Nonlinear multivariate modelling of wetland dynamics

Angesh Anupam
AAnupam@cardiffmet.ac.uk
School of Technologies, Cardiff Metropolitan University
Cardiff, UK

ABSTRACT

Wetlands are very complex yet pivotal ecosystems on Earth. They serve as habitats for various flora and fauna. Alongside, wetlands are crucial for biogeochemical exchange between the Earth's surface and its atmosphere. A large proportion of organic carbon is sequestered in wetlands and plays a substantial role in the carbon cycle. The planning and management of wetlands depend a lot upon a reliable wetland model. The underlying complex dynamics of wetlands hinder the modelling of wetland extent. This study for the first time considers multivariate nonlinear dynamical system modelling using Nonlinear Autoregressive with Exogenous Inputs (NARX) model class. The data consists of weather variables and wetland fractions for two wetland sites falling under Asia and Africa. The model is simulated using fresh testing data and can predict wetland extent satisfactorily for both sample sites. The accuracy of the models is quantified using Root Mean square Error (RMSE) and Mean Absolute Error (MAE). A transparent NARX structure reveals the dynamical elements for the potential planning and management of wetlands.

CCS CONCEPTS

• **Applied computing**;

KEYWORDS

wetlands, environmental systems, NARX, nonlinear system identification

ACM Reference Format:
Angesh Anupam. 2022. Nonlinear multivariate modelling of wetland dynamics. In *2022 4th International Conference on Advanced Information Science and System (AISS 2022), November 25–27, 2022, Sanya, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3573834.3574500

1 INTRODUCTION

Wetland is an important ecosystem responsible for harbouring a wide range of biodiversity. A typical wetland system comprises complex biogeochemical processes at both micro and macro scales. Whilst the entire planet Earth is witnessing a surge in the average temperature it becomes imperative to conserve the carbon in form of various vegetation including peatlands and wetlands [4]. Along with the sequestration of carbon, wetlands are responsible for methane exchange with the atmosphere and thus incorporation of wetland dynamics in climate projection is very important. Nonetheless, the complex and nonlinear dynamical processes complicate wetland modelling. Until now, the practitioners have mostly relied upon analytical models [8] representing the inundation dynamic of a land surface but that approach has its own limitations. In this study, a system approach is applied which considers wetlands as a complex nonlinear dynamical system having multiple inputs and a single output of interest, known as, wetland fraction (f_w), which is the fraction of inundated land area. The modelling methodology is system identification, which is often considered a grey box approach. Contrary to the other data-driven nonlinear modelling techniques, the grey box is a transparent model structure derived from the system's data and consisting of model terms. This approach enables the reasonable abstraction of the wetlands' complexities through the usage of environmental data containing the underlying dynamics of the system. The model terms are transparent enough to tease apart the characteristics of wetland systems like the extent to which a particular variable is influencing the output of interest, f_w.

For the first time, a system identification method is applied to obtain multivariate wetland models. The contribution of this study is, obtaining Multiple Input Single Output (MISO) wetland models using Nonlinear Autoregressive Moving Average with Exogenous Inputs (NARX) model class under the realm of system identification. The f_w data used in this study is prepared through remote sensing techniques whereas weather data are employed as the input variables [7]. The study takes into account two sample sites altogether from Asia and Africa. This modelling approach serves two purposes. Firstly, the identification of multivariate wetland model structure reveals the interrelationship among the weather variables and f_w. Secondly, the forecasting of f_w for the management of wetlands in the advent of changing climate.

The paper is structured into various sections. The following section reviews some related works and establishes the rationale behind this study. Section 3 details the materials such as data used in this study. The same section will also describe the used methodologies such as nonlinear system identification. Section 4 will present all the major results and discuss their significance. Finally, the paper will be concluded with some future directions for this work.

