Adaptive Filters for Processing Water Level Data

Natasa Reljin¹, Dragoljub Pokrajac¹ and Michael Reiter²
¹Delaware State University, ²Bethune-Cookman University USA

1. Introduction

Salt marshes are composed of various habitats contributing to high levels of habitat diversity and increased productivity (Kennish, 2002; Zharikov et al., 2005), making them among the most productive ecosystems on the Earth. The salt marsh consists of a halophytic vegetation community growing near saline waters (Mitsch & Gosselink, 2000) characterized by grasses, herbs, and low shrubs (Adam, 2002). Salt marshes exist between the upper limit of the high tide and the lower limit of the mean high water tide (Adam, 2002). They represent an important factor in the support of surrounding food chains, and due to the high level of productivity their economic and aesthetic value is increasing (Delaware Department of Natural Resources and Environmental Control, 2002; Zharikov et al. 2005). The survival and reproduction of many species of commercial fish and shellfish is dependent upon salt marshes (Zharikov & Skilleter, 2004). In addition, salt marshes provide critical habitat and food supply to crustaceans (Zharikov et al., 2005) and shorebirds (Potter et al., 1991). They are often considered as a primary indicator of the ecosystem health (Zhang et al., 1997). Because of their ability to transfer and store nutrients, salt marshes are an important factor in the maintenance and improvement of water quality (Delaware Department of Natural Resources and Environmental Control, 2002; Zhang et al., 1997). In addition, they provide significant economic value as a cost-effective means of flood and erosion control (Delaware Department of Natural Resources and Environmental Control, 2002; Morris et al., 2004). This economic value makes coastal systems the site of elevated human activity (Kennish, 2002).

Determining the effects of sea level rise on tidal marsh systems is currently a very popular research area (Temmerman et al., 2004). While average sea level has increased 10-25 cm in the past century (Kennish, 2002), the Atlantic coast has experienced a sea level rise of 30 cm (Hull & Titus, 1986). Local relative sea level has risen an average rate of 0.12 cm yr⁻¹ in the past 2000 years, but at Breakwater Harbor in Lewes, DE sea level is rising at the average rate of 0.33 cm yr⁻¹, nearly three times that rate (Kraft et al., 1992). According to the National Academy of Sciences and the Environmental Protection Agency, sea level rise within the next century could increase 60 cm to 150 cm (Hull & Titus, 1986).

The changes in sea level rise are particularly affecting tidal marshes, since they are located between the sea and the terrestrial edge (Adam, 2002; Temmerman et al., 2004). The prediction is that sea level rise will have the most negative effect on marshes in the areas where the landward migration of the marsh is restricted by dams and levees (Rooth & Stevenson, 2000).
If sea level rises the almost certain prediction of 0.5 m by 2100 and marsh migration is prevented, then more than 10,360 km$^2$ of wetlands will be lost (Kraft et al., 1992). If the sea level rises 1 m then 16,682 km$^2$ of coastal marsh will be lost, which is approximately 65% of all extant coastal marshes and swamps in the United States (Kraft et al., 1992).

Due to an imminent potential threat which can jeopardize the Mid-Atlantic salt marshes, it is very important to examine the effect of sea level rise on these marshes. The marshes of the St. Jones River near Dover, DE, can be considered to be typical Mid-Atlantic marshes. These marshes are located in developing watersheds characterized by dams, ponds, agricultural lands, and increasing urbanization, providing an ideal location for studying the impacts of sea level rise on salt marsh extent and location. In order to determine the effect of sea level rise on the salt marshes of the St. Jones River, the change in salt marsh composition was quantified. Unfortunately, as for most marsh locations along the Atlantic seaboard, the data on sea level rise for this area was not available for comparison with marsh condition. However, a wide data set for this area is available through a water quality monitoring program, and if it could be properly processed and analyzed it could result in sea level rise data for the location of the interest.

In this chapter, we describe the application of signal processing on the water level data from the St. Jones River watershed. The emphasis is on adaptive filtering in order to remove the influence of upstream water level on the downstream levels.

## 2. Data

The St. Jones River, in central Delaware, is 22.3 km long (Pokrajac et al., 2007a). It has an average mean high water depth (MHW) of 4 m in the main stem, and an average width of 15 feet. The site’s watershed area is 19,778 ha, and the tidal reaches are influenced by fresh water runoff from the urbanized area upstream. An aerial photo of the St. Jones River is shown in Fig. 1.

