CS230

Project Final Report for CS230-Spring 2018
Extraction and Analysis of Earth Tide Signals

Xuhua Gao*
Department of Energy Resources Engineering
Stanford University
xuhuag@stanford.edu

Category: Time-Series Analysis and Signal Processing

1 Introduction

The gravitational forces between the Earth and the Sun and between the Earth and the Moon cause tidal deformation of the earth body. If the downhole pressure is measured in a closed well, then the periodical earth tide signals could be extracted from the pressure measurement. We are interested in extracting earth tidal signals because it provides valuable information about the properties of subsurface formation. Equation (1.1) is a simplified equation that illustrates the relationship between the measured data and the tidal signal.

$$p(t) = p_d(t) + \Delta p(t) + e(t) \tag{1.1}$$

where $p(t)$ is the measured downhole pressure. $p_d(t)$ is the long-term trend(drift) in downhole pressure that is not caused by the earth tide. $\Delta p(t)$ is the periodic pressure change that is caused by the earth tide, i.e. the tidal part. $e(t)$ represents random noise or the measurement error. $p_d(t) + e(t)$ together forms the non-tidal part. All of $p(t)$, $p_d(t)$, $\Delta p(t)$, and $e(t)$ are function of time and are treated as time-series data.

The objective is to separate the tidal part from the non-tidal part, or to extract the tidal part $\Delta p(t)$ from the measured pressure data $p(t)$.

2 Related Work

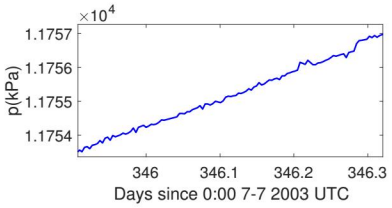
In the literature, Wada et al. (1995) studied the application of RNN for active filter to analyze electronic harmonic signals, and tidal signals in nature are also harmonic signals, so the research by Wada et al. (1995) demonstrated the feasibility of using RNN to cope with harmonic time series data including tidal signals. Mentes (2015) applied conventional feed-forward neural networks to perform barometric correction of the tidal data, and an improvement of 2-5% in tidal factors was reported. Chuan Tian and Roland Horne (2017) applied RNN for downhole pressure data analysis and pointed out an RNN method to determine the long-term drift based on injection rate.

3 Dataset and Preprocessing

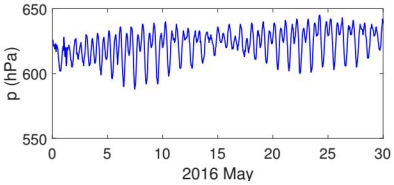
2.1 Introduction to dataset
We collected dataset from two different locations. Data from location 1 are noisy, and the amplitude of tidal signal is relatively small, thus very hard for traditional methods to perform well in term of extracting tidal signals. Data from location 2 are smooth and have very clear tidal variations, traditional methods perform well, but still could be improved. Data from location 2 is a good start point for neural network to learn tidal variation pattern. Figure 1a and Figure 1b shows two samples from location 1 and location 2 respectively.

*SUID: xuhuag; 06047680

CS230: Deep Learning, Spring 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

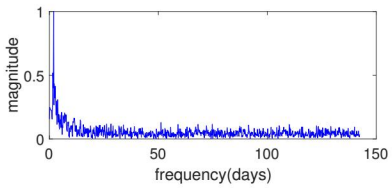


(a) sample data from location 1

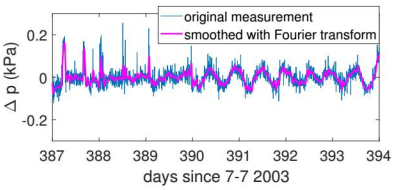


(b) sample data from location 2

Figure 1: Sample data from location 1 and location 2

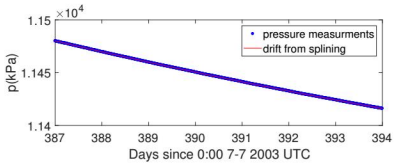


(a) frequency domain

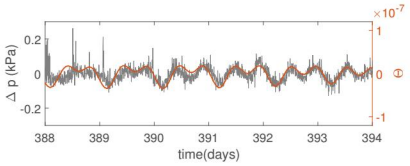


(b) smoothed data with Fourier transform

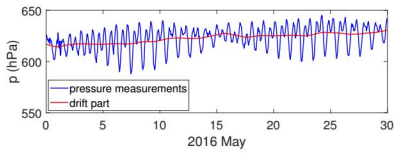
Figure 2: Preprocessing with fast Fourier transform



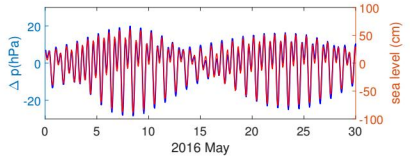
(a) drift part given by low-pass filtering



