|  |  |
| --- | --- |
| PHY 477/577: Obs. Methods & Data Analysis in Phys. Ocn. | Derek J. Grimes |
| Assignment 5: Least-Squares Tidal Regression | grimesdj@uncw.edu |

Carter Beaulieu HW#5

Goal: develop code to implement least-squares method for estimating astronomical tidal amplitudes and evaluate the errors.

Task: Astronomical tides are an excellent example of a complex periodic signal that are well represented by Fourier series. However, there are many different constituents with varying degrees of importance from one ocean basin to another. Additionally, local effects like focusing/convergence and bottom drag can also alter amplitudes and lead to non-linear generation of higher harmonics. T-tide is a robust algorithm for tidal harmonic analysis of both water-levels and currents. It uses a least-squares algorithm to fit up to 146 constituents depending on the record length. Let’s see how well you can do fitting just two common constituents, the principal lunar semidurnal (M2) period TM2 = 12*.*42h and principal solar semidurnal (S2) period TS2 = 12h. We will compare your fit amplitudes, confidence intervals, and residuals to those of the t-tide algorithm. Lastly, you’ll use the least-squares fit tidal phase for the two sites used in HW4 and compare the implied time-lag to the lag determined from the lagged-correlation analysis. You may download the matlab-package T TIDE (v1.4 or higher) from: [https://www.eoas.ubc.ca/∼rich/.](https://www.eoas.ubc.ca/~rich/) For a Python version try: [https://github.com/moflaher/ttide](https://github.com/moflaher/ttide_py) [py.](https://github.com/moflaher/ttide_py) Note, MATLAB’s t-tide assumes data is sampled at ∆*t* = 1 hr intervals, if you want to provide a different ∆*t* use parameter/value pair (’interval’, dt in decimal hours). For example if you use the six-minute verified tides, then dt in decimal hours= 0*.*1.

1. Load the water-level records *η*1 and *η*2 from HW4, or re-download them with either the same or different resolution/duration. The record duration can be 1-month for ∆*t* = 6 min and up to 1-year for ∆*t* = 1 hour. Before proceeding, remove the sample mean *η*¯ from each record, *e.g., ηj* = *ηj* −*ηj*.
2. Run the t-tide function on *η*1, specifying the ’interval’ if ∆*t* ̸= 1 hour. The MATLAB syntax is:

[NAME ,FREQ ,TIDECON ,*ηp*]=t tide(*η*1,’interval’,∆*t*); % NAME = constituent abbreviation, e.g., M2, S2, etc.

% FREQ = constitutent frequency,

% TIDECON = [Amplitude, Error, Phase, Error],

% Amplitude= fit amplitude,

% Error = confidence interval,

A blue lines on a white background

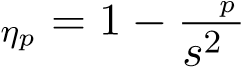
Description automatically generatedA blue line graph with white text

Description automatically generated% Phase = constituent phase relative the center of the input record, % *ηp* = tidal prediction time series.

1. Estimate the residual error *ϵp* = *η*1 − *ηp* and the variance of the error. What is the skill-score of the fit, (For the constituents M2 and S2)

*s*2*ϵ*

A screen shot of a graph

Description automatically generatedSS*, η*1

Crescent City Skill Score: M2,S2 combined

Residual Error: 0 1.866774

1 1.922119

2 1.890703

3 1.872395

4 1.907085

...

7435 0.216257

7436 0.114345

7437 -0.000958

7438 0.129998

7439 0.186870

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 1.737814

Original Data Variance: 4.705134

Skill Score: 0.630656

A screen shot of a graph

Description automatically generated

Monterey Skill Score: M2,S2 combined

Residual Error: 0 2.236869

1 2.267162

2 2.282232

3 2.312142

4 2.396969

...

7435 0.417751

7436 0.442874

7437 0.413250

7438 0.358748

7439 0.439250

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 1.473785

Original Data Variance: 2.922961

Skill Score: 0.495790

1. How many constituents were significant, *i.e.,* have an amplitude larger than the error?

Identified constituents: {'name': array(['M2', 'K1', 'O1', 'S2', 'N2', 'Q1', 'NO1', 'J1', 'MSF', 'OO1',

'2Q1', 'UPS1', 'M3', 'ETA2', 'SK3', 'MO3', 'MN4', '2SK5', 'MK3',

'S4', '2MS6', 'M4', '2MN6', 'MS4', 'M6', '3MK7', 'M8', '2MK5',

'2SM6']

Of these Constituents, there were three significant ones for both Monterey and Crescent City. These include M2, K1, and O1.

