Residual bootstrap for tidal model uncertainty analysis

S. Innocenti, P. Matte, V. Fortin, N. Bernier

vEGU 2021
Reconstruction of tidal signals

Context

Water level predictions and long term analysis

Methods

- Harmonic Analysis (HA) of water levels
- Wetland/estuary conservation
- Ecosystem functions
- Flood forecasting
- Risk mapping / mitigation
- River navigation
- Freshwater fisheries
- Calibrating/verifying altimetry products

Results

Image adapted from https://swot.nasa...
Reconstruction of tidal signals

Water level predictions and long term analysis

- wetland/estuary conservation
- ecosystem functions
- freshwater fisheries
- flood forecasting risk mapping / mitigation
- river navigation
- calibrating/verifying altimetry products

Harmonic Analysis (HA) of water levels

Image adapted from https://swot.nasa...
HA model

\[ h_t = h_0 + \sum_{k=1}^{K} \beta_{s,k} \sin(\omega_k \tau) + \beta_{c,k} \cos(\omega_k \tau) + \epsilon_t \]

- \( h_t \): surface elevation
- \( \tau \): time evaluated at time steps \( t = 1, 2, \ldots, T \)
- \( \omega_k \): harmonic frequencies of constituents \( k = 1, 2, \ldots, K \)

\( \Rightarrow \) tidal constituent selection

and estimation of \((2 \times K + 1)\) unknown parameters
HA model

\[ h_t = h_0 + \sum_{k=1}^{K} \beta_{s,k} \sin(\omega_k \tau) + \beta_{c,k} \cos(\omega_k \tau) + \epsilon_t \]

- \( h_t \): surface elevation
- \( \tau \): time evaluated at time steps \( t = 1, 2, \ldots, T \)
- \( \omega_k \): harmonic frequencies of constituents \( k = 1, 2, \ldots, K \)
- \( \epsilon_t \): error \( \Rightarrow \) unresolved component of the signal

\( \Rightarrow \) tidal constituent selection

and estimation of \( (2 \times K + 1) \) unknown parameters
HA model

\[ h_t = h_0 + \sum_{k=1}^{K} \beta_{s,k} \sin(\omega_k \tau) + \beta_{c,k} \cos(\omega_k \tau) + \epsilon_t \]

\( h_t \): surface elevation
\( \tau \): time evaluated at time steps \( t = 1, 2, \ldots, T \)
\( \omega_k \): harmonic frequencies of constituents \( k = 1, 2, \ldots, K \)
\( \epsilon_t \): error \( \Rightarrow \) unresolved component of the signal \( \Rightarrow \) autocorrelated

\( \Rightarrow \) tidal constituent selection
and estimation of \((2 \times K + 1)\) unknown parameters

Interaction between tides, weather, hydrology, and other climatological processes
\( \Rightarrow \) substantial residual energy
Conventional methods rely on simplifying assumptions to estimate the HA regression variance-covariance matrix [e.g., normal, independent, and/or homoskedastic residuals]

Analytical definition of the variance-covariance estimator for:

- UTide [Codiga, 2011]
- IRLS [Leffler and Jay, 2009; Huber and Ronchetti, 2011]
- NS_tide [Matte et al., 2013]

⇒ **Obj. 1.** Review each method statistical properties and evaluate their implications in HA uncertainty assessments:
https://youtu.be/Nqd7qMKJrw0 [4:34 min]
Residual bootstrap for estimating the HA regression uncertainty

No or weak assumptions about the error term distribution to produce random replicates of the observed residuals

[account for the coloured nature of noise signals]

Definition of two resampling algorithms:

- Moving Block Bootstrap (MBB) with fixed length blocks
- Semi-Parametric Bootstrap (SPB) based on the residual spectrum

⇒ Theoretical comparison of conventional and bootstrap methods: [https://youtu.be/voEg6NEgT8M][3:23 min]
Monte-Carlo Experiment

Simulate a large number of water level series with known statistical properties

- Two sets of simulations corresponding to **two spatial locations** in the St. Lawrence Estuary, Canada.

- 68 tidal constituents with **known amplitudes and phases** (from 11-yr records) to generate the tidal signal.

- Various series lengths at **hourly** resolution and trimmed to contain **high-low water levels**.