2 RELATED WORKS

Peatlands are considered a very important class of wetlands because of their ability to sequester a large amount of organic carbon. This results in a significant contribution of these wetlands to the global carbon cycle. A study [4] integrates the Dynamic Global Vegetation Model (DGVM), wetland model, and dynamics of peats to investigate the impact of these wetlands on the carbon cycle. The wetland model part is adapted from the TOPMODEL framework,

AISS 2022, November 25–27, 2022, Sanya, China

ACM ISBN 978-1-4503-9793-3/22/11…$15.00
https://doi.org/10.1145/3573834.3574500

which is one of the popular approaches for analysing the inundation dynamics of a land surface. The TOPMODEL [2] considers the topography for calculating the hydrological state of a given surface. The TOPMODEL needs Compound Topographic Index (CTI) to parametrise the model which induces uncertainties resulting from the approximations linked with CTIs. These uncertainties lead to incorrect estimation of wetlands and thus require a more robust approach such as obtaining high-resolution CTI parameters [6]. A project known as, WETCHIMP [8] compared ten models by simulating wetland extent across the world and comparing them with the observation data. The models within the WETCHIMP project exhibited significant inconsistency when compared to the observed values of wetland extent. The review of the above kinds of the literature suggests that process-driven analytical modelling of wetland dynamics has some limitations, and an alternative framework can be explored. Recently, a data-driven framework for wetland modelling applied NARX modelling for a Single Input Single Output (SISO) system wherein the surface temperature is considered as input and f_w is the output of interest [1]. This literature presents a satisfactory wetland model for three sample sites corresponding to the Amazon basin, Africa, and Asia as far as forecasting the wetland extent is considered. However, a univariate approach hinders the relationship of f_w with other environmental variables such as precipitation. The purpose of the proposed grey box NARX model in the literature is not fully served in the absence of a multivariate model structure. The present study is an extension of [1] wherein more than one input variable are used for obtaining NARX model structures corresponding to the sample wetland sites.

3 MATERIALS AND METHODS

3.1 Multivariate wetland modelling: a nonlinear system identification

System identification methodology has its origin in control theory but in the last few decades it has found applications in multiple areas [3]. This is essentially a data-driven dynamical system modelling approach. For nonlinear systems, this methodology offers NARX model class. A NARX structure typically consists of a combination of linear and nonlinear regressor terms representing the systems' dynamics. The data corresponding to a system contains useful information relevant to the system. Contrary to other competing data-driven modelling methodologies such as deep learning, NARX enables a grey box picture. This means, the model terms are visible and the dynamics of the system can be teased apart with the help of model structure as well as with the estimated parameter values. The overall model structure in conjunction with their parameter values represents a nonlinear function of the associated variables.

In this study, two wetland sites are considered for modelling and simulation of their wetland extent, represented by f_w. Therefore, the output of interest is f_w. The observed value of f_w is obtained using remote sensing technology [7]. Amongst the various environmental variables, precipitation and average surface temperature are considered input variables. These two variables are easily available through the various weather models and therefore the wetland model can also be used to forecast the variations in the wetland extent. The water cycle is crucial to model the inundation dynamics and hence considering precipitation values indirectly accounts for water intake and potential runoff. Similarly, the surface temperature is a key variable affecting the evaporation and eventually the extent of the wetland at a given time. Thus, two input variables and one output variable form the basis of this modelling study. In system theory terminology, this can be referred to as a Multiple Input Single Output (MISO) system. Both the input variables are shown using Fig. 1.

3.2 NARX modelling

NARX is a data-driven dynamical system modelling approach consisting of broadly two major steps, namely, model structure detection and parameter estimation. The structure of a NARX model is very transparent typically consisting of difference and differential equations. In its simplest form, it can be a Single Input Single Output (SISO) model. A SISO model despite being very simple can serve the modelling objectives especially if the point of interest is mainly forecasting. However, the purpose of the NARX modelling approach is not fully harnessed if we are unable to visualise the individual dynamical element within the system. Therefore, for many practical applications, multivariate models are employed. A general NARX structure is represented as [3],

$$\begin{aligned} y(t) = &F[y(t-1), y(t-2), ..., y(t-k), \\ &x(t-d), x(t-d-1), ..., x(t-d-l)] + e(t), \end{aligned} \tag{1}$$

where $y(t)$ represents system's output, $x(t)$ denotes input, $e(t)$ stands for noise sequences, k is the maximum lag of the system's output, l is the maximum lag of system's input, F is a nonlinear function, and d represents the time-delay. In this study, the input consists of precipitation and average surface temperature whereas the output of interest is f_w. Hence, it should be termed as a MISO system corresponding to which NARX models are obtained.