A screen shot of a graph

Description automatically generated

Most Significant Constituents based on Skill Score: ['M2', 'K1', 'O1']

Constituent: M2

Residual Error: 0 1.778445

1 1.803774

2 1.815159

3 1.842674

4 1.926401

...

7435 -0.007095

7436 0.007575

7437 -0.031312

7438 -0.093858

7439 -0.020161

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 1.595577

Original Data Variance: 2.922961

Skill Score: 0.454123

Constituent: K1

Residual Error: 0 -1.027435

1 -0.992111

2 -0.965916

3 -0.918840

4 -0.810875

...

7435 -2.735604

7436 -2.766076

7437 -2.845599

7438 -2.944174

7439 -2.901802

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 1.971050

Original Data Variance: 2.922961

Skill Score: 0.325667

Constituent: O1

Residual Error: 0 -0.577608

1 -0.538955

2 -0.509817

3 -0.460186

4 -0.350056

...

7435 -1.155237

7436 -1.207522

7437 -1.309928

7438 -1.432442

7439 -1.415051

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 2.477470

Original Data Variance: 2.922961

Skill Score: 0.152411

A screen shot of a graph

Description automatically generated

Residual Error (Significant Constituents): 3 0.335012

1 0.340844

2 0.332746

3 0.340794

4 0.405079

...

7435 -0.620788

7436 -0.613672

7437 -0.660543

7438 -0.731521

7439 -0.666718

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error (Significant Constituents): 0.160076

Skill Score: 0.903547

A screen shot of a graph

Description automatically generated

Most Significant Constituents based on Skill Score: ['M2', 'K1', 'O1']

Constituent: M2

Residual Error: 0 1.585942

1 1.610992

2 1.550162

3 1.503404

4 1.510684

...

7435 0.059781

7436 -0.075317

7437 -0.223312

7438 -0.124463

7439 -0.099025

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 1.971262

Original Data Variance: 4.705134

Skill Score: 0.581040

Constituent: K1

Residual Error: 0 -1.874533

1 -1.894917

2 -1.994445

3 -2.073104

4 -2.090882

...

7435 -2.435555

7436 -2.681541

7437 -2.936532

7438 -2.940527

7439 -3.013521

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 3.641965

Original Data Variance: 4.705134

Skill Score: 0.225959

Constituent: O1

Residual Error: 0 -1.402113

1 -1.416980

2 -1.511390

3 -1.585336

4 -1.598808

...

7435 -0.636665

7436 -0.898713

7437 -1.170973

7438 -1.193433

7439 -1.286079

Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error: 4.224755

Original Data Variance: 4.705134

A screen shot of a graph

Description automatically generatedSkill Score: 0.102097

Residual Error (Significant Constituents): 3 0.050073

1 0.072200

2 0.008820

3 -0.040177

4 -0.034885

...

7435 -0.748394

7436 -0.872875

7437 -1.010024

7438 -0.900180

7439 -0.863671

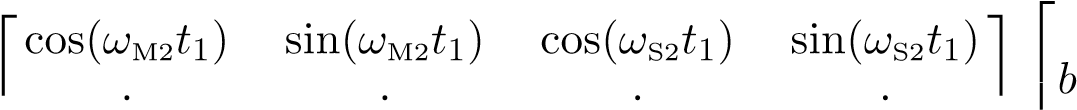
Name: Verified (ft), Length: 7440, dtype: float64

Variance of Error (Significant Constituents): 0.276570

Skill Score: 0.891081

1. In analogy to t-tide (Pawlowicz et al., 2002[[1]](#footnote-1)), we will assume the record is dominated by a truncated Fourier series of the form:

*η*(*t*) ≈ *a*M2 cos(2*πt/*TM2) + *b*M2 sin(2*πt/*TM2) + *a*S2 cos(2*πt/*TS2) + *b*S2 sin(2*πt/*TS2)*,*

*a*M2

*η* ≈ **Ha** =  .. .. .. .. *a*M2S2*,*

 cos(*ω*M2*tN*) sin(*ω*M2*tN*) cos(*ω*S2*tN*) sin(*ω*S2*tN*) *b*S2 where the second line is a matrix representation of our observed water level record vector *η* = [*η*1(*t*1);*η*1(*t*2);*...*], **H** is a matrix whose columns are unit amplitude cos() and sin() functions, and **a** is a column vector of amplitudes.