- Two **residual structures**:
  - White Noise [WN], homoskedastic uncorrelated,
  - Resampled noise [RN], heteroskedastic autocorrelated.
Monte-Carlo Experiment

Power Spectral Density (PSD) of simulated residual series

\[ \mu_{\varepsilon_1} = 0.57 \, [\text{cm}] \]
\[ \sigma_{\varepsilon_1} = 26.08 \, [\text{cm}] \]
Results
Standard errors of tidal amplitudes

The robust IRLS were used to estimate the regression coefficients for all the uncertainty assessment methods (same $\beta$s for all methods).
Standard errors of tidal amplitudes

**Context**
- Methods
- Results

**Standard errors of tidal amplitudes**

```
<table>
<thead>
<tr>
<th>Component period</th>
<th>Median A [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1yr</td>
<td>0.05</td>
</tr>
<tr>
<td>48h</td>
<td>0.1</td>
</tr>
<tr>
<td>24h</td>
<td>0.2</td>
</tr>
<tr>
<td>12h</td>
<td>0.5</td>
</tr>
<tr>
<td>6h</td>
<td>1</td>
</tr>
<tr>
<td>4h</td>
<td></td>
</tr>
</tbody>
</table>
```

**Signals with WN residuals**

- UTide: underestimation of the variance $\Rightarrow$ spectrum band-averaging and triple down-weighting of $\Sigma \beta$

**Signals with RN residuals**

- The robust IRLS were used to estimate the regression coefficients for all the uncertainty assessment methods (same $\beta$s for all methods)
Standard errors of tidal amplitudes

For the context of standard errors of tidal amplitudes, the table below shows the component period and their respective medians:

<table>
<thead>
<tr>
<th>Component Period</th>
<th>Median A [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1yr</td>
<td>0.05</td>
</tr>
<tr>
<td>48h</td>
<td>0.1</td>
</tr>
<tr>
<td>24h</td>
<td>0.2</td>
</tr>
<tr>
<td>12h</td>
<td>0.5</td>
</tr>
<tr>
<td>6h</td>
<td>1</td>
</tr>
</tbody>
</table>

The robust IRLS were used to estimate the regression coefficients for all the uncertainty assessment methods (same $\beta$s for all methods).
**Context**

Standard errors of tidal amplitudes

**Methods**

- **UTide**: underestimation of the variance
- **IRLS**: roughly constant uncertainty over the spectrum, ignores the coloured nature of residuals/parameters
- **NS-Tide (corrected)** and Bootstraps: realistic estimation of the parameter variability over the frequency spectrum

**Results**

*Innocenti et al.* Residual bootstrap for tidal HA regressions

---

**Figure**

- **signals with WN residuals**
- **signals with RN residuals**

The robust IRLS were used to estimate the regression coefficients for all the uncertainty assessment methods (same $\beta$s for all methods).
Uncertainty of reconstructed series

Conventional methods underestimate the variability of water level predictions based on HA

The spread at 95% is defined as: \( h^*_t \pm 2\sigma_{h^*_t} \)
where \( \sigma_{h^*_t} \) is the standard deviation of each hourly value obtained over \( 10^3 \) water level reconstructions (corresponding to \( 10^3 \) parameter replicates)

The hourly value spread at 95% is (on average) \(~24\) cm for MBB and \(~12\) cm for UTide

⇒ Calibration and validation of satellite altimetry products, analysis of sea-level rise, etc.
Conventional methods underestimate the variability of water level predictions based on HA

The hourly value spread at 95% is (on average)

~24 cm for MBB and ~12 cm for UTide

The difference in this spread might be weakly relevant for high or low water levels and weather-related extremes but it is significant when considering the series of hourly values and derived statistics (e.g., trends in mean water levels)

⇒ Calibration and validation of satellite altimetry products, analysis of sea-level rise, etc.
Using a large number of simulations with known constant statistical properties (i.e., same true parameters and residual structure), it is possible to evaluate the performance of the Signal-to-Noise Ratio (SNR) criterion for selecting the "significant" tidal constituents.
Constituent selection in High-Low water series analysis

SNR selection

By definition, the CI confidence level (e.g., 1 - \( \alpha = 95\% \)) must correspond to the proportion of intervals containing the true parameter value over a large number of the interval replicates.

Confidence Interval (CI) coverage

It is intended to select the constituents based on their relative uncertainty (~ CI width vs parameter value).

<table>
<thead>
<tr>
<th>True if good coverage</th>
<th>% of simulations excluding the constituent due to SNR&lt;2</th>
<th>% of simulations with CI recovering the true parameter value</th>
<th>Nominal CI coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 20 40 60 80 100</td>
<td>0 20 40 60 80 100</td>
<td>≥95%</td>
</tr>
</tbody>
</table>

Innocenti et al. Residual bootstrap for tidal HA regressions
Constituent selection in High-Low water series analysis

Using a large number of simulations with known constant statistical properties (i.e., same true parameters and residual structure), it is possible to evaluate the performance of the Signal-to-Noise Ratio (SNR) criterion for selecting the “significant” tidal constituents.

It is intended to select the constituents based on their relative uncertainty (~ CI width vs parameter value).

