 amp (m)	K_1 phase (deg)	M_2 amp (m)	M_2 phase (deg)
OLS: noise	±0.015 m	±0.014	±1.853	±0.047	±2.727
IRLS (Huber): noise	±0.015	±0.018	±2.943	±0.043	±2.471
OLS: noise + surge	±0.037	±0.108	±15.665	±0.103	±3.112
IRLS (Huber): noise + surge	±0.025	±0.013	±2.148	±0.041	±2.373

average residual power in the tidal species frequency band of each constituent, any reduction in background energy should decrease the estimated parameter error. Thus, application of IRLS fitting to this data set should generate a decrease in estimated error, and an increase in confidence in the estimate. Because the spectrum of the residual is not flat, estimates of confidence limits requires of the colored noise model to estimate parameter error. Several weighting functions were evaluated, including the bisquare, Cauchy, and Huber functions (Mathworks, 2006). Fig. 4a shows the histograms of residuals of the first iteration, overlaid with a scaled weighting curve for each method. Since the weights for the first iteration are all equal to 1, the residual itself is the residual from the OLS fit. The histograms show the data to be approximately normally distributed, with no obvious outliers. Fig. 4b shows the final residuals and the calculated weights. Examination of these results show that they are generally related to the prevailing the weather pattern. There is more down-weighting in the winter, with its frequent storms, and less down-weighting in the summer, a period of relative calm. Close examination of the weights shows differences between the methods. The Huber method gives full weight to most of the samples during the summer months. The Bisquare method imposes a less severe penalty on lower residuals and a greater penalty on higher values, compared to the Cauchy method, which penalizes the low values relatively more severely, and the higher values relatively less severely.

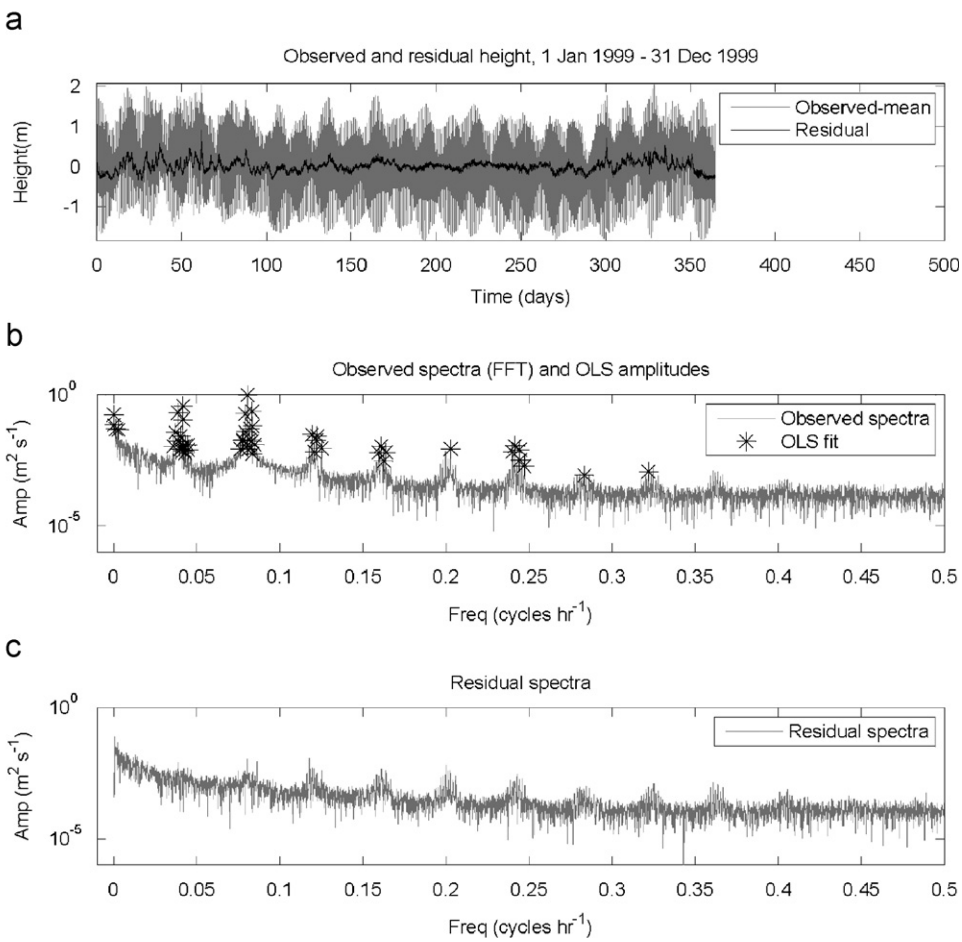

Fig. 3. Results of an OLS analysis of 1 year of data. (a) Shows the observed tidal height from Astoria, from 1 January 1999 to 1 January 2000, in gray. The black trace is the residual from the OLS fit. (b) Shows the power spectrum of the observed signal, with the OLS calculated amplitudes overlaid. (c) Shows the power spectrum of the residual.