Determine the best fit coefficients **ˆa** = [*a*ˆM2;*...*;ˆ*b*S2] using MATLAB’s built in least-squares function ‘\’,

**ˆa** = **H**\*η*

≡ (**H***T* **H**)−1(**H***T η*)

Also compute the inverse covariance matrix *C*−1 = (**H***T* **H**)−1, as we’ll use this for error estimates.

Crescent City Results:

Best-fit coefficients: [ 3.88033066e-05 7.82712587e-04 -6.03574549e-03 2.84360510e-03

-5.60077453e-04 -3.48964106e-03]

Residuals: [35005.98301258]

Inverse Covariance Matrix (C^(-1)): [[ 2.68886055e-04 9.53722800e-08 -1.41505376e-07 1.26866432e-07

-7.55435293e-07 3.11374846e-07]

[ 9.53722800e-08 2.68751532e-04 -2.84807521e-08 2.38184704e-08

1.75848097e-07 -3.00869979e-07]

[-1.41505376e-07 -2.84807521e-08 2.68656993e-04 1.37728161e-07

-6.72545131e-07 1.73436846e-07]

[ 1.26866432e-07 2.38184704e-08 1.37728161e-07 2.68981417e-04

-1.14421878e-07 -6.71772886e-07]

[-7.55435293e-07 1.75848097e-07 -6.72545131e-07 -1.14421878e-07

2.68884000e-04 -5.92070693e-09]

[ 3.11374846e-07 -3.00869979e-07 1.73436846e-07 -6.71772886e-07

-5.92070693e-09 2.68756898e-04]]

Monterey Results:

Best-fit coefficients: [-0.00226348 0.00108525 -0.00625908 0.00362393 -0.00047073 -0.00106238]

Residuals: [21746.60978738]

Inverse Covariance Matrix (C^(-1)): [[ 2.68886055e-04 9.53722800e-08 -1.41505376e-07 1.26866432e-07

-7.55435293e-07 3.11374846e-07]

[ 9.53722800e-08 2.68751532e-04 -2.84807521e-08 2.38184704e-08

1.75848097e-07 -3.00869979e-07]

[-1.41505376e-07 -2.84807521e-08 2.68656993e-04 1.37728161e-07

-6.72545131e-07 1.73436846e-07]

[ 1.26866432e-07 2.38184704e-08 1.37728161e-07 2.68981417e-04

-1.14421878e-07 -6.71772886e-07]

[-7.55435293e-07 1.75848097e-07 -6.72545131e-07 -1.14421878e-07

2.68884000e-04 -5.92070693e-09]

[ 3.11374846e-07 -3.00869979e-07 1.73436846e-07 -6.71772886e-07

-5.92070693e-09 2.68756898e-04]]

1. Estimate the residual error *ϵ*ˆ = *η*1 − **Hˆa** and the variance of the error. What is the skill-score of

the fit SS*η*ˆ?

Crescent City Residual Error:

[-0.72957616 -0.71450631 -0.78947186 ... -1.42445857 -1.42877155

-1.50255875]

Crescent City Variance of Error: 4.705105243626058

Monterey Residual Error:

[3.43096845 3.49642777 3.53967501 ... 1.96187404 1.85574045 1.89629185]

Monterey Variance of Error: 2.9229334007846592

Crescent City Skill Score: 6.120025506373139e-06

Monterey Skill Score: 9.587809528976976e-06

1. Plot the zero-mean signal as a solid black line, then the t-tide fit on top as a greed dotted line. Also, plot the least-squares fit in solid red. Make the line-widths equal to about 2 units. Label your axes, include proper units, and a legend (if necessary), and make a figure caption to accompany it (see Lecture 2 for examples). Describe any qualitative similarities/differences between the fits and original record.