(b) comparison with theoretical tides



(c) comparison with theoretical tides



(d) comparison with theoretical tides

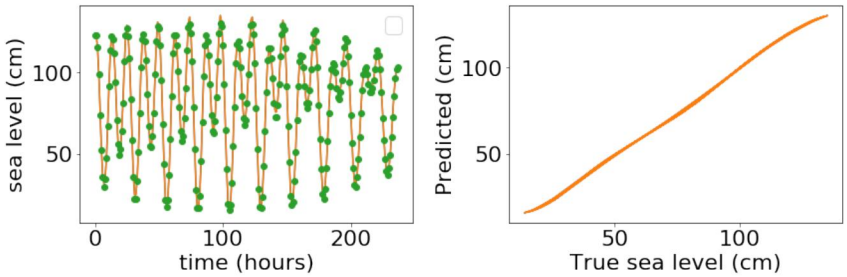
Figure 3: baseline results from low-pass filtering

2.2 Data preprocessing with Fourier transform

Due to that data from location 1 is quite noisy, we applied Fourier transform to remove the high-frequency noise and smoothed the data. The frequency domain is shown in Figure 2a, and Figure 2b shows the smoothed data after Fourier transform. Note that in Figure 2a, the largest amplitude actually corresponds to semidiurnal and diurnal tidal waves. The frequency of semidiurnal tides is half day, and the frequency of diurnal tides is one day, so the peak close to the left edge of Figure 2a reflects tidal frequencies. By removing noise whose frequency is much higher than that of tidal waves, the data become smoother and display a clearer tidal variation patten, as indicated by the comparison in Figure 2b. Data from location 2 have very high quality and are already smooth, so it could used directly for tidal signal extraction.

2.3 Baseline results obtained from low-pass filtering

In order to provide a baseline to evaluate the performance of neural network, we need to first try to extract tidal signals using the conventional low-pass filtering method. The low-passing filtering method was used in the paper by Burbey (2010). The tidal signals given by low-pass filtering were compared against theoretical tides for both location 1 and location 2. The red curves in Figure 3a and Figure 3c indicate the drift part obtained from low-pass filtering for location 1 and location 2 respectively. Figure 3b and Figure 3d show the comparison between extracted tidal signals and theoretical tides (true label). For location 1, we can see that although the background tidal signals are noisy, the trend of tidal signal still matches that of the theoretical tides (Θ, the right vertical axis). For location 2, the extracted signal matches theoretical tides very well. It can be seen from Figure 3b and 3d that the extract signal varies with theoretical tides linearly, so that we could evaluate the RMS



(a) Prediction on test set (b) True sea level vs. prediction

Figure 4: Prediction of theoretical tides

error of the low-passing filtering method by fitting a linear model (Figure 5a). The RMS error from low-passing filtering is 1.42 for location 2 , and 5.57 for location 1.

4 Methods

In this study, we adopted two neural network architectures for tidal signal extraction:
1. RNN with LSTM architecture and dropout regularization;
2. Nonlinear autoregressive neural network with external input (NARX);
For LSTM, each cell consists of a forget gate, an update gate and an output gate. The behavior of each gate is controlled by the corresponding weight matrix. The activation function is the hyperbolic tangent function, and for the recurrent step the activation function is the sigmoid function. The loss was calculated using mean squared error (mse), and Adam optimization method was adopted. For NARX, first consider an autoregressive process of order n:

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, ..., y_{t-n}) + \varepsilon_t \tag{4.1}$$

where f is a non-linear function from R^n to R, and ε_t is an i.i.d. $N(0, \sigma^2)$ random variable. With existence of external input, the structure becomes:

$$y_t = f(y_{t-1}, y_{t-2}, ..., y_{t-n}, x_{t-1}, x_{t-2}, ..., x_{t-n}) + \varepsilon_t \tag{4.2}$$

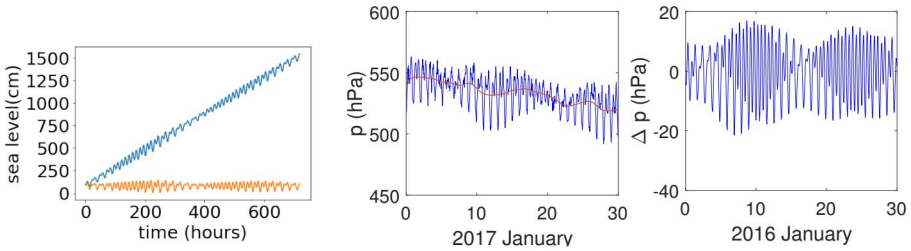
where x_t is the external input time series. In our case, the function f was approximated using a feed forward neural network (a multilayer perceptron or MLP) and the activation function was chosen to be hyperbolic tangent function. The main hyperparameters of a NARX model include the number of layers and neurons when approximating the function f in Equation (4.2) and the number of feedbacks n. The loss was calculated using the root mean squared error.