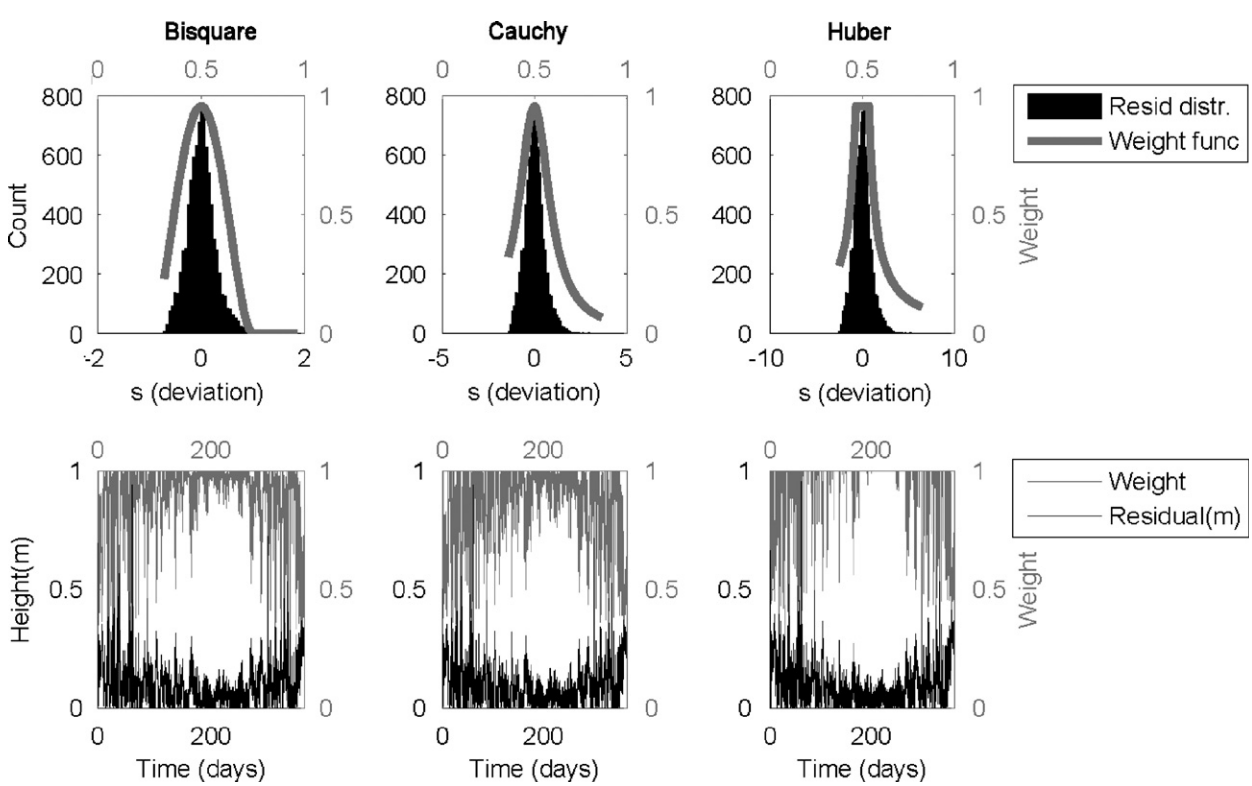

Fig. 4. A comparison of weighting functions, showing residual distribution, weighting functions, final weights and weighted residual. The upper row of graphs show the initial residual distributions, with weighting functions overlaid, for Bisquare, Cauchy and Huber functions. The lower row shows the final weights and unweighted residuals.

Please cite this article as: Leffler, K.E., Jay, D.A., Enhancing tidal harmonic analysis: Robust (hybrid L^1/L^2) solutions. Continental Shelf Research (2008), doi:10.1016/j.csr.2008.04.011

Confidence intervals were computed from the average of power spectral density within frequency bands centered on $0, 1, 2, \ldots$ cycles/day, ±0.1 cycles/day (M_0 and M_1), ±0.2 cycles/day (M_2 to M_5), ±0.21 cycles/day (M_6), 0.26− 0.29 cycles/day (M_7), and 0.3− 0.5 cycles/day (M_8) (Pawlowicz et al., 2002)). Fig. 5a shows the frequency band averages of the FFT-based power spectrum of the final weighted residuals $R_w = \omega(y - Ax)$ for the different weighting functions, using a 6 month calculation window. The band averages are systematically reduced from the OLS values across all bands. Fig. 5b shows the percent reduction in the band-averaged confidence interval width from OLS by the bisquare, Cauchy, and Huber methods, using the default tuning constants. The Cauchy method shows the greatest improvement, with reductions in the range of 30–40% for the low frequency to 5 cycles/day band. The bisquare and Huber methods reduce the confidence interval width by 10–30%, and are effective through the 6 cycles/day band.