A screen shot of a graph

Description automatically generated A graph of a signal

Description automatically generated with medium confidence

1. What is the amplitude and phase of the M2 and S2 fits, *e.g., A*ˆ2*M*2 = *a*ˆ2M2+ˆ*b*2M2 and *ϕ*ˆM2 = tan−1(ˆ*b*M2*/a*ˆM2) (use MATLAB’s *atan2*() function for the four-quadrant angles)?

Amplitude and Phase of the M2 fit and S2 fit over the detrended Utide Signal).

Amplitude and Phase for Crescent City - M2:

Amplitude: 16.145065149841987

Phase: 0.7852292834750394

Amplitude and Phase for Crescent City - S2:

Amplitude: 10.26405615979713

Phase: 0.7853431428819571

Amplitude and Phase for Monterey - M2:

Amplitude: 7.142043896881622

Phase: 0.7850873136867584

Amplitude and Phase for Monterey - S2:

Amplitude: 0.6664399366369598

Phase: 0.7822121382998488

1. Approximately what fraction of the total variance is explained by M2 and S2, that is *A*ˆ2*j/*(2*s*2*η*)?

Fraction of Variance Explained for Crescent City:

M2: 0.581040

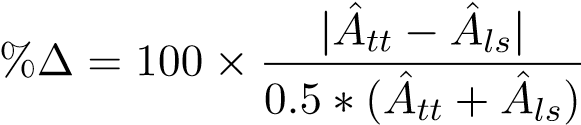
S2: 0.064433

Fraction of Variance Explained for Monterey:

M2: 0.454123

S2: 0.056014

1. Compute the percent difference in amplitude between your fit and t-tide,

*,*

where *tt* is for t-tide and *ls* is for your least-squares estimate.

Amplitude from Non-linear Regression Fit for Monterey: 1.102063147209258

Amplitude from Detrended Utide Signal for Monterey: 8.190447435475932

Percent Difference in Amplitude for Monterey: 152.56123143888624

Amplitude from Non-linear Regression Fit for Crescent City: 0.5515863074497672

Amplitude from Detrended Utide Signal for Crescent City: 8.029825581703673

Percent Difference in Amplitude for Crescent City: 174.28925148566984

1. We can approximate the error for our fit amplitudes using the student-*t* distribution assuming *ν* = *Neff* − 4 degrees of freedom. To compute *Neff* = *N/Nd* we first need to estimate the decorellation scale *Nd*. Estimate the lagged auto-correlation for both errors *ϵ*ˆand *ϵp* out to 15 days (do not compute the cross-correlation). Plot the auto-correlation versus lag and indicate the integral time-scale for both with a vertical line. What are *Nd* and *Neff* for each?

Amplitude from Non-linear Regression Fit for Monterey:

1.102063147209258

Amplitude from Detrended Utide Signal for Monterey:

8.190447435475932

Percent Difference in Amplitude for Monterey:

152.56123143888624

Confidence Intervalfor Detrended UTide Amplitudes in feet (Monterey):

5.0056727081469266

Amplitude from Non-linear Regression Fit for Crescent City:

0.5515863074497672

Amplitude from Detrended Utide Signal for Crescent City:

8.029825581703673

Percent Difference in Amplitude for Crescent City:

174.28925148566984

Confidence Interval for Detrended UTide Amplitudes in feet(CrescentCity: 5.138018850931075

Sample Size (Nd) for Crescent City: 29

Sample Size (Nd) for Monterey: 29

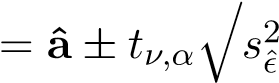
Effective Sample Size (Neff) for Monterey: 4.939156626658391

Effective Sample Size (Neff) for Crescent City: 9.478392389643775

A graph with blue and orange lines

Description automatically generated

1. The 1 − *α* confidence interval for our amplitudes {*a,*ˆ ˆ*b*} is approximately,

**a**diag(*C*−1)*,*

where diag(*C*−1) are the diagonal elements of *C*−1, which you will notice are all equal. So, to approximate the confidence interval *δα* for our amplitudes *A*ˆ we’ll use the first element on the diagonal,

*.*

*A*

=

*A*

±

*t*

*ν,α*

p

*s*

2

*ϵ*

*C*

−

1

(1

*,*

1)

|

z

{

}

*δ*

*α*

What is the ratio of our confidence limits *δα* to those of t-tide, which are based on a much more comprehensive analysis?