**SNR selection**

<table>
<thead>
<tr>
<th>UTide</th>
<th>IRLS</th>
<th>MBB</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHO1</td>
<td>O1</td>
<td>01</td>
</tr>
<tr>
<td>O1</td>
<td>P1</td>
<td>S1</td>
</tr>
<tr>
<td>P1</td>
<td>S1</td>
<td>K1</td>
</tr>
<tr>
<td>S1</td>
<td>K1</td>
<td>J1</td>
</tr>
<tr>
<td>K1</td>
<td>J1</td>
<td>OO1</td>
</tr>
</tbody>
</table>

**Confidence Interval (CI) coverage**

<table>
<thead>
<tr>
<th>UTide</th>
<th>IRLS</th>
<th>MBB</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHO1</td>
<td>O1</td>
<td>01</td>
</tr>
<tr>
<td>O1</td>
<td>P1</td>
<td>S1</td>
</tr>
<tr>
<td>P1</td>
<td>S1</td>
<td>K1</td>
</tr>
<tr>
<td>S1</td>
<td>K1</td>
<td>J1</td>
</tr>
<tr>
<td>K1</td>
<td>J1</td>
<td>OO1</td>
</tr>
</tbody>
</table>

- **True if good coverage**
- **% of simulations excluding the constituent due to SNR<2**
- **% of simulations with CI recovering the true parameter value ≥95%**

By definition, the CI confidence level (e.g., 1 - α = 95%) must correspond to the proportion of intervals containing the true parameter value over a large number of the interval replicates.

**Nominal CI coverage**

- **≥95%**
Constituent selection in High-Low water series analysis

Using a large number of simulations with known constant statistical properties (i.e., same true parameters and residual structure), it is possible to evaluate the performance of the Signal-to-Noise Ratio (SNR) criterion for selecting the "significant" tidal constituents. It is intended to select the constituents based on their relative uncertainty (~ CI width vs parameter value).

<table>
<thead>
<tr>
<th>True if good coverage</th>
<th>UTide</th>
<th>IRLS</th>
<th>MBB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>% of simulations excluding the constituent due to SNR&lt;2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% of simulations with CI recovering the true parameter value</th>
</tr>
</thead>
</table>

Nominal CI coverage ≥95%

Innocenti et al.

Residual bootstrap for tidal HA regressions
One symptom of inaccurate constituent estimation or selection is the association between the HA model parameters.

- **bootstrap methods easily provide information on the inter-dependencies among constituents without stringent parametric hypotheses.**
- **Circular statistics** to assess the association between phases ... and between phases and amplitudes.
Correlation between model constituents

One symptom of inaccurate constituent estimation or selection is the association between the HA model parameters.

Spectral leakage due to sparse sampling

⇒ physical relevance of the constituent interactions?
Summary and discussion
**Summary**

**Bootstrap methods for the inference of tidal HA regressions:** integrate the residual autocorrelation in the model evaluation.

1. Evaluate the theoretical differences between conventional variance-covariance estimations

2. Assess the impacts of ignoring the residual autocorrelation: parameter variability and CI

3. Examine the constituent selection strategies relying on these uncertainty assessments

- Multiple use of the weighted residuals: UTide and NS_Tide original.
- Inconsistent residual PSD estimation.
- Uncorrelated Gaussian residuals: IRLS.
- UTide underestimates the parameter variances for all residual structures.
- The IRLS provide adequate estimations for WN residuals and/or long records.
- NS_Tide, MBB, and SPB provide realistic estimates and adequate CI for most constituents.
- The SNR provides limited information.
- Diagnostics based on resampling can describe the inter-dependencies among the parameters (e.g., correlations).
Summary

Bootstrap methods for the inference of tidal HA regressions: integrate the residual autocorrelation in the model evaluation.

1. Evaluate the theoretical differences between conventional variance-covariance estimations

2. Assess the impacts of ignoring the residual autocorrelation: parameter variability and CI

3. Examine the constituent selection strategies relying on these uncertainty assessments

- Multiple use of the weighted residuals: UTide and NS_Tide original.
- Inconsistent residual PSD estimation.
- Uncorrelated Gaussian residuals: IRLS.
- UTide underestimates the parameter variances for all residual structures.
- The IRLS provide adequate estimations for WN residuals and/or long records.
- NS_Tide, MBB, and SPB provide realistic estimates and adequate CI for most constituents.
- The SNR provides limited information.
- Diagnostics based on resampling can describe the inter-dependencies among the parameters (e.g., correlations).
Summary

Bootstrap methods for the inference of tidal HA regressions: integrate the residual autocorrelation in the model evaluation.