To examine the effect of varying the tuning constant τ (refer to Appendix A for a full description of the IRLS algorithm), the default Cauchy tuning constant (2.385) was sequentially reduced by a factor of 2,3,4 and 5. Lower values of the tuning constant causes the algorithm to take more iterations to converge to a stable set of weights, and very low tuning constants caused the algorithm to fail to converge. The maximum iteration limit was increased from the default of 50 (Mathworks, 2006) to 250. With a tuning constant of $2.385/5 = 0.477$, many 6 month windows of hourly data did not converge. The absolute limit depends on the non-tidal component of the tide signal, and more work is necessary to find practical limits. A tuning constant of $2.385/3 = 0.795$ converged for most 6 month windows of the long-term analysis.

Fig. 6 compares the performance of OLS and IRLS fitting, using the Cauchy weighting function, with sequentially reduced tuning constant, applied to a 6 month record of hourly Astoria data. Fig. 6a shows the change in constituent amplitude, including an OLS fit, for the low-frequency, diurnal and semi-diurnal frequency bands. The major, high amplitude constituents have stable amplitudes. Smaller constituents show some change. Lower tuning constant values reduce the background energy, and allow more constituents to be resolved, based on a signal-to-noise test. For a 6 month window, 51 constituents are theoretically resolvable, using Foreman's constituent selection method (Foreman, 1977, revised 1996). The OLS fit resolved 29 constituents, based on a significance criteria of $SNR \geqslant 2$. Use of progressively smaller values of tuning constant in the Cauchy weight function resolves a greater number of the candidate constituents.

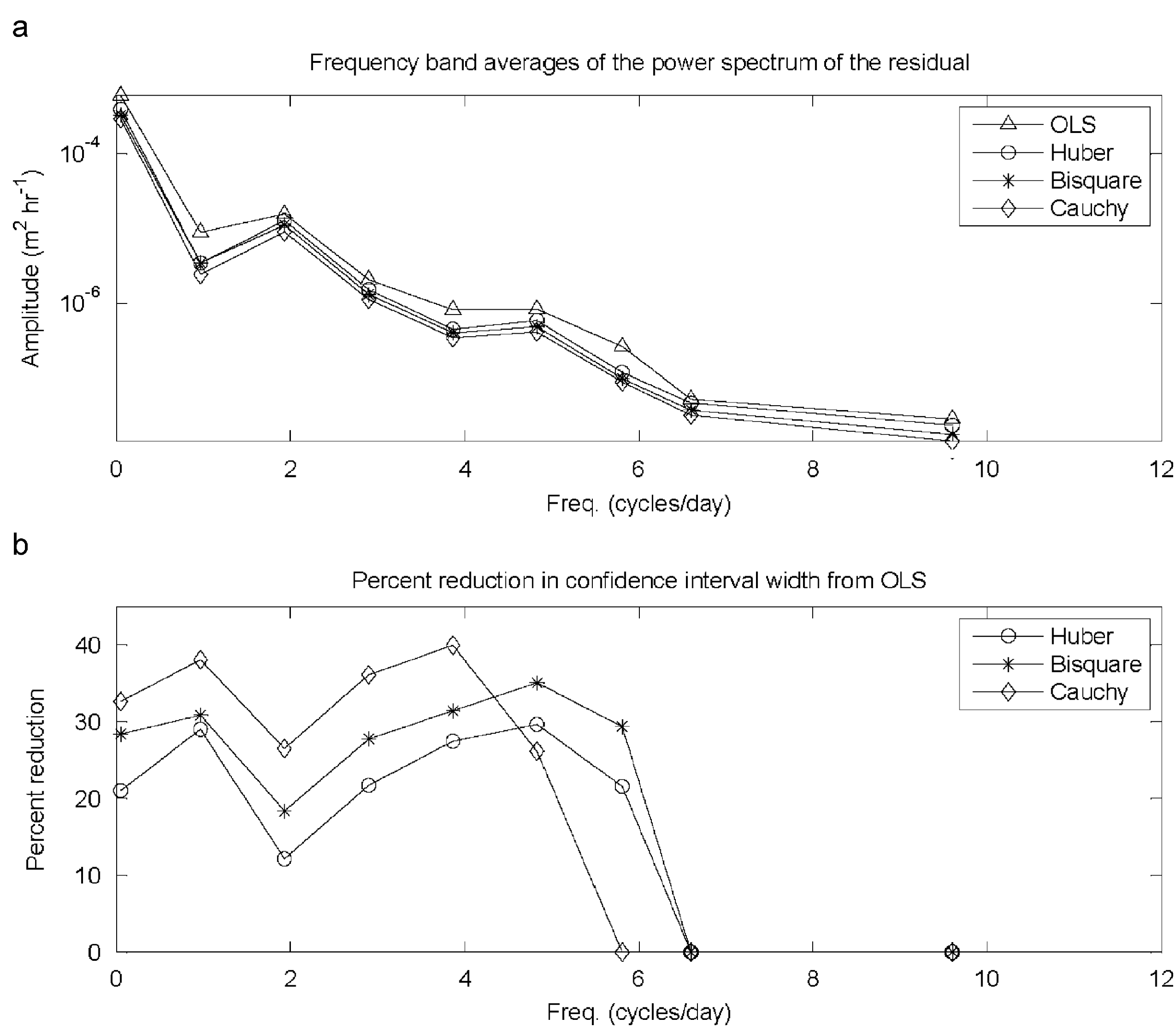

Fig. 5. A comparison of reduction in residual spectral energy and corresponding reduction in confidence interval from the OLS calculation for Huber, Bisquare and Cauchy weighting functions. (a) Shows the frequency band average of the power spectrum of the residual and (b) shows the percent reduction in confidence interval width.

