Peer-reviewed version available at Water 2020, 12, 585; doi:10.3390/w12020585

Article

Predicting the Trend of Dissolved Oxygen based on kPCA-RNN Model

Yi-Fan Zhang [1]*, Peter Fitch [2] and Peter J. Thorburn [1]

[1] CSIRO, Agriculture & Food, Brisbane, QLD, 4067, Australia
[2] CSIRO, Land & Water, Canberra, ACT, 2601, Australia
* Correspondence: yi-fan.zhang@csiro.au; Tel.: +61-452-208-587

Abstract: Water quality forecasting is increasingly significant for agricultural management and environmental protection. Enormous amounts of water quality data are increasingly being collected by advanced sensors, which leads to an interest in using data-driven models for predicting trends in water quality. However, the unpredictable background noises introduced during water quality monitoring seriously degrade the performance of those models. Meanwhile, artificial neural networks (ANN) with feed-forward architecture lack the capability of maintaining and utilizing the accumulated temporal information, which leads to biased predictions in processing time series data. Hence, we propose a water quality predictive model based on a combination of Kernal Principal Component Analysis (kPCA) and Recurrent Neural Network (RNN) to forecast the trend of dissolved oxygen. Water quality variables are reconstructed based on kPCA method, which aims to reduce the noise from the raw sensory data and preserve actionable information. With the RNN's recurrent connections, our model can make use of the previous information in predicting the trend in the future. Data collected from Burnett River, Australia was applied to evaluate our kPCA-RNN model. The kPCA-RNN model achieved R^2 scores up to 0.908, 0.823 and 0.671 for predicting the concentration of dissolved oxygen in the upcoming 1, 2 and 3 hours, respectively. Compared to current data-driven methods like ANN and SVR, the predictive accuracy of the kPCA-RNN model was at least 8 %, 17 % and 21 % better than the comparative models in these 3 cases. The study demonstrates the effectiveness of the kPAC-RNN modeling technique in predicting water quality variables with noisy sensory data.

Keywords: Water Quality; Machine learning; Recurrent Neural Network; PCA

1. Introduction

Multivariate time series data have been generated at an unprecedented speed in the field of water quality monitoring in recent years because of the widely deployed automated monitoring networks [1]. Predicting the temporal variability of water quality is essential for improving water quality management and protection activities [2].

Dissolved oxygen (DO) content is one of the most vital water quality variables as it directly indicates the status of the aquatic ecosystem and its ability to sustain aquatic life [3]. Rapid decomposition of organic materials, including manure or wastewater sources can quickly take the DO out of water in few hours, resulting in deficient DO levels that can lead to stress and death of aquatic fauna [4]. For example, DO levels that remain below 1-2 mg/l for a few hours can result in large fish kills. In the pond management, an aeration system can quickly increase dissolved oxygen levels if the decreasing of dissolved oxygen in the water can be predicted. Hence, short-term predictions of DO is critical in delivering good water quality management [5].

While some mathematical models have been applied to predict DO, they have complicated expressions, and are naturally conformed to particular ecosystems [6]. Consequently, those models are not likely to be able to be generalized without significant parameter adjustment [7].

Data-driven models have received increasing attention in predicting the concentration of DO based on the sensory data. That work involved feed-forward neural network [8], generalized regression neural network (GRNN) [9,10], radial basis neural network (RBF) [11], extreme learning machine (ELM) [12] and support vector machine (SVM) [13,14] models. Moreover, most predictive data-driven models follow a feed-forward topology structure. Under these circumstances, seasonal or diurnal patterns within the water quality time series data cannot be obtained because of the limitation of these modelling technologies [15].

The quality of input data also has a large influence on the data-driven model's performance [16]. Random errors generated by the environment, instruments or network transmission are unavoidable when monitoring water quality variables [17,18]. Though techniques such as z-score and min-max are used in preprocessing input data for data-driven models [10], those techniques aim to rescale the numeric range of water quality variables rather reduce sensor noise. Accordingly, the unwanted noise would be accepted as true by the data-driven models, which increases the challenges for generating accurate predictions for water quality variables.

In this paper, we propose a water quality predictive model based on Kernel Principal Component Analysis (kPCA) and Recurrent Neural Network (RNN) to solve the above issues. Our work differs from other comparative approaches in the following two aspects

- Kernel Principal Component Analysis (kPCA) is implemented to the input sensory data, which can filter the background noise as well as extract discriminative features effectively.
- A recurrent neural network (RNN) is designed to capture the temporal variations within water quality variables and utilize the historical changing patterns as a guide for predicting water quality in the future.

The aim of this study is to evaluate the predictive accuracy of the kPCA-RNN model by comparing it with two methods, the feed-forward neural network and support vector regression model. The evaluation is undertaken on a case study of DO concentrations in Burnett River, Australia.

2. Material and Methods

2.1. Study Area and Monitoring Data

2.1.1. Overview

The Burnett River is located on the southern Queensland coast and flows into the coral sea of the South Pacific Ocean. Cultivation of sugar cane and small crops are important lands uses in this region. The total area of the catchment is about 33,000 km^2. Figure 1 illustrates the location and extent of the catchment. Time series physiochemical water quality variables analysed in this study were obtained by a YSI 6 Series sonde sensor near the Bundaberg Co-op Wharf (Figure 1) [20]. Water quality variables such as temperature, electric conductivity (EC), pH, dissolved oxygen (DO), turbidity and chlorophyll-a (Chl-a) are recorded with 1 hour time interval for 5 months in 2015 (Table 1).