![Aerial photo of St. Jones River.](image)
The data used in this research were obtained from the Delaware National Estuarine Research Reserve (DNERR), which collected the data as part of the System Wide Monitoring Program (SWMP) under an award from the Estuarine Reserves Division, Office of Ocean and Coastal Resource Management, National Ocean Service, and the National Oceanic and Atmospheric Administration (Pokrajac et al. 2007a, 2007b). Through SWMP, researchers collect long term water quality data from coastal locations along Delaware Bay and elsewhere in order to track trends in water quality.

The original dataset contained 57,127 measurements, taken approximately every thirty minutes using YSI 6600 Data Probes (Fig. 2) (Pokrajac et al., 2007a, 2007b). The measurements were taken from January 31, 2002 through October 31, 2005. In order to determine if sea level rise is influencing the St. Jones River, the water level data were collected from two SWMP locations: Division Street and Scotton Landing (Pokrajac et al., 2007b). Probes were left in the field for two weeks at a time, collecting measurements of water level, temperature (°C), specific conductivity (mS cm⁻¹), salinity (ppt), depth (m), turbidity (NTU), pH (pH units), dissolved oxygen percent saturation (%), and dissolved oxygen concentration (mg L⁻¹). We used only the water level (depth) data for this study, which were collected using a non-vented sensor with a range from 0 to 9.1 m, an accuracy of ±0.18 m, and a resolution of 0.001 m. Due to the fact that the probes are not vented, changes in atmospheric pressure appear as changes in depth, which results in an error of approximately 1.03 cm for every millibar change in atmospheric pressure (Mensinger, 2005). However, the exceptionally large dataset (57,127 data points) overwhelms this data error.

Fig. 2. YSI 6600 Data Probe.
The downstream location, Scotton Landing, is located at coordinates latitude 39 degrees 00' 05.9160" N, longitude 75 degrees 27' 38.1049" W (Fig. 3). It has been monitored by SWMP since July 1995. The average MHW depth is 3.2 m, and the river is 12 m wide (Mensinger, 2005). This location possesses a clayey silt sediment with no bottom vegetation, and has a salinity range from 1 to 30 ppt. The tidal range is from 1.26 m (spring mean) to 1.13 m (neap mean). The data collected at the Scotton Landing site are referred as downstream data (see Fig. 4).

The water level data from the Scotton Landing site alone were not sufficient. In addition to tidal forces, this site is influenced by upstream freshwater runoff, so changes in depth could not be isolated to sea level change. However, the data from a non-tidal upstream sampling site could be used for removing the upstream influence at Scotton Landing. Therefore, the data from an upstream location, Division Street, was included in the analysis. Its coordinates are latitude 39 degrees 09' 49.4" N, longitude 75 degrees 31' 8.7" W (see Fig. 3.). The Division Street sampling site is located in the mid portion of the St. Jones River, upstream from the Scotton Landing site. At this location, the river’s average depth is 3 m and width is 9 m. The site possesses a clayey silt sediment with no bottom vegetation, and has a salinity in the range from 0 to 28 ppt. The tidal range at this location varies from 0.855 m (spring mean) to 0.671 m (neap mean). The data were monitored from January 2002 (Mensinger, 2005). The data collected at the Division Street site are referred to as upstream data (see Fig. 4).

Fig. 3. Sampling locations for St. Jones data: ▲Division Street (upstream data); • Scotton Landing (downstream data).
3. Data pre-processing

The data were sampled every $T_s = 30$ minutes, and the dataset consisted of “chunks” of continuous measurements. Some of the measurements were missing due to maintenance or malfunction of the equipment, probe replacement, etc. The length of the intervals with missing measurements varied between 1 h (1 missing measurement) and 1517.5 h (3036 missing measurements), but the majority of the intervals were shorter than 10 h.
The discrete Fourier spectra (Proakis & Manolakis, 2006) of all the chunks contained three prominent peaks, which is shown in Fig. 5 using chunk 99 from the downstream data. The first peak corresponds to lunar semi-diurnal tides with a period of approximately 12.4 h, and the diurnal tides with a period of approximately 24.8 h. In addition, there is a peak that corresponds to solar tides, which have a period of approximately 12 h. These periodicities are also shown in Fig. 6.

