State-based Regression for Modeling the Non-linear Dependency between Time Series

Bhargob Deka[1], James-A. Goulet[1]
[1]Department of Civil, Geologic and Mining Engineering, Polytechnique Montreal, Montreal, CANADA
Email: bhargobdeka11@gmail.com, james.goulet@polymtl.ca

ABSTRACT: In this paper, we present the state-based regression method for modeling the non-linear dependency between two time series. This is particularly important in the context of dam monitoring where displacement at various locations needs to be predicted depending on the reservoir water level as the input. We have developed a new state regression sub-component within the BDLM framework such that we can model the non-linear relationships between time series. The results show that the new component improves our prediction accuracy while providing valuable information regarding the type of non-linear relationships that exists between two time series.

KEY WORDS: BDLM; Time Series Forecasting; Non-linear Dependency; Probabilistic models; Dam Monitoring.

1. INTRODUCTION

Bayesian dynamic linear models (BDLM) [1] are probabilistic models used for time series analysis that are capable of online learning. BDLM consists of generic sub-components that each captures a specific pattern and that can be grouped together to model time series. Up to now, BDLM could not model non-linear relationships between two time series. This is particularly important in the context of dam monitoring where the displacement at various locations needs to be predicted depending on the reservoir water level as the input [2]. For such a scenario, learning the non-linear relationship between the water level and the dam displacement is essential. In this paper, we propose a state-based regression method to model the non-linear dependency between any two time series.

2. METHODOLOGY

The state-based regression is a kernel method relying on a set of control-points, where each point consists of a reference variable $x^{\mathtt{cp}}$ associated with a hidden state x^{ϕ}. The reference variables are defined as a fixed set of values covering the entire range of the *reference time series* used to model the dependent time series through the non-linear regression. A radial basis function (RBF) kernel is used to measure the similarity between the hidden state associated with the reference time series $x^{\mathtt{P}}$ and the reference variable $\boldsymbol{x}^{\mathtt{cp}}$ given by

$$k(x^{\mathtt{P}},\boldsymbol{x}^{\mathtt{cp}}) = \exp\left[\frac{-1}{2(\ell^{\mathtt{SR}})^2}(x^{\mathtt{P}}-\boldsymbol{x}^{\mathtt{cp}})^2\right], \qquad (1)$$

where $k(x^{\mathtt{P}},\boldsymbol{x}^{\mathtt{cp}})$ gives the kernel values as a function of the distance between the two covariates $x^{\mathtt{P}}$ and $\boldsymbol{x}^{\mathtt{cp}}$, and the kernel length $\ell^{\mathtt{SR}}$. The hidden state associated with the regression coefficient x_0^{ϕ} is computed using a weighted summation of the kernel outputs $k(x^{\mathtt{P}},\boldsymbol{x}^{\mathtt{cp}})$ and the control point's hidden states $\boldsymbol{x}^{\phi}$. Finally, the dependent time series is obtained by multiplying the independent variable with the state-dependent regression coefficient $x^{\mathtt{D}} = x_0^{\phi}\cdot x^{\mathtt{P}}$.

The non-linear regression method is applied by creating a new component called the state regression (SR) within the BDLM framework such that we can model non-linear relationships between time series and combine it with other available generic components while keeping the entire computation analytically tractable. The SR component provides the estimated values $\mu_{t|t}$ and their uncertainty bounds $\mu_{t|t}\pm\sigma_{t|t}$ for the predicted regression coefficient X_0^{ϕ} and the predicted pattern for the dependent time series $X^{\mathtt{D}}$.

3. CASE STUDY

In this section we show the application of our method for predicting the displacements for an actual dam based on its non-linear relationship with the measured water level.

3.1. *Dataset*

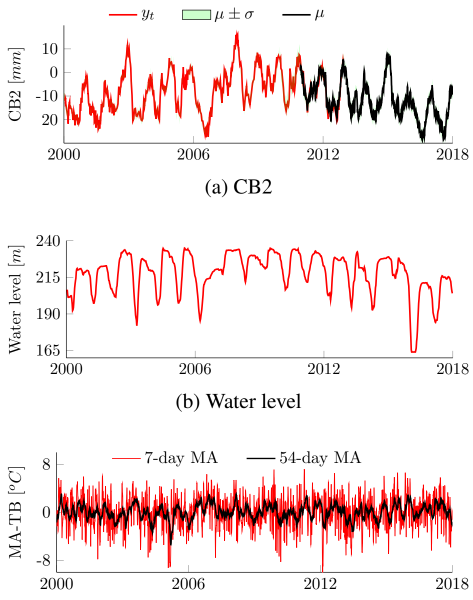
In this paper, we model the dam's radial displacement [mm] CB2 using its relationship with the available reservoir water level time series and the daily temperature recordings TB. We use daily data for the CB2 dataset, and as a result, there are missing values since the original dataset has an average frequency of acquiring of 1.5 week. We truncate the water level data at the bottom of the dam i.e., 196m to account for the physical constraint associated with the reservoir typology. For considering the thermal inertia, we consider moving averages (MA) of $\{1,7,14,28,54\}$ days for the residuals of TB obtained by removing the annual periodic pattern. The water level and the moving averages for TB are available to use for the whole period from $2000-2018$, whereas the CB2 dataset is available only upto the end of 2012.