2.1.2. Water Quality Statistical Analysis

As demonstrated in Table 1, Chl-a and turbidity have larger variability than other water quality variables (CV $>$ 50%). In the case of turbidity this is due to extreme weather events [21]. The variability of Chl-a concentration can be affected by the discharge of river, temperature and salinity variation. The high variability in turbidity and Chl-a are caused by a small number of observations with high values (Figure 2). Additionally, outliers of EC tend to have lower measurement values. These outliers can be

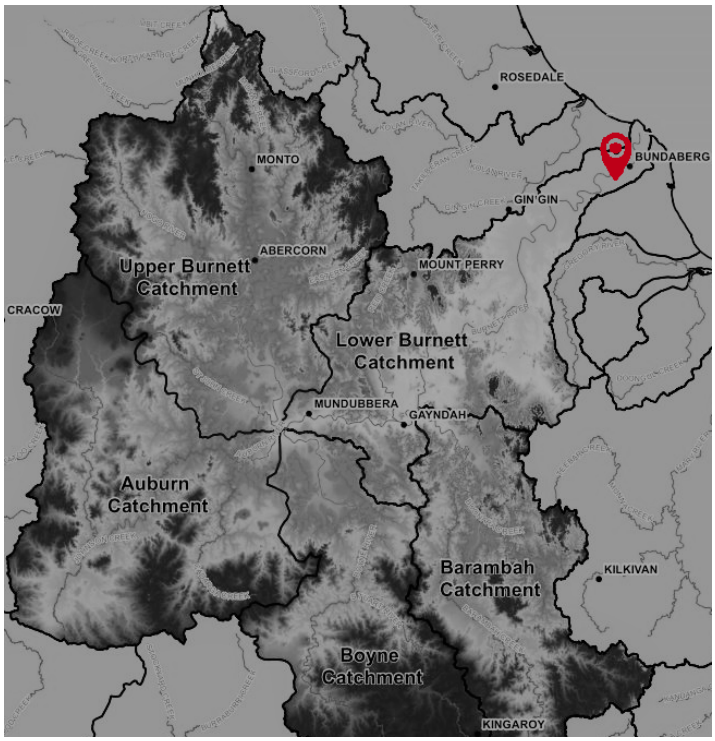

Figure 1. Burnett River catchment area and the monitoring site. This monitoring site is part of the Queensland Government's water quality monitoring network [19].

Table 1. Water quality data from 1/6/2015 to 31/10/2015.

Variables	No. of data	Unit	Min	Max	Median	Mean	SD[1]	CV[2](%)
Temperature	3672	°C	16.07	27.92	20.95	21.39	2.33	11
EC		uS cm^{-1}	613	49150	45750	44712.13	3566.75	8
pH			7.53	8.37	7.85	7.84	0.13	2
DO		mg L^{-1}	5.24	12.96	6.76	6.85	0.88	13
Turbidity		NTU	2.6	63	8.2	9.46	4.70	50
Chl-a		μg L^{-1}	0.1	137.6	2.6	3.50	3.56	102

[1] standard deviation
[2] coefficient of variation

caused by variations in river flow of other characteristics of the catchment. Ignoring those variations may cause serious information loss.

Figure 3 illustrates the changing patterns of DO both within a day and over a consecutive number of days. It is obvious that the concentration of DO follows a similar daily pattern, which makes it possible to predict the changing of DO. However, when tracking the concentration of DO in a larger time scale, it is plain to see that the mean value of the concentration of DO is increasing incrementally. Hence, the predictive model should both capture the temporal pattern within the same day and over a couple of days.

Figure 4 depicts the autocorrelation of DO for 24 time steps within a day. The present concentration of DO has a solid relationship with the concentration of DO in the previous 20 time steps. This means to predict the concentration of DO at a time, the historical DO information in the past 20 hours are very helpful. Expressly, DO concentration in the previous 1 and 2 hours can influence the current value significantly. Considering the concentration of DO varies diurnally, each input in our predictive model will involve data from 20 historical time steps.

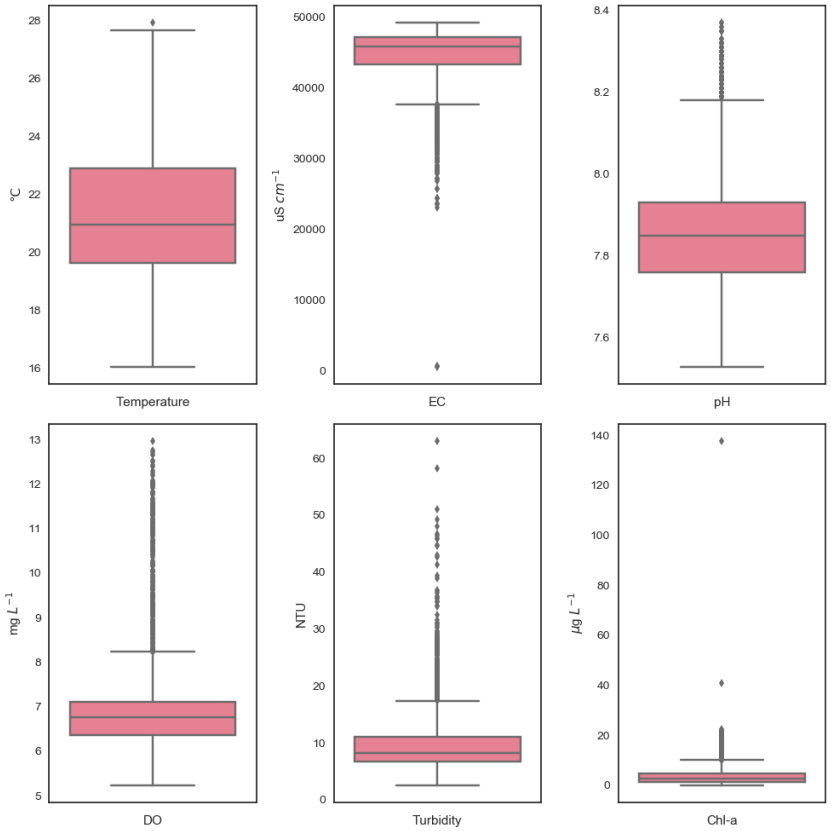

Figure 2. Data distribution for six water quality variables.