The linear-in-the-parameters representation of the NARX model is,

$$y(t) = \sum_{i=1}^{M} \theta_i \phi_i(t) + e(t) \tag{2}$$

In the eq. 2, $y(t)$ with $t = 1, 2, ..., N$ is the output, $\phi_i(t)$, with $i = 1, 2, ..., M$ represents regressor terms, θ_i, with $i = 1, 2, ..., M$ are the model parameters, $e(t)$ denotes noise sequence. From the eq. 2, it can be inferred that M represents the total number of model terms. From the implementation point of view, the eq. 2 can be transformed into,

$$\boldsymbol{y} = \Phi\boldsymbol{\theta} + \boldsymbol{e} \tag{3}$$

where, $\boldsymbol{y} = [y(1), y(2), ..., y(N)]^T$, $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_M]^T$, $\boldsymbol{e} = [e(1), e(2), ..., e(N)]^T$, $\Phi = [\phi_1, \phi_2, ..., \phi_M]$. The equation 3 allows estimation of the parameter vector $\boldsymbol{\theta}$ [3].

In this study, Forward Regression Orthogonal Least Square (FROLS) algorithm is used for obtaining a parsimonious NARX model structure. The FROLS principle allows the ranking of the terms based on their contribution to explaining the variance of the output. The contribution of each term is quantified by Error Reduction Ratio (ERR). The sum of ERR values of all the terms in a NARX model is 1, but instead of considering all the terms, a parsimonious model is preferred in most cases to avoid data overfitting. In this study,

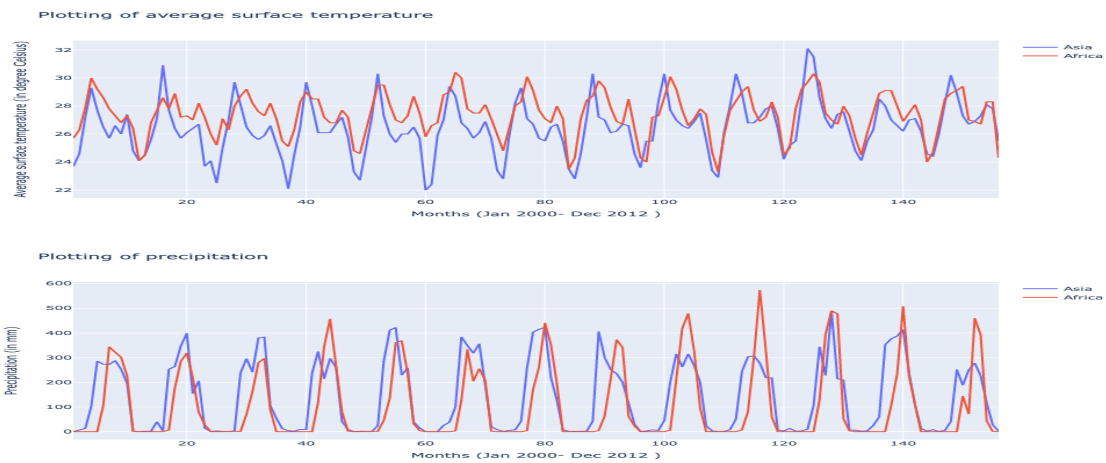

Figure 1: Time series plots of input variables – Average surface temperature and Precipitation

a Python library, SysIdentPy [5] was used to obtain satisfactory NARX models representing the dynamics of both the wetland sites.

This work is intended for performing multivariate modelling and simulation of wetland dynamics using sample sites from Asia and Africa. Surface temperature (x_1) and precipitation (x_2) were considered as inputs whereas, the output (y) is wetland fraction (f_w). The maximum possible lag for all the variables was set as 12 due to the cyclic nature of the data. All the variables are expected to show similarity after 12 data points, therefore, k and l equal to 12 is a reasonable choice. The order of polynomials in a NARX term is set to 3 for describing the nonlinearity of the system satisfactorily. All the variables are available as time series data starting from January 2000 until December 2012. The entire data was split into a training set consisting of data from January 2000 until December 2010 and the rest of the data was assigned to the testing set. The SysIdentPy library was used to train the NARX model for each sample site and the models were simulated using the test data.