Calculated Confidence Interval using inverse covariance with some extracted amplitude values from Crescent City:

Amplitude: 211.468526, Confidence Interval: 3.467607

Amplitude: 244.064979, Confidence Interval: 4.001114

Amplitude: 212.721576, Confidence Interval: 3.486668

Amplitude: 245.492162, Confidence Interval: 4.026232

Amplitude: 187.493283, Confidence Interval: 3.074455

Amplitude: 202.518464, Confidence Interval: 3.320049

These are all from utide:

Confidence Intervals for M2 in Monterey:

Amplitude Lower: 1.5976393828440534

Amplitude Upper: 1.7075703114935248

Phase Lower: 179.5797217996256

Phase Upper: 183.38571012603302

Confidence Intervals for K1 in Monterey:

Amplitude Lower: 1.1678509785926323

Amplitude Upper: 1.234006084794721

Phase Lower: 235.14129164878065

Phase Upper: 238.29823761860777

Confidence Intervals for O1 in Monterey:

Amplitude Lower: 0.7177850876470825

Amplitude Upper: 0.7839285871542109

Phase Lower: 199.41331660826992

Phase Upper: 204.46345539214704

Confidence Intervals for S2 in Monterey:

Amplitude Lower: 0.43833811496437153

Amplitude Upper: 0.5481841347628486

Phase Lower: 186.52320094171375

Phase Upper: 199.284520223789

Confidence Intervals for M2 in Crescent City:

Amplitude Lower: 2.3096944015447347

Amplitude Upper: 2.458712712418138

Phase Lower: 209.0737980194544

Phase Upper: 212.652548971469

Confidence Intervals for K1 in Crescent City:

Amplitude Lower: 1.23272789478427

Amplitude Upper: 1.3057759273421419

Phase Lower: 243.0509256658764

Phase Upper: 246.34960433949024

Confidence Intervals for O1 in Crescent City:

Amplitude Lower: 0.7533335339626189

Amplitude Upper: 0.8263812358712567

Phase Lower: 208.55651408769765

Phase Upper: 213.8573104444911

Confidence Intervals for S2 in Crescent City:

Amplitude Lower: 0.6119400123713692

Amplitude Upper: 0.7609164254980934

Phase Lower: 236.9552483348061

Phase Upper: 249.38899157523966

In sum, the utide function has the better confidence interval, as it is based on a much more comprehensive ananalysis.

1. Repeat (5) and (8) for your second water-level record *η*2. Based on the phase difference, estimate the time-lag *τ* between the sites for both constituents, *e.g.,*

*τ*M2 = (*ϕ*ˆ2*,*M2 − *ϕ*ˆ1*,*M2)TM2 2*π*

How does this compare to your lagged cross-correlation analysis of HW4?

Time-Lags:

{'M2': array([-57.87205243]), 'S2': array([-97.10980518])}

These negative values suggest that the tidal signals in Monterey led to the corresponding signals in Crescent City. Multiply by six, due to six-minute intervals, and you get -5.7 hours and -9.7 hours.

“

1. What is the maximum cross-correlation max(ˆ*ρ*1*,*2)? At what time-lag *τ*max is the cross-correlation maximal? In what direction does this imply tidal signals are propagating? From 1 to 2, or vice-versa?

This suggests that P^1 (Cresent City) tides are propagating from P^2 (Monterey), and the tide is occurring at Monterey about ten hours before the tide at Crescent City.

Maximum Cross-Correlation (max(ρ^1,2)): 0.98

Time Lag at Maximum Cross-Correlation: -10 hours”, This is my HW#4 answer.

2

1. see equation (3), <https://doi.org/10.1016/S0098-3004(02)00013-4>1 [↑](#footnote-ref-1)