2.2. kPCA-RNN Model Description

2.2.1. Kernel PCA based Input Abstraction

Principal component analysis (PCA) is routinely applied for linear dimensionality reduction and feature abstraction [22]. The diagonal of the correlation matrix transforms the original principal correlated variables into principal uncorrelated (orthogonal) variables called principal components (PCs), which are weighed as linear combinations of the original variables. The eigenvalues of the PCs are a measure of associated variances and the sum of the eigenvalues coincides with the total number of variables.

The standard PCA only allows linear dimensionality reduction. However, the multivariate water quality data have a more complicated structure which cannot be easily represented in a linear subspace. In this paper, kernel PCA (kPCA) [23] is chosen as a non-linear extension of PCA to implement non-linear dimensionality reduction for water quality variables. The kernel represents an implicit mapping of the data to a higher dimensional space where linear PCA is performed.

The PCA problem in feature space F can be formulated as the diagonalization of an $l-$sample estimate of the covariance matrix [24], which can be defined as Equation 1:

$$\hat{C} = \frac{1}{l}\sum_{i=1}^{l}\Phi(x_i)\Phi(x_i)^{\intercal}, \tag{1}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2019 doi:10.20944/preprints201912.0346.v1

Peer-reviewed version available at Water 2020, 12, 585; doi:10.3390/w12020585

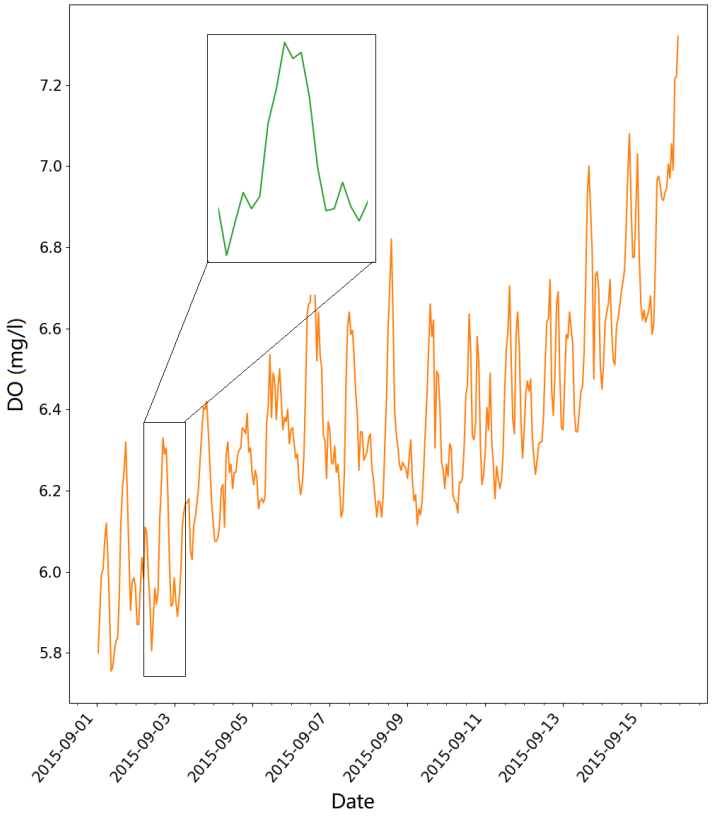

Figure 3. The changing of DO concentration among a period of time.

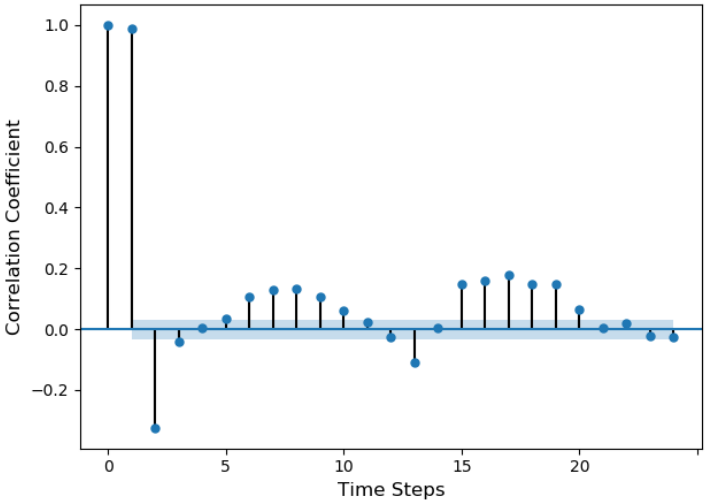

Figure 4. Partial autocorrelation of DO. The concentration of DO is collected hourly.

where $\Phi(x_i)$ are centred non-linear mappings of input variables $x_i \in \Re^n$. Then, we need to solve the following eigenvalue problem:

$$\begin{aligned} \lambda \mathbf{V} &= \hat{\mathbf{C}}\mathbf{V}, \\ \mathbf{V} &\in \mathbf{F}, \lambda \geq 0. \end{aligned} \tag{2}$$

