Keeyask Generation Project

Environmental Impact Statement

Supporting Volume

Aquatic Environment

KEEYASK
Hydropower Limited Partnership

June 2012

APPENDIX 3B

DEVELOPMENT OF RESERVOIR HABITAT AND MODELS TO ESTIMATE AQUATIC HABITAT AVAILABILITY IN THE KEEYASK RESERVOIR 30 YEARS AFTER FLOODING

3B.1 INTRODUCTION

This appendix presents a summary description of the development of aquatic habitat in riverine and lacustrine sections of reservoirs in the lower Nelson River at Year 30, and provides a summary of four spatial models used to estimate aquatic habitat availability in the Keeyask reservoir at this time step. The modelling work has built upon the results of physical environment studies that estimated post-Project shoreline, water depth, and depth averaged velocity for initial full supply level (initial FSL) conditions, and the estimated size and shape of the reservoir for Year 30 derived from models of peatland disintegration and shore erosion (PE SV, Section 4, Section 6, and Section 7).

Three models were used to estimate substrate distributions in the proposed reservoir and the fourth was used to estimate potential macrophyte habitat. For the substrate predictions, two lentic and one lotic model were used. For standing water habitat, a published deposition model was used for the offshore zone of the reservoir, and a second model was developed from flooded areas of Stephens Lake to estimate the areas where fine organic deposition would occur in bays. For lotic reservoir habitat, a deposition model was developed using data from the lower Nelson River, including the Keeyask area, Stephens Lake, and the Nelson River between the Limestone GS and the Long Spruce GS. The lotic deposition model was run twice to estimate the pattern of deposition at 95th and 5th percentile inflows. All model results were integrated to estimate the state of habitat for Year 30 post-Project. The fourth model, the predictive macrophyte model, is summarized here in brief. The full text describing the model development and validation is found in Appendix 3C.

Certainty in model predictions was assessed by means of Cross Validation and the Relative Operating Characteristic.

3B.2 DEVELOPMENT OF YEAR 30 RESERVOIR HABITAT

The evolution of reservoir habitat from flooded terrestrial and flooded aquatic habitat is complex and depends in part on the design of project-specific infrastructure, and how this intersects with the local topography. The elevation of the proposed water level on the pre-flood topography determines the shape (*i.e.*, size, depth, and geometry) of the reservoir (PE SV, Section 4), and the distribution of inundated soil types (Terrestrial Environment Supporting Volume [TE SV], Section 2.3.4.2 and Map 2.3-4. The shape of the reservoir, in turn, controls the expression of hydraulic energy (PE SV, Section 4 and Section 5), in the form of waves and currents and the relative position of water masses within the reservoir.

Over time, habitat in the reservoir develops through the interaction of these physical processes in areas of relatively high magnitude change. Habitat in reservoirs changes from flooded terrestrial or flooded aquatic habitat via the processes of erosion, transport, and sedimentation (PE SV, Section 6 and Section 7). Areas where effects are relatively small at Year 30 (*i.e.*, the habitat is altered but not markedly changed) remain similar to the initial FSL aquatic habitat and still resemble their basic pre-flood characteristics.

In the Keeyask area, the study of pre-flood soils and land cover (TE SV, Section 2.3.4.2) show that a generalized land cover sequence is apparent in the ecosite data that consists of three land cover types. The topographic sequence of mineral, thin peat, and deep peat is common in the study area (Figure 3B-1). This landcover sequence determines most of the locally available surficial and parent materials for redistribution in the reservoir.

Section 3B.2.1, Section 3B.2.2 and Section 3B.2.3 provide a basic description of the development of Shallow and Deep habitat for the Keeyask reservoir at Year 30 as understood from studies of Stephens Lake and, to a lesser extent, the Limestone reservoir. Each section below describes a major habitat type and provides a link to a model found in Section 3B.4. that predicts that type of habitat distribution.