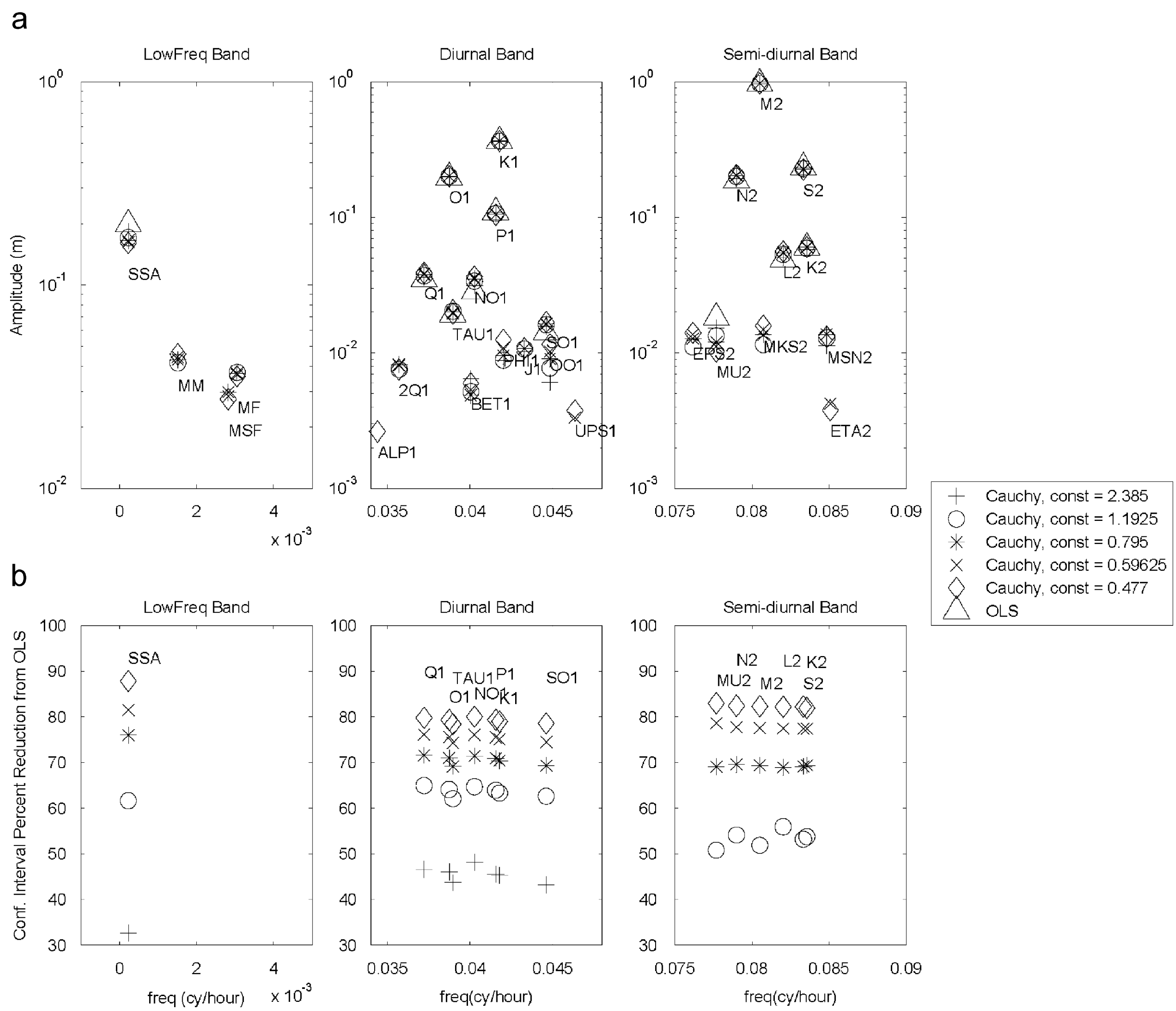

Fig. 6. A comparison of amplitudes (a) and confidence interval reductions (b), for the low-frequency, diurnal and semi-diurnal tidal frequency bands.

For a tuning constant = 0.477, all 51 candidate constituents were resolved. Fig. 6b shows the percent change in confidence interval width from OLS, for the constituents resolved by OLS, in the low frequency, diurnal and semi-diurnal frequency bands. Using IRLS with the default Cauchy weight function gives a 30–40% reduction of confidence interval width from OLS. Halving the tuning constant increases the reduction to 50–70%. Reducing the tuning constant by a factor of 5 yields an approximately 80% reduction in the confidence interval width.

