Miistakis Institute

# Modelled Aquatic Condition Index for Calgary Wetlands

A component of the Urban Wetland Conservation project

Tracy S. Lee, Irena F. Creed, Sassan Mohammady, Ken Sanderson, Kevin J. Erratt and Caroline Lesage

Document prepared for The City of Calgary

Innovative research. Engaged communities. Healthy landscapes.

## Modelled Aquatic Condition Index for Calgary Wetlands

Prepared by Tracy S. Lee, Irena F. Creed, Sassan Mohammady, Ken Sanderson, Kevin J. Erratt and Caroline Lesage

September 2023

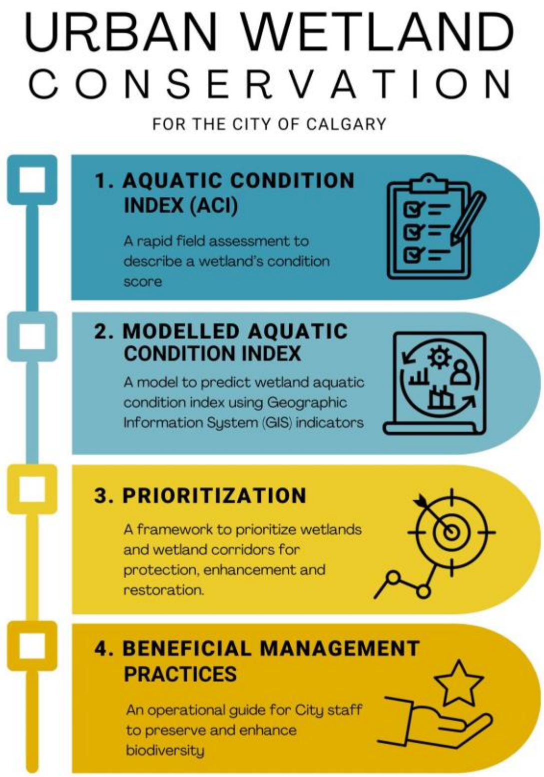
Miistakis Institute
EB3013, Mount Royal University
4825 Mount Royal Gate SW
Calgary, Alberta T3E 6K6

Phone: (403) 440–8444
Email: institute@rockies.ca
Web: www.rockies.ca

## Contents

## Acknowledgements

Funding to support development of the modelled-ACI was provided by The City of Calgary. The modelled-ACI (step 2 in the diagram below) is one component of a larger Urban Wetland Conservation program that provides support to The City of Calgary by developing products to support wetland conservation.




Thank you to the project advisory committee:

- Vanessa Carney, City of Calgary
- Lea Randall, Amphibian Biologist
- Dr. Felix Nwaishi, Mount Royal University

Caroline Lesage's involvement was supported by a grant from the NSERC-CREATE program in Leadership in Environmental and Digital Innovation for Sustainability (LEADS) during her time as a graduate student at Concordia University.

We would also like to thank Dr. Isabella Ghement for her statistical advice on the GAM approach.

# Executive Summary

*"Calgarians have an awareness, understanding and appreciation of the benefits of wetlands. As a result, wetlands have become an integral part of our city's urban fabric and they are maintained for the benefit, use and enjoyment of present and future Calgarians and visitors."*

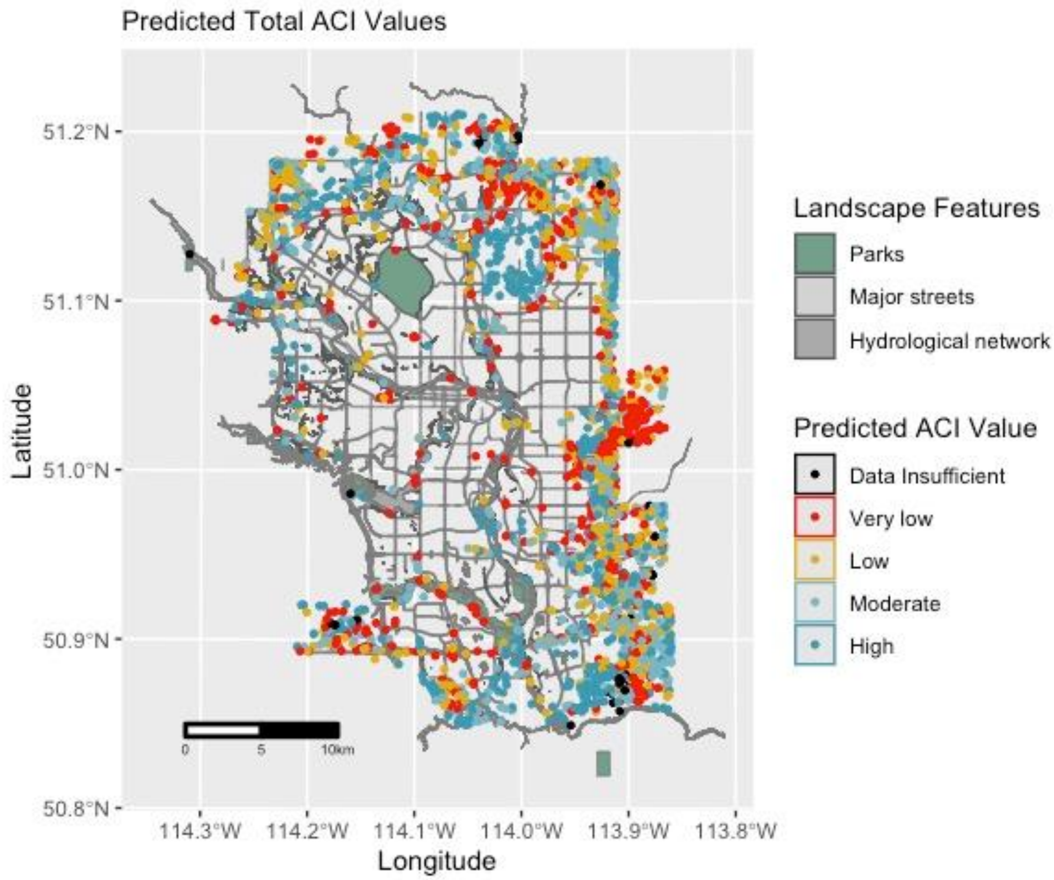
*(Wetland Conservation Plan, City of Calgary, 2004)*

Wetlands are unique ecosystems, occupying only a small percentage of the landscape but contributing a significant proportion of ecosystem services—they improve water quality and quantity, reduce flooding and soil erosion, provide biodiversity, and moderate climate conditions. Despite these benefits, urbanization greatly impacts a wetland's ability to provide ecosystem services, due to alterations to hydrology, ecology and water quality functions. Calgary has over 2700 water bodies that span a gradient from natural wetlands to constructed storm ponds. Given the large number and condition of Calgary wetlands, how does The City of Calgary identify where to invest in protection or improvements to their condition?

Here we test for the best approach to develop a modelled landscape-scale aquatic condition index for Calgary's wetlands. Once developed, we use our approach to predict the condition for all wetlands in the city of Calgary. We generated 27 predictive indicators deemed appropriate for assessing three wetland functions: hydrology, ecology, and water quality. For the model response variable, we used 74 wetland condition values generated from a rapid wetland assessment tool developed for The City of Calgary. To explore predictor-response relationships, we tested a combination of machine learning techniques, including nonlinear machine learning models such as Random Forest (RF) and Neural Network (NN), and linear and nonlinear regression models such as Generalized Additive Model (GAM). Our modelling framework not only tested different statistical approaches for the model itself, but also compared statistical and expert-driven indicator selection. Finally, we explored different training and testing data set proportions for cross-validation tests.

We assessed each modelling result using $R^2$ values, where we identified a minimum threshold of $R^2 > 0.60$ as necessary for management application. We assessed the predictive accuracy of each model using Mean Standard Error (MSE) and Unscaled Mean Bounded Relative Absolute Error (UMBRAE). A few of our modelling approaches did not meet the minimum $R^2$ value and were not considered further. Of the remaining models, we reviewed predictive accuracy as well as model approach limitations to select the best for predicting wetland condition.

We used a Neural Network approach to predict wetland condition, with a machine learning indicator selection process at a threshold of 3.5 and a 90/10 training and testing dataset. We generated modelled-ACI values for hydrology, ecology, and water quality functions and averaged condition values across these functions to generate a total modelled-ACI value for each wetland. We categorized the resulting modelled-ACI values into four classes: very low, low, moderate and high condition. All results are displayed spatially.




Landscape level spatial patterns that indicate wetland condition values tend to be higher in non-urbanized Calgary than in the inner city, the industrial area, on agriculture lands to the east, and along major transportation corridors. These general patterns can be used to support policy changes or management programs focused on specific impacts, for example to improve wetland conditions along roadsides or on industrial lands. The results can be used internally by The City of Calgary to prioritize wetlands for protection and/or restoration. Additionally, specific departments could use condition values to report trends or status. For example, the Parks Department could use condition scores for Natural Environment Parks (NEP) to track overall wetland scores in parks over time. A preliminary analysis found that 73% of wetlands in NEP have *very low* to *low* ecology condition, highlighting priority areas for field visits to identify opportunities to improve conditions.

As with all modelling approaches, there are limitations that should be acknowledged to improve future model outputs. The wetland inventory provided by The City of Calgary is limited by a highly changing landscape and data currency. We were aware that some wetlands used in our analyses no longer exist due to land use change. We recommend an updated wetland inventory could improve the modelling product. Another important limitation is the low number of response variables (field-ACI surveys) for predictor-response modelling. This may have limited the Generalized Additive Modelling approach and our ability to predict indicators of the modelled conditions. Indeed, the low number of ACI-field surveys means we likely did not have the right conditions to represent the entire city. With over 2,700 wetlands it is likely there are some in better and worse condition than those surveyed, constraining the Random Forest model that is restricted to the collected data range. We therefore recommend additional surveys (at least 180) to improve future versions of the modelled ACI.

Interpretation of predicted condition values could be further enhanced by considering The City of Calgary's new wetland typology. Wetland typologies include consideration of wetlands management goals (constructed storm pond is managed differently than a natural wetland) and would help refine prioritization of restoration activities at wetland with very low to low condition scores. We recommend that the new wetland typology is assigned to each wetland in future surveys.

Wetlands are an important component of Calgary's urban fabric, requiring investment in protection and restoration to enhance function and ultimately ecosystem services. The urban wetland conservation project, including the modelled-ACI, provides The City of Calgary with a wetland tool to improve wetland planning and management.

# Introduction

Wetlands play an important role in urban landscapes: they improve water quality, provide habitat for biodiversity, help address heat island effects, and are integral to stormwater infrastructure (Alikhani et al., 2021; Ampatzidis and Kershaw, 2020). Despite the many benefits of wetlands, as urban areas grow and intensify, wetland loss can be significant (Kentula et al., 2004; Lee et al., 2022). Furthermore, wetlands that are retained are often degraded, with increased pollutant and contaminant loads, and increasingly isolated from hydrological networks (Kometa et al., 2017). Wetlands in the urban environment often suffer from a gradient of anthropogenic impacts along a gradient from natural areas to fully constructed ponds for stormwater infrastructure (Alikhani et al., 2021). Urbanization, therefore, can greatly impact a wetland's ability to support ecosystem services such as filtering, water storage, or supporting biodiversity. Delivery of ecological services depends on specific wetland functions and can be measured by assessing condition (McLaughlin and Cohen, 2013). A better understanding of wetland condition would provide urban municipalities with the information necessary to prioritize areas for protection, restoration or management action.

To measure condition, wetland assessments can occur at three intensities: detailed, rapid, and landscape (Fennessy et al., 2007). A detailed site assessment requires a lengthy and rigorous field assessment that can be costly and unfeasible for urban municipalities with many wetlands. A rapid wetland assessment is conducted in the field and includes measuring/observing selected indicators that emphasize wetland condition. A rapid wetland assessment is intended to take less than half a day to complete (Fennessy et al., 2007). We designed a rapid wetland assessment index to be undertaken by ecologists for The City of Calgary to assess wetland condition (Nwaishi et al., 2023). This field-based assessment tool, adapted from a wetland assessment approach developed for Alberta, was termed field-aquatic condition index (field-ACI) (Creed et al., 2018). Generating scores from 0 (poor) to 1 (healthy), it measures three wetland functions: hydrology, water quality and biodiversity. Calgary has over 2,700 wetlands and it is therefore challenging for the municipality to conduct field visits to every wetland. This limits city-wide planning. We identified the need for a landscape scale assessment tool to predict wetland condition based on readily available datasets. We aimed to develop a tool that enables The City of Calgary to predict wetland condition of urban wetlands to inform wetland protection, restoration, and management.

Here we introduce a landscape modelled aquatic condition index for Calgary's wetlands referred to as modelled-ACI[1]. The modelled-ACI builds on the field-ACI[2] (Nwaishi et al., 2023) and predicts condition of all wetlands in Calgary. An important component of this project was to identify the best modelling approach for predicting wetland condition. To

[1] Modelled-ACI has been referred to internally at The City of Calgary as the predicted-ACI.
[2] Field-ACI has been referred to internally at The City of Calgary as the actual-ACI.

explore predictor-response relationships, we tested a combination of machine learning techniques, including nonlinear machine learning models such as Random Forest (RF) and Neural Network (NN) and linear and nonlinear regression models such as the Generalized Additive Model (GAM).

The modelled-ACI is an important tool for The City of Calgary as little is currently known about urban wetland conditions. Although the field-ACI is a powerful new methodology for systematically measuring aquatic condition, it requires a site visit that is unlikely to be completed in time for the full urban wetland inventory. Similar to the field-ACI, the modelled-ACI results in an overall condition score for each wetland in Calgary, as well as separate condition scores for each of the three wetland functions (hydrology, ecology, and water quality). The modelled-ACI was designed to be used by The City of Calgary to inform policy, planning and management decisions relating to urban wetlands. In addition, the modelled-ACI was developed to augment the accuracy of habitat condition ratings for natural area parks with greater than 10% aquatic environments and to inform annual City Parks and Open Spaces department reporting.

# Methods

## Study Area

Calgary, Alberta, is one of Canada's largest cities, with a population greater than 1.2 million. Typical of many North American urban areas, Calgary has a heavily developed core surrounded by residential neighbourhoods that continue to spread, currently covering 848 $km^2$. As a result of this expansion, it is estimated that Calgary has lost 90% of its wetlands since European settlement began in the $18^{th}$ century (City of Calgary, 2004), with most remaining wetlands in Calgary's urbanized areas contributing in some form to stormwater management. A current estimate indicates approximately 2,720 wetlands remain within the city limits (Figure 1), with the majority occurring in non-urbanized areas. Wetlands predominately occur in the north, east, and south of the city where densification has not yet occurred, alongside major roads within the transportation network, or within urban parks.