3B.2.1 DEVELOPMENT OF SHALLOW WATER HABITAT IN FLOODED TERRESTRIAL AREAS

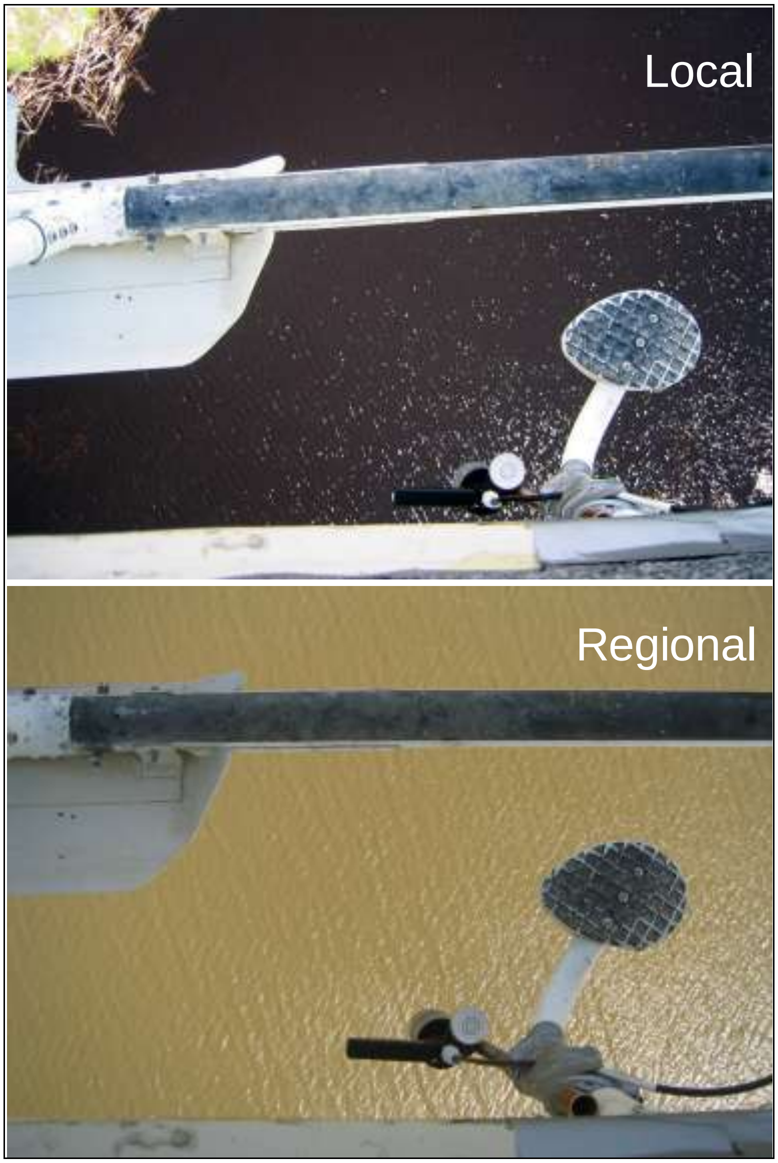
The water masses in the flooded thalweg of the Nelson River in Stephens Lake are markedly different than those of the surrounding peatland watersheds (Section 2). The water masses can be divided into local, mixed, and regional water masses (Map 3B-1) based on multivariate analysis of the water quality and light attenuation characteristics (Figure 3B-2). Local water masses appear tea-stained, are derived from peatland watersheds, and are adjacent to the Nelson River, which enters the Stephens Lake area with higher turbidity (Photo 3B-1).

Persistent deposition of fine organic material (FOM) in Stephens Lake typically occurs when a tributary of sufficient drainage area pools Local peatland water in the terminal end of a bay. The pool of Local water prevents the sediment-rich water of the main reservoir from fully diffusing through the entire bay. The ends of flooded bays in the reservoir tend to be peatlands before flooding (TE SV, Section 2.3.4.2 and Map 2.3-4). The persistent accumulation of FOM at the ends of flooded bays results from the exclusion of water with higher suspended sediment concentration from an area that has abundant sources of organic material (from local peatlands and via stream inflows).

The area where the Local and Mixed water masses meet determines the boundary between the fine organic deposition and the zone of silt deposition. The zone of silt deposition in deep water is the dominant bottom type in the Stephens Lake reservoir.

Mixed water masses tend to be more dilute than Regional water masses (Section 2) and so rates of silt deposition from mixed water masses appear to be lower than most of the main basin. Near surface Ponar grab samples in areas with Mixed water masses show a layered sample where silt has superimposed the pre-flood inundated peatland soil. The deposits of silt are relatively deep and homogenous farther offshore in areas of Regional water masses. The changes in bottom type of the flooded terrestrial bay forms a sequence where organic deposition changes to silt deposition. Silt deposition appears to cover the entire Lentic flooded terrestrial habitat that is not influenced by waves and/or slope (Photo 3B-2).

The model that predicts the depositional boundary between fine organic material and silt is described in Section 3B.4.3.



Source: North/South Consultants Inc. (P. Cooley), 2006

Photo 3B-1: Aerial view of Local and Regional water masses observed in the western end of Stephens Lake

Source: North/South Consultants Inc. (P. Cooley), 2006

Photo 3B-2: **Ponar samples taken from Shallow Lentic habitat within a flooded terrestrial bay showing (A) fine organic deposition from the end of a bay in a peatland water mass, (B) layer of silt covering pre-flood peat in a mixed water mass, and c) homogenous silt deposit in Deep habitat near the main reservoir**

3B.2.2 DEVELOPMENT OF NON-DEPOSITIONAL HABITAT IN SHALLOW LENTIC AREAS

Studies of the shallow water habitat in flooded areas of the west end of Stephens Lake about 30 years after flooding show that pre-flood thin peat soils or mineral soils, inundated in shallow water in moderate to high exposure, erode through the thin peat (if present) down to the mineral parent material (Figure 3B-3). The reworking of peat veneer or deep mineral soils by wave action over time forms a nearshore slope that is a mainly cohesive clay matrix with a smaller amount of sand to cobble surface lag found in the swash zone (Photo 3B-3). About 79% (109/137) of all samples taken in Stephens Lake from areas that were either deep mineral or peat veneer before flooding evolved into clay-based samples by 2005 or 2006. Other nearshore slopes formed from glaciofluvial deposits tend to have more sand/gravel, or infrequently, cobble in the study area. The clay substrate that forms in shallow water receives enough wave energy and is of sufficient slope to remain free of silt; further, much of this shallow water habitat is located within the IEZ of water level variation that may also move materials down slope. The clay-based bottom that forms from mineral soils, with or without a thin mantle of peat, provides a nutrient rich and cohesive fine-grained substrate that is potential macrophyte habitat. Studies of Stephens Lake showed that silt was found in Lentic habitat below the clay nearshore areas, which was deeper than the effects of waves. The shallow depth bound of the 95% confidence interval of the mean depth of sediment is 3.4 m (n = 100). This near minimum depth of silt is the same as the maximum depth observed for rooted macrophytes.

A model from the published literature with extensive validation was used to estimate the water depth at which silt deposition would occur in standing water areas of the Keeyask reservoir below the effects of waves (Section 3B.4.2).