The long-term analysis compared the performance of OLS and Cauchy weighted IRLS over time. A 190 day time window was advanced by a 28 day step across the Astoria tidal record, from 1 January 1980 to 31 December 2005, using a tuning constant of $\tau = 0.795$ with the Cauchy weighting function. The estimated amplitudes and percent reduction in confidence intervals for the constituents K_1, M_2 and M_4 are shown in Fig. 7. The results are again consistent. Both analyses show little change in the parameters estimated from one year of data, while the confidence intervals of the parameters are systematically reduced. The mean reduction for K_1, M_2 and M_4 was 64.5, 67.1, and 58.9%, respectively, with standard deviations of 7.58, 6.25 and 7.01, respectively.

As a final comparison, the effects on the analysis of barometric pressure fluctuation was considered. Barometric pressure causes sea level changes via the inverse barometer effect, and deviation in barometric pressure may be considered a proxy for general storminess. Winter storms at the entrance of the Columbia may generate winds in excess of 80 km/h, with combined seas in excess of 12 m, so storms play an obvious role in observed water level. Fig. 8 shows the observed water level, the residual of the IRLS predicted tide, the deviation from mean barometric pressure $\hat{bp} = bp - \bar{bp}$, and the final weights from the Cauchy function. It is notable that $\hat{bp}$ and the final weight are roughly inversely proportional. Intuitively, this makes sense. The greatest deviation from the harmonic model is expected during storm periods and during periods of strong river outflow. It is gratifying that low weights usually correspond to a physical phenomenon, and that periods of strong storms are algorithmically down-weighted, without any direct input indicating presence of storms.

5. Discussion and further work

Robust methods have been shown to be a more efficient method of identifying non-stationary components of the tidal signal, reducing the influence of noise on harmonic constants derived from non-stationary data and increasing confidence in the resulting estimates. This is a clear improvement over standard