![Downstream data](image)

**Fig. 6.** The periodicities of the downstream data.

The dataset had several problems that had to be rectified before further processing. One data sample (Sep 28, 2004, 09:00:00) had an incorrect time, which was located sometime between Sep 27, 2004, 23:30:00 and Sep 28, 2004, 00:30:00, and was corrected. Four data samples (Jul 24, 2003, 07:30:00; Jun 10, 2005, 09:00:00; Aug 11, 2005, 15:00:00; Aug 11, 2005, 15:30:00) had missing values. In addition, the number of intervals with no measurements (total of 99 “gaps” in experiment) represented a problem for signal processing (for example, for filtering). Fig. 7 shows the number of chunks as a function of the duration of the missing measurements. Due to the properties of the used data and the shortest period of 12 h, we decided to interpolate intervals shorter than 12 h. Also, we interpolated all the above mentioned samples with missing data values. The treatment of the missing values is shown in Fig. 8.

In order to interpolate data for each interval of missing measurements, first we approximated the existing data within 20 samples from the interval. We used a least squares approximation followed the combination of the 4th order polynomial and trigonometric functions:

$$x(t) = \sum_{j=0}^{4} a_j t^j + \sum_{j=1}^{3} A_j \sin \left( \frac{2\pi t}{T_j} + \theta_j \right)$$

(1)
Fig. 7. The number of chunks as function of the duration of missing measurements.

Fig. 8. The treatment of the missing values.
where $T_1 = 12.4$ h, $T_2 = 24.8$ h and $T_3 = 12$ h. Then, we interpolated missing values using the computed approximation functions. The interpolation was performed on 866 samples, which represented less than 2% of the original number of samples. One example of the interpolated intervals is depicted in Fig. 9.

![Interpolation example](image)

Fig. 9. An example of interpolated intervals.

The interpolation resulted in the merging of the majority of chunks, thus giving us only 11 chunks. The sizes of the new chunks were as follows: 4105, 5422, 4, 4, 7154, 14357, 10750, 5, 4491, 9423, and 2278. Three of those chunks (3, 4 and 8) have very small size, which made them suitable for discarding. Therefore, the interpolation process left us with only 8 chunks.

**4. Filtering of the tidal components**

We performed discrete filtering of both upstream and downstream data using the Filter Design and Analysis (FDA) Tool in Matlab Signal Processing Toolbox, v.6.2 in order to remove the tidal periodic components from the data. The first idea was to create and use the infinite impulse response (IIR) filter (Proakis & Manolakis, 2006), because it can potentially meet the design specifications with lower order than the corresponding finite impulse response (FIR) filter, which would also result in shorter time to buffer the data. However, several attempts (using the Yule-Walker method, notch or elliptic filters) didn’t achieve the expected results - the order was too high and the attenuation was less than specified (Pokrajac et al., 2007a). Hence, we designed the FIR filter. Since the spectrum of the data had peaks in two bands (see Fig. 5), two stopband filters were designed. Both of them had a passband ripple of 0.05, and the sampling frequency $f_s = (1/30) \text{min}^{-1} = 0.556$ mHz (Pokrajac et al., 2007a). In order to have a stopband attenuation of at least 20 dB in the $11 - 11.4$ μHz band, which corresponds to a 24.8 h period, the first created filter was of order 168. The attenuation of 40 dB in the $22.401 - 23.148$ μHz band (which corresponds to periods of 12 and 12.4 h) was achieved with the second filter of order $N_{\text{filter}} = 354$. Here, more attenuation was needed due to the very high corresponding peak in the spectrum. In Figs. 10 and 11,
magnitude responses of the first and the second filters are shown. The result of applying both filters on chunk 99 and downstream data is illustrated in Fig. 12. At the beginning of each chunk, we had to discard \( N_{\text{filter}} - 1 \) data samples in order to perform filtering. This led to discarding less than 5% of the data. The standard deviation of the downstream data after the filtering was \( \text{std}(y_{\text{FIR}}(t)) = 0.200 \). Also, we tried the alternative approach by applying a moving average (MA) filter of length \( Q = 25 \), which corresponds to a period of 12.4 h. Standard deviation of the downstream data after the MA filter was \( \text{std}(y_{\text{MA}}(t)) = 0.223 \). The result of filtering the downstream data is shown in Fig. 13.

Fig. 10. Magnitude response of the first filter.

Fig. 11. Magnitude response of the second filter.
Fig. 12. Spectrum after filtering (chunk 99 and downstream data).