Note that all the solutions $\mathbf{V}$ with $\lambda \geq 0$ lie in the span of $\Phi(x_1), \Phi(x_1), ..., \Phi(x_l)$. An equivalently problem is defined below:

$$n\lambda\alpha = \mathbf{K}\alpha, \tag{3}$$

where α denotes the column vector such that $\mathbf{V} = \sum_{i=1}^{l} \alpha_i \Phi(x_i)$, and $\mathbf{K}$ is a kernel matrix which satisfies the following conditions:

$$\begin{aligned} &\iint K(x,y)g(x)g(y)dxdy > 0, \\ &\int g^2(x)dx < \infty, \end{aligned} \tag{4}$$

where $K(x,y) = \sum_{i=1}^{\infty} \alpha_i \psi(x)\psi(y), \alpha_i \geq 0$. Then we can compute the kth non-linear principal component of x as the projection of $\Phi(x)$ onto the eigenvector $\mathbf{V}^k$:

$$\beta(x)_k = \mathrm{V}^k \Phi(x) = \sum_{i=1}^{l} \alpha_i^k K(x_i, x). \tag{5}$$

Then the first $p < l$ non-linear components are chosen, which have the desired percentage of data variance. By doing this, the complex of the original data series can be greatly reduced.

2.2.2. Recurrent Neural Network

Recurrent Neural Network (RNN) is able to exhibit dynamic temporal behavior by establishing connections between units form a directed cycle. In contradistinction to feed-forward neural network, RNN has information traveling in both directions. Computations derived from the earlier input are fed back into the network, which is critical in learning the non-linear relationships between multiple water quality variables.

The general input to RNN model is a variable-length sequence $x = \{x_1, x_2, ..., x_T\}$ where $x_i \in \mathbb{R}^d$. At each time step, RNN maintains its internal hidden state h, which results in a hidden sequence of $\{h_1, h_2, ..., h_k\}$. The operation of RNN at time step t can be formulated as:

$$h_t = f(w_{xh}x_t + w_{hh}h_{t-1}), \tag{6}$$

where $f()$ is an activation function, w_{hx} is the matrix of conventional weights between an input layer and a hidden layer and w_{hh} is the matrix between a hidden layer and itself at adjacent time steps.

The output of RNN is computed by:

$$y_t = w_{hy}h_t, \tag{7}$$

where w_{hy} is the matrix of weights between the hidden layer and output.

As exhibited in Figure 5, the structure of the RNN model across time can be expressed as a deep neural network with one layer per time step. Because this feedback loop occurs at every time step in the series, each hidden state contains traces not only of the previously hidden state but also of all those that preceded h_{t-1} for as long as memory can persist.

Comparing to transitional feed-forward neural network, RNN networks can have information travelling in both directions by introducing loops in the network. Computations derived from earlier input are fed back into the network, which gives them a kind of memory. Beside this, RNN networks are dynamic and their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found.

The features of RNN networks are especially suitable for processing time series water quality data because of the following reasons: Firstly, water quality data are periodically collection from different sensors and the previous values have strong relationship with the following changing. Secondly,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2019 doi:10.20944/preprints201912.0346.v1

Peer-reviewed version available at Water 2020, 12, 585; doi:10.3390/w12020585

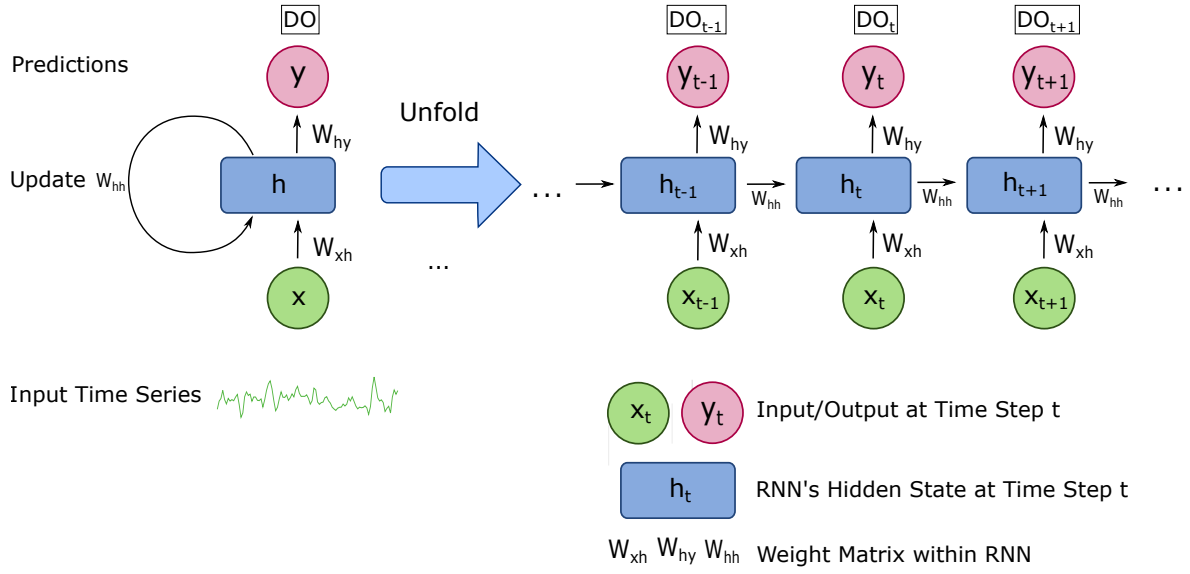

Figure 5. Recurrent neural network for predicting DO.

the pattern of many water quality variables can only be recognized when enough historical data are involved and analysed.