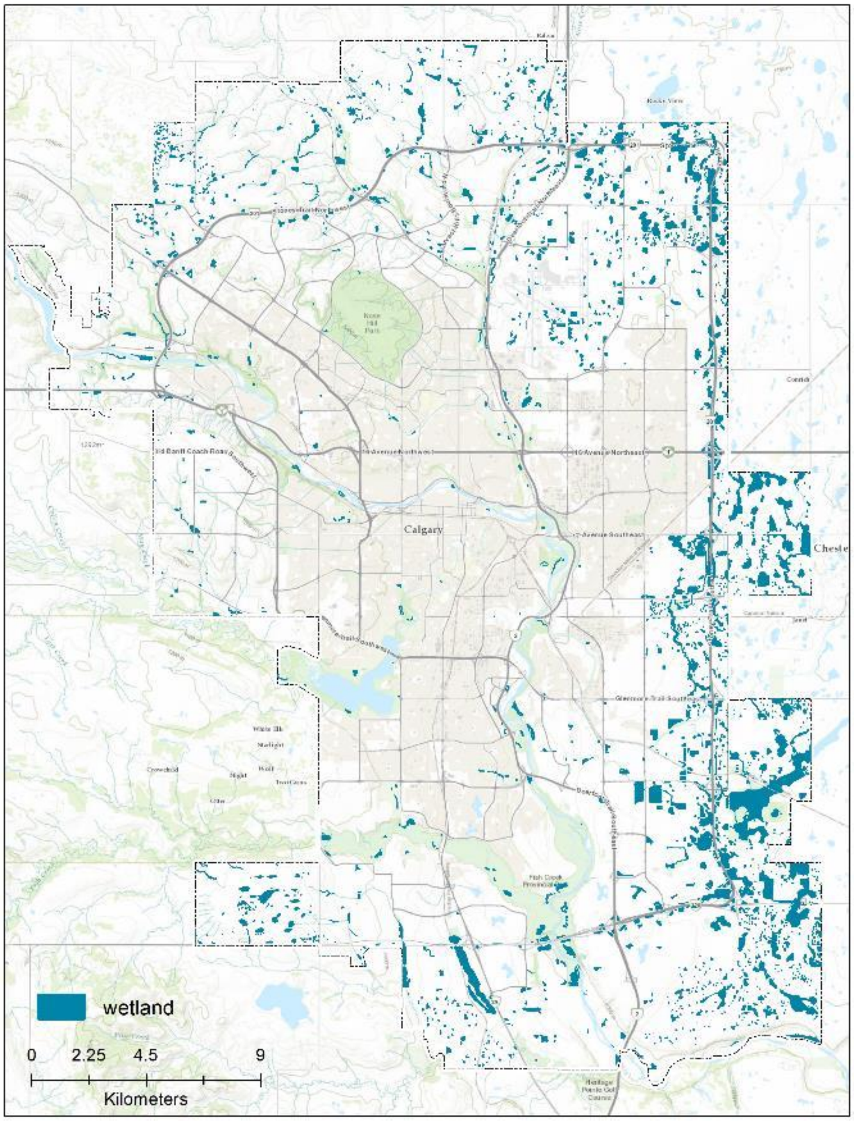



*Figure 1: Wetland inventory (dark blue), in Calgary, Alberta with natural areas (green) and roads (light grey).*

The wetland inventory used in the modelled-ACI was provided by Calgary's Parks and Open Spaces Department and was merged with The City's storm pond asset management inventory. We removed duplicates and inventory types recorded as reservoirs, dry ponds, historic wetlands, and community lakes. A new wetland inventory is under development by Calgary and partners and the modelled-ACI product would benefit from an updated analysis once the new wetland inventory is complete.

## Modelling Framework

### *Overview*

To predict wetland condition for The City of Calgary wetland inventory, we developed 27 predictive indicators and used field-ACI condition values for 74 wetlands surveyed in 2023 as the response variable (Nwaishi, 2023). To determine the best approach for predicting wetland condition, three modelling frameworks (A-C) were developed that tested three different statistical models and two different approaches to selecting indicators, resulting in seven predictive approaches (Figure 2). Each of the seven approaches was tested using three different divisions of training and testing data, and each of the 21 model results assessed for accuracy using three measures: $R^2$, Mean Squared Error (MSE), and Unscaled Mean Bounded Relative Absolute Error (UMBRAE) (Chen et al., 2017). Each of the three accuracy measures provides different information and have different strengths and limitations, so all three were considered to provide the most complete picture.

Each modelled-ACI resulted in an indexed condition value from 0 to 1 (where 0 indicates low condition and 1 indicates high condition) for each wetland function as well as an overall condition value based on average of the three wetland functions.

Predictive indicator selection, modelling approach, and predictive accuracy measures are described in more detail below.

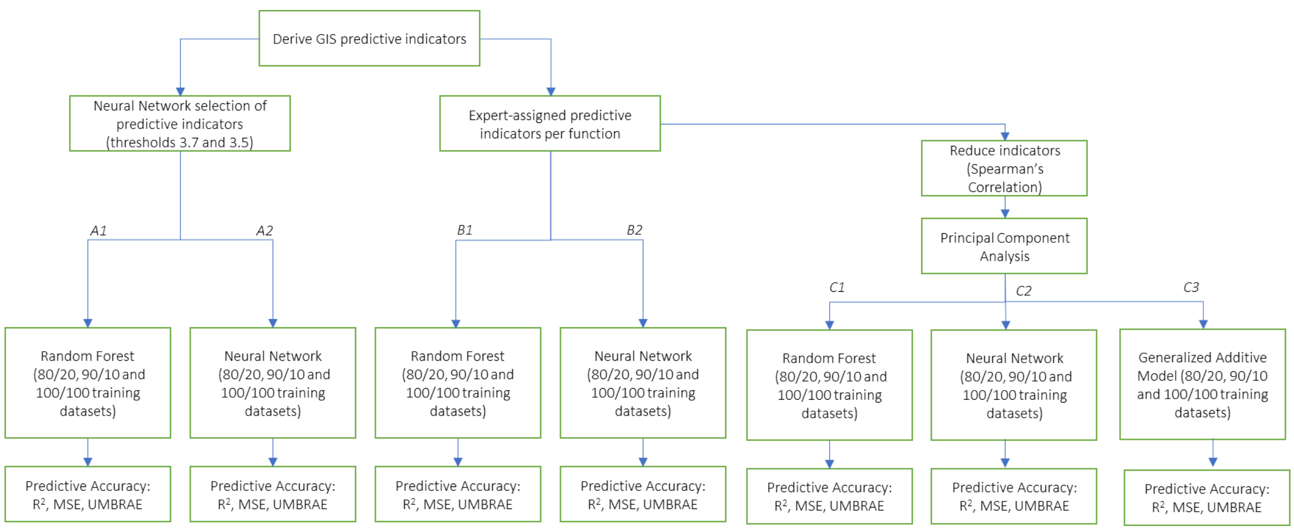



*Figure 2: Modelling Framework (A-C) to predict urban wetland condition for hydrology, water quality, and ecology functions.*

## *Predictive Indicators*

Through discussions with City staff and an Advisory Committee established for the project, we identified 27 predictive indicators to best represent urban wetland condition. All datasets were provided by The City of Calgary with the exception of amphibian core habitat, movement pathways that were generated by the Miistakis Institute in partnership with The City of Calgary (Lee et al., 2022), and the Aquatic Vulnerability Index (AVI)[3] derived by Government of Alberta.

To estimate quantitative indicators generated through GIS layers, we buffered each wetland by 500 m and calculated the percent area as a continuous variable. The 500 m buffer represents the mean wetland catchment area within the urban environment based on a catchment spatial layer developed by The City of Calgary to approximate hydrological connection. The City's catchment layer was not used for our analysis due to the quantity of piped infrastructure between catchments and treatment of the larger transportation networks as long linear catchments independent of the neighbouring landscape. As defined, an urban catchment may not represent water movements expected in a more natural catchment. Because we chose to use a 500 m buffer around each wetland, our modelling approach may not represent the full extent of hydrological connection for a wetland or may over-represent neighbouring catchment impact on a wetland.

We used two modelling methods to select predictive indicators. First, we included all 27 predictive indicators as inputs in a neural network to find the best indicators to represent each wetland function (Figure 2; Modelling Framework *A*). Second (Modelling Frameworks *B* and *C*) we asked experts to assign each predictive indicator to one or more of the wetland functions.

We further refined expert assignments of the number of predictive indicators for each function through a correlation analysis. We performed Spearman's rank correlations using the *rcorr* function in the R package Hmisc (Harrell, 2023) to identify highly correlated indicators. Indicators that were highly correlated based on a threshold of +/− 0.60 were considered for removal. To select an indicator for removal we assessed three factors: 1. the currency of the data used to generate the predictive indicator (e.g., impervious surface was recently updated, whereas the land cover dataset used to generate human modified landcover was from 2015); 2. if the indicator was highly correlated in other wetland functions; or 3. if expert opinion suggested the indicator was less important for predicting wetland condition. We used this reduced list of predictive indicators to inform Modelling M Framework *C*.

### *Predictive Approaches*

To explore predictor-response relationships, a combination of machine learning techniques was used. The implemented methodologies included nonlinear machine learning models using Random Forest (RF), Neural Network (NN), and linear and nonlinear regression models such as Generalized Additive Model (GAM).

#### RANDOM FOREST

Random Forest (RF) is an ensemble machine learning method that incorporates the prediction of many experts (trees) into the decision (Fox et al., 2020). We selected this

[3] https://open.alberta.ca/opendata/gda-3a2a8bb2-aaaa-4741-8f4c-bd148bfcbc80

modelling approach for its ability to complete both regression and classification tasks, robustness to outliers or noisy data, and for its ease in measuring an indicator's importance in contributing to the model. RF is used where there is a high number of predictors and complex data. One of the biggest challenges in machine learning is model overfitting, but RFs are less prone to overfitting especially if many trees are included in the model (Breiman, 2001).

### NEURAL NETWORK

Neural Network (NN) is a machine learning method, mimicking human brain neural connections, and is widely used to explore the nonlinear predictor-response relationship (Gholamiangonabadi et al., 2020; Graesser, 2016). The advantage of the NN approach is it can complete more complex tasks than other machine learning methods, and it can process unorganized data (by considering data subsets to find similar organizational patterns). NNs have three layers: input (i.e., in our case predictive indicators and response indicator derived from field-ACI), hidden (i.e., where predictive response relationship is interpreted and processed) and output (i.e., response of predicted wetland condition) (Figure 3).

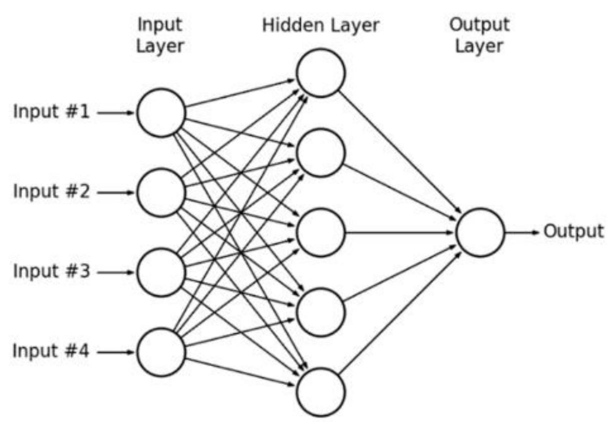



*Figure 3: A simple Neural Network architecture with one hidden layer. Arrows show the connections and circles represent neurons.*

Each neuron (collection of a set of inputs, weights, and activation functions) in the net was processed using the following formula:

$$Net_{\ q} = \sum_{p=1}^{n} w_{pq}\ I_p$$

where $p$ is the sender neuron in the sender (previous) layer and $q$ is the receiver neuron in the next layer. $w_{pq}$ is the connection weight from neuron p in the sender layer to neuron $q$ in the next (receiver) layer and $I_p$ is the signal coming from the sender layer (Yeh and Li, 2003). $Net_q$ is the accumulated signal received in neuron $q$ that will be used in the activation function (often a sigmoidal curve) in the hidden layers, and the activation function value will be passed to the next layer neurons (Yeh and Li, 2003). For the last (output) layer, a linear function is widely used:

$$\frac{1}{1+e^{-Net_q}}$$

NNs are stochastic models. This means that every program run will produce different results because initial weights and parameters are randomly chosen for each iteration. We ran the NNs many times (especially since our training dataset was small) and we used cross-validation techniques (testing and training) to improve model results. For NN, predictive indicators were normalized before training neural networks to avoid initialization of weight with very small numbers (Hagan et al., 2014).

### GENERALIZED ADDITIVE MODELLING

A Generalized Additive Model (GAM) is a linear model that can identify both linear and nonlinear components in the predictors-responses space (Wood, 2006). GAMs use splines that are fitted to the numeric predictor variables. GAM modelling has an advantage of being able to predict outcomes while examining interactions between outcomes and indicators.

To support GAM we first conducted a Principal Component Analysis (PCA) to reduce dataset dimensionality. This is achieved by transforming predictive indicators into smaller datasets that retain the information in the original dataset. The PCA generates dimensions that can be input into the GAM as predictive indicators (see Appendix A for more details).

We fitted a GAM using R package MGVC (Ghosal and Kormaksson, 2019; Wood, 2023) based on five PCA dimensions that explained 70% of the input data and a response variable of the field-ACI values for the 74 surveyed wetlands for each wetland function. We used a specific type of statistical model (Gaussian family), which included both controlled and uncontrolled factors (mixed model). We selected the restricted maximum likelihood (REML) method to estimate model parameters. Additionally, we used REML to help determine the appropriate level of smoothness in the data. This helped us understand if fixed (i.e., the predictive indicators we defined) and random effects (factors influencing wetland condition that we did not define) are important for drawing meaningful conclusions.

### *Predictive Accuracy*

To assess model accuracy, we considered three metrics: $R^2$, MSE and UMBRAE. Considering all three accuracy measures provides a more robust understanding of model performance as each of these measures improves our understanding of different components of model accuracy.

$R^2$, the coefficient of determination, is a statistical metric that evaluates the model goodness of fit and tells us how much variability the model explains. $R^2$ values range between 0 and 1 whereby 1 represents a perfect model fit (City of Calgary, 2004). The advisory committee recommended an $R^2$ of 0.6 as the minimum acceptable threshold for the modelled-ACI to be applied for Calgary's wetland management.

MSE evaluates the difference or magnitude of the prediction error between the predicted and actual values. Better model predictions have smaller MSE values.

UMBRAE is a relatively new measure that evaluates the absolute difference between predicted and actual values divided by the range of observed values (Chen et al., 2017). UMBRAE is less sensitive to outliers, is informative, scale independent, and easy to understand (Aguirre-Larracoechea and Borges, 2021). We compared predicted and actual result values to a naïve model in which all wetlands were given an ACI value of 0.5. The threshold UMBRAE is 1; the departure of UMBRAE from 1 indicated how much better (>1) or worse (<1) the method performed compared to the naïve model presented as a percent, as follows:

- when UMBRAE <1, (1−UMBRAE) × 100 indicates model improvement.
- when UMBRAE >1, (UMBRAE−1) × 100 indicates model worsening.

### *Modelling framework*

To determine the best model for predicting wetland condition we developed a research statistical framework (Table 1).

*Table 1: Research statistical framework*

| | Statistical indicator selection (Machine learning) | | | Heuristic indicator selection (Expert) | | | Heuristic indicator Selection (Expert /PCA process) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Model method** | **R2** | **MSE** | **UMBRAE** | **R2** | **MSE** | **UMBRAE** | **R2** | **MSE** | **UMBRAE** |
| Random Forest | A1 | A1 | A1 | B1 | B1 | B1 | C1 | C1 | C1 |
| Neural Network | A2 | A2 | A2 | B2 | B2 | B2 | C2 | C2 | C2 |
| GAM | n/a | n/a | n/a | n/a | n/a | n/a | C3 | C3 | C3 |

For Modelling Framework *A,* we used a neural network to select predictive indicators for each function and then developed a statistical model using a Random Forest approach (*A1*) and a stochastic model using a Neural Network approach (*A2*). For evaluating model predictive performance, especially when complex models are used, it is important to separate the dataset into training and testing segments. We ran three different combinations of training/testing segments (80/20, 90/10, 100/100).

For Modelling Framework *B,* we used a heuristic (expert assigned) approach to indicator selection for each function, then developed a statistical model using a Random Forest approach (*B1*) and a stochastic model using a Neural Network approach (B2). Response variables were cross-validated for both modelling approaches using 80/20, 90/10 and 100/100 of the data for both training and testing.

For Modelling Framework *C,* we used a heuristic (expert assigned) approach to indicator selection for each function comparing methods with Spearman's correlation and mixed Principal Component Analysis (PCA) tests. Using dimensions from the mixed PCA we developed statistical models with RF (*C1*), and a GAM (*C3*), and stochastic model using a NN approach (*C2*). The response variables (field-ACI values) were cross-validated for all three modelling approaches using 80/20, 90/10 and 100/100 segmentations for both training and testing.

## Modelling Selection and Predicting Wetland Condition

To select the best model for predicting wetland condition, we reviewed $R^2$, MSE and UMBRAE for the three wetland functions.

We used the selected predictive model to determine modelled-ACI values for hydrology, ecology, and water quality for the remaining 2,646 wetlands in the inventory. Modelled-ACI values for each function were normalized to 0 and 1and then averaged to develop an integrated modelled-ACI value for each wetland[4]. We categorized the modelled-ACI into four categories using quantiles to represent very low, low, moderate, and high wetland condition. Results were displayed spatially to identify spatial patterns.

# Results

## Predictive indicators

For the heuristic (expert based) models (*M*odel *F*rameworks B and C) there were 27 *predictive* indicators used to *model* wetland condition (18 for hydrology, 15 for water quality, and 17 for ecology function*)* (*Table 2*).

[4] To normalize the data we used the function (x−min(x))/(max(x)−min(x))

*Table 2: Predictive indicators for three wetland functions, where 1 denotes whether a function (hydrology (H), water quality (WQ) and/or ecology (E)) was considered for each indicator.*

| Indicator | H | E | WQ |
|---|---|---|---|
| Amphibian movement pathway | 0 | 1 | 0 |
| Aquifer Vulnerability Index (AVI) | 1 | 0 | 0 |
| Aspect in 100m buffer | 0 | 0 | 1 |
| Catchment size | 1 | 0 | 0 |
| Channel Connections - surface water connection | 1 | 1 | 1 |
| Core amphibian habitat | 0 | 1 | 0 |
| Elevation relative to the catchment | 1 | 0 | 0 |
| Floodways or riparian area | 1 | 0 | 1 |
| Human modification level - disturbed | 1 | 1 | 1 |
| Human modification level - natural | 1 | 1 | 1 |
| Human modification level - semi natural | 1 | 1 | 1 |
| Impervious area | 1 | 1 | 1 |
| Industrial | 1 | 1 | 1 |
| Nearest wetland | 0 | 1 | 0 |
| Percent wooded (forest and canopy) | 1 | 1 | 1 |
| Perimeter - area ratio | 1 | 1 | 1 |
| Residential/Commercial | 1 | 1 | 1 |
| Roads | 1 | 1 | 1 |
| Slope in 100 m buffer | 0 | 0 | 1 |
| Storm water infrastructure | 1 | 0 | 1 |
| Stream crossings/culverts | 1 | 0 | 0 |
| Trails (gravel, paved and unofficial) | 1 | 1 | 1 |
| Wetland area | 1 | 0 | 0 |
| Wetland coverage/density | 1 | 1 | 0 |
| Within ecological network | 0 | 1 | 0 |
| Within provincial or natural environment park | 0 | 1 | 0 |

For statistical modelling (Model Framework *A*), the 27 predictive indicators and wetland function response variables were evaluated using machine learning (NN) to identify important predictors for the designated response variable. The predictors with higher feature importance (tested models using both 3.5 and 3.7 thresholds for inclusion) derived from NN are outlined in Figure 4 and Table 3. Although we ran models using predictive indicators derived from two thresholds, the best model for prediction was based on the 3.5 threshold.

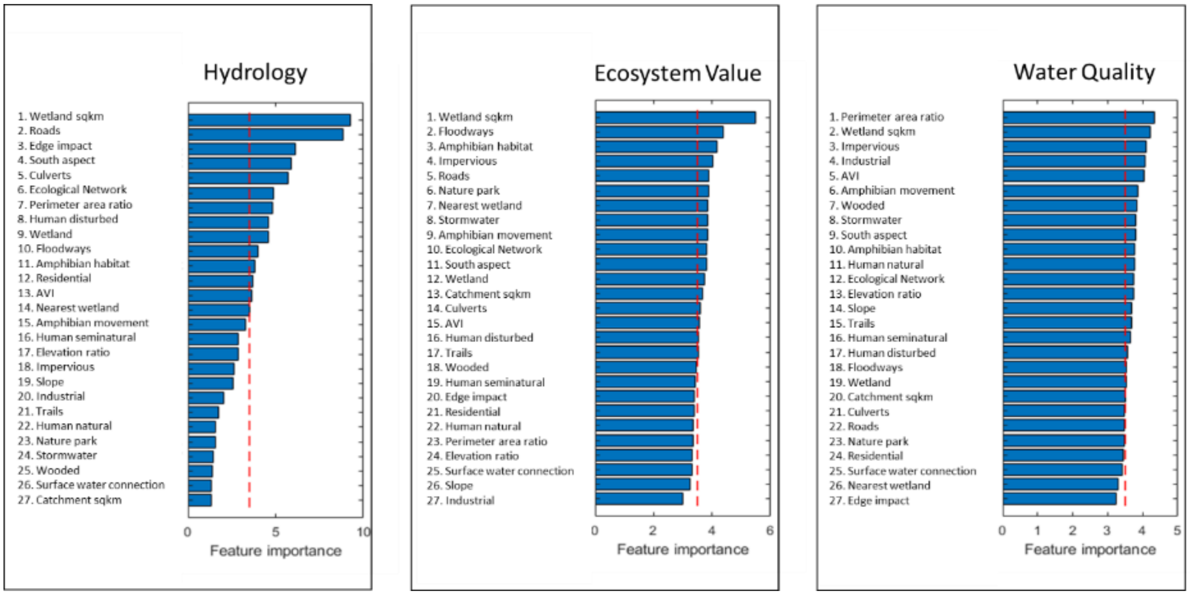



*Figure 4: Statistically derived (Neural Network) predictor indicators. Indicators over a threshold of 3.5 were selected for final model.*

*Table 3: Statistically derived (Neural Network) predictor indicators included in the model for each wetland function.*

| Indicator | H | E | WQ |
|---|---|---|---|
| Amphibian movement pathway | 0 | 1 | 1 |
| Aquifer Vulnerability Index (AVI) | 1 | 1 | 1 |
| Catchment size | 0 | 1 | 0 |
| Channel Connections - surface water connection | 0 | 0 | 0 |
| Core amphibian habitat | 1 | 1 | 1 |
| Edge impact | 1 | 0 | 0 |
| Elevation relative to the catchment | 0 | 0 | 1 |
| Floodways or riparian area | 1 | 1 | 0 |
| Human modification level - disturbed | 1 | 1 | 0 |
| Human modification level - natural | 0 | 0 | 1 |
| Human modification level - semi natural | 0 | 0 | 1 |
| Impervious area | 0 | 1 | 1 |
| Industrial | 0 | 0 | 1 |
| Nearest wetland | 0 | 1 | 0 |
| Percent wooded (forest and canopy) | 0 | 0 | 1 |
| Perimeter area ratio | 1 | 0 | 1 |
| Residential/Commercial | 1 | 0 | 0 |
| Roads | 1 | 1 | 0 |
| Slope in 100 buffer | 0 | 0 | 1 |

| | | | |
|---|---|---|---|
| South aspect in 100 buffer | 1 | 1 | 1 |
| Storm water infrastructure | 0 | 1 | 1 |
| Stream crossings/culverts | 1 | 1 | 0 |
| Trails (gravel, paved and unofficial) | 0 | 0 | 1 |
| Wetland area | 1 | 1 | 1 |
| Wetland coverage/density | 1 | 1 | 0 |
| Within ecological network | 1 | 1 | 1 |
| Within provincial or natural environment park | 0 | 1 | 0 |

## Predictive Model Results

We compared outcomes ($R^2$, MSE, and UMBRAE) for predictive Modelling Frameworks A, B, and C using the three modelling approaches (RF, NN, and GAM) for hydrology, ecology, and water quality functions (Tables 4–6). Raw UMBRAE values are given where lower scores represent better fits. We also calculated how much better our models are when compared to a naïve model. For example, an UMBRAE value of 0.33 indicates the model performed 67% better than a naïve model where all wetlands were assigned a value of 0.5 on an index of 0 to 1.

*Table 4: Performance summary among models trained with different training/testing splits, for hydrology function. Red text depicts models selected for predicting ACI values for Calgary's wetland inventory.*

| **Hydrology Function** | | **Statistical (ML = F1>3.5)** | | | **Heuristic (expert)** | | | **PCA on Heuristics** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Test/Train | $R^2$ | MSE | UMBRAE | $R^2$ | MSE | UMBRAE | $R^2$ | MSE | UMBRAE |
| RF | 80/20 | -0.9 | 0.014 | 0.48 | 0.04 | 0.014 | 0.39 | 0.04 | 0.01 | 0.35 |
| | 90/10 | -0.76 | 0.013 | 0.55 | 0.14 | 0.02 | 0.55 | 0.3 | 0.0005 | 0.25 |
| | 100/0 | 0.81 | 0.003 | 0.28 | 0.8 | 0.003 | 0.29 | 0.82 | 0.002 | 0.15 |
| NN | 80/20 | 0.49 | 0.008 | 0.4 | 0.46 | 0.005 | 0.35 | 0.28 | 0.01 | 0.47 |
| | 90/10 | 0.74 | 0.005 | 0.33 | 0.55 | 0.008 | 0.41 | 0.41 | 0.013 | 0.38 |
| | 100/0 | 0.91 | 0.91 | 0.11 | 1 | 0 | 0.00027 | 0.41 | 0.008 | 0.37 |
| GAM | 80/20 | | | | | | | -0.22 | 0.015 | 0.75 |
| | 90/10 | | | | | | | -0.19 | 0.010 | 0.80 |
| | 100/0 | | | | | | | 0.23 | 0.010 | 0.68 |

*Table 5: The summary of performances of different models, trained with different training/testing splits, for ecology function. Red text depicts the model selected for predicting ACI values for Calgary's wetland inventory.*

| **Ecosystem Function** | | **Statistical (ML = F1>3.5)** | | | **Heuristic (expert)** | | | **PCA on Heuristics** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Test/Train | $R^2$ | MSE | UMBRAE | $R^2$ | MSE | UMBRAE | $R^2$ | MSE | UMBRAE |
| RF | 80/20 | 0.03 | 0.016 | 0.43 | 0.26 | 0.005 | 0.19 | 0.01 | 0.011 | 0.32 |
| | 90/10 | 0.07 | 0.011 | 0.39 | 0.65 | 0.002 | 0.18 | 0.09 | 0.007 | 0.25 |
| | 100/0 | 0.83 | 0.002 | 0.14 | 0.78 | 0.002 | 0.16 | 0.79 | 0.002 | 0.16 |
| NN | 80/20 | 0.7 | 0.004 | 0.19 | 0.59 | 0.006 | 0.2 | 0.42 | 0.006 | 0.21 |
| | 90/10 | 0.72 | 0.006 | 0.23 | 0.64 | 0.002 | 0.16 | 0.39 | 0.009 | 0.3 |
| | 100/0 | 0.97 | 0.0002 | 0.05 | 0.92 | 0.0008 | 0.22 | 0.58 | 0.004 | 0.19 |
| GAM | 80/20 | | | | | | | -0.45 | 0.012 | 0.62 |
| | 90/10 | | | | | | | 0.12 | 0.006 | 0.61 |
| | 100/0 | | | | | | | 0.33 | 0.007 | 0.54 |

*Table 6: The summary of performances of different models, trained with different training/testing splits, for water quality function. Red text represents the model selected for predicting ACI values for Calgary's wetland inventory.*

| **Water Qual. Function** | | **Statistical (ML = F1>3.5)** | | | **Heuristic (expert)** | | | **PCA on Heuristics** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Test/Train | $R^2$ | MSE | UMBRAE | $R^2$ | MSE | UMBRAE | $R^2$ | MSE | UMBRAE |
| RF | 80/20 | -0.27 | 0.014 | 0.51 | -0.15 | 0.01 | 0.49 | 0 | 0.01 | 0.34 |
| | 90/10 | -0.05 | 0.01 | 0.54 | -0.42 | 0.012 | 0.68 | -0.2 | 0.009 | 0.32 |
| | 100/0 | 0.82 | 0.001 | 0.18 | 0.78 | 0.002 | 0.19 | 0.78 | 0.002 | 0.16 |
| NN | 80/20 | 0.47 | 0.003 | 0.22 | 0.5 | 0.003 | 0.24 | 0.27 | 0.005 | 0.29 |
| | 90/10 | 0.64 | 0.002 | 0.19 | 0.51 | 0.008 | 0.31 | 0.39 | 0.005 | 0.27 |
| | 100/0 | 0.95 | 0.0004 | 0.07 | 1 | 0 | 0.012 | 0.45 | 0.004 | 0.24 |
| GAM | 80/20 | | | | | | | 0.06 | 0.008 | 0.67 |
| | 90/10 | | | | | | | -0.51 | 0.011 | 0.61 |
| | 100/0 | | | | | | | 0.17 | 0.006 | 0.59 |

Overall, Modelling Framework *C* resulted in weaker $R^2$ values (< 0.5), and lower predictive ability based on poorer MSE and UMBRAE values. Modelled ACI values did not meet the desired threshold of an $R^2$ > 0.60 that we set for management use.

Predictive models (100 training and testing) fitted by RF (Modelling Framework *A1* and *B1*) resulted in $R^2$ values above the 0.60 threshold and high predictability based on UMBRAE values. In addition, the statistical and heuristic approach to indicator selection provided similar predictability based on UMBRAE values.

Predictive models for all three ecosystem functions fitted by NN (modelling framework *A2*) using 100 training and testing cross validation runs resulted in $R^2$ values above the 0.60 threshold and high predictability based on UMBRAE values. The statistical approach to indicator selection provided better predictability based on $R^2$, MSE and UMBRAE values. However Neural Networks are often associated with overfitting, especially when large-scale data is used (Shen and Lin, 2022). Overfitting happens when the model is too sensitive to small variation or fluctuations in the training dataset. One way to avoid overfitting is to stop the training early (Prechelt, 1998; Ying, 2019). This technique prevents the model from learning more than what is expected from a training dataset and stopping the model when it starts to degrade, avoiding overfitting. To avoid overfitting, we considered NN for 90/10 and 80/20testing training datasets for predictive modelling. The NN approach based on 90% training and 10% testing with a statistical indicator selection approach and feature importance threshold of 3.5 was selected as the best prediction model.

We plotted the Modelling Framework *A2* actual *vs.* predicted ACI values from the NN for the 74 wetlands based on the three functions for the selected modelling approach (Figure ).

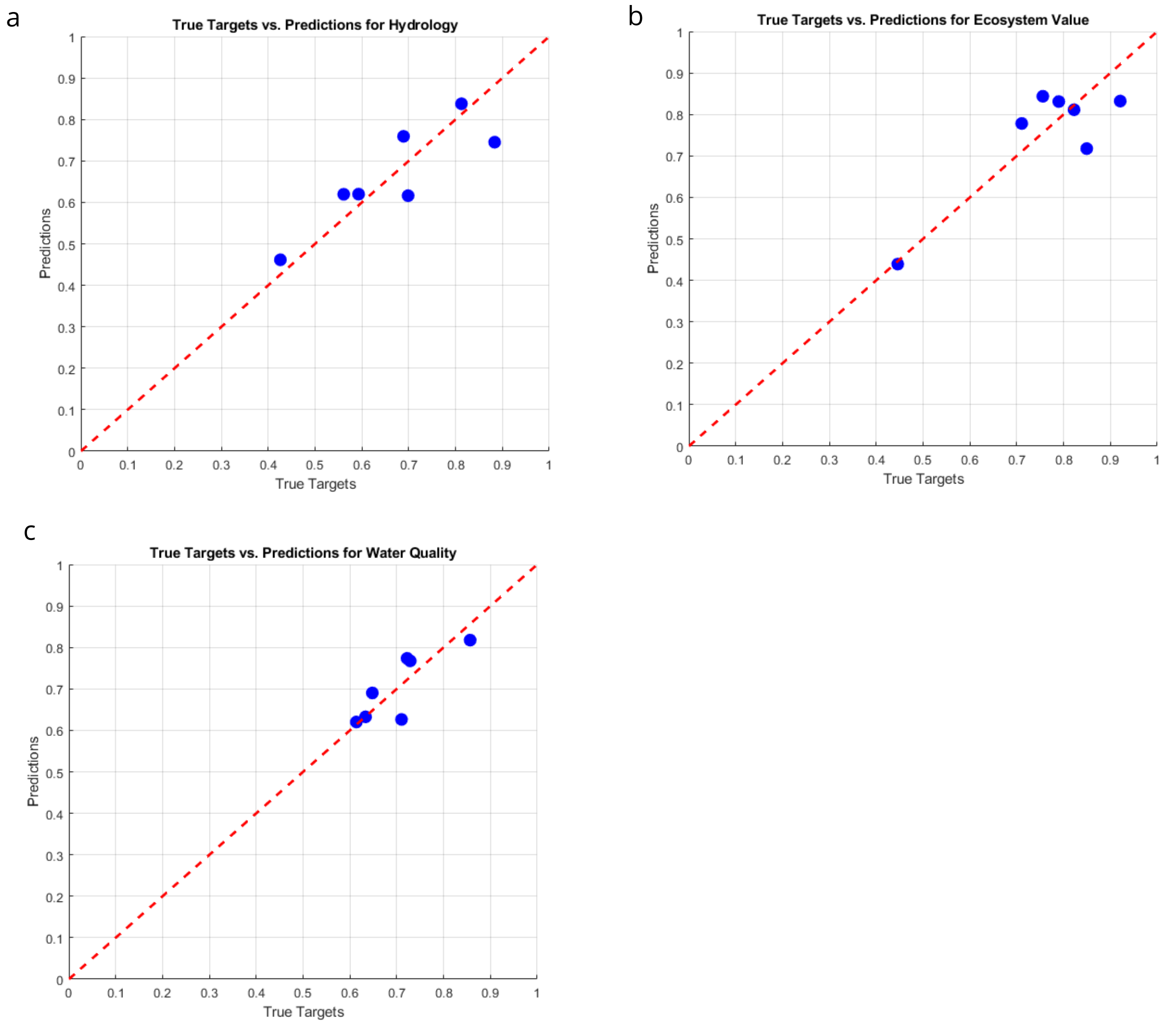



Figure 5: Modelled ACI predictive values compared to field ACI values (true target) for a. hydrology, b. ecology and c. water quality. The red-dashed line is the 1:1 fit.

Predictive indicators importance to the NN approach is plotted for hydrology, ecology, and water quality (Figure ).

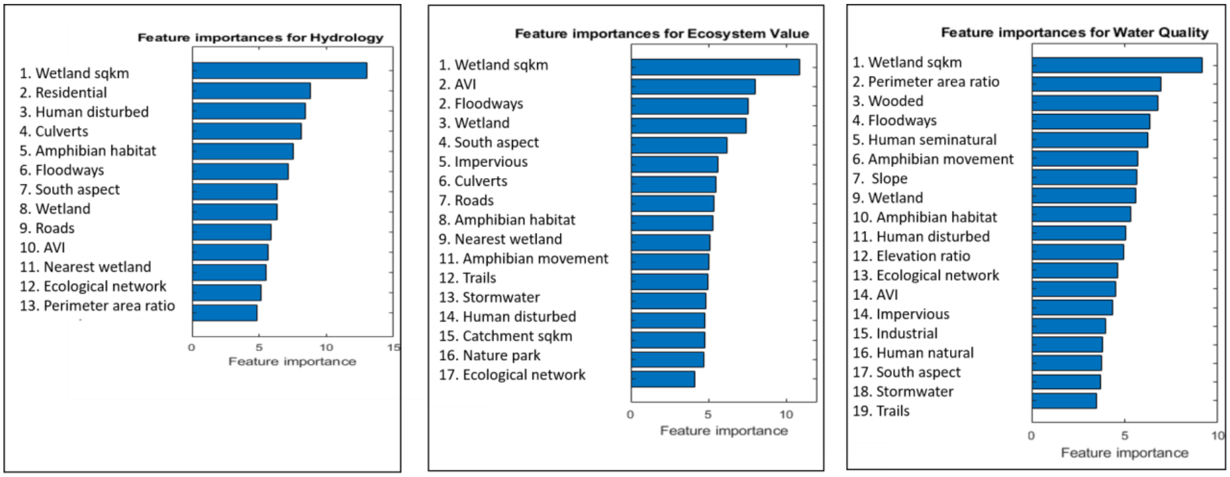



*Figure 6: Neural Network predictive indicators for hydrology, ecology, and water quality functions.*

## Modelled ACI Values

We mapped modelled-ACI values derived from Modelling Framework *A2* (NN, 90/10, stochastic predictive indicator selection) and found a spatial pattern of lower wetland condition values for inner city wetlands and along major transportation corridors, and higher ACI values within natural or non-urbanized areas (Figure 8). See Appendix B for modelled-ACI values for each wetland function.

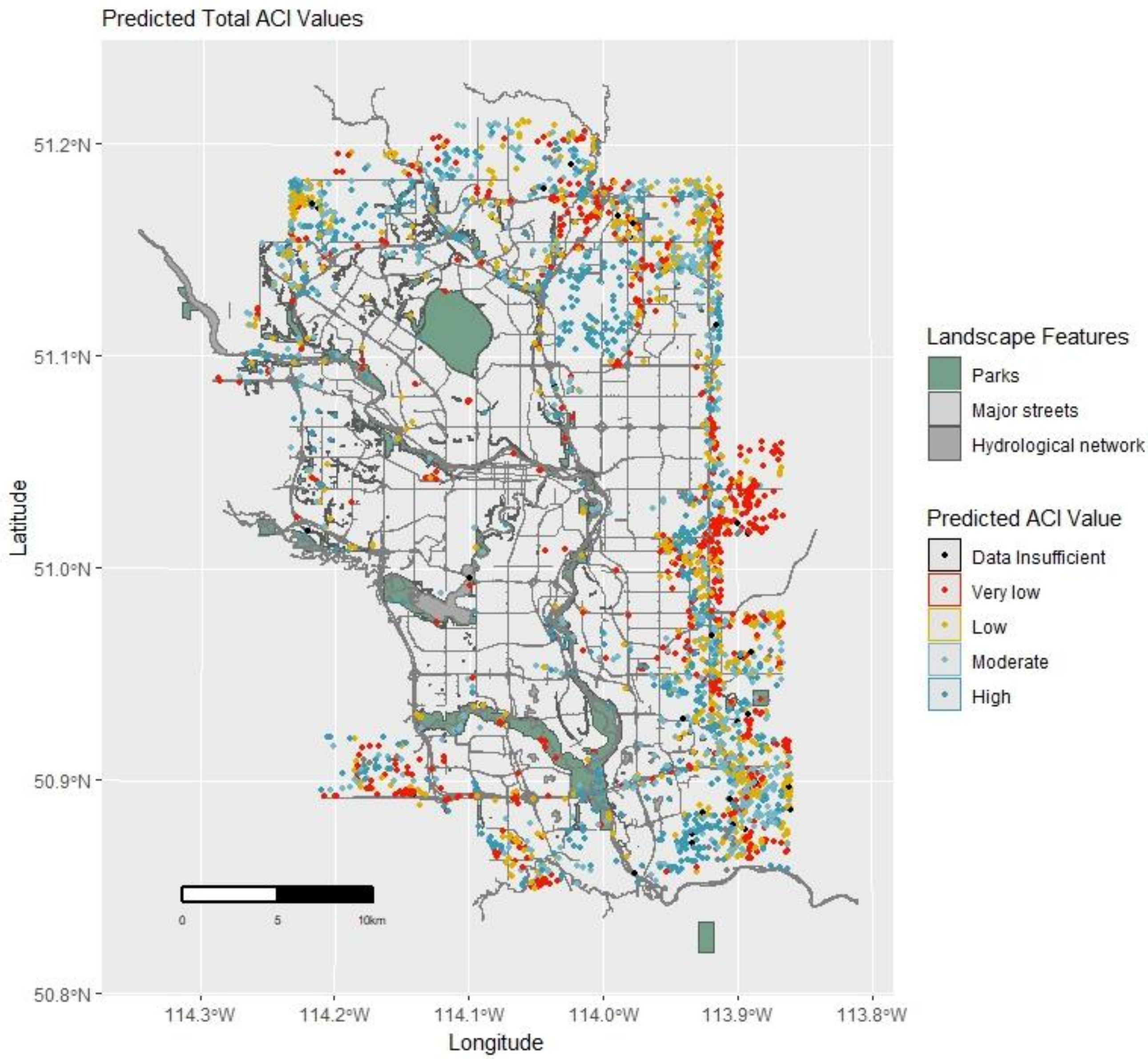



*Figure 8: Wetland modelled-ACI values across a gradient from very low (red dots) to high (blue dots) value in Calgary.*

## Natural Environment Parks

We selected the 230 wetlands occurring in Natural Environment Parks (NEP) managed by The City of Calgary and identified wetlands with very low modelled-ACI values (Figure ). The ecology function had lower modelled-ACI values than both hydrology and water quality (Figure  and maps in Appendix B). For example, of the NEP wetlands, 73% were in very low to low condition categories for ecology function.

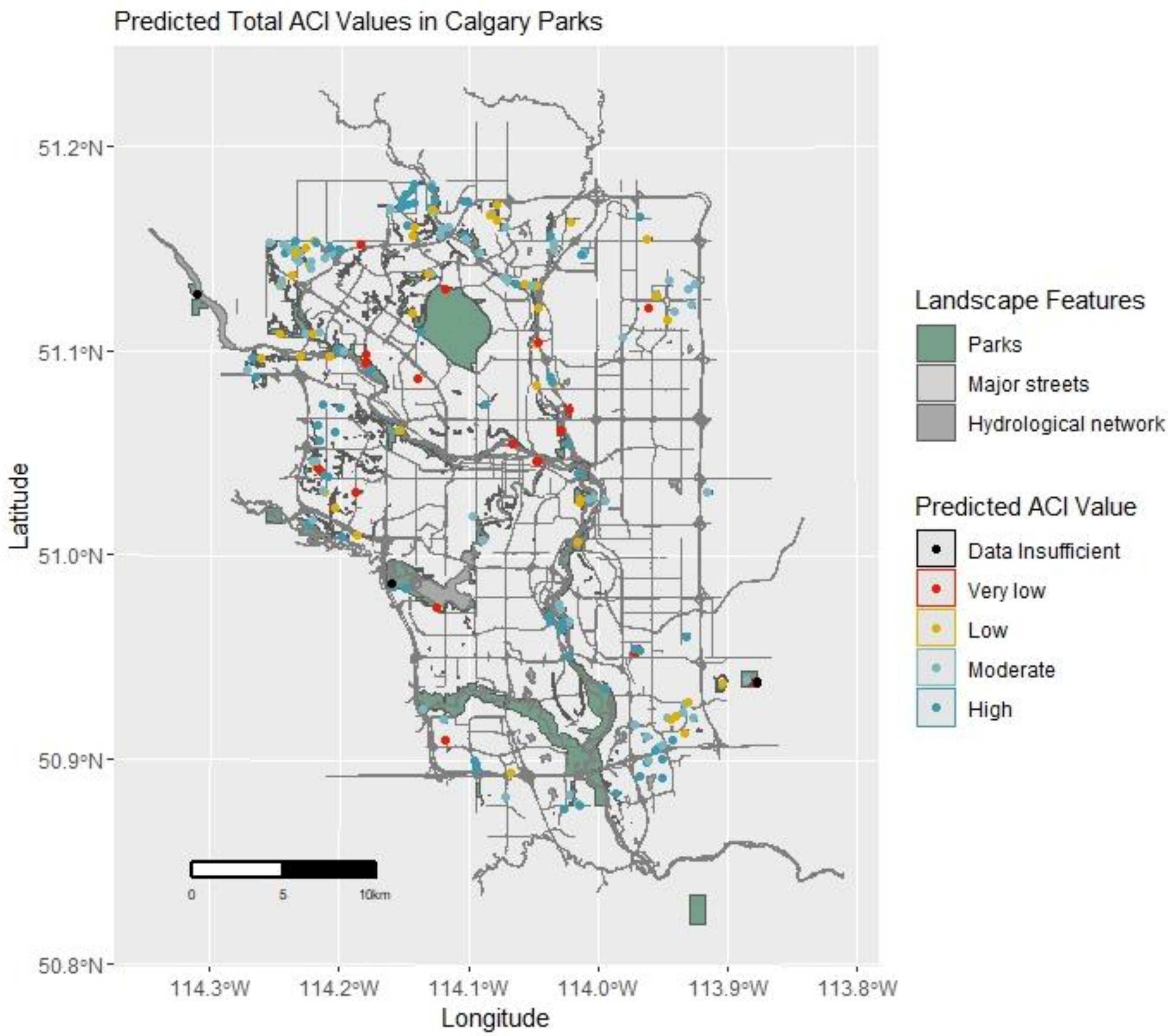



*Figure 9: Modelled ACI values for Natural Environment Parks, ranging from very low (red dots) to high (blue dots) value.*

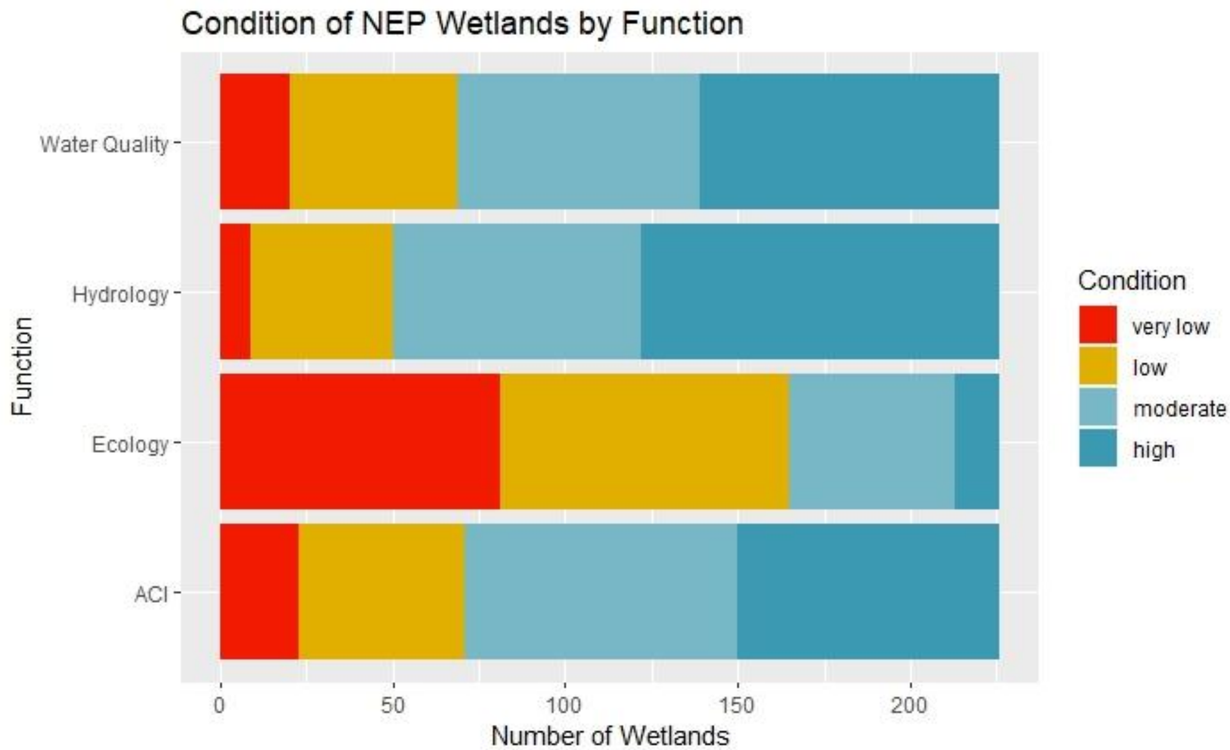



*Figure 10: Number of wetlands in Natural Environment Parks (NEP) per prediction category for hydrology, ecology, and water quality functions and modelled ACI. Four wetlands had insufficient data and were not given a modelled condition value.*

# Discussion

We designed a modelled Aquatic Condition Index (modelled-ACI) for urban wetlands in Calgary. The modelled-ACI was based on estimates of condition generated in the field for hydrology, ecology, and water quality functions. We generated 27 prediction-based indicators to depict wetland condition. Using three modelling approaches (Random Forests (RF), Neural Networks (NN), and Generalized Additive Models (GAM)), we assessed model predictability using $R^2$, MSE and UMBRAE. Here we describe model selection to predict wetland condition, discuss management implications of the resulting modelled-ACI wetland values, outline model limitations, and recommend next steps for The City of Calgary.

## Model Selection

We found that all three wetland functions (hydrology, water quality, and ecology) were more reliably predicted using a machine learning approach (RFs and NNs) as opposed to the more traditional statistical approach (GAMs). We showed that the statistical approach to indicator selection resulted in better performing models than the heuristic approach for

NNs but resulted in similar performing models for RFs. The modelling frameworks that used 100% training data consistently resulted in higher predictability; however, we discarded these from consideration since they are prone to overfitting. Taking a balanced approach maximizing predictability while limiting overfitting, we selected Modelling Framework *A2* as the most appropriate to predict wetland condition; *A2* used an NN and statistically selected indicator and a split of 90% training data and 10% test data.

GAM has many strengths, such as ease of interpretability between predictor and response variables, ability to shrink coefficients of less important variables toward zero, and robustness to data outliers. In our process, however, GAMs did not perform well, with $R^2 < 0.60$ and UMBRAE values for all three functions predicting only 45% (or less) better than a naïve model. We deemed the GAM modelled-ACI results as inadequate for management purposes, and it was not selected to predict wetland condition. The lower $R^2$ value indicates the GAM was not able to explain most of the variation in model inputs. This could be due to our use of Principal Component Analysis (PCA) to reduce the large number of predictor indicators variables into dimensions that can result in a loss of information. We carried forward a large number of dimensions, which can result in issues with GAMs. In addition, it is possible that the relationship between predictors is complex and nonlinear and is not represented well with GAM smoothed functions.

The RFs with 100% training and no testing met the required threshold ($R^2 > 0.6$) as did both the statistical (Modelling Framework *A1*) and heuristic (Modelling Framework *C1* and *B1*) models ($R^2 > 0.6$). Both had good predictive accuracy values. However, we did not select Modelling Framework *C1* as its performance was poorer when we used training and testing datasets. To prevent overfitting in machine learning techniques we preferred to select a model that used training and testing datasets. In addition, while RFs provide a robust modelling framework, they cannot predict beyond the range of data in the training dataset (Fox et al., 2020; Meyer and Pebesma, 2021; Takoutsing and Heuvelink, 2022). As such, all possible range of values in the training dataset, including extreme field conditions, should be included. However, we could not ensure that extreme condition sites were among the 74 wetlands selected for field surveys (Nwaishi, 2023). It is likely that, if we increased the number of wetland field surveys, our range of what is poor to high would grow. We therefore did not select an RF model to predict wetland condition.

The NN for Modelling Framework *C2* did not perform well, potentially due to noise in the PCA derived elements. It is possible the transformation of predictive indicators to PCA dimensions made the NN training difficult because the model was trying to fit already transformed predicted indicators to the response variables.

## Management Applications

The selected Modelling Framework *A2* (NN using 90% training and 10% testing dataset) was used to predict ACI values for Calgary's urban wetland inventory. Overall, the ecology function model was better associated with its predictor indicators, while the water quality function had the poorest association with its predictor indicators. This is likely because many of the predictive indicators for the ecology function are designed to be sensitive to ecological condition. For example, change in roads, residential development, or wooded habitat can be directly measured and directly impact ecology condition. For the water quality function, the predictive indicators were not a direct measure of water quality, rather are indirect influencers of water quality. It was therefore expected that water quality models would not perform as well.

Spatial patterns of modelled ACI values for the hydrology function indicate poorer condition at inner-city wetlands and in non-urbanized areas such as agricultural and industrial land to the east and south (Figure 17, appendix C). Wetland condition values were higher in parks and in non-urban areas in the southern portion of Calgary from Priddis slough to South Seton. Strong predictors of hydrology function included wetland size, the percent of human features such as residential developments, human disturbance and culverts, presence of amphibian core habitat (Lee et al., 2022), or floodways (riparian areas).

Spatial patterns of modelled-ACI values indicate ecology function had poorer condition along roads and the inner city and better condition in the non-urbanized areas, as expected (Figure 18, appendix C). Strong predictive indicators for the ecology function included wetland size and the percent of natural features around a wetland (e.g., whether the wetland is in a floodway, in an area with high wetland density, or if the wetland has a southerly aspect). Wetland loss in Calgary has been high and remaining wetlands tend to be located along river edges (floodways), in natural areas or in non-urbanized areas.

Spatial patterns of modelled-ACI values of the water quality function indicate poorer wetland condition along roads, north of the airport, in the northwestern corner of the city, and on industrial and agricultural area on the eastern side of Calgary (Figure 19, appendix C). Predictive indicators of the water quality function included wetland size, perimeter to area ratio, and natural features around a wetland including if the wetland occurred near wooded areas, within a floodway, or a semi-natural area. Wetlands more associated with natural features would be expected to have higher ACI values for water quality.

Overall, modelled-ACI values indicate a spatial pattern of poorer condition in the inner city, along roads, or in the eastern portion of the study area on agriculture and industrial areas. The modelling results provide guidance and identify potential areas where wetland protection or restoration could be further explored. For example, modelled-ACI results are being used in an ecology prioritization process to identify wetlands where protection or restoration could be considered. Calgary continues to develop into non-urbanized lands where wetlands are prevalent, and the modelled-ACI can provide direction on where wetland complexes are in better condition. These areas could be considered in land use planning or zoning to support biodiversity strategies to retain natural wetlands and connections among them. Wetlands prioritized for restoration can then be assessed using the field-ACI methods to evaluate management actions for improving wetland condition.

With over 2,700 wetlands in the city, identifying areas of focus is an important management tool.

The modelled-ACI values could also be used to strategically support existing or new policies related to wetland protection and restoration. For example, modelling results indicate wetlands along major road rights of way exhibited lower modelled-ACI values. Wetlands in green spaces along road rights of way play an important role in biodiversity as there are limited green spaces in urban areas. Therefore, improving wetland ecology function at wetlands along roadsides may be an important biodiversity strategy. What are the policy drivers that influence wetland construction, modification of a natural wetland or wetland management actions along road rights of ways? Results could indicate the need for policy change and inform discussion.

Lastly, the modelled-ACI results can be used by specific city departments as indicators to track wetland conditions over time. For example, a condition value can be generated for each Natural Environment Park with more than 10% wetlands or for the entire Calgary park system as a city-wide or individual-park indicator. As well, different business units can focus on their issues by selecting the specific functions of interest. For example, Water Services may want to look at the water quality function to track wetland water quality over time.

## Modelling limitations and recommendations

A key challenge in this work is to compare urban wetlands to a natural reference condition. In our assessment, wetland values were classed relative to scores generated for the full inventory. A low wetland classification is low relative to the full wetland inventory in the city, but this does not necessarily identify the actions that are required to improve wetland condition. To better understand the modelled-ACI values and their application to management actions we recommend field visits to a range of modelled-ACI values to verify where, and what, interventions need occur.

Another limitation is our ability to understand the relationships among predictive indicators that derive the modelled-ACI, limiting our ability to determine how a predictive indicator influences condition values. The lack of a clear, interpretable relationship is a limitation of the NN and interpretation of predictive indicators and modelled-ACI values could potentially be improved by using additional modelling frameworks that assess predictive indicator relationships. In our analysis, other methods did not perform as well, but this may improve with the addition of more field-ACI data.

Another challenge for informing management actions is assigning factors responsible for low condition values to predictive indicators that are difficult to address (i.e., built infrastructure). There are likely wetlands with low modelled-ACI values where management is unlikely to create improvements due to the permeance of urbanization. Therefore, field visits to very low and low value wetlands may be necessary to identify if there are opportunities to improve condition scores.

Wetlands in the urban environment span a gradient from natural to fully constructed storm ponds, and within the urban landscape, wetlands are managed primarily for stormwater control. The City of Calgary is in the process of finalizing the development of a new wetland typology that outlines management goals for wetland types (e.g., natural, modified natural wetland, constructed storm pond). The typology has not yet been applied to Calgary's wetland inventory and so we were not able to consider our modelled-ACI results in relation to City wetland typologies. Future iterations of the modelled-ACI could include wetland management typologies, which would inform the feasibility of management actions relative to wetland types. For example, management actions such as dredging, changes to wetland shape, size, or slope, and/or addition of piped infrastructure could impact condition values. But if the wetland's intended function is to manage stormwater, there may be limited opportunity to improve condition. Therefore, adding a wetland typology attribute to the wetland inventory could help refine a prioritized list of wetland sites where improvements to condition are considered.

We developed the modelled-ACI for a small number (74 sites) of field assessed wetlands. Future ACI field surveys should increase the number of responses and enable refined modelling. Wetland functions were considered using 15 to 18 GIS-based indicators and ideally for each indicator 10 response variables or a total of 180 surveys could be included to better inform predictive modelling. We recommend that The City undertake additional field-ACI surveys and, once a more robust response dataset is available, refine the predictive model. This process should include further testing of modelling approaches. Doing this will permit a better understanding of how each predictor influences wetland condition which will be important to understand integrated wetland function.

## References


Aguirre-Larracoechea, U., Borges, C.E., 2021. Imputation for repeated bounded outcome data: Statistical and machine-learning approaches. Mathematics 9, 1–27. https://doi.org/10.3390/math9172081

Alikhani, S., Nummi, P., Ojala, A., 2021. Urban wetlands: A review on ecological and cultural values. Water (Switzerland) 13, 1–47. https://doi.org/10.3390/w13223301

Ampatzidis, P., Kershaw, T., 2020. A review of the impact of blue space on the urban microclimate. Sci. Total Environ. 730, 139068. https://doi.org/10.1016/j.scitotenv.2020.139068

Chen, C., Twycross, J., Garibaldi, J.M., 2017. A new accuracy measure based on bounded relative error for time series forecasting. PLoS One 12, 1–23. https://doi.org/10.1371/journal.pone.0174202

City of Calgary, 2004. Calgary Wetland Conservation Plan. Calgary, AB. https://doi.org/10.1002/ejoc.201200111

Creed, I.F., Aldred, D.A., Serran, J.N., Accatino, F., 2018. Maintaining the Portfolio of Wetland Functions on Landscapes: A Rapid Evaluation Tool for Estimating Wetland Functions and Values in Alberta, Canada, Wetland and Stream Rapid Assessments. Elsevier Inc. https://doi.org/10.1016/b978-0-12-805091-0.00027-x

Fox, E.W., Ver Hoef, J.M., Olsen, A.R., 2020. Comparing spatial regression to random forests for large environmental data sets. PLoS One 15, 1–22. https://doi.org/10.1371/journal.pone.0229509

Ghosal, I., Kormaksson, M., 2019. plsmselect: Linear and Smooth Predictor Modelling with Penalisation and Variable Selection.

Harrell, J.F., 2023. Hmisc: Harrell Miscellaneous.

Kentula, M.E., Gwin, S.E., Pierson, S.M., 2004. Tracking changes in wetlands with urbanization: Sixteen years of experience in Portland, Oregon, USA. Wetlands 24, 734–743. https://doi.org/10.1672/0277-5212(2004)024[0734:TCIWWU]2.0.CO;2

Kometa, S.S., Kimengsi, J.N., Petiangma, D.M., 2017. Urban Development and its Implications on Wetland Ecosystem Services in Ndop, Cameroon. Environ. Manag. Sustain. Dev. 7, 21. https://doi.org/10.5296/emsd.v7i1.12141

Lee, T.S., Randall, L.A., Kahal, N.L., Kinas, H.L., Carney, V.A., Rudd, H., Baker, T.M., Sanderson, K., Creed, I.F., Moehrenschlager, A., Duke, D., 2022. A framework to identify priority wetland habitats and movement corridors for urban amphibian conservation. Ecol. Solut. Evid. 3, 1–15. https://doi.org/10.1002/2688-8319.12139

McLaughlin, D.L., Cohen, M.J., 2013. Realizing ecosystem services: Wetland hydrologic function along a gradient of ecosystem condition. Ecol. Appl. 23, 1619–1631. https://doi.org/10.1890/12-1489.1

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12, 1620–1633. https://doi.org/10.1111/2041-210X.13650

Nwaishi, F., 2023. Aquatic Condition Index Field Testing and Sensitivity Analysis. Calgary, AB.

Nwaishi, F., Dennett, J., Lee, T.S., Allison, A., Bartlett, K., Kinas, H., Duke, D., 2023. Aquatic

Health Indicator Manual for the City of Calgary. Calgary, AB.

Prechelt, L., 1998. Early Stopping - But When?, in: Orr, G.B., Müller, K. (Eds.), Tricks of the Trade. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Germany.

Shen, X., Lin, J., 2022. Consistency of Neural Networks with Regularization 1–26.

Takoutsing, B., Heuvelink, G.B.M., 2022. Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. Geoderma 428, 116192. https://doi.org/10.1016/j.geoderma.2022.116192

Wood, S., 2023. Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.

Wood, S.N., 2006. Generalized Additive Models: An Introduction With R., Second. ed. CRC Press., Boca Raton, FL.

Ying, X., 2019. An Overview of Overfitting and its Solutions. J. Phys. Conf. Ser. 1168. https://doi.org/10.1088/1742-6596/1168/2/022022

# Appendix A: Modelling Framework C – GAM

## *Testing for correlations*

Spearman's rank correlation matrices for the hydrology (Figure1), ecology (Figure ), and water quality (Figure 13).

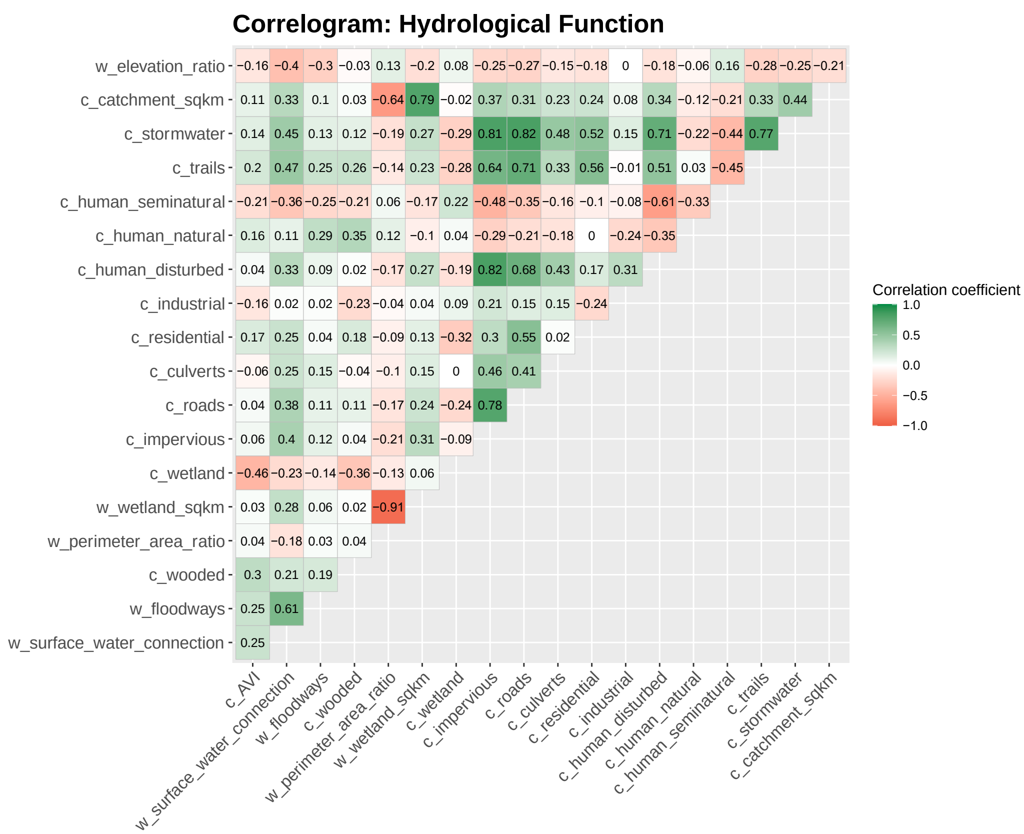



*Figure 11: Hydrology function: Spearman's rank correlation matrix, where 1.0 is a strong positive correlation, −1.0 is a strong negative correlation and 0 is no correlation.*

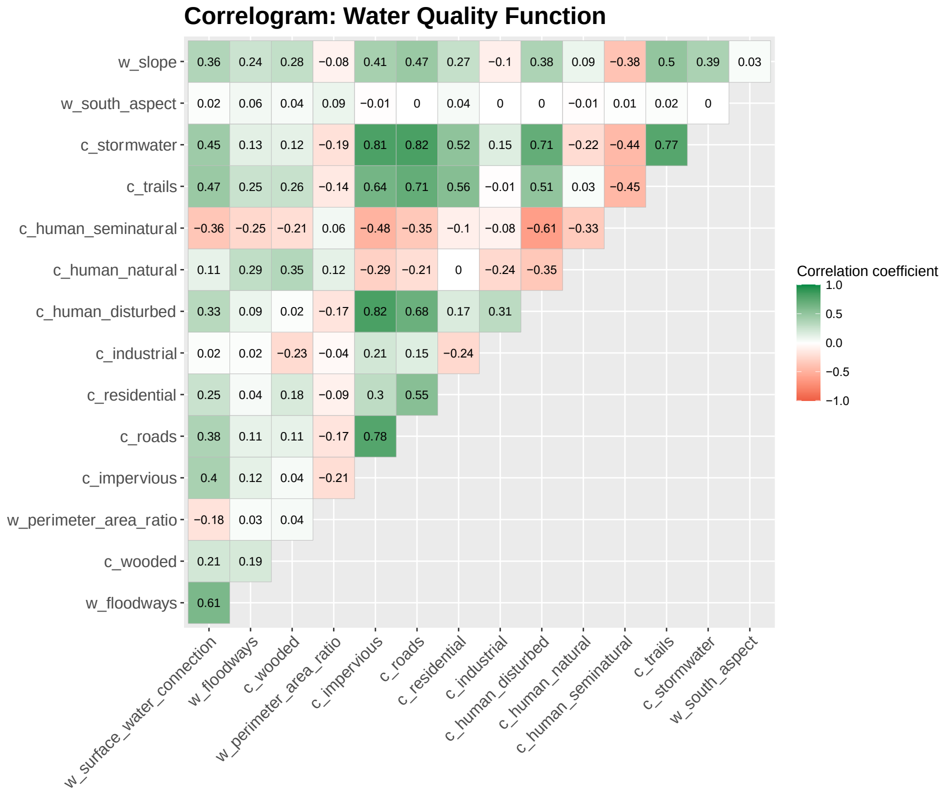



*Figure 12: Water quality function: Spearman's rank correlation matrix.*

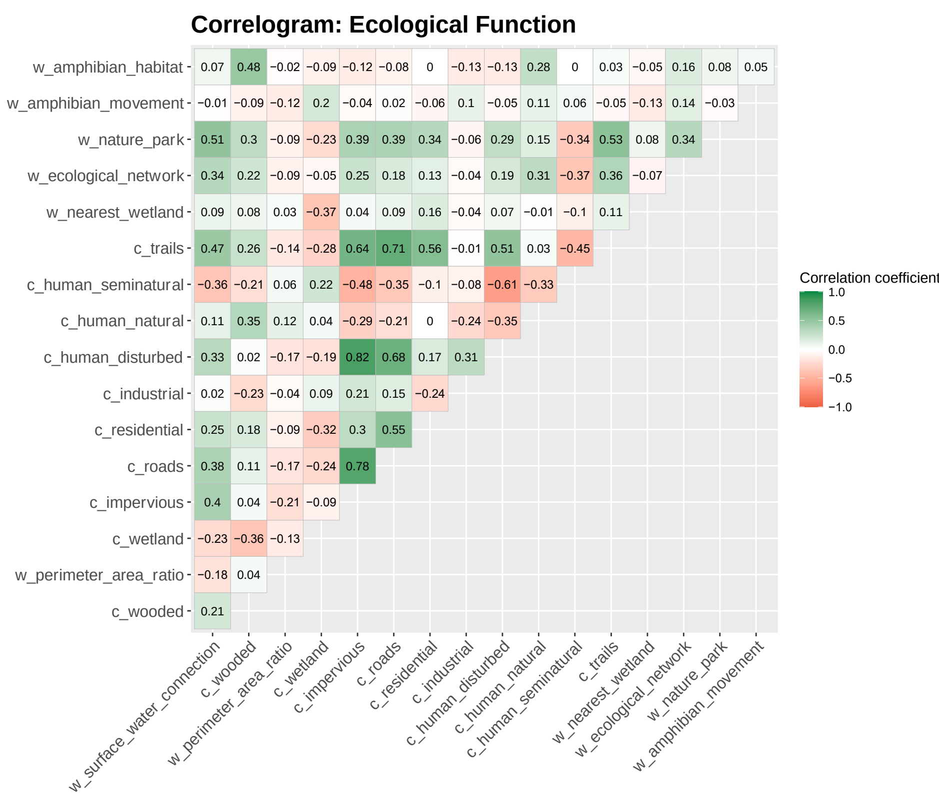



*Figure 13: Ecology function: Spearman's rank correlation matrix.*

The correlation testing process led to the removal of six indicators for the hydrology function, five indicators for the water quality function, and three indicators for the ecology function (Table 7). Trail, road, and disturbed human modification level indicators were removed for all three functions. The trails indicator was strongly and positively correlated to roads, impervious area, and stormwater infrastructure. The roads indicator was strongly and positively correlated to the disturbed human modification level, trails, impervious area, and stormwater infrastructure indicators. Finally, the disturbed human modification indicator was strongly and positively correlated to impervious area, roads, and stormwater infrastructure.

Additionally, surface water connection, perimeter-area ratio, and stormwater infrastructure were removed for the hydrology function, and surface water connection and storm water infrastructure were removed for the water quality function. Stormwater infrastructure was strongly and positively correlated to the following indicators: impervious area, roads, disturbed human modification level, and trails. Perimeter-area ratio was strongly and negatively correlated with catchment size and wetland area.

The remaining indicators were used for model streams B and C.

*Table 7 - Highly correlated indicators removed for each function at the +/− 0.60 threshold.*

| **Function** | **H** | **WQ** | **E** |
|---|---|---|---|
| **Indicators removed** | Trails (gravel, paved and unofficial) | Trails (gravel, paved and unofficial) | Trails (gravel, paved and unofficial) |
| | Roads | Roads | Roads |
| | Channel Connections - surface water connection | Channel Connections - surface water connection | Human modification level – disturbed |
| | Perimeter - area ratio | Human modification level – disturbed | |
| | Storm water infrastructure | Storm water infrastructure | |
| | Human modification level – disturbed | | |

## Principal Component Analysis Results

For modelling stream C the results of the PCA led to the selection of the first five dimensions to be used in the GAM modelling process. The optimal minimum threshold for the percentage of variation represented by the dimensions was 70%, and the fifth dimension represented 65.78% of the variation for the hydrology function, 71.0% for the water quality function, and 60.69% for the ecology function (Table 8).

*Table 8 - Proportion of variation represented by the first five dimensions for each function.*

| **Dimension (rotated)** | Proportion of data represented (%) | | |
|---|---|---|---|
| | **H** | **WQ** | **E** |
| Dimension 1 | 15.79 | 21.60 | 17.48 |
| Dimension 2 | 14.27 | 15.90 | 15.10 |
| Dimension 3 | 12.30 | 13.27 | 10.54 |
| Dimension 4 | 11.44 | 10.21 | 10.17 |
| Dimension 5 | 11.98 | 10.02 | 7.40 |
| Total Proportion | 65.78 | 71.0 | 60.69 |

### *Indicator Contributions to PCA Dimensions*

The results of the PCA process allowed for the identification of the indicators that have the greatest influence on the five dimensions for each function. The contributions of these indicators to the hydrology function PCA dimensions are displayed in Table 9. These results indicate that the three variables that contribute most to the first dimension, which accounts for the most variation in the hydrology function, are human modification level – natural, human modification level – semi natural, and percent wooded (Table 9). The indicators with the largest contributions to the second dimension for this function are wetland area, catchment size, and wetland coverage (Table 9).

*Table 9 - Hydrology function: Indicators contributions to each dimension*

| **Indicator** | Contribution to Dimension | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| Human modification level - natural | 0.71 | 0.00 | 0.08 | 0.01 | 0.01 |
| Human modification level - semi natural | 0.62 | 0.00 | 0.22 | 0.00 | 0.00 |
| Percent wooded (forest and canopy) | 0.35 | 0.00 | 0.00 | 0.02 | 0.08 |
| Aquifer Vulnerability Index | 0.16 | 0.00 | 0.00 | 0.12 | 0.33 |
| Floodways or riparian area | 0.09 | 0.00 | 0.01 | 0.01 | 0.49 |
| Wetland coverage | 0.05 | 0.07 | 0.03 | 0.29 | 0.02 |
| Stream crossings/culverts | 0.04 | 0.00 | 0.34 | 0.06 | 0.05 |
| Elevation relative to the catchment | 0.02 | 0.01 | 0.01 | 0.00 | 0.55 |

| | | | | | |
|---|---|---|---|---|---|
| Residential/Commercial | 0.01 | 0.00 | 0.02 | 0.71 | 0.00 |
| Industrial | 0.01 | 0.00 | 0.13 | 0.23 | 0.03 |
| Wetland area | 0.00 | 0.89 | 0.00 | 0.00 | 0.00 |
| Impervious area | 0.00 | 0.01 | 0.74 | 0.04 | 0.00 |
| Catchment size | 0.00 | 0.86 | 0.02 | 0.00 | 0.01 |

The contributions of these indicators to the water quality function dimensions are displayed in Table 10. The three variables that contribute most to the first dimension, which accounts for the most variation in the water quality function, are human modification level – natural, human modification level – semi natural, and percent wooded (Table 10). The indicators with the largest contributions to the second dimension for this function are impervious area, human modification level – semi natural, and slope in 100 meter buffer (Table 10).

*Table 10 - Water quality function: Indicators contributions to each dimension*

| **Indicator** | Contribution to Dimension | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| Human modification level - natural | 0.72 | 0.07 | 0.00 | 0.00 | 0.00 |
| Human modification level - semi natural | 0.46 | 0.24 | 0.01 | 0.01 | 0.00 |
| Percent wooded (forest and canopy) | 0.43 | 0.00 | 0.04 | 0.01 | 0.00 |
| Floodways or riparian area | 0.28 | 0.02 | 0.00 | 0.05 | 0.00 |
| Slope in 100 m buffer | 0.21 | 0.19 | 0.12 | 0.01 | 0.00 |
| Residential/Commercial | 0.03 | 0.14 | 0.5 | 0.00 | 0.00 |
| Industrial | 0.03 | 0.06 | 0.65 | 0.00 | 0.00 |
| Perimeter - area ratio | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Impervious area | 0.00 | 0.87 | 0.00 | 0.00 | 0.00 |
| Aspect in 100m buffer | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 |

The contributions of these indicators to the ecology function dimensions are displayed in Table 11. The three variables that contribute most to the first dimension, which accounts for the most variation in the ecology function, are impervious area, within provincial or natural environment park, and surface water connection (Table 11). The indicators with the largest contributions to the second dimension for this function are human modification level – natural, human modification level – semi natural, and percent wooded (Table 11).

*Table 11 - Ecology function: Indicators contributions to each dimension*

| **Indicator** | Contribution to Dimension | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| Impervious area | 0.70 | 0.02 | 0.01 | 0.03 | 0.00 |
| Within provincial or natural environment park | 0.51 | 0.04 | 0.00 | 0.08 | 0.01 |
| Channel Connections - surface water connection | 0.49 | 0.04 | 0.00 | 0.04 | 0.00 |
| Human modification level - semi natural | 0.22 | 0.44 | 0.05 | 0.06 | 0.00 |
| Within ecological network | 0.22 | 0.20 | 0.14 | 0.02 | 0.00 |
| Residential/Commercial | 0.15 | 0.06 | 0.05 | 0.39 | 0.00 |
| Nearest wetland | 0.07 | 0.01 | 0.35 | 0.02 | 0.01 |
| Wetland coverage | 0.03 | 0.02 | 0.26 | 0.08 | 0.07 |
| Percent wooded (forest and canopy) | 0.02 | 0.37 | 0.01 | 0.07 | 0.03 |
| Industrial | 0.02 | 0.03 | 0.00 | 0.43 | 0.07 |
| Human modification level - natural | 0.01 | 0.80 | 0.00 | 0.00 | 0.00 |
| Core amphibian habitat | 0.01 | 0.08 | 0.07 | 0.19 | 0.17 |
| Perimeter - area ratio | 0.00 | 0.00 | 0.01 | 0.02 | 0.66 |
| Amphibian movement pathway | 0.00 | 0.00 | 0.51 | 0.00 | 0.01 |

The graphical representations of the PCA results for the hydrology function, water quality function, and ecology function are displayed in Figure 14-16, respectively. Each of the four plots provides valuable information about the results of the PCA process and the influence of various indicators on the hydrology, water quality and ecology functions.

The *Observations* plot shows all the observations based on the first two principal components (PCs) calculated during the PCA process. Using the *Levels* plot, we can identify the influence of specific levels of categorical variables on the first and second principal components. For example, in the *Levels* plot in Figure 16, we see that the “No” level of the within-ecological network indicator has a strong positive influence on PC2 for the ecology function while the “Yes” level has a strong negative influence on this principal component.

The numerical variables and all variable plots indicate the influence that indicators have on the first two principal components. The numerical variables plot provides information on the influence of numerical indicators on the first two PCs and the all variables plot provides information on the influence of numerical and categorical indicators on PC1 and PC2. The longer the vector, or the further away they are from the center, the larger the influence the indicator has on the PC. A small angle between vectors for two indicators indicates a positive correlation while a large angle suggests a negative correlation. A 90-degree angle suggests no correlation. Further, the direction of the vector indicates the dimension(s) it most strongly influences.

For example, in the all variables plot of 10, the vector for human modification level - natural is further from the center and aligned with the y-axis, which indicates it has a strong influence on PC2 and a negligible influence on PC1 for the ecology function. Further, the small angle between the vector for human modification level - natural and the vector for percent wooded, indicates that these indicators are correlated. The 90-degree angle between the vectors for human modification level – natural and impervious indicate no correlation between these indicators.

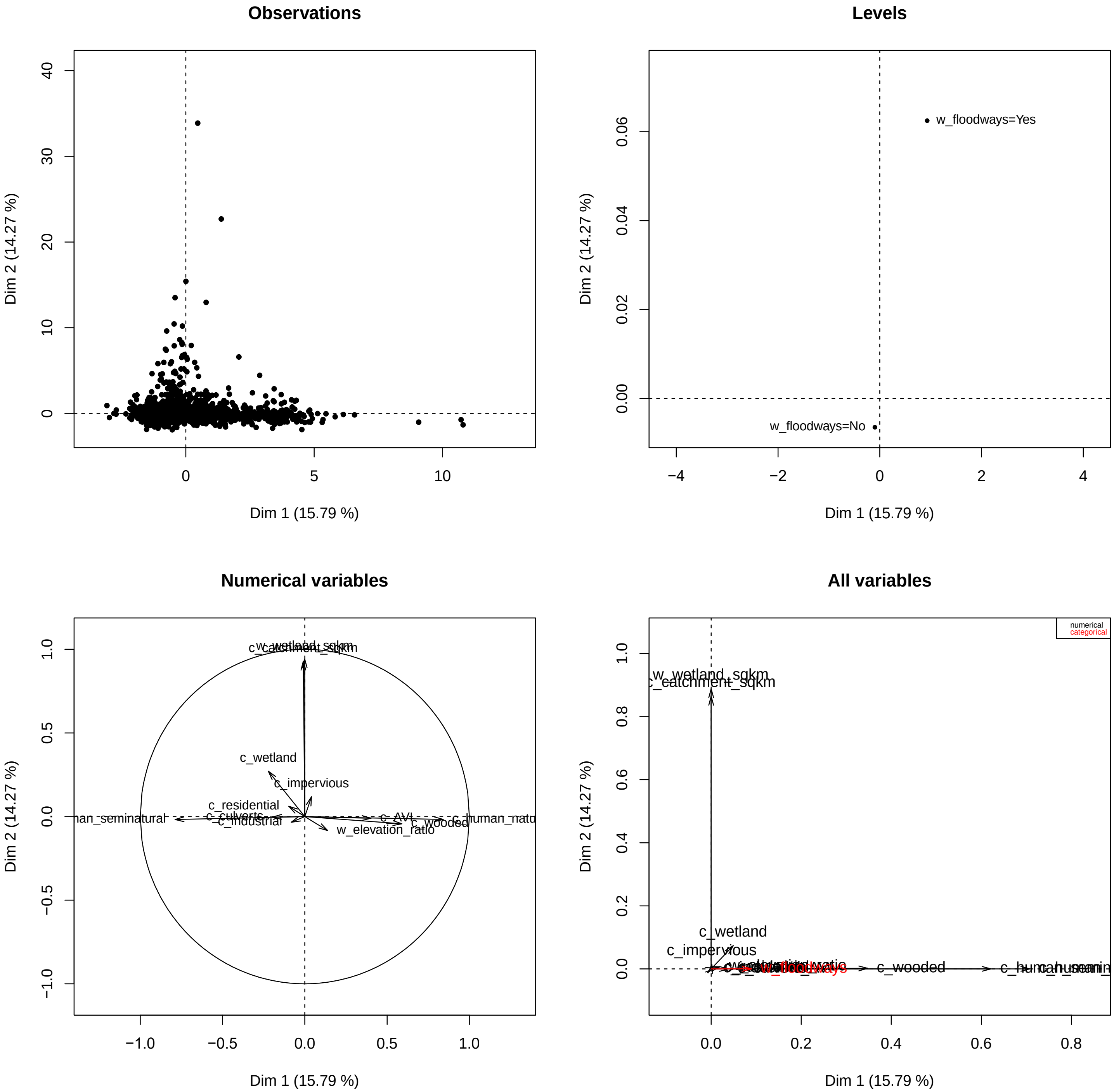



*Figure 14 Plots of results for PCAMix on hydrology function.*

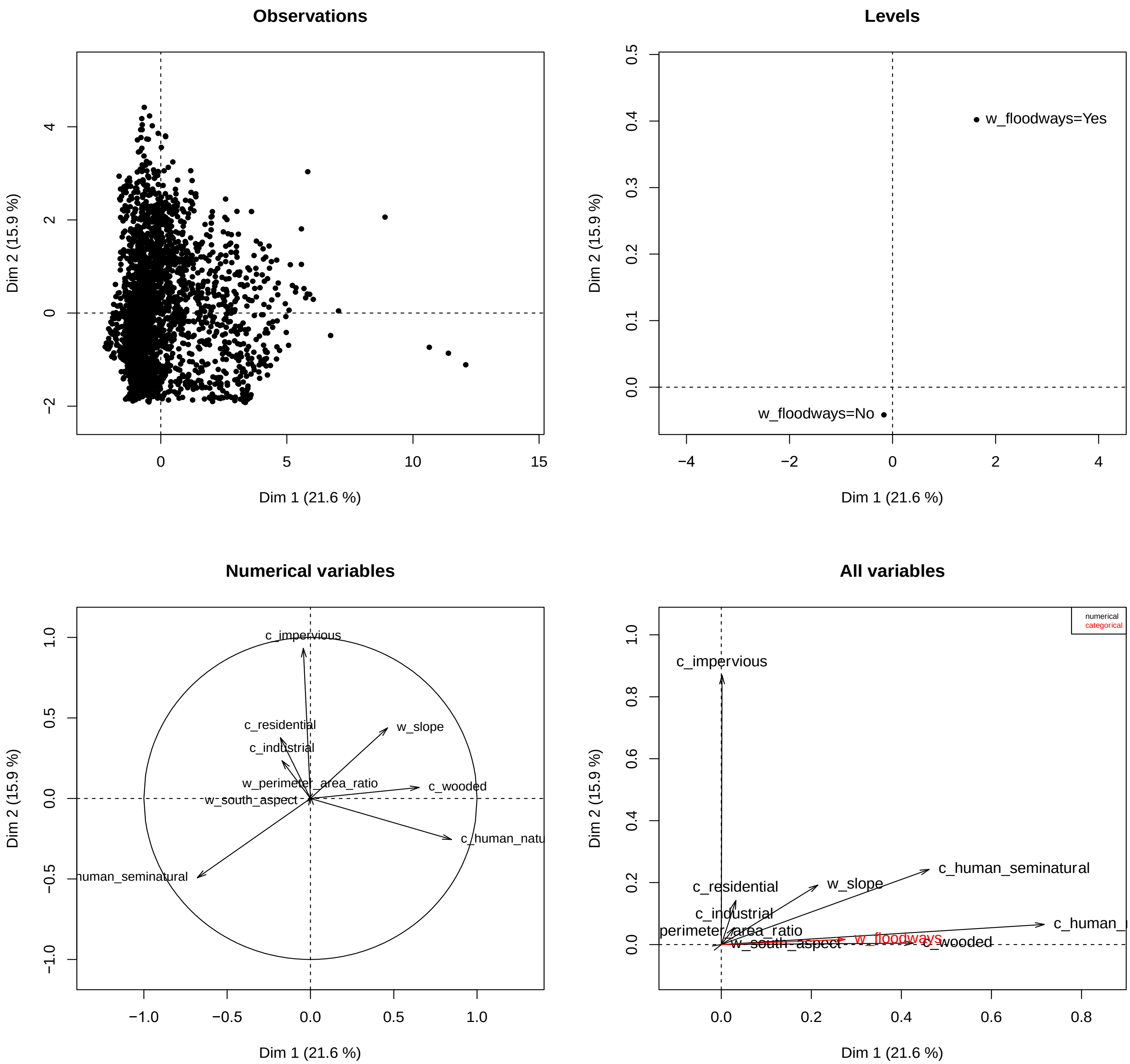



*Figure 15: Plots of results for PCAMix on water quality function.*

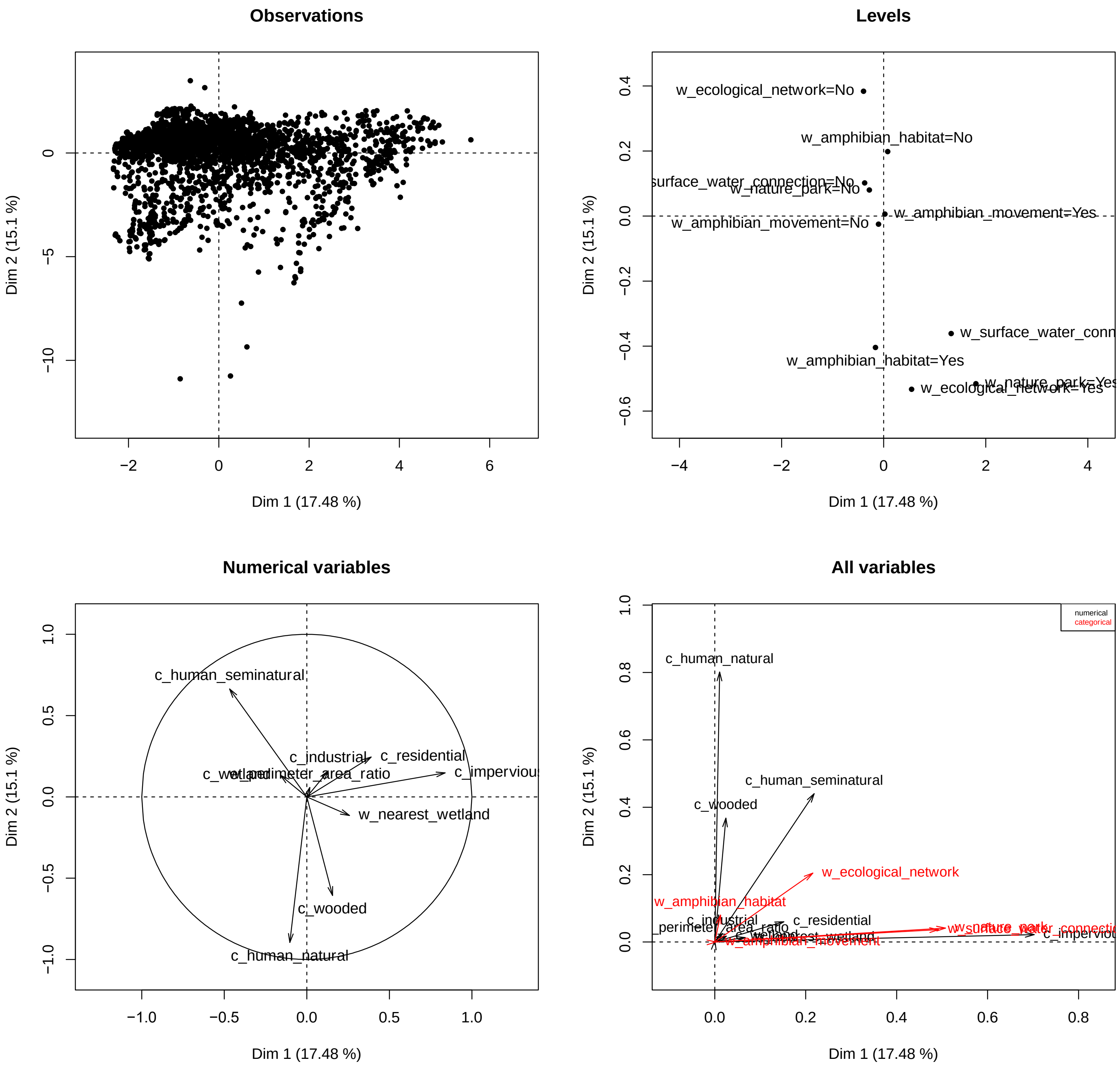



*Figure 16: Plots of results for PCAMix on ecology function.*

# Appendix B: Modelled-ACI value maps for each wetland function

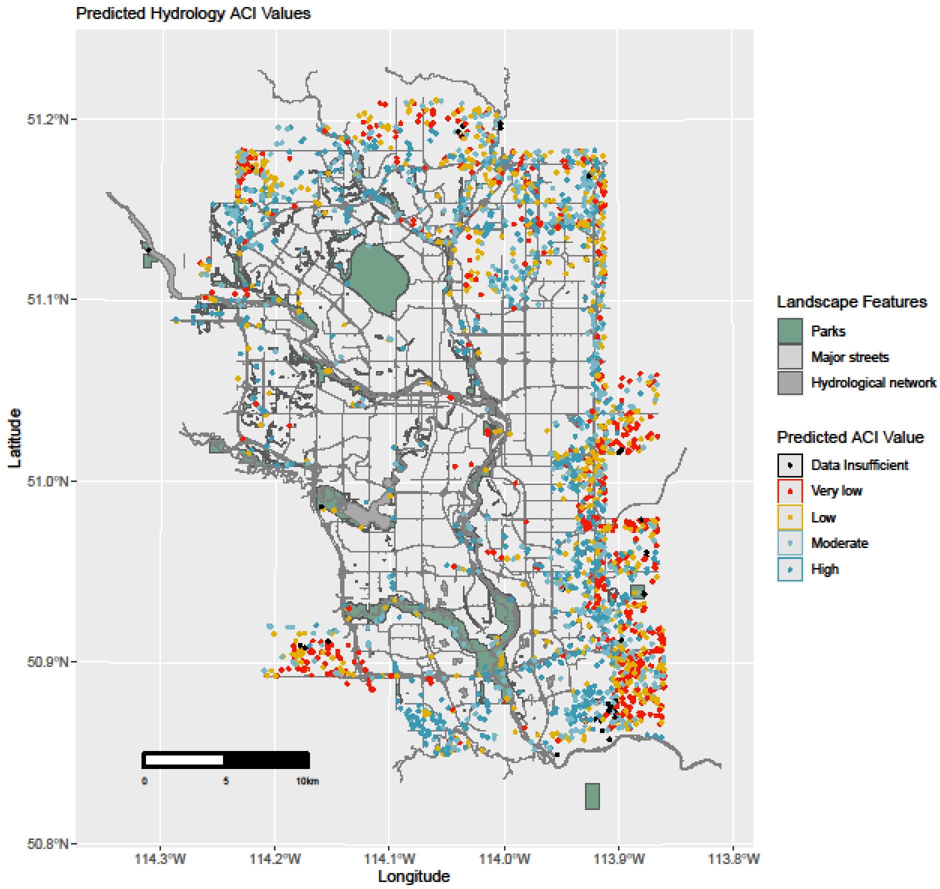



*Figure 17: Modelled-ACI values for hydrology function, displayed as a gradient based on quartiles from very low (red) to high (blue).*

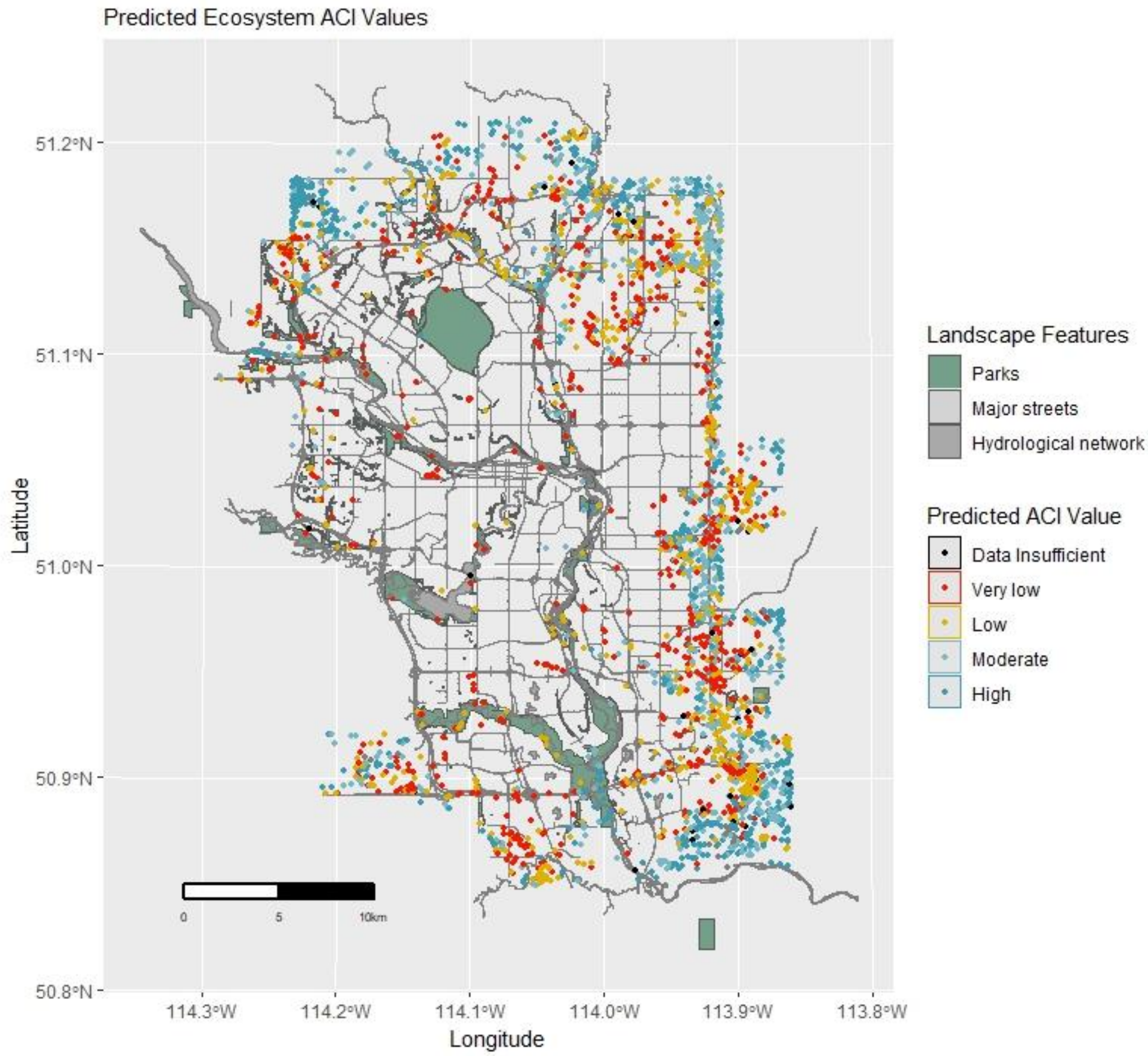



*Figure 18: Modelled ACI values for ecology function, displayed as a gradient based on quartiles from very low (red) to high (blue).*

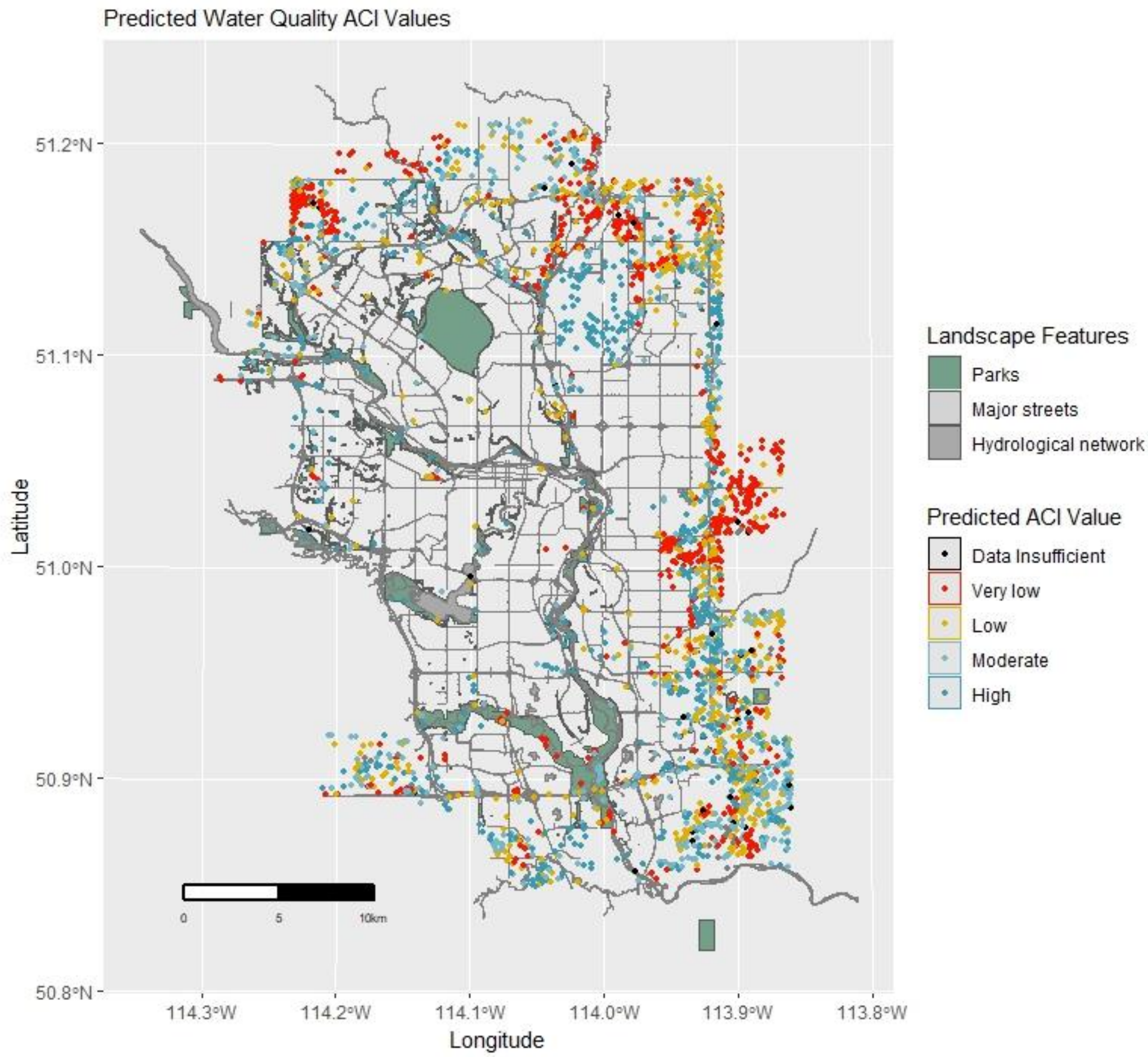



*Figure 19: Modelled-ACI values for water quality function, displayed as a gradient based on quartiles from very low (red) to high (blue).*

## *Modelled ACI for Natural Environment Parks*

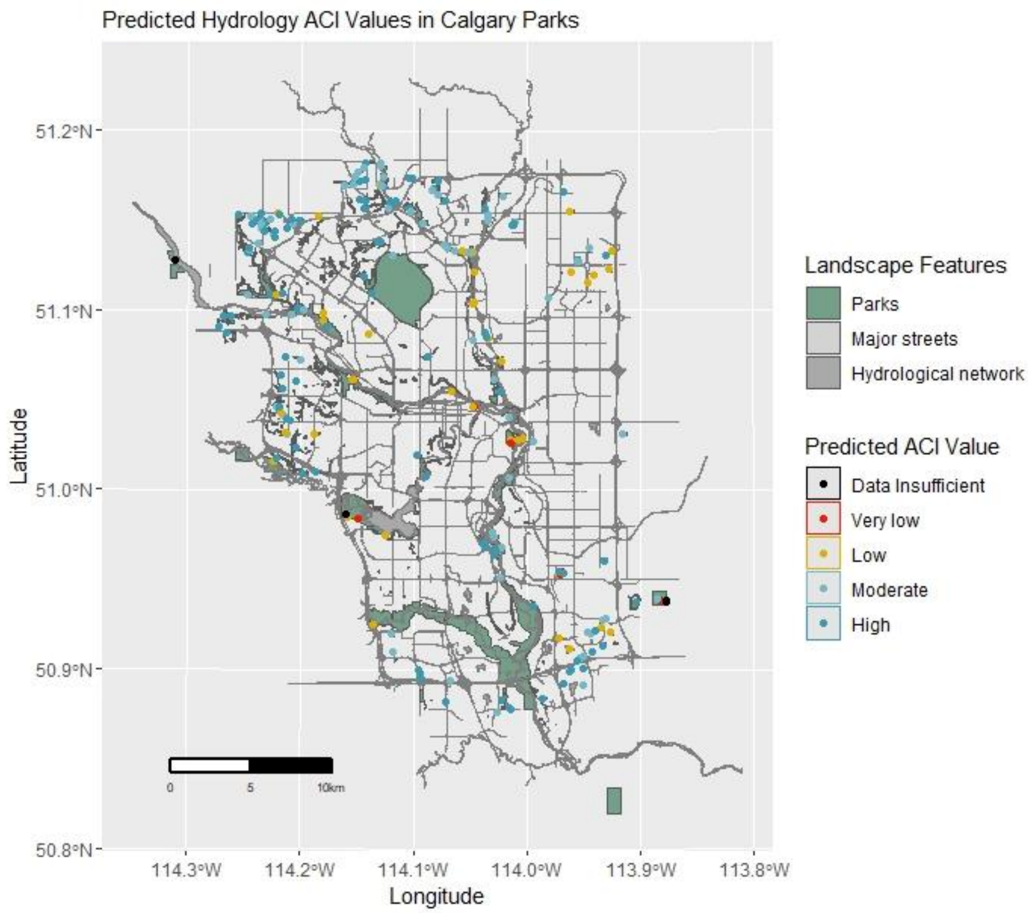



*Figure 20:: Modelled-ACI values for water quality function for wetlands in Calgary's Natural Environment Parks, displayed as a gradient based on quartiles from very low (red) to high (blue).*

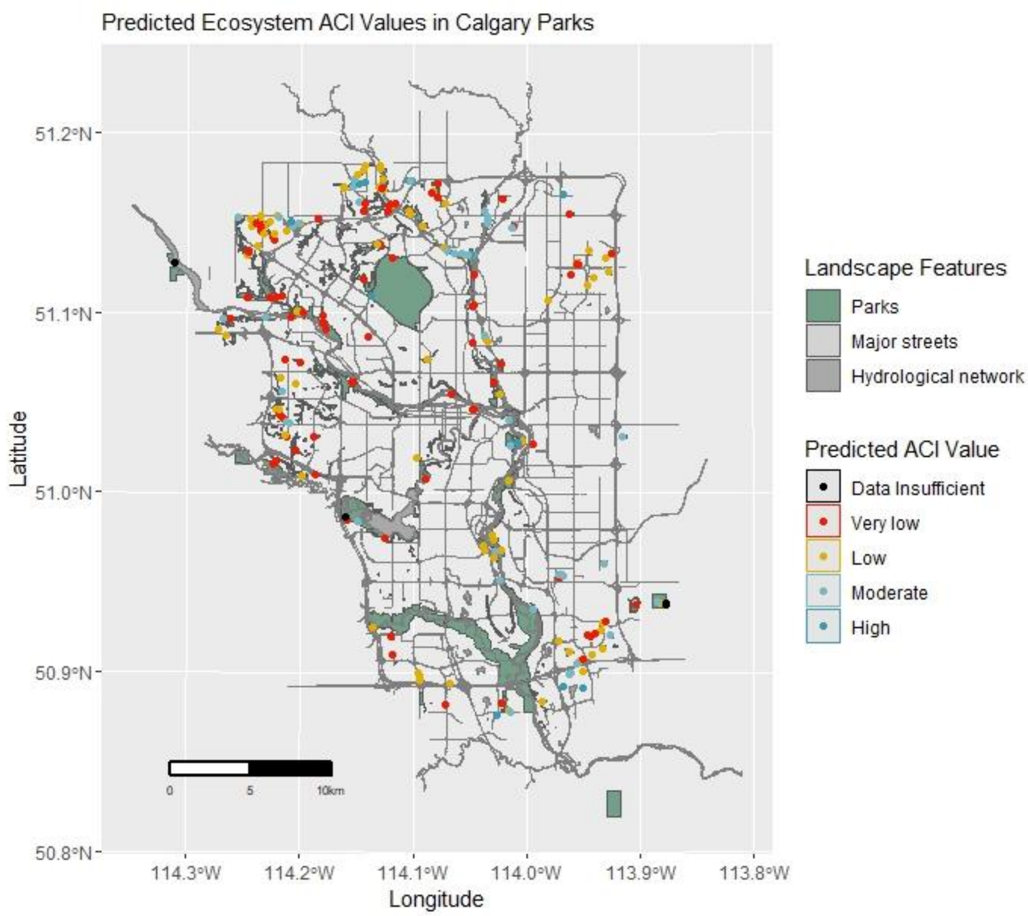



*Figure 5: Modelled-ACI values for ecosystem function for wetlands in Calgary's Natural Environment Parks, displayed as a gradient based on quartiles from very low (red) to high (blue).*

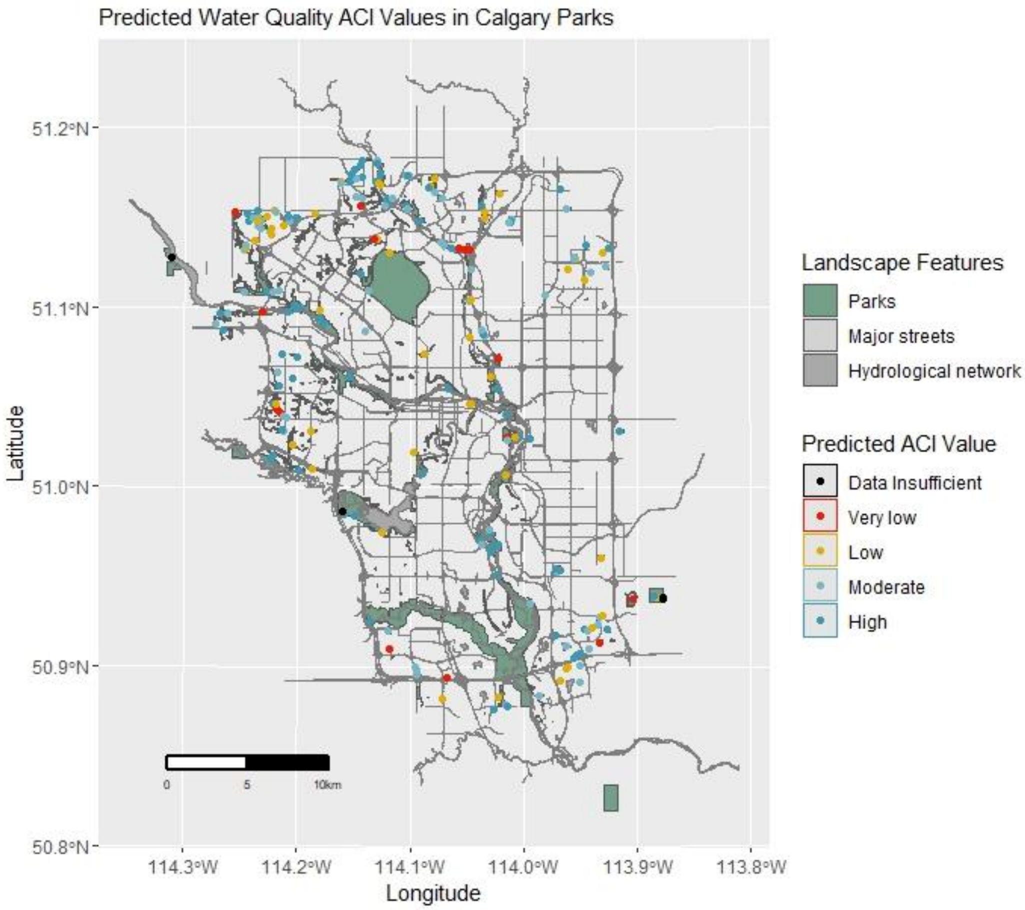



*Figure 21: Modelled-ACI values for water quality function for wetlands in Calgary's Natural Environment Parks, displayed as a gradient based on quartiles from very low (red) to high (blue).*

Miistakis Institute
EB3013, Mount Royal University
4825 Mount Royal Gate SW
Calgary, Alberta T3E 6K6

www.rockies.ca

Miistakis Institute