Fig. 13. Filtered downstream data.
5. Application of the adaptive filters

The downstream data $y_t$ can be considered as a non-stationary function of the delayed upstream data $x_t$ (see Fig. 14) (Pokrajc et al., 2007a, 2007b). It can be described as the discrete model $y_t = f_t(x_t, x_{t-Ts}, ..., x_{t-(L-1)Ts}) + r_t$, where $L$ is the maximal delay of the model and $r_t$ is the residual corresponding to the portion of the downstream data which cannot be explained by the upstream data.

Fig. 14. Removal of the upstream data influence.

If a function $f_t$ is linear, the adaptive linear model can be represented as follows:

$$y_t = w_t^T x_t + r_t$$

(2)

where $w_t = [w_{0,t}, ..., w_{L-1,t}]^T$ are coefficients and $x_t = [x_t, x_{t-Ts}, ..., x_{t-(L-1)Ts}]^T$ is the upstream data vector. A linear regression model could be obtained if the coefficients $w$ are held constant (Devore, 2007):

$$y_t = w^T x_t + r_t$$

(3)

The coefficient of determination, $R^2$, is usually used to measure the accuracy of the model, (Devore, 2007). It is defined as a function of averaged squared residuals and the standard deviation of the response:

$$R^2 = 1 - \frac{\overline{r_t^2}}{std(y_t)^2}$$

(4)
where the residuals are estimated with:

\[ \hat{r}_t = y_t - w_t^T x_t \]  
(5)

The updating of the coefficients \( w_t \) in Eq. (2) is performed using the Widrow-Hoff least mean squares (LMS) algorithm (Widrow & Stearns, 1985):

\[ w_{t+1} = w_t + 2\mu \hat{r}_t x_i \]  
(6)

where \( \mu \) represents the adjustable learning rate, and \( \hat{r}_t \) is estimated using Eq. (5). In addition to the Widrow-Hoff LMS algorithm, we applied time notching by adjusting the coefficients only when all the time instants, \( t, \ldots, t-(L-1)T_s \), belonged to the same chunk of interpolated data (Pokrajac et al., 2007a, 2007b), see Fig. 15.

![Time notching in adaptive filtering.](image)

Fig. 15. Time notching in adaptive filtering.

Using the linear regression given with Eq. (3) on the data \( y_{MA}(t) \), which is processed by the MA filter, we were able to explain only 6% of the variance, i.e. \( R^2 = 0.06 \) for \( L = 55 \). In Table 1 are shown the results of obtained \( \text{std}(\hat{r}(t)) \), for different values of the learning rate and the filter delay, when the adaptive filter given with Eqs. (2), (5) and (6) is used. Useful models were obtained when \( \text{std}(\hat{r}(t)) < \text{std}(y_{MA}(t)) = 0.200 \), and are shown in the shaded boxes in Table 1. As can be seen, the best results were obtained for \( L = 55, \mu = 0.015 \), which yielded to \( R^2 = 0.37 \). Small values of the learning rate, combined with small filter length, lead to unsatisfactory results. On the other hand, the learning becomes unstable if the filter length and learning rates are getting large.
Table 1. Std of residuals for different values of learning rate $\mu$ and the filter length $L$, for MA filter. Useful models are in shaded boxes.

When the designed FIR filter was used on the same data $y_{MA}(t)$, we received the results shown in Table 2. As can be seen, the results given in Table 1 are better. However, if $L = 45$ and $\mu = 0.015$ (which yields to $R^2 = 0.24$), the best performance of the designed FIR filter is achieved.

Table 2. Std of residuals for different values of learning rate $\mu$ and the filter length $L$, for designed FIR filter. Useful models are in shaded boxes.

Fig. 16 shows the residuals in time of the three useful adaptive filters, applied on data $y_{MA}(t)$ and processed by using the MA filter. The particular combination of the filter parameters was different, but the residuals showed similar behavior. Table 3 provides the time intervals when the relative residuals are larger than four standard deviations for the adaptive filter with $L = 55$, $\mu = 1.5 \times 10^{-2}$ (Pokrajac et al., 2007b). At the beginning of the learning process, the filter

![Figure 16](image-url)
coefficients were not adapted fully, which caused the identified peaks. These peaks corresponded to observations from February 2002. In addition, there are two other peaks, corresponding to Sep 4, 2002, and Oct 25, 2004, which can be explained by a transient behavior after the notching interval.