1. Evaluate the theoretical differences between conventional variance-covariance estimations

2. Assess the impacts of ignoring the residual autocorrelation: parameter variability and CI

3. Examine the constituent selection strategies relying on these uncertainty assessments

- Multiple use of the weighted residuals: UTide and NS_Tide original.
- Inconsistent residual PSD estimation.
- Uncorrelated Gaussian residuals: IRLS.
- UTide underestimates the parameter variances for all residual structures.
- The IRLS provide adequate estimations for WN residuals and/or long records.
- NS_Tide, MBB, and SPB provide realistic estimates and adequate CI for most constituents.
- The SNR provides limited information.
- Diagnostics based on resampling can describe the inter-dependencies among the parameters (e.g., correlations).
Summary

Bootstrap methods for the inference of tidal HA regressions: integrate the residual autocorrelation in the model evaluation.

1. Evaluate the theoretical differences between conventional variance-covariance estimations

2. Assess the impacts of ignoring the residual autocorrelation: parameter variability and CI

- Multiple use of the weighted residuals: UTide and NS_Tide original.
- Inconsistent residual PSD estimation.
- Uncorrelated Gaussian residuals: IRLS.
- UTide underestimates the parameter variances for all residual structures.
- The IRLS provide adequate estimations for WN residuals and/or long records.
- NS_Tide, MBB, and SPB provide realistic estimates and adequate CI for most constituents.

Differences up to 50% in the spread of the distributions of predicted water levels
Summary

Bootstrap methods for the inference of tidal HA regressions: integrate the residual autocorrelation in the model evaluation.

1. Evaluate the theoretical differences between conventional variance-covariance estimations

2. Assess the impacts of ignoring the residual autocorrelation: parameter variability and CI

3. Examine the constituent selection strategies relying on these uncertainty assessments

- Multiple use of the weighted residuals: UTide and NS_Tide original.
- Inconsistent residual PSD estimation.
- Uncorrelated Gaussian residuals: IRLS.
- UTide underestimates the parameter variances for all residual structures.
- The IRLS provide adequate estimations for WN residuals and/or long records.
- NS_Tide, MBB, and SPB provide realistic estimates and adequate CI for most constituents.
- The SNR provides limited information.
- Diagnostics based on resampling can describe the inter-dependencies among the parameters (e.g., correlations).
Resampling methods can be applied to different noise structures and account for several aspects of the error distribution

Provide a more accurate estimation of the HA uncertainty for applications that need cm or higher accuracy water level reconstructions:

- Tides into future climate and sea-level trends
- Calibrate and validate satellite measurements → future research
- Support the analysis of sparsely sampled records

Limitations and Future Work:

- Correlated residuals may result in biased estimators of the point regression coefficients \(\beta\)
- Resampling methods are not suitable for the analysis of extreme events and/or with strong non-stationarity of the residuals
Resampling methods can be applied to different noise structures and account for several aspects of the error distribution

Provide a more accurate estimation of the HA uncertainty for applications that need cm or higher accuracy water level reconstructions:

- Tides into future climate and sea-level trends
- Calibrate and validate satellite measurements → future research
- Support the analysis of sparsely sampled records

Limitations and Future Work:

- Correlated residuals may result in biased estimators of the point regression coefficients [i.e., the $\beta$’s]
- Resampling methods are not suitable for the analysis of extreme events and/or with strong non-stationarity of the residuals
Thank you
silvia.innocenti@canada.ca
Bibliography

- Codiga [2011], *Unified tidal analysis and prediction using the UTide Matlab functions*, Graduate School of Oceanography.


- Leffler and Jay [2009], “Enhancing tidal harmonic analysis”, *Continental Shelf Research*.

- Matte et al. [2013], “Adaptation of Classical Tidal Harmonic Analysis to Nonstationary Tides, with Application to River Tides”, *Journal of Atmospheric and Oceanic Technology*.

- Pawlowicz et al. [2002], “Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE”, *Computers & Geosciences*.

- Innocenti et al., Residual bootstrap for tidal HA regressions.
Unresolved variability $\Rightarrow$ Autocorrelated HA residuals

1. Review the statistical properties of commonly used HA methods, e.g.,
   - UTide\(^1\) and T_Tide\(^2\)
   - IRLS
   - NS_Tide\(^3\)

   Unified notation

   - Band-averaged and "line-decimated" residual spectrum
   - Triple down-weighting of the variance-covariance matrix

2. Compare the methods with two resampling strategies using simulated series [known properties]
   - Moving Block Bootstrap
   - Semi-Parametric Bootstrap

   Parameter variances
   Confidence Interval coverage

3. Performance with sparsely sampled series
   - High-Low Water Levels
     - Constituent selection, variance of water levels

   The bootstraps make better estimation and use of:
   - Parameter correlations
   - Derived statistics (e.g., variance of predicted values)

---

https://youtu.be/Ye4f7x9QjtA [2:00 min]