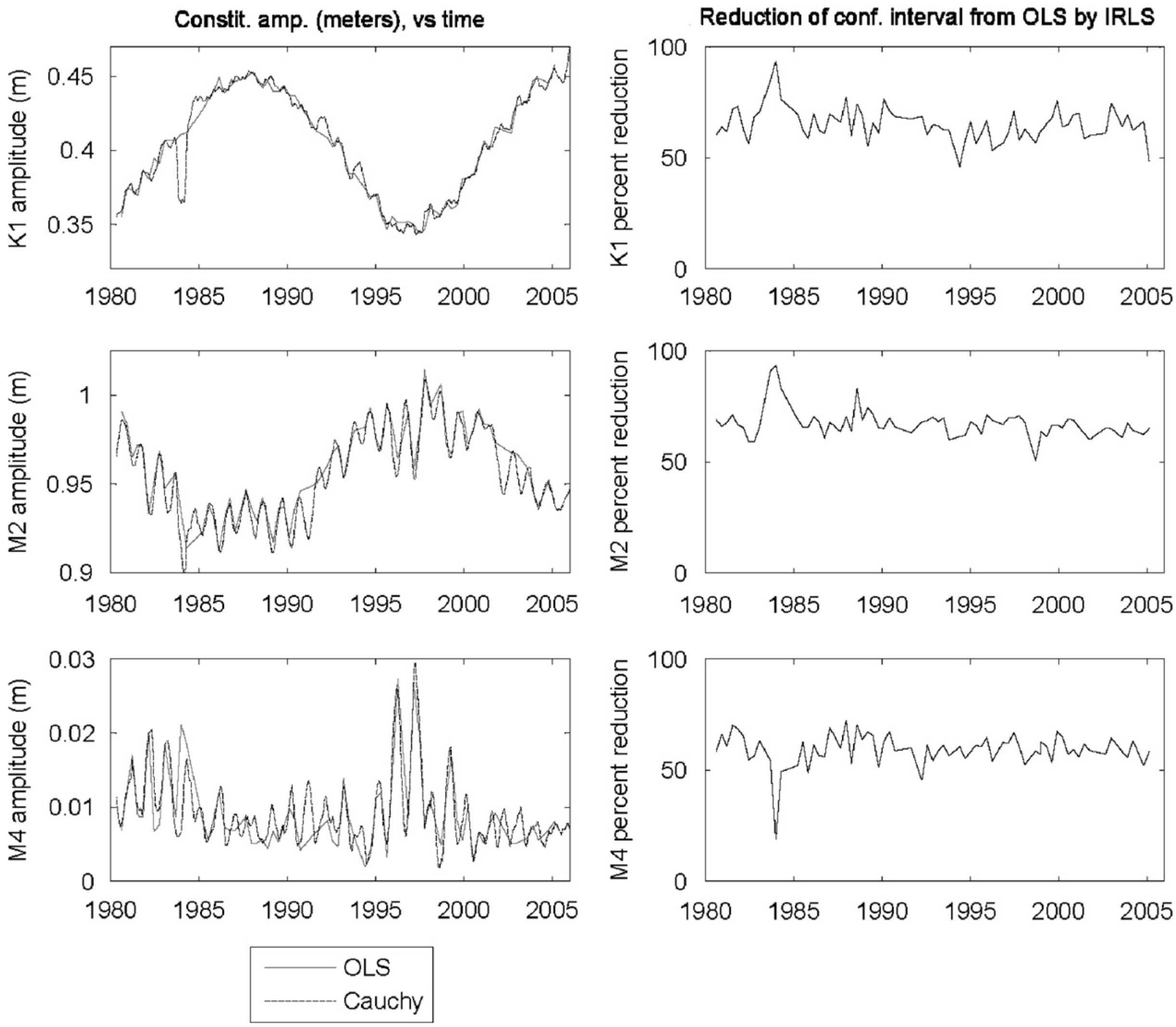

Fig. 7. A comparison of amplitude (left column) and confidence interval reduction (right column), computed by OLS and Cauchy-weighted IRLS, for K_1, M_2 and M_4.

harmonic analysis with OLS. Still, there is one potential issue that arises from the seasonality of storminess; does down-weighting of winter data systematically discriminate against certain small constituents? Work to date does not answer this question. Clearly, reliable determination of the smaller constituents requires analysis of several years (preferably 19 years) of data, whether one uses a Robust fit or OLS. Over a number of years, storm occurrence will be variable, and impacts of weighting on individual constituents will likely be averaged out. It is still possible, however, that a Robust fit may achieve systematically different answers in situations involving specific non-linear interactions.

Further work is needed in several areas. Both OLS and Robust analyses exhibit a characteristic failure mode as LOR is reduced, holding the constituent ensemble constant. As the Rayleigh ratio (the number of data points per frequency analyzed (Foreman, 1977, revised 1996) decreases, closely spaced constituent pairs increase unrealistically in amplitude, yielding a solution that (a) has more energy than the original signal and (b) is useless for prediction. This failure occurs because, over a limited LOR, a closely spaced constituent pair can partially cancel. While there are real limits on frequency resolution set by the Heisenberg uncertainty principle, a tidal analysis is normally conservative relative to limits set by the uncertainty principle for the major tidal constituents (Jay and Flinchem, 1999). Thus, this failure mode is in part a function of the inverse methods used. We had hoped that the resistance of Robust analyses to the adverse influence of noise would allow postponement of the failure to somewhat smaller LOR, but this was not the case. It appears that tidal analysis would benefit from introduction of total energy constraints, an approach used in image processing, where (like tidal analysis) the frequency-domain support of the signal is usually limited (Candés et al., 2006).

Experiments with various weight functions, especially the Huber and Cauchy functions, have been performed in this paper. However, a rigorous evaluation of weighting methods across many tide stations should be conducted to determine an optimal weighting method for tidal data. As noted in Section 4, barometric pressure plays both direct and indirect roles in water level variation. Cartwright (1968) investigated atmospheric effects in the context of a response analysis. Inclusion of atmospheric forcing in the basis functions of a harmonic model, with the idea of avoiding the down-weighting of all but the most extreme periods, might well be worthwhile. This approach may be preferable to filtering out atmospheric disturbances with a high-pass filter, both for short records, and in cases where a consistent analysis of variance in all frequency bands is desired.

Please cite this article as: Leffler, K.E., Jay, D.A., Enhancing tidal harmonic analysis: Robust (hybrid L^1/L^2) solutions. Continental Shelf Research (2008), doi:10.1016/j.csr.2008.04.011