3.2. *Model Architecture*

The CB2 displacement is modeled using a local level component to capture the long-term average value, two state regression components to model the non-linear dependency with respect to 1) the average long-term pattern for the water level, and 2) the mean-centered raw water level, a linear dependency with the moving averages for TB, and an autoregressive term to model the residuals. The mean centered water level is modeled using an autoregressive component, and the long-term pattern is modeled using a local trend component. The moving averages for TB are modeled using autoregressive components.

3.3. Results

Figure 1 provides the raw data and the model forecasts for CB2 along with the water level and two of the moving averages for TB. The true values are shown in red, the model predictions in black and their uncertainty bound using a green shaded region. To validate our model's performance, we further divide the dataset into a training set and a validation set. Figure 2 shows the validation result for the period 2010 – 2013 which confirms that our method has a competitive performance. The method has a total training time in the order of an hour. But once trained, can produce predictions in the order of a minute.



(a) CB2

(b) Water level

(c) 7 & 54 days moving average TB

Figure 1. The raw values and the forecasts for CB2 displacements, water level, and 7 and 54 day moving averages for the temperature TB.

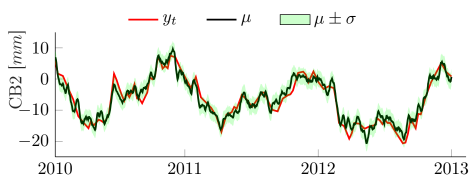
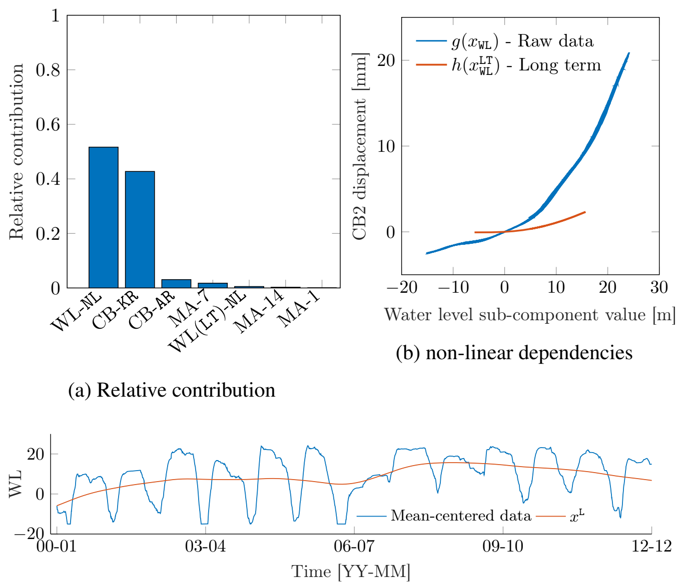


Figure 2. Performance of the state-based regression method on a validation set.

3.3.1. Model Interpretation

Figure 3 shows the different information that can be derived from our BDLM model; (a) presents the relative contribution of each component used for creating the model, (b) presents the non-linear dependencies extracted with respect to both the mean-centered water level (in blue) and the long-term pattern (in red), respectively, and (c) presents the mean-centered water level and its long-term pattern. The results shows that the component modeling the non-linear dependency WL-NL is the most important factor followed by other components modeling the periodic pattern CB-KR and the residuals CB-AR. The long-term pattern for the water level WL(LT)-NL has a low contribution but remains useful for detecting smaller magnitude anomalies that evolves over time. Except for the MA of 7 days TB-MA7, the temperature time series has the least contribution overall but it affects the daily fluctuations in the forecast for CB2.



(a) Relative contribution

(b) non-linear dependencies

(c) Mean-centered raw water level data and its long-term pattern

Figure 3. Model interpretation for the CB2 dataset made from the BDLM components.

4. CONCLUSION

In this paper, we present a novel approach of modeling non-linear dependency using a state-based regression (SR) method to be used for time series forecasting. The SR method is built on the BDLM framework such that it provides a probabilistic approach and is easier to combine with other generic sub-components. The results show that such a method is important for not only improving prediction accuracy but also understanding the relationships between different structural responses and other physical quantities affecting the structure.

ACKNOWLEDGEMENTS

This work is financially supported by research grants from Hydro-Quebec/IREQ, and the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Goulet, J.-A. (2017). "Bayesian dynamic linear models for structural health monitoring." *Structural Control and Health Monitoring*, 24(12), e2035.

[2] Salazar, F., Morán, R., Toledo, M. Á., and Oñate, E. (2017). "Data-based models for the prediction of dam behaviour: a review and some methodological considerations." *Archives of computational methods in engineering*, 24(1), 1–21.