4 RESULTS

In this section, the NARX model structure for both the sample sites, their estimated parameters, and ERR value of each term as well as a model simulation based on testing data will be presented. The NARX-MISO model for the Asian wetland site is presented through Table 1. The NARX model for the Asian site consists of six terms and interestingly only precipitation is used as input in the model. The FROLS algorithm picks up a term based on their contribution and in this case, none of the terms consisting of average temperature was selected by the algorithm. The first term is an autoregressive term having a time lag equal to 12 and is the most significant as well with ERR approximately equal to 0.96. However, to explain the overall dynamics of the Asian wetland system, other remaining terms are also included.

Term No.	Regressors	Parameters	ERR
1	$y(t-12)$	1.1232	0.9613
2	$y(t-12)^2 y(t-7)$	-45.1572	0.0040
3	$x_2(t-12)y(t-11)y(t-5)$	0.0292	0.0046
4	$y(t-12)y(t-7)y(t-1)$	34.8029	0.0030
5	$y(t-12)^2 y(t-4)$	-17.7382	0.0039
6	$x_2(t-12)y(t-11)y(t-6)$	-0.0118	0.0011

Table 1: NARX-MISO model for Asian wetland

For a simpler interpretation of the Asian NARX model structure (Table 1), the eq. 4 can be referred.

$$\hat{y}(t) = 1.1232 \cdot y(t-12) - 45.1572 \cdot y(t-12)^2 y(t-7), ..., \\ -0.0118 \cdot x_2(t-12)y(t-11)y(t-6) \tag{4}$$

Similarly, the NARX model corresponding to the African wetland is shown in Table 2. In this case, also, the model is mainly driven by an autoregressive term having a time lag equal to 12. However, unlike the Asian wetland system, both input variables, namely, average surface temperature and precipitation constitute the model structure. The ERR value of the first term suggests a very strong contribution to the term. Nonetheless, to explain the wetland dynamic of this site satisfactorily, the rest of the terms are also included by the FROLS algorithm. A simple representation of the model can be made through the eq. 5.

$$\hat{y}(t) = -0.6350 \cdot y(t-12) + 0.0909 \cdot x_1(t-12)y(t-1), ..., + \\ 537.3041 \cdot y(t-10)y(t-9)y(t-5) \tag{5}$$

The performance of both the models is summarised in Table 3 using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The model predicted output of both the NARX-MISO models are shown through simulation in Fig. 2.

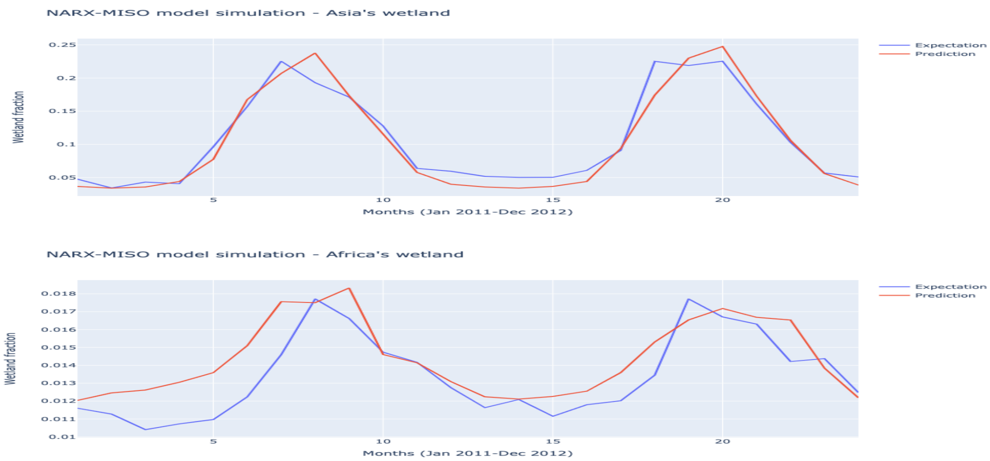

Figure 2: Model predicted output (MPO) simulation for both the sample wetland sites −Asia and Africa