The training process followed two major steps: First, the model was trained to learn theoretical tidal variation pattern. After that, a drift part was added to the input data, and the model was trained to differentiate tidal signal from the drift.

5 Results and Discussion

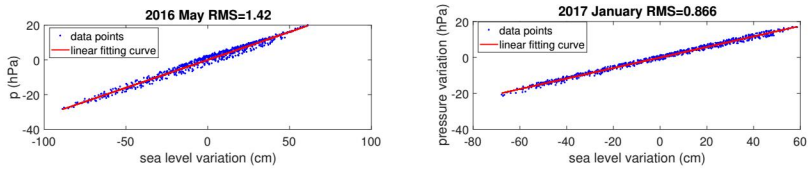
5.1 Learn theoretical tides
In the first step, the model was trained to predict theoretical tidal variation based on history information. One month of theoretical tides was used as input, and the model predicted the theoretical tidal change for the following week. This is a relatively simple task and both LSTM and NARX performed well. As an example, the results from LSTM model are shown in Figure 4a. The orange curve in Figure 4a is the true sea level data, and the green dots are predicted values. The results were obtained after around 300 epochs. In the error analysis, the predicted value was compared with true value through a linear fitting process, and the results are shown in Figure 4b. The RMS error is 0.99 for training data, and 0.86 for test data, so the model is not overfitting and is able to predict theoretical tides with reasonable accuracy.

5.2 Tidal signal extraction
Based on the model describe in 5.1, a drift part (or long-term trend part) was added to theoretical tides and the ability of the network to differentiate tidal signal from the drift was tested both based on the LSTM model and the NARX model.
For the LSTM model, in early experiment cycles, the network was trained using a simple data



(a) Idealized case: linear trend

(b) real data case LSTM result



(c) real data case error: low-pass filtering

(d) real data case error: LSTM

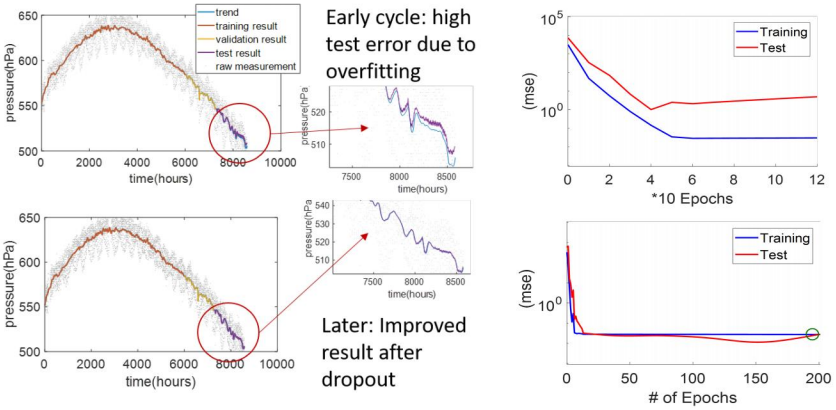
Figure 5: LSTM performance

structure where the drift part followed a linear pattern, as shown in Figure 5a by the blue curve, and the LSTM network provided a good result (the output is shown by the orange curve in Figure 5a). In later experimental cycles, real data from location 2 were fed into the model and the results from LSTM are shown in Figure 5b, where the blue cyclic curve in the left figure of Figure 5b is the input, while the red curve in the left figure is the output drift part, and the right figure shows the extracted tidal signal. The performance of LSTM network (Figure 5d) was compared with that of low-pass filtering (Figure 5c), and it can be seen that the RMS error was reduced from 1.42 (low-pass filtering) to 0.85 (test error of LSTM).

For the NARX model, when the network was trained using data from location 2, there was a overfitting problem in early experiment cycles, as shown by Figure 6. In Figure 6a, we can see that before applying regularization, although the training performance and validation performance were good, the test performance was not satisfactory, indicating that the model was overfitting. The zoom-in version of Figure 6a clearly shows an improve in test accuracy. The problem of overfitting was alleviated in later experiment cycles when the number of neurons were reduced after regularization. As shown in Figure 6b, the training history before regularization displays a large gap between training performance and the test performance, and after regularization, the gap was reduced and training performance approached test performance when the training ended after around 200 epochs. Figure 6b also shows that the test error was reduced after applying regularization. After fine tuning the hyperparameters, the NARX model is also tested for data from location 1, and the results are shown in Figure 7. As described in Section 3, the data from location 1 are noisy and thus smoothed by Fourier transform, and the input data for the neural network are smoothed data. The relative amplitude of tidal variation was amplified so that the tidal signal became more detectable (note that when the amplitude of tidal variation is smaller than the amplitude of random error, it is very difficult to extract tidal signals). The drift part of data from location 1 is also more complex, consisting of several sharp corners. However, the NARX model still extracted out tidal signals with lower error than baseline result. The RMS error in Figure 7b is 0.48, which is better than 5.57 given by low-pass filtering. The performance of LSTM is not as good as NARX in the case of location 1.

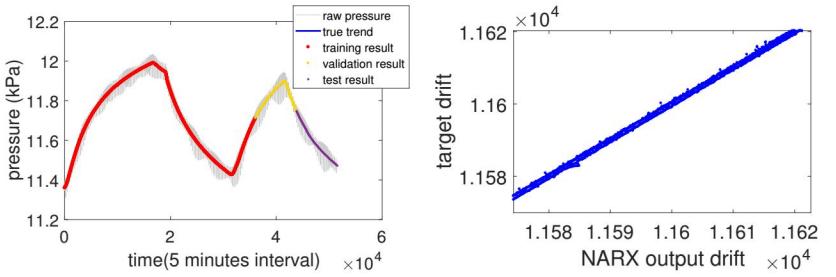
5.3 Discussion on hyperparameter tuning

The main hyperparameters tuned for LSTM model are the number of epochs and the number of neurons. The best number of epochs for the LSTM model were found when the validation error becomes flat and stops decreasing significantly. The best number of epochs for the NARX model were determined in the same way. An example is provided in Figure 6b, where the green circle denotes the optimized number of epochs. For LSTM, the model tends to underfit in early cycles, so number of neurons were gradually increased until training performance stopped improving and test performance was approximately the same with the training performance. For NARX, the process was reversed, as the model overfitted in the beginning, and after the number of neurons were reduced after regularization, the test performance increased. An additional important hyperparameter that needs to be fine tuned for NARX is the feedback delays (n in Equation (4.2)). Similar to the number



(a) NARX time series response location 2 (b) Training history location 2

Figure 6: NARX performance for location 2



(a) NARX time series response location 1 (b) NARX test error location 1

Figure 7: NARX performance for location 1

of neurons, the feedback delays is fined tuned so that the model neither underfits nor overfits. The optimal level of feedback delays was found to be around 8 time steps.

5.4 Justification of choices and decisions

Major choices made in this study include: 1. choice of the model/architecture (LSTM and NARX); 2. choice of activation functions. LSTM and NARX models were chosen due to their advantages in dealing with time series data, especially time series forecasting. Activation functions are chosen to be hyperbolic tangent function due to our requirement of stronger gradient for tidal analysis. In addition, the neural network does not need to be very deep for our purpose (we have less than five layers) so that we do not need to worry about vanishing or exploding gradient problems. Although LSTM is more well-known than NARX, for this particular problem, we found that the performance of NARX is even better. For theoretical tide prediction and cases with simple data structure (e.g. linear drift), both LSTM and NARX performed well. Nevertheless, for cases with more complicated data structure (e.g. real data with any random drift), the performance of NARX surpassed that of LSTM. Siegelmann et. al. (1997) demonstrated that NARX is as much computationally powerful as a fully connected RNN. The success of NARX reported in the literature is one of the motivations behind our choice of the architecture.

6 Conclusion and Future Work

In this study, we examined the application of sequential deep learning models in tidal signal analysis and extraction. Two deep learning architectures were investigated, including RNN with LSTM and NARX. Both models showed the capability of extracting tidal signals accurately and provided better results than conventional data filtering methods. The performance of NARX model is better than the performance of LSTM when the data structure is relatively complicated. In the future, the model could be further developed to analyze tidal measurement data with even smaller amplitude and more complex drift behavior. Increasing the training data size could be considered in future work as well.

References

[1] Burbey, T.J. (2010) Fracture characterization using Earth tide analysis *Journal of Hydrology* **380**(3-4): 237-246.

[2] Mentes, G. (2015). Artificial neural network model as a potential alternative for barometric correction of extensometric data. *MARÉES TERRESTRES: BULLETIN D INFORMATIONS, 149*, 12001-12012.

[3] Siegelmann, H. T., Horne, B. G., & Giles, C. L. (1997). Computational capabilities of recurrent NARX neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 27(2), 208-215.

[4] Tian, C., & Horne, R. N. (2017). Recurrent neural networks for permanent downhole gauge data analysis. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

[5] Wada, Y., Pecharanin, N., Taguchi, A., Iijima, N., Akima, Y. and Sone, M. (1995). Application of recurrent neural network for active filter. In *Neural Networks Proceedings.*, IEEE International Conference(Vol. 1, pp. 488-491).