<table>
<thead>
<tr>
<th>Year</th>
<th>Begin</th>
<th>End</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date</td>
<td>Time</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>02/08</td>
<td>23:00</td>
<td>02/10</td>
</tr>
<tr>
<td></td>
<td>02/17</td>
<td>06:30</td>
<td>02/18</td>
</tr>
<tr>
<td></td>
<td>07/20</td>
<td>00:30</td>
<td>07/20</td>
</tr>
<tr>
<td></td>
<td>08/07</td>
<td>01:00</td>
<td>08/07</td>
</tr>
<tr>
<td></td>
<td>08/07</td>
<td>16:30</td>
<td>08/07</td>
</tr>
<tr>
<td></td>
<td>08/21</td>
<td>14:30</td>
<td>08/21</td>
</tr>
<tr>
<td></td>
<td>09/04</td>
<td>16:00</td>
<td>09/04</td>
</tr>
<tr>
<td></td>
<td>09/11</td>
<td>15:00</td>
<td>09/12</td>
</tr>
<tr>
<td></td>
<td>12/21</td>
<td>09:30</td>
<td>12/21</td>
</tr>
<tr>
<td>2003</td>
<td>01/09</td>
<td>02:00</td>
<td>01/09</td>
</tr>
<tr>
<td></td>
<td>10/22</td>
<td>12:00</td>
<td>10/24</td>
</tr>
<tr>
<td>2004</td>
<td>04/01</td>
<td>15:30</td>
<td>04/02</td>
</tr>
<tr>
<td></td>
<td>04/09</td>
<td>17:00</td>
<td>04/09</td>
</tr>
<tr>
<td></td>
<td>05/29</td>
<td>12:30</td>
<td>05/29</td>
</tr>
<tr>
<td></td>
<td>07/12</td>
<td>20:30</td>
<td>07/13</td>
</tr>
<tr>
<td></td>
<td>09/19</td>
<td>00:00</td>
<td>09/19</td>
</tr>
<tr>
<td></td>
<td>09/29</td>
<td>20:30</td>
<td>09/30</td>
</tr>
<tr>
<td></td>
<td>10/05</td>
<td>15:30</td>
<td>10/05</td>
</tr>
<tr>
<td></td>
<td>10/25</td>
<td>03:00</td>
<td>10/25</td>
</tr>
<tr>
<td></td>
<td>11/26</td>
<td>09:00</td>
<td>11/26</td>
</tr>
<tr>
<td>2005</td>
<td>09/27</td>
<td>20:00</td>
<td>09/28</td>
</tr>
</tbody>
</table>

Table 3. Identified intervals of large residuals (adaptive filtering on $y_{MA}(t)$ data, $L = 55$, $\mu = 1.5 \ e^{-2}$

6. Conclusion

We have described the application of the adaptive filtering for analyzing river hydrographic data. When determining the portion of the downstream data that is not influenced by the upstream data, the numerical results show that adaptive filtering is superior to linear regression.

7. Acknowledgement

This work was partially supported by the US Department of Commerce (award #NA06OAR4810164), NOAA (#NA06OAR4810164), NIH (#2P20RR016472-04), DoD/DoA (#45395-MA-ISP, #54412-CI-ISP, W81XWH-09-1-0062), and NSF (#0320991, CREST grant #HRD-0630388, #HRD-0310163).

8. References


Delaware Department of Natural Resources and Environmental Control (DNREC) (2002). Technical Background Report Silver Lake Watershed, pp. 81, Dover, Delaware, USA
Adaptive Filters for Processing Water Level Data


Adaptive filtering is useful in any application where the signals or the modeled system vary over time. The configuration of the system and, in particular, the position where the adaptive processor is placed generate different areas or application fields such as: prediction, system identification and modeling, equalization, cancellation of interference, etc. which are very important in many disciplines such as control systems, communications, signal processing, acoustics, voice, sound and image, etc. The book consists of noise and echo cancellation, medical applications, communications systems and others hardly joined by their heterogeneity. Each application is a case study with rigor that shows weakness/strength of the method used, assesses its suitability and suggests new forms and areas of use. The problems are becoming increasingly complex and applications must be adapted to solve them. The adaptive filters have proven to be useful in these environments of multiple input/output, variant-time behaviors, and long and complex transfer functions effectively, but fundamentally they still have to evolve. This book is a demonstration of this and a small illustration of everything that is to come.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: