




Article

Assessing Water Temperature and Dissolved Oxygen and Their Potential Effects on Aquatic Ecosystem Using a SARIMA Model

Samuel Larance ^{1,2}, Junye Wang ^{1,*} , Mojtaba Aghajani Delavar ¹  and Marwan Fahs ³ 

¹ Faculty of Science and Technology, Athabasca University, 1 University Drive, Athabasca, AB T9S 3A3, Canada; samuel.larance@engees.fr (S.L.); maghajanidelavar@athabascau.ca (M.A.D.)

² National School for Water and Environmental Engineering of Strasbourg (ENGEES), The University of Strasbourg, 1 cour des cigarières CS 61039, 67070 Strasbourg, France

³ Centre National de la Recherche Scientifique (CNRS), École Nationale du Génie de l'Eau et de l'Environnement de Strasbourg (ENGEES), Institute de Terre et Environnement de Strasbourg (ITES), The University of Strasbourg, UMR 7063, 67000 Strasbourg, France; fahs@unistra.fr

* Correspondence: junyew@athabascau.ca

Abstract: Temperature and dissolved oxygen (DO) are of critical importance for sustainable aquatic ecosystem and biodiversity in the river systems. This study aims to develop a data-driven model for forecasting water quality in the Athabasca River using a seasonal autoregressive integrated moving average model (SARIMA) for forecasting monthly DO and water temperature. DO and water temperature observed at Fort McMurray and Athabasca from 1960 to 2023 were used to train and test the model. The results show the satisfied model performance of DO with a coefficient of determination (R^2) value of 0.76 and an RMSE value of 0.79 for training and 0.67 and 0.92 for testing, respectively, at the Fort McMurray station. At the Town of Athabasca station, the RMSE and R^2 of DO were 0.92 and 0.72 for training and 0.77 and 0.86 for testing, respectively. For the modeled temperature, RMSE and R^2 were 2.7 and 0.87 for training and 2.2 and 0.95 for testing, respectively, at Fort McMurray and were 2.0 and 0.93 for training and 1.8 and 0.97 for testing, respectively, in the Town of Athabasca. The results show that DO concentration is inversely proportional to the temperature. This implies that the DO could be related to water temperature, which, in turn, is correlated with air temperature. Therefore, the SARIMA model performed reasonably well in representing the dynamics of water temperature and DO in the cold climate river. Such a model can be used in practice to reduce the risk of low DO events.

Keywords: SARIMA model; water quality; dissolved oxygen; water temperature; aquatic ecosystem; machine learning



Academic Editors: Sergio Ulgiati, Pengxiao Zhou, Qianqian Zhang, Fei Zhang and Zoe Li

Received: 13 November 2024

Revised: 24 December 2024

Accepted: 13 January 2025

Published: 14 January 2025

Citation: Larance, S.; Wang, J.; Delavar, M.A.; Fahs, M. Assessing Water Temperature and Dissolved Oxygen and Their Potential Effects on Aquatic Ecosystem Using a SARIMA Model. *Environments* **2025**, *12*, 25. <https://doi.org/10.3390/environments12010025>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Water quality is of critical importance for sustainable economic development, aquatic ecosystems, and biodiversity. However, human activities, economic development, and climate change can change river systems, leading to water pollution, and loss of biodiversity [1,2]. Dissolved oxygen (DO) and water temperature are two key water quality parameters for aquatic ecosystems and biodiversity. Aquatic species have a favored range of temperature and DO that they favor as their habitats [3]. Any changes in stream temperature and DO may have an adverse impact on aquatic biodiversity [4–6]. For example, no fish can live below a dissolved oxygen concentration of 4 mg/L [7]. Furthermore, DO plays a crucial role in the decomposition process of river organic matter. Therefore, DO is an important indicator of water quality and the health of rivers. The water temperature

significantly influences the timing of life cycle events, such as breeding, spawning, hatching, growing, and mortality in aquatic species [4,8]. It also impacts on the metabolic rates of aquatic organisms. Warmer temperatures generally increase metabolic rates, leading to higher growth rates and reproduction for many species. However, excessively high temperatures can be stressful or lethal to some species, especially those adapted to cooler environments. Therefore, sustainable management of river systems requires an understanding of how water quality in river systems responds to climate change and natural and anthropogenic activities [9,10]. The ability to predict future trends and patterns of DO and temperature is important for aquatic ecosystems and data-driven decision.

Modeling is a useful and powerful tool for predicting stream temperature and water quality. Large efforts have been made in the literature to develop various water quality models [11–13]. There are three main classes of water quality models: physics-based models, statistical models, and machine learning models. Physics-based models are sophisticated tools that are based on partial differential equations, such as QUAL2E, MAKE11, and the Soil & Water Assessment Tool (SWAT) [11,14–16]. These models can simulate processes of hydrodynamics, dispersion, and pollutant kinetics in streams and their impacts on DO levels. However, running most of these models requires many input parameters, such as topography, soil and vegetation properties, and meteorological data. In some situations, the highly intensive data requirement limits the applications of the physics-based models [11].

The statistical models have been traditionally used for estimating stream temperature and assessing water pollution [17–19]. These models are based on empirical regression relationships between stream temperatures and meteorological parameters, such as the widely used regression models between air and stream temperatures. However, with the advent of remote sensing and monitoring networks, the datasets have become more complex in amount and structures. Statistical models are difficult to tackle with increased data complexity. The state-of-the-art statistical models of water quality are enhanced using artificial intelligence (AI), such as genetic algorithms (GAs) and artificial neural networks (ANNs) [20–22]. Statistical models are especially beneficial at sites and in regions where deterministic models are inapplicable or impossible to implement. As a result, machine learning and artificial intelligence have been increasingly applied for estimating water quality. Different machine learning techniques have been used in the literature, such as recurrent neural network (RNN) [23], long short-term memory (LSTM) [24,25], autoregressive integrated moving average (ARIMA) [26,27], and quantile regression forest (QRF) [28]. Khan and Valeo [29] (2016) used a fuzzy neural network (FNN) to analyze DO and temperature. They found that that low DO events are better captured by FNN compared to a non-fuzzy network. These data-driven models for water quality analysis and forecasting are usually implemented to analyze time series data produced at monitoring stations. The ARIMA is a long-established time series predicting model and is promising to analyze seasonality, periodicity, and trends in time series data. Particularly, the seasonal autoregressive integrated moving average model (SARIMA) is an extended ARIMA for modeling seasonal data. Due to its powerful feature extraction capability, it has been widely utilized for various time series analyses and forecasting [30–32]. However, it is still insufficient to evaluate the effects of temperature on DO. Although a saturated DO may inversely be correlated with water temperature in pure water, the inversed relation may not be correct in some situations. DO in water bodies depends not only on temperature but also on other factors, such as pressure, pH, phosphorus, or organic matter [33–35]. For example, phosphorous can accelerate eutrophication, leading to a reduction in DO in rivers and lakes. Insufficient oxygen is often caused by the decomposition of organic matter and/or nutrient pollution [36]. Waterways depleted of DO can damage ecosystems or even result in “dead zones” in larger bodies of water and fish death [37]. He et al. [38] showed

that abiotic and biotic factors could affect biomass. This could lead to variation in DO from river to river. It was reported that the spatial-temporal distribution of DO is heterogeneous in the Athabasca River Basin [39,40]. Therefore, in some environments, there may not be an inverse relation.

Despite many successes, it is still unclear how the trends of water temperature and DO change due to the effects of snow melting, glacial retreat, permafrost, and freeze thaw cycles in cold river basins. Particularly, these changes may impact the aquatic ecosystems and biodiversity. The sustainability of river basins depends on assessments and predictions of water quality from river to river. The time series data not only trace the past variation, but also arrest specific non-stationary, seasonal, and time-based events. There exists a knowledge gap in cold river basins, with an inadequate understanding of temperature and DO concentration. Therefore, several important issues remain to be addressed in improving our understanding of the temperature and DO changes in high latitude cold regions: (1) the dynamic performance forecast ability in river systems, and (2) data-driven model performance uncertainties in forecasting stream temperature and DO monthly, seasonally, annually under environmental variations.

Data-driven time series analysis and forecasting models would be beneficial for analyzing the temperature and DO concentration. This study aims to predict the trends and evolution of dissolved oxygen concentration and temperature in the Athabasca River. To reach this goal, the SARIMA is used to obtain a longitudinal profile of dissolved oxygen and temperature in the river. It is trained and tested on the time series data from sampling stations. The SARIMA model's performance is statistically assessed and used for forecasting. Finally, the impacts of water temperature and DO on fishes are analyzed to provide an insight into the general mitigation strategies and policy recommendations. Data-driven modeling and forecasting will improve understanding of datasets and inform a solution for water management.

2. Materials and Methods

2.1. Study Area and Input Data

The Athabasca River is located in the Alberta province, west of Canada (Figure 1). This river originates from the Columbia icefields and extends around 1500 km from the southwest to the northeast into the Lake of Athabasca. It is drained by an area of about 160,000 km², including several various urban regions, forestry, agriculture, coal, and oil and gas industries. Average annual precipitation ranges from 1000 mm in the headwaters to about 300 mm in the downstream area of the river basin. The mean annual temperature is 1.8–5.1 °C [41]. As one of the largest rivers in Alberta, the Athabasca River plays a pivotal role in the economic development and ecosystem services of the northern Alberta region. Water quality in the river is monitored through two stations. These stations were located at U/S Fort McMurray (AB07CC0030: 56.72028° N latitude and –111.40556° W longitude) over the period of 1976 to 2021 and in the Town of Athabasca (AB07BE0010: 54.72222° N latitude and –113.28611° W longitude) over the period of 1987 to 2021. These two stations are across a land use gradient (i.e., forest, oil and mining, built-up, and agriculture).

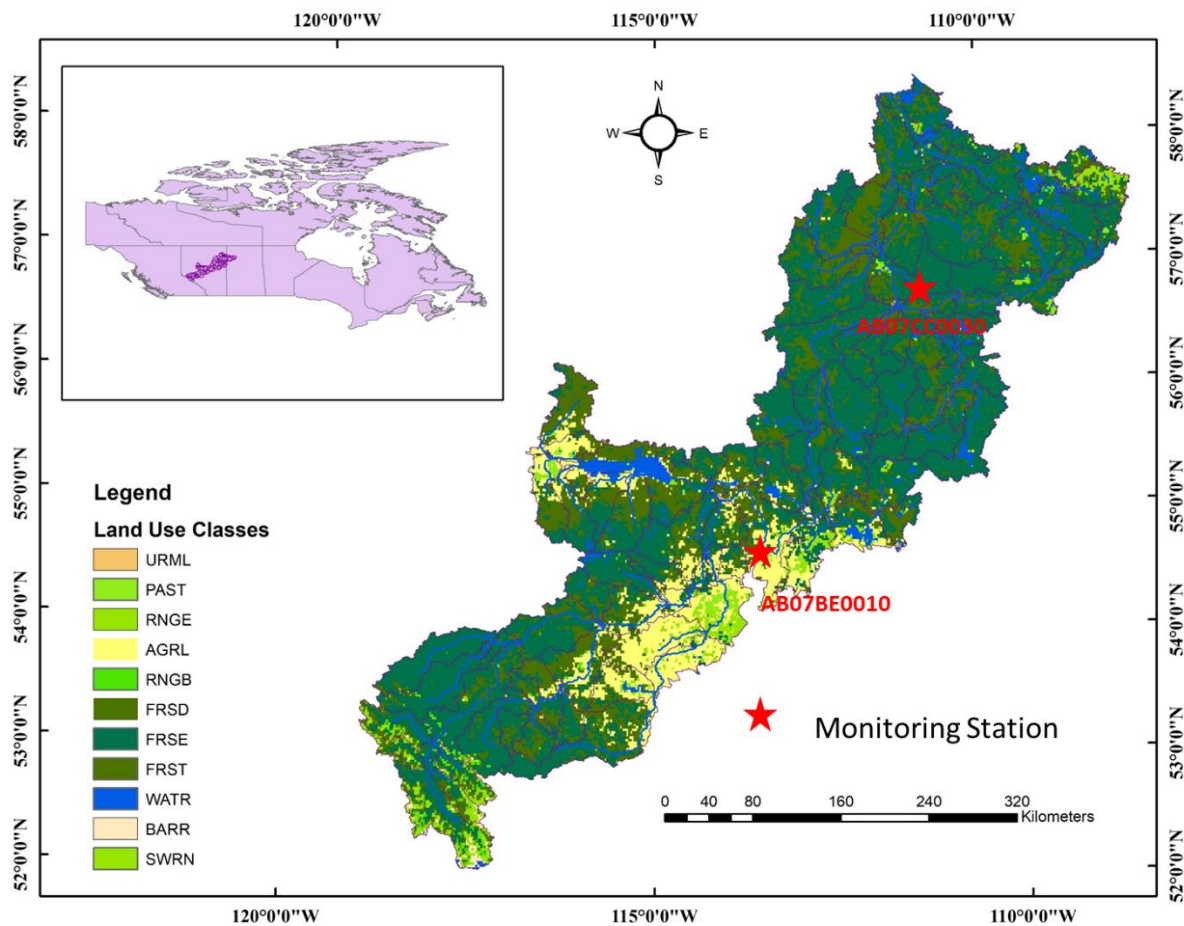


Figure 1. Athabasca River and monitoring stations: Fort McMurray and the Town of Athabasca. Land use classes: URML: urban medium density, PAST: pasture; RNGE: range shrub land; AGRL: agricultural land; RNGB: grassland/herbaceous; FRSD: deciduous forest; FRSE: evergreen forest; FRST: mixed forest; WATR: water; BARR: barren land; SWRN: southwestern rangeland.

Data were downloaded from the Long-Term River Network for water quality monitoring in Alberta [42]. Real-world environmental data usually require preprocessing. Once the data were loaded on the Python interface, temperature and dissolved oxygen were extracted from the dataset and plotted. There were missing observations, and sampling, sometimes, was irregular for temperature and DO concentration before 1990. These can cause analytical problems in data-based analyses. For example, Figure 2 shows a set of raw data of DO at U/S Fort McMurray. For Fort McMurray, the dataset covered from 1960 to 2023. For Athabasca, the dataset covered from 1987 to 2023. The sampling dates were irregular between 1978 and 1990, and some data were missing. Particularly, in 1960, only a sample was collected. There were no samples between 1960 and 1978. There were missing data between 1980 and 1990. The irregularity of sampling frequency can cause a broken integrity of the time series and a poor fit with the models. Therefore, irregular data must be preprocessed before applying any machine learning models. Here, the interpolate function in Pandas identified missing values in the time series. The time intervals between adjacent existing values were calculated. These intervals were used to determine the distance between the missing values and the existing values. Secondly, there was a pattern repeated each year (mostly obvious for the temperature). These time series were seasonal (12 months). However, by looking at the dataset, there was more than a value sampled per year. Hence, to get a fully seasonal time series, a monthly mean of values was realized during the preprocessing step. To get a fully seasonal time series, one sample per month

was used for the same month. So, the group of values sampled in the same month was replaced by their means. A linear interpolation function was used to fill the missing values:

$$x_i = x_{i-1} + (x_{i+1} - x_{i-1}) \cdot \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}} \tag{1}$$

where x_i is a value between x_{i+1} and x_{i-1} , and t_i is time at i . The missing values were replaced with the interpolated values obtained in the previous step.

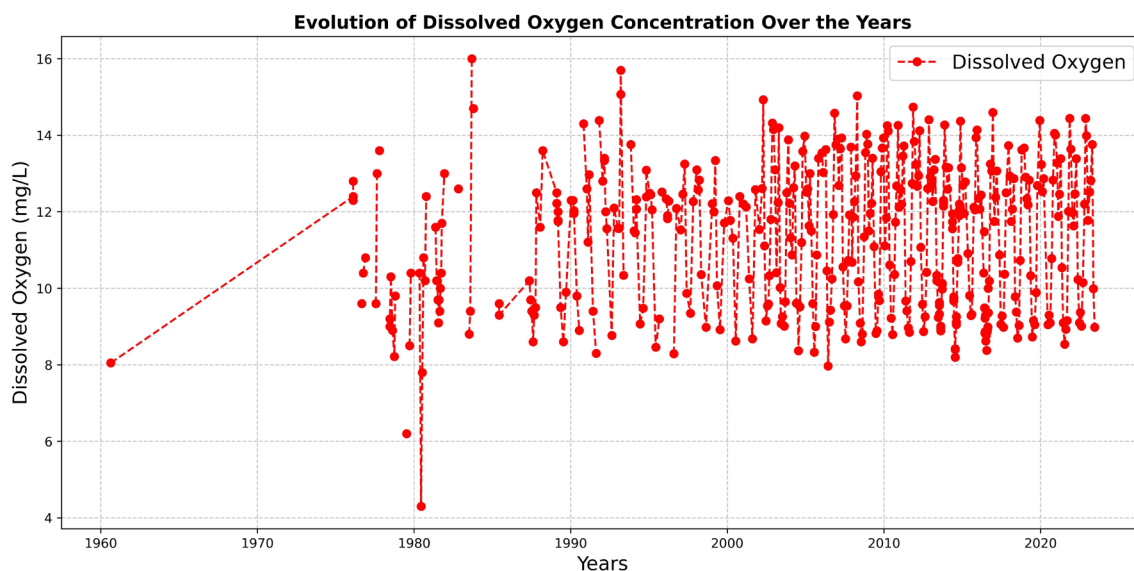


Figure 2. A raw dataset of DO at U/S Fort McMurray.

2.2. Seasonal Autoregressive Integrated Moving Average Model (SARIMA)

The seasonal autoregressive integrated moving average (SARIMA) model is a predictive machine learning model used to analyze and forecast time series and extend basic ARIMA models [43]. An ARIMA considers the past values using autoregressive (AR), integrative differentiation (I), and moving average (MA) and forecasts future values. AR utilizes observed data from previous time steps for a regression and predicts the value at the subsequent step. An MA model is the forecast related to a linear combination of previous forecast errors. Integrative differentiation (I) makes the time series stationary. The SARIMA also uses past values but allows the incorporation of seasonal patterns. It is suitable for analyzing time series with trends and/or seasonal components.

The parameters in the SARIMA model can be divided into two categories: seasonal parameters ($p, q, \text{ and } d$) and non-seasonal parameters (P, Q, D, S). p represents the maximum lag in the model. This value was determined by a partial autocorrelation plot. P is the value of autoregressive order which represents the number of previous data points taken to realize the autoregression on the seasonal part of the model. This value was determined by the number of non-negligible multiples of the season on the partial autocorrelation plot. q is the number of past errors taken to realize the moving average on the non-seasonal part of the model. This value was determined by an autocorrelation plot. Q is the number of past errors taken to realize the moving average on the seasonal part of the model. This value was determined by the number of non-negligible multiples of the season on the autocorrelation plot. Uppercase D is a Boolean value which represents the order of seasonal integration.

The SARIMA (p, d, q) (P, D, Q) model can be expressed as [44]:

$$\Phi_p(B^s)\varphi(B)\nabla_s^D\nabla^d(x_t) = \theta_Q(B^s)\theta(B)(w_t) \tag{2}$$

where x_t is the nonstationary time series value at time t , w_t the usual Gaussian error at time t , S the seasonality (12 months in this study), and B a backshift operator. ∇_s^D and ∇^d are the ordinary and seasonal difference operators.

$$B^k(x_t) = x_{t-k} \quad (3)$$

Φ_P is the polynomial related to the autoregressive seasonal component (P).

$$\Phi_P(B^S)(x_t) = x_t - \Phi_1 \cdot x_{t-1.S} - \Phi_2 \cdot x_{t-2.S} - \dots - \Phi_P \cdot x_{t-P.S} \quad (4)$$

φ is the polynomial related to the autoregressive non-seasonal component (p).

$$\varphi(B)(x_t) = x_t - \varphi_1 \cdot x_{t-1} - \varphi_2 \cdot x_{t-2} - \dots - \varphi_p \cdot x_{t-p} \quad (5)$$

∇_s^D is the operator for the polynomials related to the seasonal difference component. When $D = 1$, a seasonal difference was applied to make the series stationary. When $D = 0$, no seasonal difference was applied.

$$\nabla_s^D(x_t) = (1 - B^S)^{D=0}(x_t) = 1 \quad (6)$$

$$\nabla_s^D(x_t) = (1 - B^S)^{D=1}(x_t) = x_t - x_{t-s} \quad (7)$$

∇^d is the polynomial operator related to the number of non-seasonal differences applied to make the series stationary. In our case, $d = 0$. Hence:

$$\nabla^{d=0}(x_t) = x_t \quad (8)$$

$\theta_Q(B^S)$ is the polynomial related to the seasonal moving average component (Q).

$$\theta_Q(B^S)(w_t) = w_t + \theta_1 \cdot x_{t-1.S} + \theta_2 \cdot x_{t-2.S} + \dots + \theta_Q \cdot x_{t-Q.S} \quad (9)$$

$\theta(B)$ is the polynomial related to the non-seasonal moving average (q).

$$\theta(B)(w_t) = w_t + \theta_1 \cdot w_{t-1} + \theta_2 \cdot w_{t-2} + \dots + \theta_q \cdot w_{t-q} \quad (10)$$

2.3. Statistical Evaluation of Model Performance

To assess the performance of the SARIMA model, a statistical performance analysis is commonly used. There are different statistical model performance evaluations. The coefficient of determination (R^2) is one the most widely used metrics but is oversensitive to high extreme values (outliers) and insensitive to additive and corresponding changes between predicted values and actual data [45]. On the other hand, the root mean squared error (RMSE) may give a small error variance but at the expense of significant model bias. Thus, the statistical model performance evaluations should be incorporated into at least one dimensionless statistic, such as R^2 , and one absolute error index, such as RMSE and mean absolute error (MAE), with additional data. In this study, five metrics were used, including R^2 , mean absolute percentage error (MAPE), RMSE, mean absolute percentage error (MAPE), median absolute error (MDAE), and MAE.

The MAE is an index of closeness between predicted and actual values, regardless of their direction, and can be expressed as:

$$MAE = \frac{1}{n} * \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

where \hat{y}_i and y_i are the actual and predicted values at the time t_i , respectively, and n is the number of observations. These metrics can be formulated as follows.

MDAE is the median of all the absolute differences between predicted and actual values:

$$MDAE = \text{median}(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|) \quad (12)$$

RMSE is the square root of the sum of square errors between predicted and actual values:

$$RMSE = \text{SQRT} \left(\frac{1}{n} * \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) \quad (13)$$

Generally, the lower the RMSE, the better the model performance.

MAPE is the difference between modeled and actual values and can be calculated using Equation (14):

$$MAPE = \frac{100}{n} * \sum_{i=1}^n |y_i - \hat{y}_i| / y_i \quad (14)$$

MAPE is effective because it is unaffected by actual and modeled values as dimensionless values.

The Nash–Sutcliffe efficiency coefficient (NSE) is an index to measure the goodness of fit of a model. The NSE can be expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2} \quad (15)$$

NSE indicates the deviation between predicted and observed values relative to the total variation in the response variable. It ranges from 1 to $-\infty$ with a value of 1 indicating no difference between predicted and observed values, whereas lower values represent a poorer predictive accuracy.

2.4. Implementation of the SARIMA Model

For the SARIMA model, we must determine both non-seasonal parameters (p , q , and d) and seasonal parameters (P , Q , D , S). The initial values of several parameters, i.e., p , d , q , P , D , and Q , should be determined to represent the autoregressive order and the maximum lag, order of integration, the moving average order for the error of the time series, the order of autoregression for the seasonal component of the model, the order of seasonal integration, and the similar logic, respectively. The Dickey–Fuller (DF) test was performed to determine these parameters [46]. First, parameters must be established by using the Dickey–Fuller parameter. The autocorrelation function (ACF) is a metric of linear dependence between observed data in a time series that are separated by a lag q . The partial autocorrelation function (PACF) helps determine how many autoregressive terms p are necessary. We used an autocorrelation plot and a partial autocorrelation plot to examine the correlation and partial correlation between the time series and its different lags. Figure 3 shows the autocorrelation plot and the partial autocorrelation plot. In this figure, the Dickey–Fuller parameter p is equal to 0. The p value between 0 and 1 indicates a positive autocorrelation, while the value between -1 and 0 denotes a negative autocorrelation. The closer the parameter p to 0, the more stationary the time series can be considered. A variance stabilization should be applied when the variance grows with time. The null hypothesis of the DF test is that there is a unit root in an AR model, implying that the data series is not stationary. However, there were still too many significant peaks on the correlation and partial correlation plots.

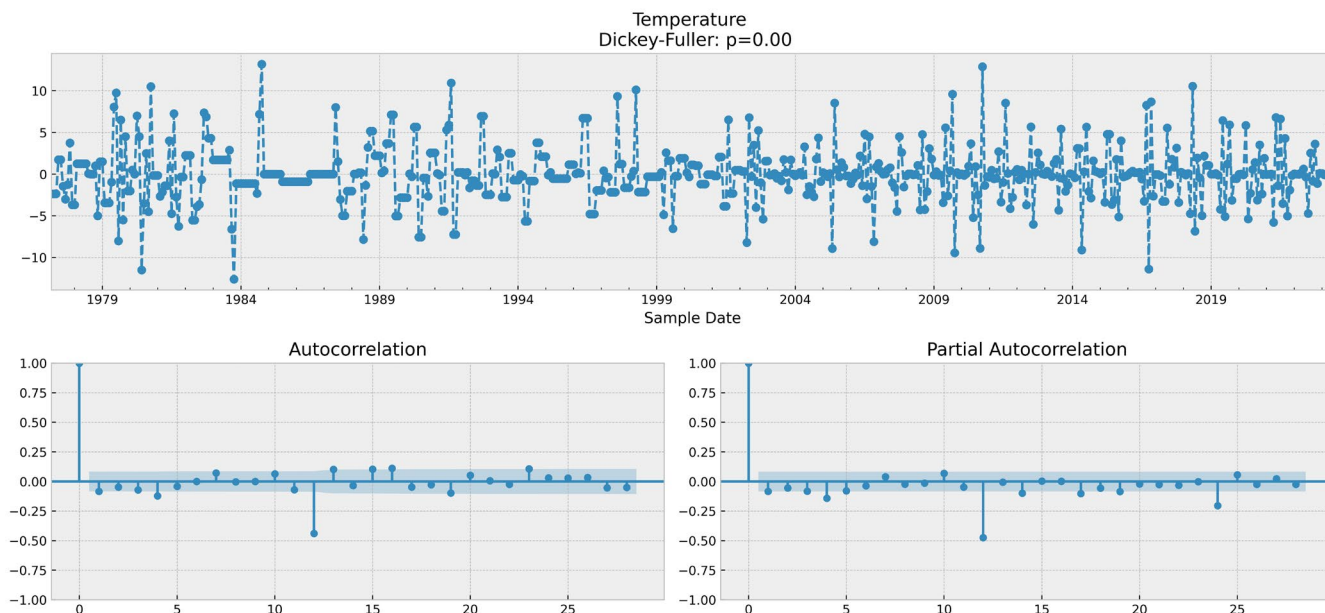


Figure 3. Dickey–Fuller value, autocorrelation plot, and partial autocorrelation plot for the temperature time series with a non-seasonal differentiation.

To reduce the number of significant peaks, it is possible to apply to the time series a non-seasonal integration. Significant peaks on autocorrelation and partial autocorrelation plots allow us to determine the seasonal parameters P and Q , respectively, and the non-seasonal parameters p and q , respectively. Hence, to decrease the p -value and to limit the number of significant peaks, a seasonal integration was considered. The parameter D is a Boolean value which represents the order of seasonal integration. If a seasonal differentiation existed, $D = 1$. Otherwise, $D = 0$. Lowercase d is the number of non-seasonal differentiations made. S is the season of the time series (12 months). It means subtracting the time series by itself shifted of 12 months (the seasonality). On the autocorrelation and partial autocorrelation figures, there are now less significant peaks. As one non-seasonal integration was realized, the parameter d was equal to 1. The values of all parameters are summarized in Table 1.

Table 1. Parameters of the SARIMA model.

Features/Parameters	p	q	d	S	P	Q	D
Temperature (°C)	3	2	1	12	1	0	1
Dissolved oxygen concentration (mg/L)	4	4	1	12	1	0	1

It was not necessary that these parameters (p, q, P and Q) in the models were considered. To compute the optimum combination of parameters, a grid search method was used. First, a list was made for parameters p, q, P , and Q . The value of each parameter in this list was filled with integers from 0 or 1. For example, if $p = 4$ for the DO concentration, the list would be $[1, 2, 3, p = 4]$. The Akaike’s Information Criterion (AIC) was applied further for determining the values of these parameters. It is used to determine which model is best for a dataset. The AIC is a model selection tool that measures the quality of a statistical model for a given data set. A model with a lower AIC is more parsimonious or has a better balance of model fit and generalizability. Every combination of parameters $[p, d, q, P, D, Q, s]$ was tested on 96% of the dataset. Model forecasting demonstrated reliability in the short term (3 to 4 years). Hence, temperature and dissolved oxygen predictions were carried out

for a span of approximately two years, representing roughly 4% of the complete dataset. The best parameter combination is the one associated with the lowest AIC. Table 2 shows the optimal parameters to fit and forecast temperature and DO time series (trained on 96% of the dataset). It can be found that the temperature model fitted better with $p = 1, q = 1, P = 1, Q = 1$ and the DO model fitted better with $p = 4, q = 2, P = 0, \text{ and } Q = 1$ (Table 2). Regarding parameters D, d , and s, d was equal to 1 because one non-seasonal integration had been realized. D was equal to one because a seasonal integration had been realized (if no seasonal integration is realized, D is equal to 0). The season parameter, s , was equal to 12 (months). Finally, of these models, we plugged the set of the best values into Equation (1) for the fittest model in this study.

Table 2. Optimal parameters to fit and forecast temperature and DO time series (trained on 96% of the dataset).

Rank	Parameters (p, q, P, Q)	AIC
Temperature		
0	(1, 1, 1, 1)	2593.3
1	(3, 1, 1, 1)	2593.4
2	(2, 1, 1, 1)	2593.9
3	(1, 2, 1, 1)	2594.1
Dissolved oxygen		
0	(4, 2, 0, 1)	1664.7
1	(4, 2, 1, 1)	1665.1
2	(3, 1, 0, 1)	1665.2
0	(4, 2, 0, 1)	1664.7

3. Results and Discussion

The modeled outputs were compared with the observed values. Therefore, the temperature and DO from the monitoring station, Fort McMurray, were employed for the model training and testing to evaluate the model performance. The model performance was evaluated using the metrics of RMSE, MAE, and MDAE during model training and testing. Then, the model with the same values of parameters was examined using the temperature and DO data from the Town of Athabasca for the validation.

3.1. Model Training and Testing

To assess the efficiency of the model, datasets were separated into training and testing parts. Several groups of data ratios were tested for training/testing: 70%/30%, 80%/20%, 90%/10%, and 96%/4%. For 80%/20% and 90%/10%, the RMSE of the temperature were 1.99 and 2.92 in training and 2.07 and 1.81 in testing, respectively, in Athabasca. Furthermore, the RMSE of DO were 0.78 in training and 0.59 and 0.51 in testing, respectively, in Athabasca. Finally, 96%/4% was the best ratio and the RMSE of DO was 0.92 for training and 0.77 for the testing at Athabasca. Different types of errors were calculated on these training and testing parts (Table 3). Table 3 summarizes the performance statistics of the model for the temperature simulations for the two monitoring stations. Table 3 shows training and testing metrics for temperature. It can be observed that the RMSE and NSE were 2.7 and 0.87 for training and 2.2 and 0.95 for testing, respectively, at Fort McMurray. In contrast, the RMSE and NSE were 2.0 and 0.93 for training and 1.8 and 0.97 for testing, respectively, in the Town of Athabasca. The testing performances were better than the training counterparts at both stations. This is because the testing data were of higher quality.

Therefore, the resulting model performance for the temperature simulations was satisfied in both training and testing.

Table 3. Errors of temperature forecasting model.

Metrics	RMSE °C	MDAE °C	MAE °C	NSE
Fort McMurray				
Training part error	2.7	1.3	1.9	0.87
Testing part error	2.2	0.8	1.5	0.95
Town of Athabasca				
Training part error	2.0	0.7	1.3	0.93
Testing part error	1.8	0.5	1.2	0.97

Table 4 shows the performance statistics of the model for the DO simulations for the two monitoring stations. It can be observed that the RMSE and NSE were 0.76 and 0.79 for training and 0.67 and 0.92 for testing, respectively, at Fort McMurray. In contrast, the RMSE and NSE were 0.92 and 0.72 for training and 0.77 and 0.86 for testing, respectively, in the Town of Athabasca. It is obvious that the testing values were better than the training ones. This demonstrates that the SARIMA model was general once its parameters had been optimized. After the parameters of the model had been obtained/adjusted using the dataset from Fort McMurray, the model could also perform well in the Town of Athabasca. The testing performances were better than the training ones at both stations.

Table 4. Errors of the DO forecasting model.

Metrics	RMSE	MDAE	MAE	NSE
Fort McMurray				
Training part error	0.76	0.3	0.47	0.79
Testing part error	0.67	0.47	0.58	0.92
Town of Athabasca				
Training part error	0.92	0.49	0.68	0.72
Testing part error	0.77	0.43	0.58	0.86

3.2. Temperatures Effects

Figure 4 shows a comparison of the temperature between the observed time series and the modeled results at the Fort McMurray station. The blue curve represents the observed temperature time series, while the red one represents the modeled temperature. There are three bands delimiting different portions. The green band represents the training period. The red band indicates the testing. Finally, the gray part represents the prediction via the extrapolated values. It can be found that the model fitted well the observed data in both the training and the testing parts. Figure 4 shows the correlation plots of the modeled results and observed data in training and testing. It can be found that the R^2 value was 0.87 for the training and 0.95 for the testing, respectively. It means that the model accounted for a major part of the variation in the observed values and captured the seasonality, the trend, and the variation of the values well.

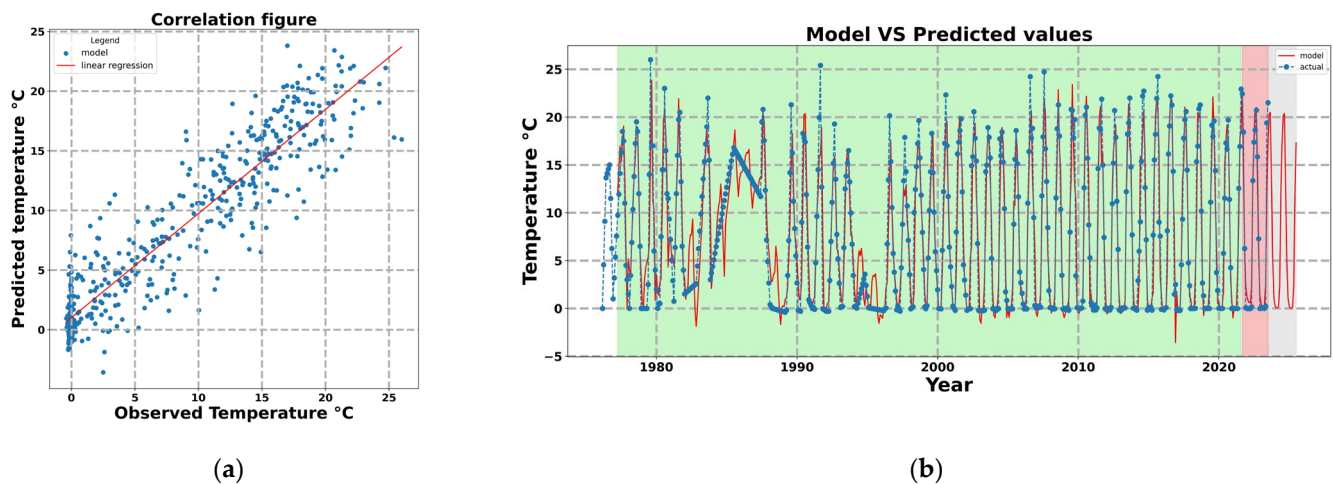


Figure 4. Performance of the SARIMA model between actual and forecasted temperature at Fort McMurray (trained on 96% of the dataset): (a) correlation figure and linear regression of the temperature time series (training portion) (NSE = 0.87, slope = 0.87, intercept = 1.01) and (b) time series.

The model fitted well the observed data in both the training and the testing parts. This is confirmed by the low errors encountered during training and testing. Regarding the prediction, the trend was constant for the next two years. Indeed, the model did not catch any increasing or decreasing trend in the training. The maximal annual predicted temperatures were slightly lower than the previous year's temperature peaks. It is because the model had slightly overestimated the observed values at the end of the training. Hence, it had reduced the predicted peak values. Regarding the model transition values (the increasing or decreasing parts of the model curves), they overall fitted well the actual values. The reason for this is that, from one year to the other, the monthly average temperature had approximately the same variation. Thus, by giving the right season parameter (12 months) to the model, it will catch the transition values and place them in the right period.

Without loss of generality, we checked if the model was potentially suitable for other sampling stations or not; the sampling station used to test this generalization was in the Town of Athabasca. The temperature from the observed data and the simulated results was compared in Athabasca from 1985 to 2023. It can be found that the model fitted well the observed data in both the training and the testing parts. Figure 5 shows a comparison of temperature between the observed time series and the modeled results at the Town of Athabasca station. It can be found that NSE value was 0.95 for the training and 0.95 for the testing, respectively. It should be noted that the R^2 values were better in the Town of Athabasca than those at Fort McMurray. This can be explained by the fact that the observed data from the Town of Athabasca were more regular than those from Fort McMurray. It means that the model accounted for a major part of the variation of the observed values and captured the seasonality, the trend, and the variation in the values well. Thus, the quality of the data is important for the training and testing of the SARIMA model. The model efficiency on another station was due to a similar temperature amplitude and seasonal variation. Again, the model accounted for a major part of the variation of the observed values and captured the seasonality, the trend, and the variation in the observed values well. Hence, at least between the two stations studied, stream temperature was scarcely influenced by external factors.

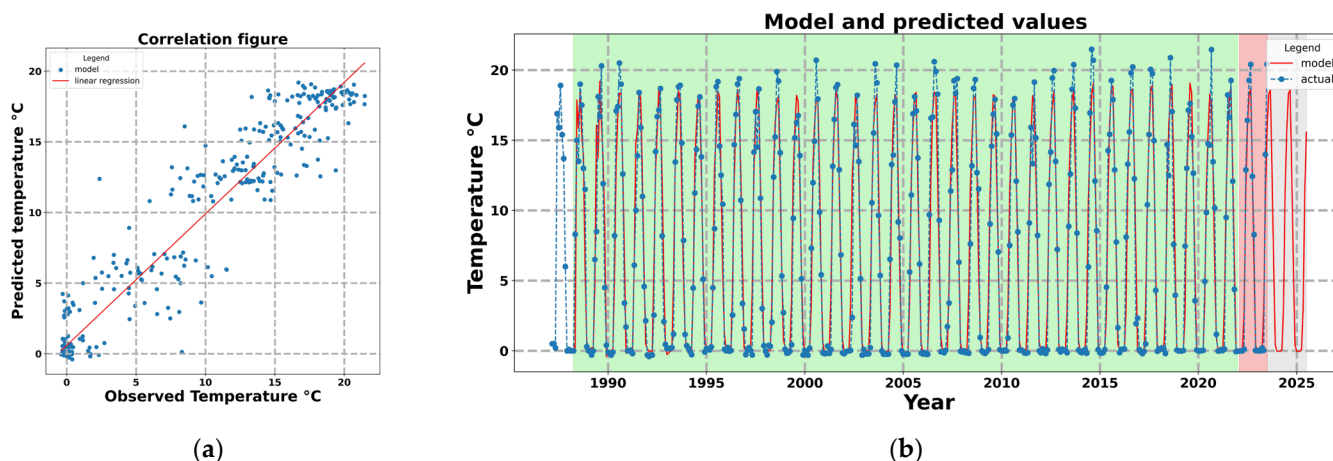


Figure 5. A comparison of actual and forecasted temperature in the Town of Athabasca (trained on 96% of the dataset): (a) correlation figure and linear regression of the temperature time series (training portion) (NSE = 0.95, slope = 0.86, intercept = 0.91) and (b) time series.

Stream water temperature can affect a wide range of aquatic organisms from invertebrates [47] to salmonids [48,49]. Many cold-blooded organisms and fishes prefer a specific range of temperature, affecting their habitats and distribution within streams [4,50]. Shifting thermal regimes from the preferred thermal range can inversely affect the habitats of the aquatic organisms [10,51]. For example, stream temperature can directly affect the development of salmonid eggs [52]. Stream temperatures show that the seasonal changes tend to be higher during summer and lower in winter. However, some yearly changes can be found. From 1982 to 1997, there were much lower temperatures in some years at Fort McMurray (Figure 4). This is because of the exceedingly small number of samples and big data gaps. However, the model did capture the valleys in temperature. This implies that the model would be useful for ecosystem assessment in a changing climate.

3.3. Effects of Dissolved Oxygen

Figure 6 shows a comparison of the dissolved oxygen between the observed data and the modeled results at Fort McMurray. The blue curve represents the observed DO time series while the red one represents the modeled DO time series. There are three bands delimiting different portions: Training (green), testing (red) and prediction (grey). It can be found that the model fits well the observed data with NSE values of 79% for the training and 92% for the testing, respectively. It can be found that the DO is very irregular due to missing samples, particularly from 1976 to 1990. This leads to the modeling difficulty of fitting the observed values. However, the model catches well the pattern and the trend of the seasonality with slightly low amplitudes. This is consistent with the results of the process-based models [14]. Figure 6a shows the correlation plot of the modeled results and the observed data in the training and testing sets that confirms the good performances. Furthermore, it can be found that DO ranged from 8 mg/L to 14 mg/L. However, DO showed a high peak of 16 mg/L in 1983 and a low valley of 6 mg/L in 1980.

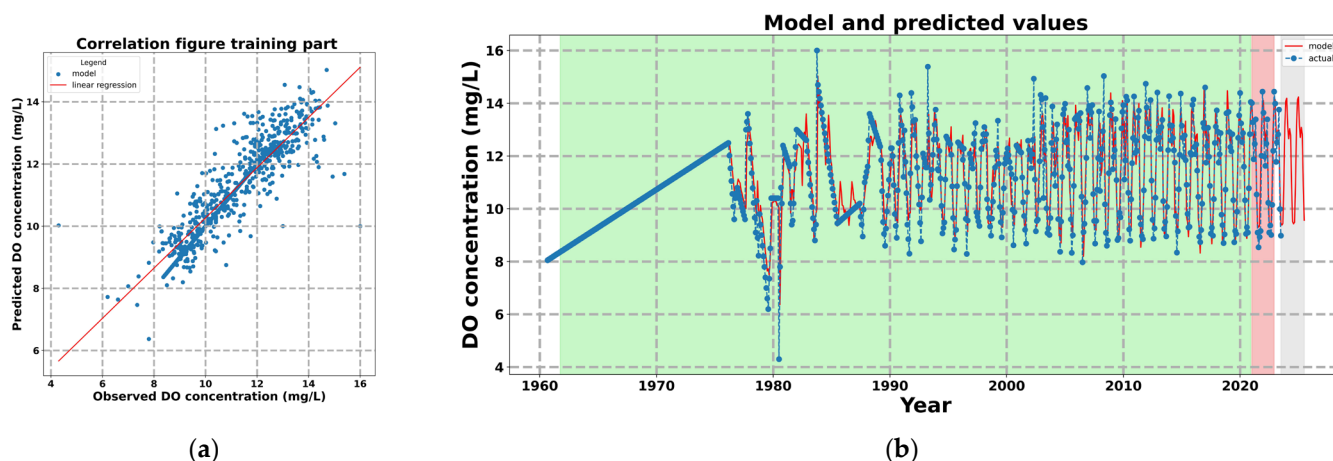


Figure 6. A comparison of actual and forecasted DO at Fort McMurray (training on 96% of the dataset): (a) correlation figure and linear regression of the temperature time series (training portion) (NSE = 0.79, slope = 0.81, intercept = 2.19) and (b) time series.

DO is an indicator of the most studied water quality in freshwater ecosystems because it frequently limits the survival and growth of fish in rivers and lakes. DO determines the suitability of the stream for the habitat and survival of the fishes. River water having DO concentrations < 2 mg/L and a BOD >15 mg/L are considered severely polluted, making the survival of aquatic organisms difficult [53]. For fishes, about 4 mg/L of minimum DO is necessary for the survival of the fishes in water [54–56]. Oxygen-depleting substances reduce the available DO. During winter, the river becomes covered with meters of ice, implying that these waters might have a lower oxygen level, possibly affecting some aquatic organisms. This might explain that environmental protection was insufficient in the 1960s–1980s. It can be observed that the lower valley of DO was in the 1980s. Figure 7 shows the effects of stream temperatures on DO at Fort McMurray from 2016 to 2022. The DO level was inversely proportional to water temperature. During summer months, the rate of biological oxidation is highly increased, and warm water holds less DO than cool water. The DO concentration could be at its minimum due to the higher temperature.

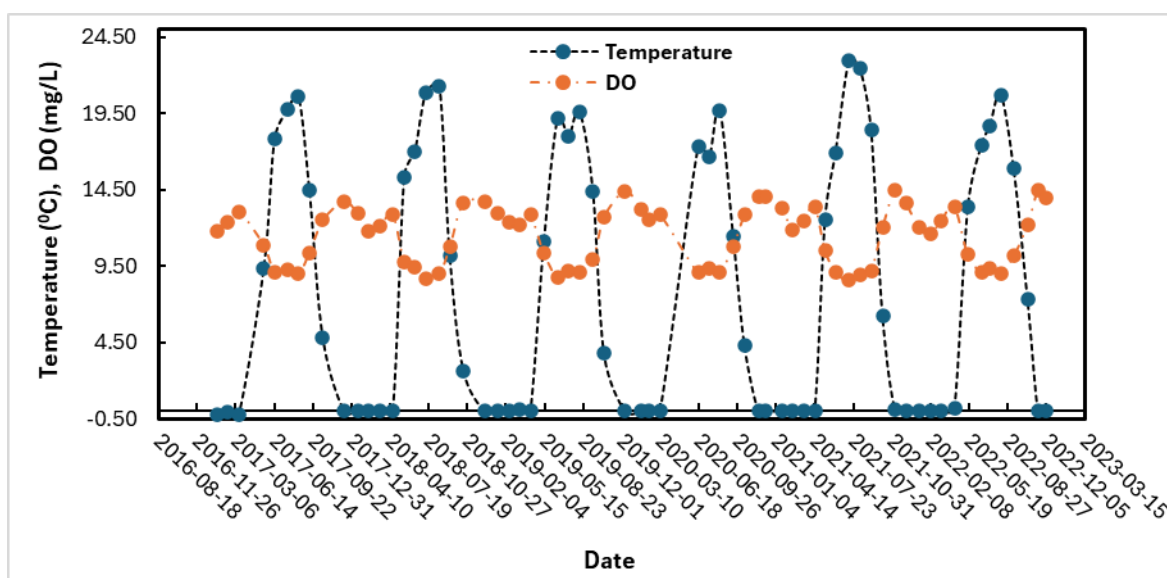


Figure 7. Effects of water temperature on DO at Fort McMurray (2016–2022).

The model was examined against the data of the dissolved oxygen concentration in the Town of Athabasca using the same values for the parameters from Fort McMurray station. Concerning the model generalization, the average errors in the training and testing parts were, respectively, around 0.7 mg/L and 0.6 mg/L. Figure 8 shows a comparison of the DO between the modeled results and the observed data. It can be found that the model fitted well the observed data in both the training and the testing parts. The model caught well the pattern and the trend of the seasonality with slightly low amplitudes. According to the correlation plots made with the modeled results and the observed data, the model performed well during the training and testing. Indeed, the NSE value was 0.72 for the training and 0.86 for the testing. Furthermore, it can be found that DO ranged from 8 mg/L to 14 mg/L. However, DO showed a high peak of 14 mg/L in 1983 and a low level of 7 mg/L in some years. This implies that the DO levels were better in the Town of Athabasca than at Fort McMurray.

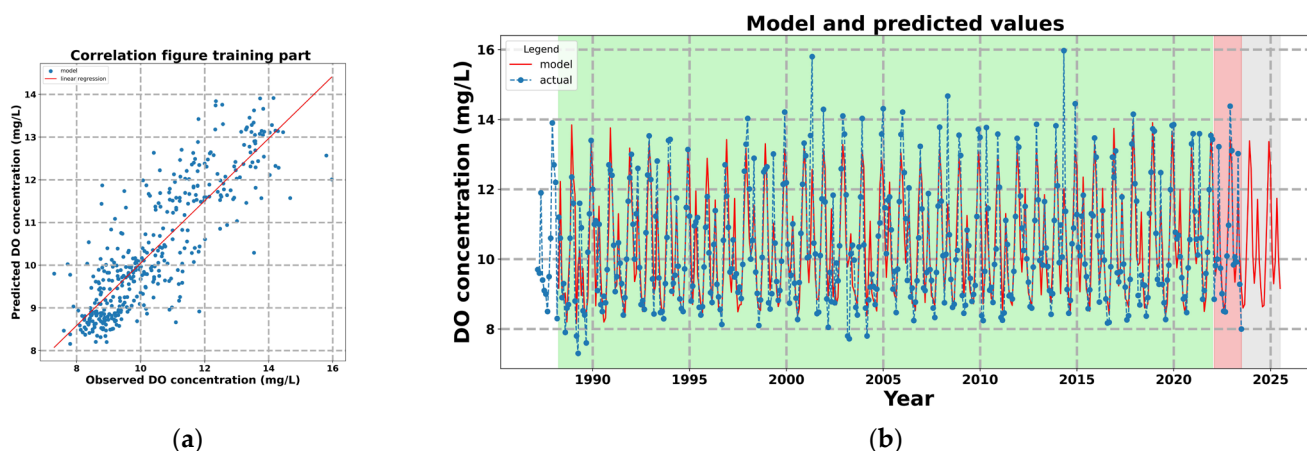


Figure 8. A comparison of actual and forecasted DO in the Town of Athabasca (trained on 96% of the dataset): (a) correlation figure and linear regression of the temperature time series (training portion) (NSE = 0.72, slope = 0.73, intercept = 2.74) and (b) time series.

3.4. Discussion

In these regions, atmospheric temperature has a great gradient, and its difference in this region can change from 35 °C in summer to −40 °C in winter [57]. The high difference can impact on DO levels. The two monitoring stations are geographically far away and have different land use and land cover regimes. The Fort McMurray station is surrounded by industries, homes, and residences. In contrast, the Town of Athabasca station is in the middle of the stream, in a rural and agricultural area. This might affect the evolution of DO concentrations. While the model parameters, e.g., p, q, d, S, P, Q, D , were calibrated only using the data from Fort McMurray, it still captured the pattern of temperature and DO using the same model parameters at Athabasca. This demonstrates its ability and robustness in predicting temperature and DO in cold climate regions if the model parameters are calibrated carefully.

In contrast to process-based models, the SARIMA model requires relatively few input data points and less technical users’ expertise, and, therefore, requires less computing power. It has benefits for temperature and DO forecasting in data-scarce cases, since it can make reasonably accurate forecasts even with limited data. The data-driven models are relatively more adaptable and dynamic. To create river-specific models for watershed management, the model covers a wide range of input variables, allows the use of pre-existing water quality statistics, and captures nonlinear interactions. However, the data-driven model has limitations since it has no explicit representation of the underlying physical processes. This makes it hard to understand how variations in land use or climate affect

the underlying physical processes and identify the main factors in management and research, a phenomenon which may limit their interpretability and generalizability in certain cases. Monitoring instrument malfunctions and sampling difficulties might introduce data anomalies or losses, especially after extreme cold weather events, a phenomenon which may lead to significant data gaps, such as the one for the Athabasca River. Furthermore, measured data are not error-free due to measurement uncertainty. The SARIMA model could manipulate irregularities either to fill or exclude the corresponding data points using the Pandas library. However, this is not an optimal process for training if water quality data follow a time series pattern. Furthermore, the length and frequency of data are often limited by sampling and resources in many situations. Nowadays, the monthly sampling or the lower frequency sampling need to be improved to be spatially and temporally representative. Due to the low sampling frequency, data-driven models are sensitive to the quality and representativeness of the training data. The SARIMA model performance may be affected due to biases or outliers in the time series data. For example, the SARIMA model missed seasonal changes in temperature due to missing data at Fort McMurray in the 1980s and 1990s. Furthermore, it did not capture the seasonal pattern of the DO at Fort McMurray from 1960 to 1990. Annual peaks, valleys, and patterns were not real due to data gaps. Therefore, for insufficient datasets, it is more difficult to train the model for each component, requiring more data for training. Hence, the model must be regularly trained on further data. However, with a larger dataset, the model performed well at this station when these irregular sections of the data corresponded to a lower percentage of the total data during the testing period.