3B.2.3 DEVELOPMENT OF DEEP WATER HABITAT IN FLOODED TERRESTRIAL AND FLOODED AQUATIC HABITAT

Studies were undertaken to understand the character of deep-water habitat in the lower Nelson River. The studies from Stephens Lake and Limestone reservoir contrast two different types of reservoirs. The Limestone reservoir represents a riverine reservoir contained mainly within the original river valley (*i.e.*, a large increase in depth relative to area) whereas Stephens Lake is mainly lacustrine as it was formed over a wide area of low relief (large increase in area relative to depth). While both reservoirs have similar maximum depths (*i.e.*, about 32 m) they have notably different thalweg habitats after flooding. The shape of the flooded topography appears to determine if the thalweg substrate will be changed (*e.g.*, cobble to silt) or altered (generally similar bottom composition but with areas of change like bank materials).

Source: Lynden Penner, J.D. Mollard and Associates, 2005

Photo 3B-3: Clay nearshore substrate with granular surface lag, formed in Stephens Lake

Information on depth, substrate, velocity, and exposure (see Appendix 3C 2.1.1) in the central thalweg of the Limestone reservoir showed that the substrate remained hard (*i.e.*, rock) with some finer infill materials. In the Limestone reservoir it appears the thalweg habitat was altered mainly by an increase depth and decrease in velocity given that river currents remained confined to a U-shaped channel; this appears to have maintained the dominant composition of the pre-flood thalweg character except in the area immediately upstream of the dam. In contrast, the habitat type of most of the flooded thalweg within Stephens Lake changed to silt deposition (Section 3.3.2.4) even in areas that were observed as lotic habitat during acoustic Doppler surveys (Appendix 3A) due to the increase in depth and loss of channel confinement. Substrate in some areas of the thalweg below Gull Rapids did not become depositional given currents remained in a riverine channel and depth changes due to the Kettle GS were relatively small.

A lotic deposition model was developed from these studies to estimate the distribution of deposition in the thalweg of the proposed Keeyask reservoir, which is expected to have both riverine and lacustrine-like reaches (Section 3B.4.1). This model extends the results that describe the rate of mineral sedimentation studies (PE SV, Section 7) by being spatially explicit for Year 30.

3B.3 MODELLING APPROACH

3B.3.1 MODELS TO ESTIMATE AQUATIC HABITAT AVAILABILITY IN THE KEEYASK RESERVOIR 30 YEARS AFTER FLOODING

Four spatial aquatic habitat models were developed to estimate habitat availability in the Keeyask reservoir (Table 3B-1) at about 30 years after flooding. Three models were derived to predict the presence or absence of deposition that underlies either lentic (standing) or lotic (flowing) water masses. Two of the depositional models estimated the distribution of mineral deposition (*i.e.*, silt) and the third estimated the distribution of fine organic material. The fourth model predicted the presence of absence of suitable habitat for *Potamogeton richardsonii* and *Myriophyllum sibiricum*, the two dominant species of macrophyte found in flooded habitat of Stephens Lake. The development of a predictive macrophyte model designed for application on the proposed Keeyask reservoir is described here in brief, and in detail in Appendix 3C.

3B.3.1.1 Data Sources and Uncertainty

Initial FSL datasets representing depth, maximum fetch, exposure, and slope were used to estimate the aquatic habitat distributions. The initial FSL data represents the existing environment topography with the only ecological change being the addition of the full supply water level. This approach was adopted when the results of the physical environment studies suggested that changes in the shape of the reservoir over time are relatively small when compared to changes due to initial FSL. Changes in bottom topography in the nearshore zone are expected to occur between initial FSL and Year 30 and are due mainly to peat resurfacing, mineral shore erosion, and mineral sedimentation. A summary follows that describes these changes in order explain the applicability of the initial FSL topography as a proxy for Year 30.

3B.3.1.1.1 Peat Resurfacing

Any effect of peat resurfacing on the bottom topography of the Keeyask reservoir between Initial FSL and Year 30 would be most marked in shallow water where hydrostatic pressure does not keep the peat on the bottom (PE SV, Section 6).

Laboratory analysis of peat resurfacing potential after flooding reveals that the fibrous surface layer (O_f) has the lowest specific gravity and is typically the only layer that floats to the surface after separating from the mesic or humic layers below(PE SV, Section 6). The composition of the dominant peatlands within the predicted flooded area would be 40% veneer bog, 26% blanket bog, and 23% peat plateau bog. These peatland types have O_f thicknesses that average 0.22 m, 0.37 m, and 0.25 m, respectively (PE SV, Section 6). These O_f thicknesses are less than half the 1 m contour interval and therefore are within the error of the post-Project initial FSL depth map. Therefore, the effect of peat resurfacing on bottom topography is small given that almost 90% of the flooded area has layers of peat that have some potential to uplift, which are thinner than the error inherent in the Initial FSL elevation model.

3B.3.1.1.2 Mineral Soil Erosion

Bank recession distances projected over the 30-year modelling period for the Keeyask Project average 4.8 m/year (y), with a maximum of 40.8 m at highly exposed sites (PE SV, Section 6). Maximum bank recession distances without the Project were estimated at 0.4 m/y, or a maximum total recession of about 12 m over the 30-year period. A maximum incremental bank recession of 29 m can be attributed to the Project after 30 years. The changes in the shape of the reservoir over time are therefore relatively small when compared to changes incurred from initial FSL. For example, when the depth of fine mud deposition is estimated (Model # 2 in Table 3B-1) for a 7 km fetch common to the lower reservoir with a 4% slope and then again with the additional 29 m attributed to the Project, the estimated water depth where deposition begins changes from 1.59 m to 1.60 m.