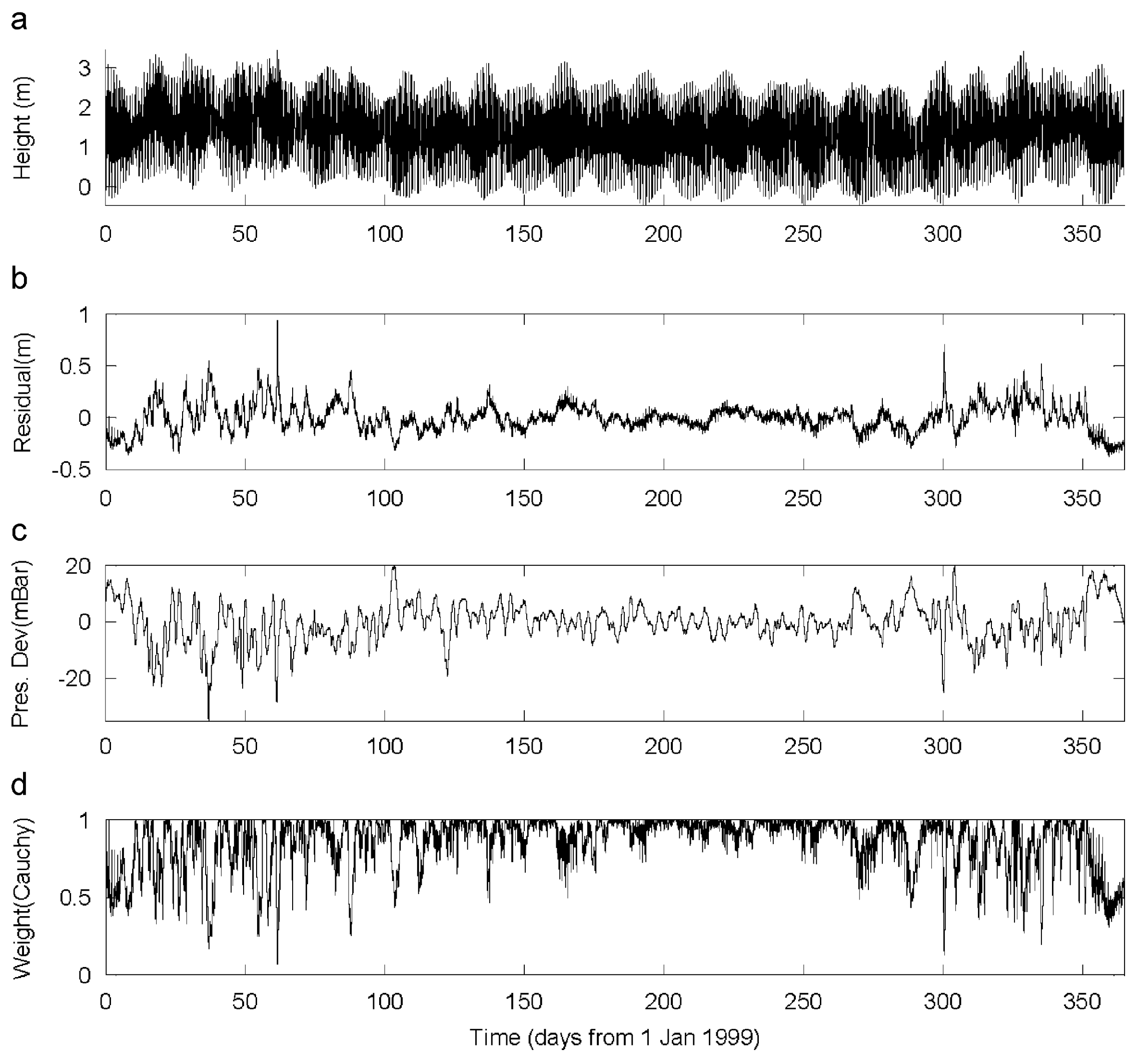

Fig. 8. A comparison of observed tidal signal (a), residual (b), deviation from mean barometric pressure (c) and final Cauchy weights (d) for Astoria Tongue Point, 1 January 1999–31 December 1999.

6. Summary

Traditional harmonic analysis of tides is highly sensitive to omnipresent instrumental and environmental noise. Hybrid L^1/L^2 fitting via an IRLS algorithm, an extension of the standard OLS methodology of harmonic analysis, has been shown to reduce the influence of broad-spectrum noise, including that caused by storms and other atmospheric processes. Use of an algorithm eliminates the subjective and labor intensive process of manually separating 'good' data from 'bad', and incorporates all the available data in weighted form into the parameter estimates. Admittedly, this moves the point of judgment. Instead of subjectively removing 'bad' data, the investigator is left with the more subtle choice of algorithm and tuning parameters that weight the data. On the whole, the authors believe an algorithmic approach adds consistency and reduces overall effort, while maximizing data utilization.

Experiments were performed using 6 month calculation windows, on records from an estuarine tidal station that contain influences from varying river flow and strong winter storm surges. Use of IRLS with the Cauchy weighting function and default tuning parameter reduces the confidence interval width from the OLS value by 30–40%, with a 20% increase in resolved constituents. Reducing the tuning parameter further decreases the confidence interval width and increases constituent resolution. Decreases in confidence interval width of 80%, and an increase in constituent resolution of 75% were demonstrated. Decreasing the tuning parameter eventually causes the calculation to fail to converge, and more work is necessary to define practical limits. Reductions of confidence intervals by 65% were consistently demonstrated in a long-term analysis. The current work demonstrates that hybrid L^1/L^2 fitting via the IRLS algorithm successfully increases confidence in tidal parameter estimates and resolution of low-amplitude constituents from background noise, which should be helpful for studies of long-term tidal evolution and sea level rise, while remaining consistent with existing analysis methods.

Acknowledgments

This research was funded by the Bonneville Power Administration and NOAA-Fisheries (project—"Estuarine habitat and juvenile salmon—current and historic linkages in the lower Columbia River"), and the NSF (project RISE—River Influences on Shelf Ecosystems OCE 0239072). The authors thank Rich

Please cite this article as: Leffler, K.E., Jay, D.A., Enhancing tidal harmonic analysis: Robust (hybrid L^1/L^2) solutions. Continental Shelf Research (2008), doi:10.1016/j.csr.2008.04.011

Pawlowicz, Bob Beardsley and Steve Lentz for developing and distributing the t-tide Matlab package, which is the basis for the analyses in this paper. The authors thank the two anonymous reviewers, who raised several suggestions leading to improvement of this paper.