In the proposed water quality predictive model, we apply the RNN structure with LSTM cell [25]. To predict the concentration of DO at time step $t+1$, the input time series include data in previous m time steps. Additionally, each time step owns n water quality variables. Consequently, each input of the RNN model can be interpreted as a $m \times n$ matrix. The explicit hyperparameters of our RNN model will be outlined in the following subsection 3.2.

2.3. Model Evaluation

We compared the kPCA-RNN model with the following two machine learning methods:

1. Feed-forward neural network (FFNN). FFNN has been broadly adopted for water quality analysis due to its capability in capturing non-linear relationships within the short-term period [8].
2. Support vector regression (SVR). SVR is a classic machine learning technique which can mapping inputs into higher dimensional space and interpret the problem as linear regression [26].

The following performance indicators were applied to evaluate the predictive results. Those are the mean absolute error (MAE), the coefficient of determination (R^2), the root mean square error (RMSE) and FA1.1.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - \hat{f}_i|. \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(f_i - \hat{f}_i)^2}{\sum_{i=1}^{n}(f_i - \overline{f_i})^2}. \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|f_i - \hat{f}_i|)^2}. \tag{10}$$

$$FA1.1 = \frac{m}{n}, \qquad m = |0.9 < \frac{\hat{f}_i}{f_i} < 1.1|. \tag{11}$$

where f_i, $\hat{f}_i$, n and m represent the observed value, the predicted value, the number of observations and the number of predictions within a factor of 1.1 of the observed values, respectively. Additionally, $\overline{f_i} = \frac{1}{n}\sum_{i=1}^{n} f_i$.

2.4. WorkFlow of Predicting DO

Figure 6 depicts the workflow of predicting the concentration of DO by using the kPCA-RNN model.

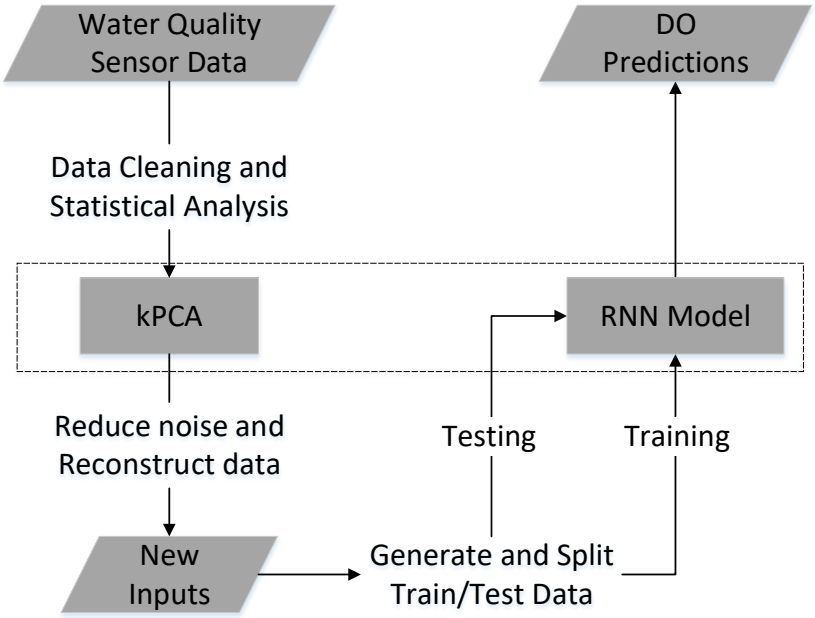

Figure 6. Workflow for predicting DO by applying the kPCA-RNN model.

Firstly, the kPCA method is implemented on the tabulated water quality data (Table 1). Principal components are constructed and elected as new inputs. Next, the input data is formed as $m \times n$ matrix as we explained in subsection 2.2.2. After training and testing the RNN model, the concentration of DO in the upcoming time steps can be estimated.

3. Model Application

3.1. Applying kPCA on the Water Quality Data

We applied the kPCA method to the water quality dataset (Table 1) and obtained 5 principal components (Table 2).

Table 2. Descriptive statistics of 5 principal components.

Principal Components	Eigenvalue	Cumulative Variance Proportion (%)
PC1	466.60944806	44.372466
PC2	285.28307768	71.501607
PC3	129.81274444	83.846217
PC4	114.77119055	94.760442
PC5	55.0978458	100.0

Five principal components (Table 2) are ordered by their corresponded eigenvalue. The first principal component is the linear combination of all the variables that has maximum variance, so it accounts for as much variation in the data as possible. After that, each succeeding component, in turn,

has the highest variance possible under the constraint that it is orthogonal to the preceding components. The cumulative variance proportion of the first 4 principal components is 94.76 %. This indicates by retaining only the first 4 principal components, one can explain 94.76 % of the full variance.