Term No.	Regressors	Parameters	ERR
1	$y(t-12)$	-0.6350	0.9884
2	$x_1(t-12)y(t-1)$	0.0909	0.0038
3	$y(t-11)$	0.2971	0.0011
4	$y(t-11)y(t-10)y(t-7)$	-515.1535	0.0005
5	$y(t-12)y(t-10)y(t-7)$	3165.6270	0.0003
6	$x_2(t-12)y(t-5)^2$	0.0333	0.0002
7	$x_1(t-12)y(t-1)^2$	-2.9164	0.0007
8	$x_1(t-12)y(t-10)y(t-7)$	-1.7832	0.0001
9	$x_2(t-1)y(t-9)y(t-3)$	0.0219	0.0001
10	$y(t-10)y(t-9)y(t-5)$	537.3041	0.0001

Table 2: NARX-MISO model for African wetland

Site	**RMSE**	**MAE**
Asia	0.0183	0.0138
Africa	0.0015	0.0011

Table 3: Summary of the performance of the NARX models in predicting f_w for the years 2011-2012.

5 CONCLUSION

This study presents a multivariate data-driven dynamical system modelling of wetland sites. Altogether two sample sites were considered, one from Asia and another from Africa. A monthly data comprising average surface temperature, precipitation, and wetland fraction were used in this work. The purpose of modelling was to obtain a simple transparent dynamic model able to incorporate the nonlinearity of the system as well as possess a reasonable prediction accuracy. Moreover, the model was also required to be transparent to exhibit the interrelationship among the variables to augment the management of wetlands. Therefore, NARX model class was chosen for modelling the wetland dynamics. The model was trained using the data points starting from January 2000 until December 2010 and the data from January 2011 until December 2012 were used for model testing purposes. The model simulation and their performance measures are very satisfactory considering the constraints such as the availability of limited data. In future, a more holistic picture would be presented by combining more wetland sites across different continents. In addition to the weather variables, vegetation, and soil inputs such as net primary productivity and soil water content could be considered for developing a more robust NARX model. This modelling approach could also be harnessed to simulate the impact of climate change on wetland dynamics.

REFERENCES

[1] Angesh Anupam, David J Wilton, and Visakan Kadirkamanathan. 2022. Data driven modelling and simulation of wetland dynamics. *International Journal of Modelling and Simulation* 42, 3 (2022), 450–463.

[2] Keith Beven. 1997. TOPMODEL: a critique. *Hydrological processes* 11, 9 (1997), 1069–1085.

[3] Stephen A Billings. 2013. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains.* John Wiley & Sons, New York, USA.

[4] T Kleinen, V Brovkin, and RJ Schuldt. 2012. A dynamic model of wetland extent and peat accumulation: results for the Holocene. *Biogeosciences* 9, 1 (2012), 235–248.

[5] Wilson Rocha Lacerda, Luan Pascoal Costa da Andrade, Samuel Carlos Pessoa Oliveira, and Samir Angelo Milani Martins. 2020. SysIdentPy: A Python package for System Identification using NARMAX models. *Journal of Open Source Software* 5, 54 (2020), 2384.

[6] TR Marthews, SJ Dadson, Bernhard Lehner, Simon Abele, and Nicola Gedney. 2015. High-resolution global topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences* 19, 1 (2015), 91–104.

[7] Benjamin Poulter, Philippe Bousquet, Josep G Canadell, Philippe Ciais, Anna Peregon, Marielle Saunois, Vivek K Arora, David J Beerling, Victor Brovkin, Chris D Jones, et al. 2017. Global wetland contribution to 2000–2012 atmospheric methane growth rate dynamics. *Environmental Research Letters* 12, 9 (2017), 094013.

[8] Rita Wania, JR Melton, Elke L Hodson, Benjamin Poulter, Bruno Ringeval, Renato Spahni, Ted Bohn, CA Avis, Guangsheng Chen, Alexey V Eliseev, et al. 2013. Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (WETCHIMP). *Geoscientific Model Development* 6, 3 (2013), 617–641.