The forecasting of the SARIMA model is only reliable in the short to medium term, but not in the long term. However, such SARIMA models are still useful for aquatic ecosystem management, since the model can capture patterns and connections directly from the available data if DO concentrations fall below a threshold value. Short-term forecasting can provide insights into outcomes and potential risks that can potentially occur in a given situation in advance and make data-based decisions optimal for specific risks. This offers opportunities for cost-effective solutions to prevent water quality crises, such as reestablishing oxic conditions, as part of risk management. With the rapid accumulation of monitoring data and the advancements in computer techniques, there is the potential to enhance the performance of data-driven models for dissolved oxygen (DO) modeling in the near future.

4. Conclusions

Temperature and DO are two of the most important indicators of water quality in water resource management, owing to their extensive effects on aquatic ecosystems. In this study, we established a SARIMA model, optimized the parameters of the model, and performed the preprocessing and purification of misleading data. The two monitoring stations with different land uses, Fort McMurray and the Town of Athabasca, were employed for the model training and testing to evaluate the model performance. It was found that data uncertainty and irregularity did affect the model performance and accuracy. Particularly, if the data are insufficient, this could lead to biased results. With the data increased, the SARIMA model fitted the data well and the stochastic seasonal fluctuation and patterns were successfully captured, except for some irregular extreme values after training. The results show that, for the modeled temperature, the RMSE and R^2 were 2.7 and 0.87 for training and 2.2 and 0.95 for testing, respectively, at Fort McMurray. In contrast, the RMSE and R^2 were 2.0 and 0.93 for training and 1.8 and 0.97 for testing, respectively, in the Town of Athabasca. It can be observed that, for the modeled DO, the RMSE and R^2 were 0.76 and 0.79 for training and 0.67 and 0.92 for testing, respectively, at Fort McMurray. In

contrast, the RMSE and R^2 were 0.92 and 0.72 for training and 0.77 and 0.86 for testing, respectively, in the Town of Athabasca. Therefore, the resulting model performance for the temperature simulations was satisfying during both training and testing. The proposed SARIMA model was satisfying in terms of capturing non-linear temporal correlations for the trend of temperature and DO change. After the parameters of the model were obtained/adjusted using the dataset from Fort McMurray, the model could also perform well in the Town of Athabasca. The testing performances were better than the training counterparts at both stations. This demonstrates that the SARIMA model was general once its parameters had been optimized.

The application of time series forecasting models and analysis methods can be employed to extract essential information concerning temperature and dissolved oxygen (DO) which is crucial for the well-being of aquatic ecosystems. These techniques enable the creation of site-specific predictive models for temperature and DO, facilitating cost-effective monitoring of river health and offering insights for region-specific management strategies across various land use and land cover gradients. Moreover, the model and methodology are transferable to the prediction of temperature and DO in other aquatic environments like lakes and rivers.

Author Contributions: S.L.: conceptualization, data collection, methodology, investigation, analysis and interpretation, visualization, original draft preparation; J.W.: conceptualization, data collection, original draft writing, review and editing, supervision; M.A.D.: review and editing, supervision; M.F.: conceptualization, review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Nazemi, A.; Wheeler, H.S.; Chun, K.P.; Elshorbagy, A. A stochastic reconstruction framework for analysis of water resource system vulnerability to climate-induced changes in river flow regime. *Water Resour. Res.* **2013**, *49*, 291–305. [[CrossRef](#)]
2. Reid, A.J.; Carlson, A.K.; Creed, I.F.; Eliason, E.J.; Gell, P.A.; Johnson, P.T.; Kidd, K.A.; MacCormack, T.J.; Olden, J.D.; Ormerod, S.J.; et al. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biol. Rev.* **2019**, *94*, 849–873. [[CrossRef](#)] [[PubMed](#)]
3. Caissie, D.; Satish, M.G.; El-Jabi, N. Predicting water temperatures using a deterministic model: Application on Miramichi River catchments (New Brunswick, Canada). *J. Hydrol.* **2007**, *336*, 303–315. [[CrossRef](#)]
4. Du, X.; Shrestha, N.K.; Wang, J. Assessing climate change impacts on stream temperature in the Athabasca River Basin using SWAT equilibrium temperature model and its potential impacts on stream ecosystem. *Sci. Total Environ.* **2019**, *650*, 1872–1881. [[CrossRef](#)]
5. Meshesha, T.W.; Wang, J.; Melaku, N.D. Modelling spatiotemporal patterns of water quality and its impacts on aquatic ecosystem in the cold climate region of Alberta, Canada. *J. Hydrol.* **2020**, *587*, 124952. [[CrossRef](#)]
6. Ouellet, V.; St-Hilaire, A.; Dugdale, S.J.; Hannah, D.M.; Krause, S.; Proulx-Ouellet, S. River temperature research and practice: Recent challenges and emerging opportunities for managing thermal habitat conditions in stream ecosystems. *Sci. Total Environ.* **2020**, *736*, 139679. [[CrossRef](#)]
7. Franklin, P.A. Dissolved oxygen criteria for freshwater fish in New Zealand: A revised approach. *N. Z. J. Mar. Freshw. Res.* **2014**, *48*, 112–116. [[CrossRef](#)]
8. Du, X.; Shrestha, N.K.; Ficklin, D.L.; Wang, J. Incorporation of the equilibrium temperature approach in a Soil and Water Assessment Tool hydroclimatological stream temperature model. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 2343–2357. [[CrossRef](#)]
9. Harvey, R.; Lye, L.; Khan, A.; Paterson, R. The Influence of Air Temperature on Water Temperature and the Concentration of Dissolved Oxygen in Newfoundland Rivers. *Can. Water Resour. J.* **2011**, *36*, 171–192. [[CrossRef](#)]