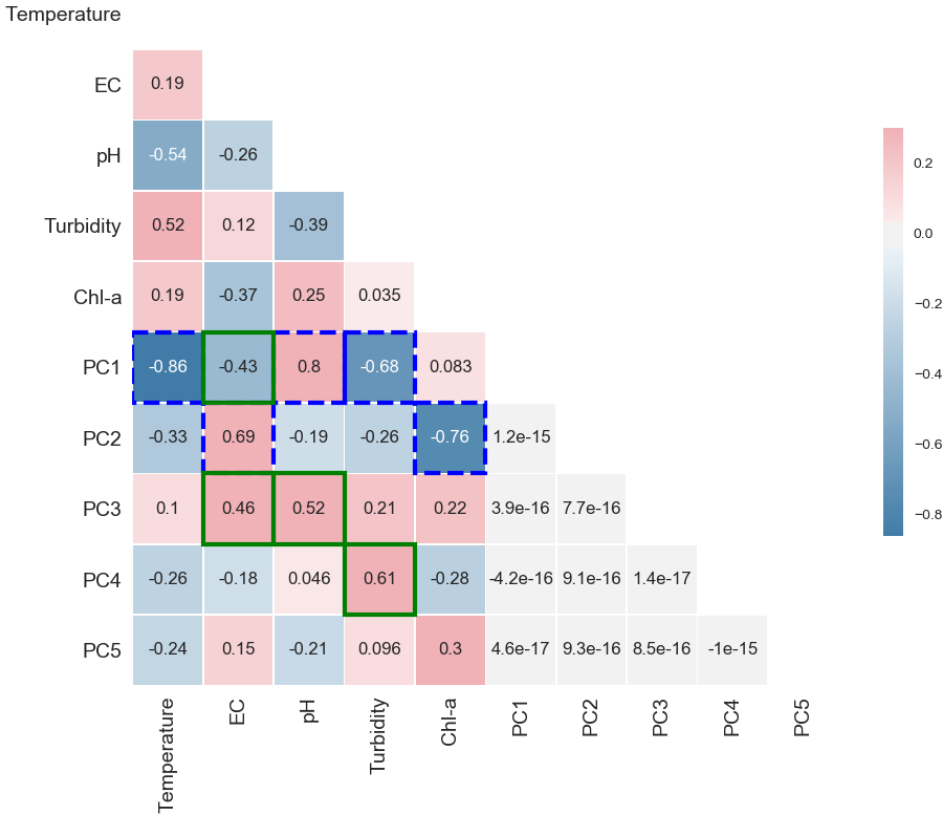

Figure 7. Correlations between water quality variables and principal Components.

As has been pointed out, the first principal component (PC1) has the highest correlation (dotted box, Figure 7) with variables like temperature, pH and turbidity. Furthermore, the second principal component (PC2) has the highest correlation (dotted box) with the remaining variables EC and Chl-a. This indicates that by utilizing only principal components PC1 and PC2, most information involved in those five water quality variables can be presented. Furthermore, PC3 and PC4 also have a strong correlation with EC, pH, and Turbidity (solid box). On the contrary, PC5 has a low value of correlation coefficient to all water quality variables, which means it carries much noise information [27]. Accordingly, we accept the first 4 principal components as new inputs. The kPCA method can reduce the input size by 20 % while still keep the most valuable information.

3.2. RNN Hyperparameters Settings

Three RNN models were designed to predict next 1, 2 and 3 hour's DO concentration independently. Each RNN model has various parameters and they all accept 4 months' data (2928 samples) for training and 1 month's data (744 samples) for testing. Based on the partial autocorrelation analysis in subsection 2.1.2, data from the previous 20 time steps were accepted as the model's input when predicting the concentration of DO in each future step.

The hyperparameters of the three RNN models were defined in Table 3.

Table 3. Experiments settings.

Model Settings	Experimental Cases		
	1 Hour Ahead	2 Hours Ahead	3 Hours Ahead
No. of Hidden Layers	1	2	3
No. of Hidden Units	40	30	20
Recurrent Cell	LSTM[1]		
Optimizer	Adam[2]		
No. of Historical Time Steps	20		
No. of Input in each time step	4		
No. of Training Data	2928		
No. of Testing Data	744		

[1]Long short-term memory [25]
[2]Adam [28]

3.3. Results and Discussion

The model predictions have R^2 values of 0.908 and 0.823 for 1 hour ahead and 2 hour ahead predictions (Figure 8). Those R^2 values indicate our RNN model can gain high accuracy in predicting the trend of DO in the near future.

In the 3 hours ahead prediction (bottom subfigure in Figure 8), around 93 % prediction results are within ± 10 % range of the original observations (FA1.1). The model did not utilize the information at time step $t+1$ and $t+2$ to predict the concentration of DO at time step $t+3$. Based on the partial autocorrelation analysis (Figure 4), the concentration of DO at time step t has a much stronger relationship with the concentration in the previous 1 and 2 time steps. Hence, this model did not utilize the most relevant information and it reveals why the predictive performance declined in this case.

In the water monitoring reports published by Queensland Government, there was a large amount of discharge for total nutrients, dissolved and particulate nutrients during 10/2015. By contrast, the discharge in the previous months was low. It indicates that the trend of concentration of DO was changing more frequently and heavily in October, while the kPCA-RNN model is trained based on the concentration of DO obtained from historical months with regular DO change. Consequently, there are some predictions below the high points of the observations, for example, the predictions around 22/10/2015. To resolve this problem, it is necessary to involve extra water quality data to cover a longer time period.

We additionally compared the performance of the kPCA-RNN model with two models stated in subsection 2.3. Same data set described in subsection 1 was applied in all cases. For FFNN, we set the same neural network size as in the kPCA-RNN model. For SVR, the Radial Basis Function kernel (RBF) is taken as the non-linear kernel. The corresponding results are listed in Table 4.

Table 4. Performance comparison with the FFNN and SVR.