Appendix A. IRLS algorithm

Variables:

- t Sample times
- θ Known constituent frequencies
- A Basis matrix
- R Residual
- r Normalized residual
- τ Tuning constant for the weighting function
- ω Weights
- x Tidal parameters
- h Observed heights
- $MAD(R)$ The median absolute deviation of R
- s Deviation statistic
- l Leverage vector

The basis matrix A is defined as:

$$A = \begin{bmatrix} \cos(2\pi\theta_1 t_1) & \cdots & \cos(2\pi\theta_n t_1) & \sin(2\pi\theta_1 t_1) & \cdots & \sin(2\pi\theta_n t_1) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \cos(2\pi\theta_1 t_m) & \cdots & \cos(2\pi\theta_n t_m) & \sin(2\pi\theta_1 t_m) & \cdots & \sin(2\pi\theta_n t_m) \end{bmatrix}$$

The leverage vector l is a measure of the influence of each point in the sampled time series. Each value in the leverage vector can be thought of as a measure of the change in response variables resulting from the deletion of the corresponding sample from the observed variable and is defined (Belsley et al., 1980)

$$l = diag(h(h^{\mathrm{T}}h)^{-1}h^{\mathrm{T}})$$

For iteration 0,

$$w_0 = [1 \ \ldots \ 1]', \quad \text{so the first iteration is the OLS solution.}$$

Then repeating the following steps until the residual R converges:

$$x = (A^{\mathrm{T}}\omega A)^{-1}A^{\mathrm{T}}\omega h$$

$$R = \omega(h - Ax)$$

$$s = \frac{MAD(R)}{0.6745}$$

$$r = \frac{R}{\tau s\sqrt{1 - l^2}}$$

$\omega = f(r)$, with $f(r)$ and default values for τ given by

Method	ω	Default τ
Andrews	$\omega_{andrews} = \begin{cases} \frac{\sin(r)}{r}, & \lvert r\rvert < \pi \\ 0, & \lvert r\rvert \geqslant \pi \end{cases}$	1.339
Bisquare	$\omega_{bisquare} = \begin{cases} (1 - r^2)^2, & \lvert r\rvert < 1 \\ 0, & \lvert r\rvert \geqslant 1 \end{cases}$	4.685
Cauchy	$\omega_{cauchy} = \frac{1}{1 + r^2}$	2.385
Fair	$\omega_{fair} = \frac{1}{1 + \lvert r\rvert}$	1.400
Huber	$\omega_{huber} = \frac{1}{\max(1,\lvert r\rvert)}$	1.345
Logistic	$\omega_{logistic} = \frac{\tanh(r)}{r}$	1.205
Talwar	$\omega_{talwar} = \begin{cases} 1, & \lvert r\rvert < 1 \\ 0, & \lvert r\rvert \geqslant 1 \end{cases}$	2.795
Welsch	$\omega_{welsch} = e^{-r^2}$	2.985

The tuning constant τ controls the width of the weight function, and may be adjusted from the default value. Lower values of τ penalize outliers more heavily, while higher values penalize outliers less severely. From experiments, the authors found that conservative (higher values of τ) filters produce better output, while severe filters essentially over-filter the data. In general, the default tuning value constants suggested by Matlab have been used. Further work is needed to define the optimal IRLS weight function for tidal data.

References

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics, Identifying Influential Data and Sources of Collinearity. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Bube, K.P., Langan, R.T., 1997. Hybrid $l1/l2$ minimization with applications to tomography. Geophysics 62 (4), 1183–1195.

Candés, E.J., Romberg, J.K., Tao, T., 2006. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics 6, 227–254.

Cartwright, D., 1968. A unified analysis of tides and surges round north and east Britain. Philosophical Transactions of the Royal Society London A 263, 1–55.

Darche, G., 1989. Iterative l1 deconvolution. Technical Report SEP Report 61, Stanford Exploration Project.

Foreman, M., 1977, revised 1996. Manual for tidal heights analysis and prediction. Technical Report Pacific Marine Report 77-10, Institute of Ocean Sciences, Patricia Bay Victoria, BC, Canada.

Fox, J., 2002. Web Appendix to An R and S-PLUS Companion to Applied Regression. first ed. URL ⟨http://cran.r-project.org/⟩.

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.

Mathworks, 2006. Matlab R2006a Documentation, r2006a ed. The Mathworks, Inc.

Moler, C., 2004. Numerical Computing with Matlab. SIAM, pp. 1–27 (Chapter 5).

Munk, W., Cartwright, D., 1966. Tidal spectroscopy and prediction. Philosophical Transactions of the Royal Society of London, Series A, Mathematical and Physical Sciences 259, 533–581.

Munk, W., Hasselmann, K., 1964. Super-resolution of tides. In: Studies in Oceanography. University of Washington Press, Seattle, pp. 339–344.

Pawlowicz, R., Beardsley, B., Lentz, S., 2002. Classical tidal harmonic analysis with errors in matlab using t-tide. Computers & Geosciences 28, 929–937.

Please cite this article as: Leffler, K.E., Jay, D.A., Enhancing tidal harmonic analysis: Robust (hybrid L^1/L^2) solutions. Continental Shelf Research (2008), doi:10.1016/j.csr.2008.04.011