10. Larocque, S.M.; Tang, R.W.K.; Doka, S.E. *Water Temperature and Dissolved Oxygen Monitoring in Areas of Concern in the St. Clair-Detroit River System*; Canadian Technical Report of Fisheries and Aquatic Sciences; Department of Fisheries and Oceans: Vancouver, BC, Canada, 2020; 120p.
11. Cox, B.A. A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. *Sci. Total Environ.* **2003**, *314–316*, 335–377. [[CrossRef](#)]
12. Zhi, W.; Ouyang, W.; Shen, C.; Li, L. Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers. *Nat. Water* **2023**, *1*, 249–260. [[CrossRef](#)]
13. Feigl, M.; Lebiezinski, K.; Herrnegger, M.; Schulz, K. Machine-learning methods for stream water temperature prediction. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 2951–2977. [[CrossRef](#)]
14. Shrestha, N.K.; Wang, J. Water Quality Management of a Cold Climate Region Watershed in Changing Climate. *J. Environ. Inform.* **2020**, *35*, 56–80. [[CrossRef](#)]
15. Wang, J.; Li, Y.; Bork, E.W.; Richter, G.M.; Eum, H.-I.; Chen, C.; Shah, S.H.H.; Mezbahuddin, S. Modelling spatio-temporal patterns of soil carbon and greenhouse gas emissions in grazing lands: Current status and prospects. *Sci. Total. Environ.* **2020**, *739*, 139092. [[CrossRef](#)]
16. Zhu, S.; Heddam, S. Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: Extreme learning machines (ELM) versus artificial neural network (ANN). *Water Qual. Res. J.* **2020**, *55*, 106–118. [[CrossRef](#)]
17. Bustillo, V.; Moatar, F.; Ducharme, A.; Thiéry, D.; Poirel, A. A multimodel comparison for assessing water temperatures under changing climate conditions via the equilibrium temperature concept: Case study of the middle Loire River, France. *Hydrol. Process.* **2014**, *28*, 1507–1524. [[CrossRef](#)]
18. Piccolroaz, S.; Calamita, E.; Majone, B.; Gallice, A.; Siviglia, A.; Toffolon, M. Prediction of river water temperature: A comparison between a new family of hybrid models and statistical approaches. *Hydrol. Process.* **2016**, *30*, 3901–3917. [[CrossRef](#)]
19. Van Vliet, M.T.; Franssen, W.H.; Yearsley, J.R.; Ludwig, F.; Haddeland, I.; Lettenmaier, D.P.; Kabat, P. Global river discharge and water temperature under climate change. *Glob. Environ. Change* **2013**, *23*, 450–464. [[CrossRef](#)]
20. Herzog, S.; Wörgötter, F.; Parlitz, U. Data-Driven Modeling and Prediction of Complex Spatio-Temporal Dynamics in Excitable Media. *Front. Appl. Math. Stat.* **2018**, *4*, 60. [[CrossRef](#)]
21. Abba, S.I.; Linh, N.T.T.; Abdullahi, J.; Ali, S.I.A.; Pham, Q.B.; Abdulkadir, R.A.; Costache, R.; Anh, D.T. Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration. *IEEE Access* **2020**, *8*, 157218–157237. [[CrossRef](#)]
22. Tiyasha Tung, T.M.; Yaseen, Z.M. A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J. Hydrol.* **2020**, *585*, 124670. [[CrossRef](#)]
23. Csábrági, A.; Molnár, S.; Tanos, P.; Kovács, J. Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube. *Ecol. Eng.* **2017**, *100*, 63–72. [[CrossRef](#)]
24. Dabrowski, J.J.; Rahman, A.; George, A. Prediction of dissolved oxygen from pH and water temperature in aquaculture prawn ponds. In Proceedings of the Australasian Joint Conference on Artificial Intelligence—Workshops (AIW'18), Wellington, New Zealand, 11–14 December 2018; Deng, J.D., Rahman, A., Eds.; ACM: New York, NY, USA, 2018; pp. 2–6.
25. Qiu, R.; Wang, Y.; Rhoads, B.; Wang, D.; Qiu, W.; Tao, Y.; Wu, J. River water temperature forecasting using a deep learning method. *J. Hydrol.* **2021**, *595*, 126016. [[CrossRef](#)]
26. Adamowski, J.; Chan, H.F.; Prasher, S.O.; Ozga-Zielinski, B.; Sliusarieva, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour. Res.* **2012**, *48*, W01528. [[CrossRef](#)]
27. Abdul Wahid, A.; Arunbabu, E. Forecasting water quality using seasonal ARIMA model by integrating in-situ measurements and remote sensing techniques in Krishnagiri reservoir, India. *Water Pract. Technol.* **2022**, *17*, 1230–1252. [[CrossRef](#)]
28. Ahmed, M.H.; Lin, L.S. Dissolved oxygen concentration predictions for running waters with different land use land cover using a quantile regression forest machine learning technique. *J. Hydrol.* **2021**, *597*, 126213. [[CrossRef](#)]
29. Khan, U.T.; Valeo, C. Dissolved oxygen prediction using a possibility theory based fuzzy neural network. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2267–2293. [[CrossRef](#)]
30. Kobiela, D.; Krefta, D.; Król, W.; Weichbroth, P. ARIMA vs LSTM on NASDAQ stock exchange data. *Procedia Comput. Sci.* **2022**, *207*, 3836–3845. [[CrossRef](#)]
31. Han, J.; Lin, H.; Qin, Z. Prediction and Comparison of In-Vehicle CO₂ Concentration Based on ARIMA and LSTM Models. *Appl. Sci.* **2023**, *13*, 10858. [[CrossRef](#)]
32. Long, B.; Tan, F.; Newman, M. Forecasting the Monkeypox Outbreak Using ARIMA, Prophet, Neural Prophet, and LSTM Models in the United States. *Forecasting* **2023**, *5*, 5. [[CrossRef](#)]
33. Khan, U.T.; Valeo, C. A new fuzzy linear regression approach for dissolved oxygen prediction. *Hydrol. Sci. J.* **2015**, *60*, 1096–1119. [[CrossRef](#)]
34. Hoang, T.H.T.; Nguyen, V.D.; Van, A.D.; Nguyen, H.T.T. Decision tree techniques to assess the role of daily DO variation in classifying shallow eutrophicated lakes in Hanoi, Vietnam. *Water Qual. Res. J.* **2020**, *55*, 67–78. [[CrossRef](#)]

35. Hong, P.T.; Hien, N.T.T.; Chung, N.T.; Hung, N.Q.; Cuong, N.C.; Huong, H.T.T. Combination impact of pH and temperature on the toxicity of lead on zooplankton in the context of global warming. *Vietnam. J. Sci. Technol.* **2020**, *58*, 105–114. [[CrossRef](#)]
36. Liu, C.; Sun, X.; Su, L.; Cai, J.; Zhang, L.; Guo, L. Assessment of phytoplankton community structure and water quality in the Hongmen Reservoir. *Water Qual. Res. J.* **2021**, *56*, 19–30. [[CrossRef](#)]
37. Aklakur, M.; Bakli, S.; Deo, A.D.; Singh, D.K.; Pailan, G.H. Cyanobacteria toxicity in aquaculture system and its impact on fish physiology. *J. Aquac. Mar. Biol.* **2023**, *12*, 28–33. [[CrossRef](#)]
38. He, J.; Chu, A.; Ryan, M.C.; Valeo, C.; Zaitlin, B. Abiotic influences on dissolved oxygen in a riverine environment. *Ecol. Eng.* **2011**, *37*, 1804–1814. [[CrossRef](#)]
39. Bhanja, S.N.; Wang, J.; Shrestha, N.K.; Zhang, X. Microbial kinetics and thermodynamic (MKT) processes for soil organic matter decomposition and dynamic oxidation-reduction potential: Model descriptions and applications to soil N₂O emissions. *Environ. Pollut.* **2019**, *247*, 812–823. [[CrossRef](#)]
40. Bhanja, S.N.; Wang, J. Influence of environmental factors on autotrophic, soil and ecosystem respirations in Canadian boreal forest. *Ecol. Indic.* **2021**, *125*, 107517. [[CrossRef](#)]
41. Meshesha, T.W.; Wang, J.; Melaku, N.D. A modified Hydrological model for assessing effect of pH on fate and transport of Escherichia coli in the Athabasca River Basin. *J. Hydrol.* **2020**, *582*, 124513. [[CrossRef](#)]
42. LTRN. Long Term River Network for Water Quality Monitoring in Alberta. 2024. Available online: <http://environment.alberta.ca/apps/EdwReportViewer/LongTermRiverStation.aspx> (accessed on 3 March 2024).
43. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2016.
44. Chang, X.; Gao, M.; Wang, Y.; Hou, X. Seasonal autoregressive integrated moving average model for precipitation time series. *J. Math. Stat.* **2012**, *8*, 500–505. [[CrossRef](#)]
45. Moriasi, D.N.; Gitau, M.W.; Pai, N.; Daggupati, P. Hydrologic and water quality models: Performance measures and evaluation criteria. *Trans. ASABE* **2015**, *58*, 1763–1785.
46. Dickey, D.A.; Fuller, W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431.
47. Cox, T.J.; Rutherford, J.C. Predicting the effects of time-varying temperatures on stream invertebrates mortality. *N. Z. J. Mar. Freshw. Res.* **2000**, *34*, 209–215. [[CrossRef](#)]
48. Caissie, D. The thermal regime of rivers: A review. *Freshw. Biol.* **2006**, *51*, 1389–1406. [[CrossRef](#)]
49. Lee, R.M.; Rinne, J.N. Critical thermal maximum of five trout species in the southwestern United States. *Trans. Am. Fish. Soc.* **1980**, *109*, 632–635. [[CrossRef](#)]
50. Wichert, G.A.; Lin, P. A species tolerance index of maximum water temperature. *Water Qual. Res. J. Can.* **1996**, *31*, 875–893. [[CrossRef](#)]
51. Colby, B.R.; Niles, J.M.; Persons, M.H.; Wilson, M.J. Shifting thermal regimes influence competitive feeding and aggression dynamics of brook trout (*Salvelinus fontinalis*) and creek chub (*Semotilus atromaculatus*). *Ecol. Evol.* **2022**, *12*, e9056. [[CrossRef](#)]
52. Beer, W.N.; Anderson, J.J. Effect of spawning day and temperature on salmon emergence: Interpretations of a growth model for Methow River chinook. *Can. J. Fish. Aquat. Sci.* **2001**, *58*, 943–949. [[CrossRef](#)]
53. Giri, I.; Ritika, K.C.R.; Khadka, U.R. Water quality status in Bagmati river of Kathmandu valley, Nepal. In *Ecological Significance of River Ecosystems*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 481–502.
54. EPA 440/5-86-001; Quality Criteria for Water 1986. USEPA (United States Environmental Protection Agency): Washington, DC, USA, 1986.
55. Smith, A.J.; Delorme, L.D. Ostracoda. In *Ecology and Classification of North American Freshwater Invertebrates*; Thorp, J.H., Covich, A.P., Eds.; Academic Press: Cambridge, MA, USA, 2010; Chapter 19; pp. 725–771.
56. Patel, H.; Vashi, R.T. *Characterization and Treatment of Textile Wastewater*; Elsevier Inc.: Amsterdam, The Netherlands, 2015.
57. Delavar, M.A.; Wang, J. Simulation of a hybrid system of solar-microturbines in cold climate regions. *Appl. Therm. Eng.* **2021**, *182*, 116080. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.