Predictive Models	Evaluation Criteria			
	MAE	R^2	RMSE	FA1.1
1 Hour Ahead Prediction				
kPCA-RNN Model	0.149	0.908	0.208	0.995
FFNN	0.175	0.893	0.224	0.989
SVR	0.219	0.810	0.299	0.962
2 Hour Ahead Prediction				
kPCA-RNN Model	0.211	0.823	0.288	0.973
FFNN	0.258	0.757	0.338	0.958
SVR	0.314	0.594	0.437	0.890
3 Hour Ahead Prediction				
kPCA-RNN Model	0.303	0.671	0.394	0.926
FFNN	0.455	0.358	0.550	0.756
SVR	0.358	0.515	0.478	0.858

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2019 doi:10.20944/preprints201912.0346.v1

Peer-reviewed version available at Water 2020, 12, 585; doi:10.3390/w12020585

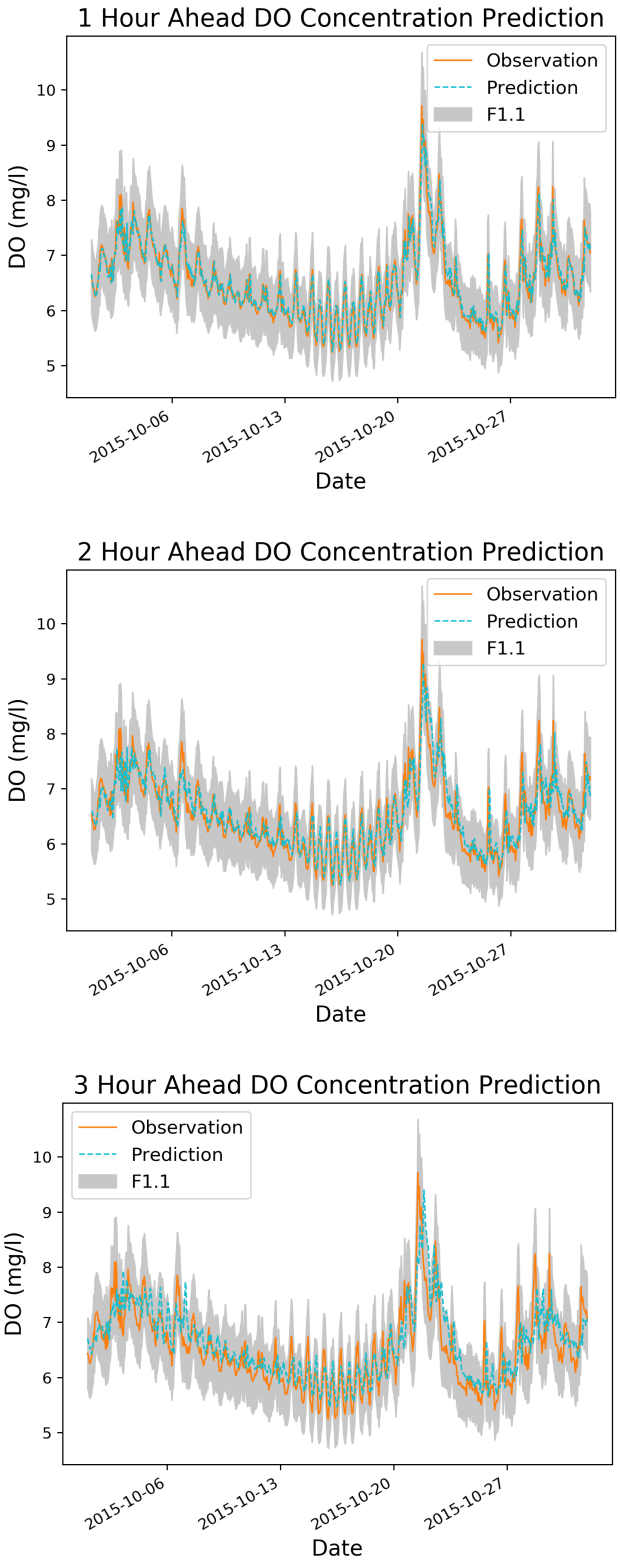

Figure 8. 1, 2 and 3 hours ahead predicting for the concentration of DO. The gray shadow, solid line and dotted line represent the FA1.1 range, DO observations and predictions, respectively.

The kPCA-RNN models offer the best performance in all the 3 prediction cases (Table 4). For example, in the 1 hour ahead prediction, 99.5% of the predictions are within the FA1.1 range, which demonstrates the model has a stable accuracy for most predictions.

The kPCA-RNN model has 8 %, 17 % and 40 % improved performance on the RMSE than the FFNN in all the three cases, respectively. Similarly, the kPCA-RNN model achieves 43 %, 52 % and 21 % improved performance on the RMSE than the SVR. The improvement in predictive performance of the kPCA-RNN relative to the FFNN and SVR increase as the predictive time step increases. The FFNN and SVR are ineffective in predicting long-term water quality changes because their model structures make them cannot utilize the information learned in previous time steps.

Hence, the kPCA-RNN model can perform as an early warning predictor for DO in application areas such as aquaculture ponds. By providing the DO significant changing alarm, farmers can consider appropriate actions to maintain the DO on a suitable level for the health of the aquatic ecosystem.

4. Conclusion

To summarize, the kPCA-RNN model was able to successfully predict the trend of DO in the following 1 to 3 hours. We evaluated our model by applying the water quality data from Burnett River, Australia and compared with the FFNN and SVR methods. The results demonstrate our method is more accurate and stationary to the alternative methods, especially when the prediction lead time is increasing. Furthermore, as a data-driven modeling method, the kPCA-RNN model is not limited to a specific hydrological area and can be extended to predict various water quality variables.

For future work, inputs can be improved to include extra information such as rail fall and cover longer periods of time. In addition, the water quality predictive model can be extended to support multiple targets prediction tasks.

Author Contributions: Methodology, writing–original draft preparation, Yi-Fan Zhang; writing–review and editing, Peter Fitch; project administration, writing–review and editing, Peter J. Thorburn.

Acknowledgments: This work was conducted within the CSIRO Digiscape Future Science Platform.

Conflicts of Interest: The authors declare no conflict of interest.