3B.3.1.1.3 Mineral Deposition in Lentic and Lotic Habitat

Sediment coring and ground penetrating radar were used to study sedimentation processes in the lower Nelson River in 2006 for studies in support of PE SV, Section 6. In Stephens Lake, sampling was undertaken at eight sites along transects from the shoreline to about 200 m in the offshore direction. Sampling was directed to study nearshore processes, and did not target the main depositional basins of the reservoir or the pre-flood thalweg of the river. Results demonstrated a general fining of grain sizes with increasing water depth and distance from shore, except where slope was sufficiently high to refocus materials downward. Sediment thickness above pre-flood strata was lower in lotic sites than lentic sites, and the proportion of organic material deposited with the mineral sediment was greater in lentic sites. Glacial deposits that lack either mineral or organic deposition from Stephens Lake were observed on the upper beach slope at some sites, indicating that the upper beach slope is primarily an erosive environment with little fine-grained sediment deposition occurring above the wave base depth. Average sedimentation rates in Stephens Lake since impoundment and below the effect of waves were often 1 cm/y, but can be as high as 2.4 cm/y.

An average nearshore depositional rate of 1 cm/y multiplied by 35 years (1971–2006) equates to an average deposit thickness of 35 cm. In deep water, this change in depth is small relative to the depth in the Initial FSL map and would therefore not have a measurable effect on the results of any model applied. In lentic habitat, silt deposition would be expected below the effects of waves, water levels, or where slope also increases the depth of the depositional boundary. This, however, would not change the location of the silt boundary and only marginally decreases the depth. If peat uplift at a site below the effects of waves occurred shortly after flooding then the silt deposition would likely result in filling the "crater" to an elevation similar to that of the surface of the substrate at initial FSL.

3B.3.1.2 Analysis Methods

Logistic regression and Linear Discriminant Analysis (LDA) are multivariate methods that are suited to the multi-variable data required to estimate reservoir habitat based on conditions before and/or after flooding. Each method has different data requirements and employs a different analytical method, but both result in a predicted outcome that is classified (*i.e.*, nominal) based on a probability value. Classification of objects into groups enables an assessment of the performance of the model by comparison of the agreement between observed and predicted classes, referred to as cross-validation. Both methods also support "block entry" or "stepwise" methods of analysis that describe how the variables are analyzed. Block entry methods analyze all variables together as a group, whereas stepwise methods evaluate the contribution of each variable to the model, and conditionally, drop the variables that do not improve the model significantly.

3B.3.1.2.1 Logistic Regression

Logistic regression is used for predicting the probability of an event by fitting data to a logistic curve, which is sigmoid in shape. Logistic regression is a generalized linear model used for binomial regression. It makes use of several predictor variables that may either be numerical or categorical to predict a binary response variable (0 or 1, presence or absence). Classification of each observation into one of two binary response variables typically is undertaken at a probability of 0.5 "cut" threshold. This type of regression is often used in ecological studies to determine what factors are responsible for the presence or absence of a species.

3B.3.1.2.2 Linear Discriminant Analysis

Linear discriminant analysis (Manley 1994) considers a division of objects into groups by reducing the number of dimensions in the data, develops a predictive model, and supports cross validation to assess the agreement between observed and predicted classes. A predictive model is constructed for known groups (k) that are known *a priori* based on the linear combinations of the available environmental variables (p) that best discriminate among the groups. The number of discriminant axes is the smaller of k – 1 or p. Like the one way Analysis of Variance (ANOVA), the LDA maximizes the F ratio by forming linear composites that maximize the inter- to intra-class variation over k. Each observation in the LDA results with a probability of being assigned to each of the groups; the class with the highest probability is assigned to the observation. The relative contribution of each variable to the LDA can be examined by review of standardized discriminant function coefficients (Legendre and Legendre 2004). The equations of the LDA analysis are provided by the Fishers Discriminant Function coefficients. LDA may be preferred over logistic regression when the number of groups required of the predictive model is greater than two (Pohar *et al.* 2004).

3B.3.1.2.3 Cross-Validation and the Relative Operating Characteristic

Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent dataset. This method is generally used when the goal of the analysis is prediction, and an estimate is needed that shows how well predictive model will perform in practice. Cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset used to develop the model (*i.e.*, the model set), and validating the analysis on the other subset (*i.e.*, the test set). In this manner, predicted classifications generated from the model subset are compared against the test subset for which the group association is already known. The overall agreement, in percent, is used to suggest how well the model would run under similar conditions on a different dataset.

Selection of observations into model and test groups for each of the models was undertaken by selecting one in every three or four records (depending on the size of each dataset) in the database, which was considered the test validation group and was not used in model building. This systematic sampling was undertaken to ensure all ecotypes were represented in the test group.

The relative operating characteristic (ROC) is a comparison of true positive responses and false positive responses of a classification (Egan 1975; Swets 1988). The ROC may be may be reduced to a single value to facilitate comparison of expected classification performance. A common measure of ROC is that of the area under the curve where the values range from 0.5 to 1.0. Relative operating characteristic values of 0.5 infer the model classifies only about as well as a random model and ROC values approaching 1.0 indicate a perfect fit (*i.e.*, only true positive classification results).

3B.4 PREDICTIVE HABITAT MODELS FOR THE KEEYASK RESERVOIR 30 YEARS AFTER FLOODING

3B.4.1 ESTIMATING THE DISTRIBUTION OF DEPOSITION WITHIN LOTIC HABITAT

An empirical model to estimate the presence or absence of deposition in lotic areas of the proposed Keeyask reservoir was derived from depth, velocity, and exposure data (n = 171) (Table 3B-2) from data collected during habitat survey (Appendix 3A). The depth averaged velocity data were those introduced in Section 4 of the PE SV.

The range in model estimates was assessed under low and high flows by substituting either the 95th FSL or 5th minimum operating level depth-averaged velocity and exposure percentile conditions for these data (n=60; Table 3B-2). Samples finer than sand (*i.e.*, mostly clay and silt) were considered depositional.

A binary logistic regression model was fitted to the lotic deposition/no deposition data. The logistic model was derived from 75% (n = 130) of the available data, referred to as the Model group, by entering the data in a forward stepwise procedure using the variables: 1) site depth (m), 2) exposure (m), and 3) depth-averaged velocity (m/s). Likelihood-ratio tests were used to determine the statistical significance of explanatory variables. Classification agreement and performance was assessed using cross-validation and by means of the ROC.

Cross-validation was undertaken by running the lotic deposition logistic model on the remaining 25% (n = 41) of the data for which class membership was known, but was excluded during model building. These validation samples are referred to as the Test group.

Logistic regression equations to estimate deposition in lotic habitat:

i) 5th percentile inflow

$$\text{Lotic deposition}_5 = 0.7336 + 0.182479\,\mathit{depth} + 0.000836\,\mathit{exposure} - 22.429063\,\mathit{velocity}$$

ii) 95th percentile inflow

$$\text{Lotic deposition}_{95} = 1.6099 + 0.052980\,\mathit{depth} + 0.001530\,\mathit{exposure} - 17.42653\,\mathit{velocity}$$

Forward stepwise logistic regression results show that velocity, depth, and exposure together provided the best model for the 5th and 95th percentile model runs (Table 3B-3). As expected, depth-averaged velocity was a highly significant variable (Table 3B-4) for describing the presence or absence of deposition in both models. For the 5th percentile model, the contribution of depth was also significant; whereas, in the 95th percentile run, the role of exposure was important (*i.e.*, nearly significant).

Logistic regression results do not lend well to graphical presentation so trends in the data are shown using principal component analysis (PCA). The PCA results are visually similar for the 5th and 95th percentile inflows and the first two component axes in each trial explained about 85% of the variance in the data. For the 95th percentile PCA (Figure 3B-1), the first principal component represented contributions from both exposure and depth, which combined explained most of the variance along that axis (85%); whereas, the second component was dominated by depth averaged velocity (83%). Figure 3B-1 shows that depositional sites tend to be those that had relatively high exposure and water depth at moderate velocity (*e.g.*, Kettle reservoir), or low velocity at moderate exposures and depth (*e.g.*, Stephens Lake thalweg). Sites without deposition tended to occur where velocity is relatively high and where depth and exposure is moderate or low (*e.g.*, central thalweg of Limestone reservoir or the lotic areas of the Keeyask Study Area). In particular, sites upstream of Gull Lake (where the Nelson River flows are fast, the channel is narrow and relatively shallow) are readily apparent in the upper left corner of the biplot.

Cross-validation results employed in the logistic regression analysis showed that classification agreement for the 5th and 95th percentile inflow scenarios was excellent, ranging between 82% to 91% (Table 3B-5). The Test group was not included in model building and provided agreement slightly lower than the

Model group (5–6% lower), as could be expected from a relatively small sample size when compared to the Model group. The deposition class in the Test group achieved 73% agreement. Cross-validation results suggest that the lotic deposition model can correctly classify depositional sites 73% of the time, but can be as high as 83%.

The area under the ROC curve for these according to the 95th percentile inflow and 5th percentile inflow is: $ROC_{95} = 0.967$, $ROC_{5} = 0.948$

where ROC = 1 indicates a perfect fit; and ROC = 0.5 indicates a random fit.

According to the ROC assessment approach, the lotic deposition model has a probability of assigning a true positive result about 95–97% of the time.

The cross-validation and ROC methods of assessment both suggest a strong predictive capacity is achieved in the lotic deposition model. This is evident in Map 3B-2, which compares the predicted bottom type (deposition/no deposition) to the data observed in the field. Map 3B-3 shows the modelled distribution within the lotic habitat area.

3B.4.2 ESTIMATING THE MUD DEPOSITION BOUNDARY DEPTH

Equation 25 of Rowan *et al.* (1992) predicts the presence or absence of deposition in standing water. The boundary between depositional and non-depositional areas is referred to as the mud depositional boundary depth (Mud DBD) that results due to waves and/or from slope due to the tractive force of gravity creating sheer stress. The presence or absence of deposition is estimated using the variables: site depth (m), maximum fetch (km), and slope (%). Equation 25 was derived from empirical data gathered from 54 lakes over a wide range in size in temperate Canada and the northern United States. In a reservoir drawdown study, Cooley and Franzin (2008) conducted a detailed validation of this equation in a drawdown experiment and found remarkable agreement between the observed and modelled deposition distributions.

Deposition is defined as particles that are smaller than 23 um or 5.5 phi, or greater than 60% water content.

Logistic regression equation to estimate the mud DBD:

$$\text{Mud DBD} = -0.107 + 0.742 \log \textit{Maximum Fetch} + 0.0653\ \textit{slope}$$

Rowan *et al.* (1992) show that this equation correctly classifies 683 out of 783 (87%) of fine grained sites and 344 out of 477 (70%) of coarse grained sites from which the model was built.

The extent of silt deposition below the effects of waves was estimated for lentic areas of the lower reservoir using Equation 25 (Rowan *et al.* 1992), the initial FSL depth map, a slope map (%), and a map of maximum fetch distance. This model predicts a zone of no deposition that often appears as a band that follows the perimeter of the reservoir and islands due to wave energy or slope. The lower extent of this zone is delineated as the Mud DBD, below which deposition was predicted for all of the remaining lentic areas.

To assess the validity of the results modelled by Equation 25 for the Keeyask reservoir, the extent of the wave-washed zone mapped according to Equation 25 was scrutinized further by comparison to empirical data from Stephens Lake. Keeyask aquatic studies show that both of the two dominant species of rooted macrophyte are found in shallow water areas above the silt boundary. The upper extent to the distribution of deposition estimated by Equation 25 appears as a band along the shoreline. The width of this band was compared visually to empirical data describing the distance from each plant stand to the shoreline (Figure 3B-4).

The width of the zone between the shoreline and the silt boundary estimated by Equation 25 was 60–75 m wide for most of the lower reservoir, but was as wide as 300 m in a few areas. These distances are in good agreement with measured distances between each plant stand and the shoreline at Stephens Lake. The average distance between the shoreline and stands of *Potamogeton richardsonii*, the most abundant species, was 60 m but ranged as far as 352 m.

The mapped results of this model are provided with those of the next model, described below.

3B.4.3 ESTIMATING THE DISTRIBUTION OF DEPOSITION BY FINE ORGANIC MATERIAL

A binary logistic regression model was fitted to FOM and silt substrate data (n = 238) obtained from flooded areas of Stephens Lake collected during the Keeyask aquatic studies to predict the boundary between FOM and silt substrata in peatland bays. The logistic model was derived from 75% (n = 179) of the available data, referred to as the Model group, by entering the data using a forward stepwise procedure using the variables: site depth (m), exposure (m), and slope (%). Likelihood-ratio tests were used to determine the statistical significance of explanatory variables. Classification agreement and performance was assessed using cross-validation and by means of the ROC.

Cross-validation was undertaken by running the FOM logistic model on the remaining 25% (n = 59) of the data for which class membership was known, but was excluded during model building. These validation samples are referred to as the Test group.

Logistic regression equation for estimating the distribution of FOM:

$$FOM = 5.008 - 0.710\,Depth - 0.003\,Exposure - 0.438\,slope$$

3B.4.3.1 Logistic regression statistics

The explanatory variables bring significant information to the model when compared to the model using only a constant (Table 3B-6). Stepwise results demonstrated three variables provided the best model fit, with each step forming a significant improvement to the model (Table 3B-7). The contribution each variable to the model (in the form of standardized coefficients) is shown in Table 3B-8. Exposure contributed most to the model with a highly significant effect on model form. The effect of depth and slope was also significant, but the role of slope was only about half as important as that of exposure.

3B.4.3.2 Classification agreement and ROC performance

Cross-validation results (Table 3B-9) show that overall classification agreement of the Model group and Test group is 79.9% and 79.7%, respectively. Similarity in percent agreement suggests that the sample size of the Model group was sufficiently large and likely represents all of the data found in the Test group. Results suggest that the FOM model will correctly classify FOM sites about 83% of the time.

The area under the ROC curve for these data is: ROC = 0.908

where ROC = 1 indicates a perfect fit; and ROC = 0.5 indicates a random fit.

According to the ROC assessment approach, the FOM model has a probability of assigning a true positive result about 91% of the time.

The cross-validation and ROC methods of assessment both suggest a strong predictive capacity is achieved in the FOM logistic regression model. This is evident in Map 3B-4, which compares the predicted bottom type (FOM or silt) to the data observed in Stephens Lake. The model results show that most of the error in classification agreement arose due to prediction of FOM in areas lacking inflows from peatland streams.

Application of the FOM logistic model in the Keeyask reservoir (Map 3B-5) was restricted to areas where peatland tributaries drain into flooded bays. Consequently, the results presented here are considered conservative (*i.e.*, the model results, in terms of true positives, would be expected to be higher than documented). The uncertainty of this model is relatively easy to assess given that FOM deposition occurs in bays that co-occur with tributaries, which are readily identified in maps. The precise position of the boundary between silt and FOM in a bay is less certain, but would be considered moderate due to the strong control of bay shape (*i.e.*, exposure) on model results.

Map 3B-6 shows the integration of all models described above to estimate the distribution of deposition in the Year 30 post-Project reservoir.

3B.4.4 MODEL 4 – ESTIMATING THE POTENTIAL DISTRIBUTION OF *POTAMOGETON RICHARDSONII* AND *MYRIOPHYLLUM SIBIRICUM*

This section provides an overview of the main results found in Appendix 3C that details: 1) the development of a predictive reservoir (PR) model to estimate the distribution of potential habitat for *Potamogeton richardsonii* and *Myriophyllum sibiricum* in the proposed Keeyask reservoir; 2) analyses to indicate which of the select environmental variables best accounts for the observed distribution of each species; and 3) documents the use of potential habitat by macrophytes.

The predictive macrophyte model was derived from field data collected from Stephens Lake in mid-summer 2005 and 2006 that described species, location, depth, slope, exposure, and substrate (n = 471) from the existing environment (EE) and the pre-flood (PF) landcover variables distance to mineral soil and peat depth (described in Appendix 3C). The pre-flood variables are key inputs to the model as this allows the presence or absence of aquatic plants in Stephens Lake today to be associated also with pre-flood conditions. Pre-flood soils information also provides an option for the model to work without the need for detailed substrate information, which may not be known *a priori*, when the model is applied in a future scenario.

3B.4.4.1 Assessing the Relative Importance of Existing Environment and Pre-flood Variables

The objectives supporting the development of the predictive macrophyte model was to compare and contrast the EE variables with the PF variables to determine the strengths and weaknesses of the available data to better critique the model.

The first of three LDA analyses included all EE and PF variables and explained 79% of the variance in the data. Substrate type from the EE was the dominant variable discriminating between both species from areas where they were absent. The second LDA analysis was constrained to EE variables only. As expected, the amount of variance increased relative to the first trial (87% explained) and showed that substrate grain size and water depth primarily determined macrophyte distribution in the EE. The third LDA trial, the PR model, aimed to learn which EE and PF variables would be most important when the PF surrogate variables (*i.e.*, distance to mineral soils and peat depth) were used in place of the EE substrate, which was assumed to be unavailable. The third trial aimed to determine if removal of the most important variable in the first two trials resulted in a decrease of model performance. Results of the PR model (Table 3B-10) confirmed that, like the two previous trials, information on bottom type (*i.e.*, either EE substrate or PF soils) was most important to discriminate between the presence of each species from absence. As shown in Figure 3B-6, the PF soil variables dominated discrimination and so comprise most of the weight along the axis of function 1, whereas the EE variables dominated function 2. On function 1, the minimum distance to mineral soil variable weighted the axis nearly twice that of peat depth. The second function was weighted most by slope and exposure, which were weighted similarly, and to a lesser extent by depth.

The PR model classification results explained 67% of the variance in the PF and EE data, which is a decrease of 20% relative to the EE model. This may suggest that the classification performance of the PR model might have dropped drop notably. The cross-validation results, however, showed clearly that this was not the case (Table 3B-11). Both trials on the Test group, not used to build the model, achieved high and equal classification agreement (81%).

The LDA analysis, unlike the logistic models above, supports the discrimination of more than two groups; this enables the two main species of macrophytes in Stephens Lake to be discriminated and the performance of the predictive model for each species to be assessed. Cross-validation results by species or absent show that *M. sibiricum* and *P. richardsonii* can be predicted with about equal confidence about

84 to 86% of the time; whereas, sites where these species are absent is slightly lower, about 74 to 76%. The results also show the Test and Model group each had a classification agreement that was about equal, and that estimates of agreement by species or absent were within 2%. This reveals that the PR model can be used with a high degree of confidence that is equal to that of the EE model (which operated with the benefit of contemporary field data) The PR model results are shown in Map 3B-7. Analyses of habitat preferences by each species are provided in Appendix 3C.

3B.4.4.2 Accounting for Deep Peat in Exposed Areas

The Keeyask reservoir has a few relatively large areas of deep peat in exposed locations, which was a site condition not observed in the macrophyte study area of Stephens Lake. This suggests that the LDA macrophyte model results could be improved if constrained by deep peatland type. Year 30 potential habitat in the lower reservoir was inspected visually to exclude relatively large areas of deeper peatlands, which included peatland plateau bogs and blanket bogs. These peatland types often have surface organic layers than can be up to 2 m thick (PE SV, Section 6). Soil profile information from the areas along the future Year 30 shoreline was reviewed to confirm relatively thick peat (PE SV, Section 6). Based on studies of more than 500 sites in Stephens Lake, these areas would not be suitable for macrophyte growth given peat is abundant on the bottom (intact or inundated peat), detritus, and other small woody debris, and/or water depths in areas of peat uplift that exceed the photic zone. It was therefore assumed that all of the relatively large peatland areas found above the silt boundary (modelled by Equation 25 in Rowan *et al.* 1992) were not potential macrophyte habitat.

3B.5 REFERENCES

3B.5.1 LITERATURE CITED

Cooley, P.M., and Franzin, W.G. 2008. Predicting the spatial mud energy and mud deposition boundary depth in a small boreal reservoir before and after draw down. Lake and Reservoir Management 24(3): 261–272 pp.

Egan, J.P. 1975. Signal detection theory and ROC analysis. Academic Press, New York, N.Y.

Legendre, P., and Legendre, L. 2004. Numerical ecology: developments in environmental modelling. Second English Ed. Elsevier Scientific Publishing Company, Amsterdam.

Manly, B.F.J. 1994. Multivariate statistical methods: a primer. Second Ed. Chapman and Hall, London.

Pohar, M., Blas, M., and Turk, S. 2004. Comparison of logistic regression and linear discriminant analysis: a simulation study. Metodološki Zvezki 1(1): 143–161 pp.

Rowan, D.J., Kalff, J., and Rasmussen, J.B. 1992. Estimating the mud deposition boundary depth in lakes from wave theory. Canadian Journal of Fisheries and Aquatic Sciences 49: 2490–2497 pp.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. Science. 240: 1285–1293 pp.