1. Adu-Manu, K.S.; Tapparello, C.; Heinzelman, W.; Katsriku, F.A.; Abdulai, J.D. Water Quality Monitoring Using Wireless Sensor Networks. *ACM Transactions on Sensor Networks* **2017**, *13*, 1–41.
2. Hrachowitz, M.; Benettin, P.; van Breukelen, B.M.; Fovet, O.; Howden, N.J.; Ruiz, L.; van der Velde, Y.; Wade, A.J. Transit times-the link between hydrology and water quality at the catchment scale. *Wiley Interdisciplinary Reviews: Water* **2016**, *3*, 629–657.
3. Tomić, A.Š.; Antanasijević, D.; Ristić, M.; Perić-Grujić, A.; Pocajt, V. A linear and non-linear polynomial neural network modeling of dissolved oxygen content in surface water: Inter- and extrapolation performance with inputs' significance analysis. *Science of The Total Environment* **2018**, *610-611*, 1038–1046.
4. King, A.J.; Tonkin, Z.; Lieshcke, J. Short-term effects of a prolonged blackwater event on aquatic fauna in the Murray River, Australia: considerations for future events. *Marine and Freshwater Research* **2012**, *63*, 576–586.
5. Zhang, Y.; Thorburn, P.J.; Fitch, P. Multi-Task Temporal Convolutional Network for Predicting Water Quality Sensor Data. 26th International Conference on Neural Information Processing (ICONIP2019); , 2019; Vol. 1142, pp. 122–130.
6. Ay, M.; Özgür Kişi. Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing techniques. *KSCE Journal of Civil Engineering* **2016**, *21*, 1631–1639.
7. Zhang, Y.; Thorburn, P.J.; Wei, X.; Fitch, P. SSIM -A Deep Learning Approach for Recovering Missing Time Series Sensor Data. *IEEE Internet of Things Journal* **2019**, *6*, 6618–6628.
8. Zhang, Y.; Fitch, P.; Vilas, M.P.; Thorburn, P.J. Applying Multi-Layer Artificial Neural Network and Mutual Information to the Prediction of Trends in Dissolved Oxygen. *Frontiers in Environmental Science* **2019**, *7*, 46.
9. Heddam, S. Generalized regression neural network-based approach for modelling hourly dissolved oxygen concentration in the Upper Klamath River, Oregon, USA. *Environmental Technology* **2014**, *35*, 1650–1657.

10. Antanasijević, D.; Pocajt, V.; Perić-Grujić, A.; Ristić, M. Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis. *Journal of Hydrology* **2014**, *519*, 1895–1907.
11. Yunus, M.A.M.; Faramarzi, M.; Ibrahim, S.; Altowayti, W.A.H.; San, G.P.; Mukhopadhyay, S.C. Comparisons between radial basis function and multilayer perceptron neural networks methods for nitrate and phosphate detections in water supply. 2015 10th Asian Control Conference (ASCC); , 2015; pp. 1–6.
12. Heddam, S.; Kisi, O. Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. *Environmental Science and Pollution Research* **2017**, *24*, 16702–16724.
13. Ji, X.; Shang, X.; Dahlgren, R.A.; Zhang, M. Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China. *Environmental Science and Pollution Research* **2017**, *24*, 16062–16076.
14. Kisi, O.; Parmar, K.S. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology* **2016**, *534*, 104–112.
15. Sundermeyer, M.; Oparin, I.; Gauvain, J.L.; Freiberg, B.; Schlüter, R.; Ney, H. Comparison of feedforward and recurrent neural network language models. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; , 2013; pp. 8430–8434.
16. Karami, J.; Alimohammadi, A.; Seifouri, T. Water quality analysis using a variable consistency dominance-based rough set approach. *Computers, Environment and Urban Systems* **2014**, *43*, 25–33.
17. Liu, S.; Che, H.; Smith, K.; Chang, T. A real time method of contaminant classification using conventional water quality sensors. *Journal of Environmental Management* **2015**, *154*, 13–21.
18. chuan Wang, W.; wing Chau, K.; Qiu, L.; bo Chen, Y. Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. *Environmental Research* **2015**, *139*, 46–54.
19. Great Barrier Reef Catchment Loads Monitoring Program. https://www.reefplan.qld.gov.au/measuring-success/paddock-to-reef/catchment-loads/. Accessed: 2018-06-01.
20. Ambient estuarine water quality monitoring data. https://data.qld.gov.au/dataset. Accessed: 2017-11-20.
21. Macdonald, R.K.; Ridd, P.V.; Whinney, J.C.; Larcombe, P.; Neil, D.T. Towards environmental management of water turbidity within open coastal waters of the Great Barrier Reef. *Marine Pollution Bulletin* **2013**, *74*, 82–94.
22. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **1987**, *2*, 37 – 52. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.
23. Schölkopf, B.; Smola, A.; Müller, K.R., Kernel principal component analysis. In *Artificial Neural Networks — ICANN'97: 7th International Conference Lausanne, Switzerland, October 8–10, 1997 Proceeedings*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1997; pp. 583–588.
24. Ince, H.; Trafalis, T.B. Kernel principal component analysis and support vector machines for stock price prediction. *IIE Transactions* **2007**, *39*, 629–637. doi:10.1080/07408170600897486.
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
26. Granata, F.; Papirio, S.; Esposito, G.; Gargano, R.; de Marinis, G. Machine learning algorithms for the forecasting of wastewater quality indicators. *Water* **2017**, *9*, 105.
27. Verbanck, M.; Josse, J.; Husson, F. Regularised PCA to denoise and visualise data. *Statistics and Computing* **2013**, *25*, 471–486.
28. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.