Table 3B-1: List of approaches used to estimate aquatic habitat availability at Year 30 according to: model application, type of water mass (lentic/lotic), modelling method, area of data source

Model #	Application	Water Mass	Model	Method	Area
1	Substrate	Lotic	Presence/absence of deposition	Logistic regression	Nelson River between Birthday Rapids to Limestone GS
2	Substrate	Lentic	Presence/absence of mineral deposition - Equation 25 in Rowan *et al*. (1992)	Logistic regression	Ontario/Quebec
3	Substrate	Lentic	Presence/absence of fine organic material	Logistic regression	Stephens Lake
4	Macrophyte	Lentic	Presence/absence of two dominant macrophyte species from absent	Linear discriminant analysis	Stephens Lake

Table 3B-2: Areas on the Nelson River where substrate or velocity samples were obtained according to time period, daily average discharge (Q) was measured in cubic meters per second in the lower Nelson River, and the presence or absence of deposition of materials finer than sand

Area	Substrate		Velocity		Deposition	
	Date	Q	Date	Q	Yes	No
Limestone reservoir	19 Jun 2006	6305	12 Jul 2007	4285	17	23
Kettle reservoir	21 Jun 2006	6305	19 Sep 2007	3520	15	0
Stephens Lake thalweg	15 Sep 2007	4285	17 Sep 2007	3520	15	13
Stephens Lake	02 Jul 2006	4561	02 Jul 2006	4561	2	0
Keeyask	13 Sep 2006	3423	-	Modelled	14	46
Keeyask 2008	28 Sep 2008	4090	28 Sep 2008	4090	1	25
Total					64	107

Table 3B-3: Forward stepwise logistic regression results for 5th and 95th percentile inflows using 2 or 3 explanatory variables. Decreasing values for the -2Log (likelihood) and Akaike's Information Criterion (AIC) as the number of variables increases indicates an improvement to the model. Values in bold show the best model

Percentile Inflow	No. of Variables	Variables	-2 Log (Likelihood)	Pr > Wald	AIC
5	2	Velocity/depth	72.319	0.000	80.319
	3	Velocity/exposure/depth	**71.463**	0.000	81.463
95	2	Velocity/exposure	60.915	0.000	68.915
	3	Velocity/exposure/depth	**57.971**	0.000	67.971

Table 3B-4: Standardized coefficients for 5th and 95th percentile inflows from a three variable logistic regression model for predicting deposition/no deposition in lotic areas of the lower Nelson River

Percentile Inflow	Source	Value	Standard Error	Wald Chi-Square	Pr > Chi²
5	Velocity	-4.946	0.812	37.145	<0.0001
	Exposure	0.225	0.330	0.467	0.494
	Depth	1.637	0.427	14.727	0.000
95	Velocity	-6.315	1.147	30.293	<0.0001
	Exposure	0.721	0.374	3.719	0.054
	Depth	0.596	0.363	2.700	0.100

Table 3B-5: Cross-validation results for 5th and 95th percentile inflow simulation showing classification agreement (%) of the Model and Test groups for predicting depositional and non-depositional substrata in lotic water masses. Model group n = 130; Test group n = 41

	5th Percentile Inflow		95th Percentile Inflow	
	Model Group	Test Group	Model Group	Test Group
No deposition	88.8	88.4	93.8	88.5
Deposition	87.7	73.3	87.8	80.0
Overall	88.4	82.9	91.5	85.4

Table 3B-6: Likelihood-ratio test demonstrating the effect of the explanatory variables against that of a model using only a constant

Statistic	DF	Chi-square	Pr > Chi²
-2 Log(Likelihood)	3	107.453	<0.0001

Table 3B-7: Forward stepwise logistic regression results using 1, 2, or 3 explanatory variables. Decreasing values for the -2Log (likelihood) and Akaike's Information Criterion (AIC) as the number of variables increases indicates an improvement to the model

No. of Variables	Variables	-2 Log (Likelihood)	Pr > Wald	AIC
1	Exposure	156.970	0.000	166.970
2	Depth/ Exposure	143.989	0.000	153.989
3	Depth/ Exposure/Slope	135.974	0.000	145.974

Table 3B-8: Standardized coefficients for a three variable logistic regression model for predicting the boundary between fine organic material and silt in flooded peatland bays

Source	Value	Standard Error	Wald Chi-Square	Pr > Chi²
Depth	-1.088	0.392	7.705	0.006
Slope	-0.727	0.274	7.036	0.008
Exposure	-1.511	0.306	24.315	< 0.0001

Table 3B-9: Cross-validation results showing classification agreement (%) of the Model and Test groups for predicting the boundary between fine organic material (FOM) and silt in flooded peatland bays. Model group n = 179; Test group n = 59

	Model Group	Test Group
Silt	74.7	76.0
FOM	83.7	82.4
Overall	79.9	79.7

Table 3B-10: Fishers discriminant function coefficients derived for the predictive reservoir model to estimate potential macrophyte habitat derived using linear discriminant analysis (LDA) representing the existing environment (EE) and pre-flood (PF) data. Model number is consistent with Appendix 3C. Model 3 assumes the EE substrate variable phi is unavailable in this future scenario. Adapted from Appendix 3C

Model #	LDA Model	Number of Variables	Class	Constant	EE				PF	
					Slope	Exposure	Depth	Phi	Mineral $Soil_{dist}$	Peat Depth
3	Predictive Reservoir	5	*M. sibiricum*	-13.3283	0.0622	0.0034	1.4159	-	0.0035	0.0923
			P. richardsonii	-11.5641	0.4847	0.0057	1.5413	-	-0.0022	0.0796
			Absent	-17.1007	0.7616	0.0053	1.9736	-	0.0054	0.0949

Table 3B-11: Classification agreement (%) for the predictive reservoir model (PR) to estimate potential macrophyte habitat using linear discriminant analysis (LDA). The Model group represents 75% of the available data (n = 471) and was cross-validated using the remaining Test data not used to build the model. Model number is consistent with the numbering of Appendix 3C

Model #	LDA Model	Number of Variables	Model Agreement (%)	Test Agreement (%)	Test (%)		
					M. sibiricum	***P. richardsonii***	**Absent**
2	EE	4	80.0	81.0	86.0	84.0	76.0
3	PR	5	78.0	81.0	86.0	86.0	74.0

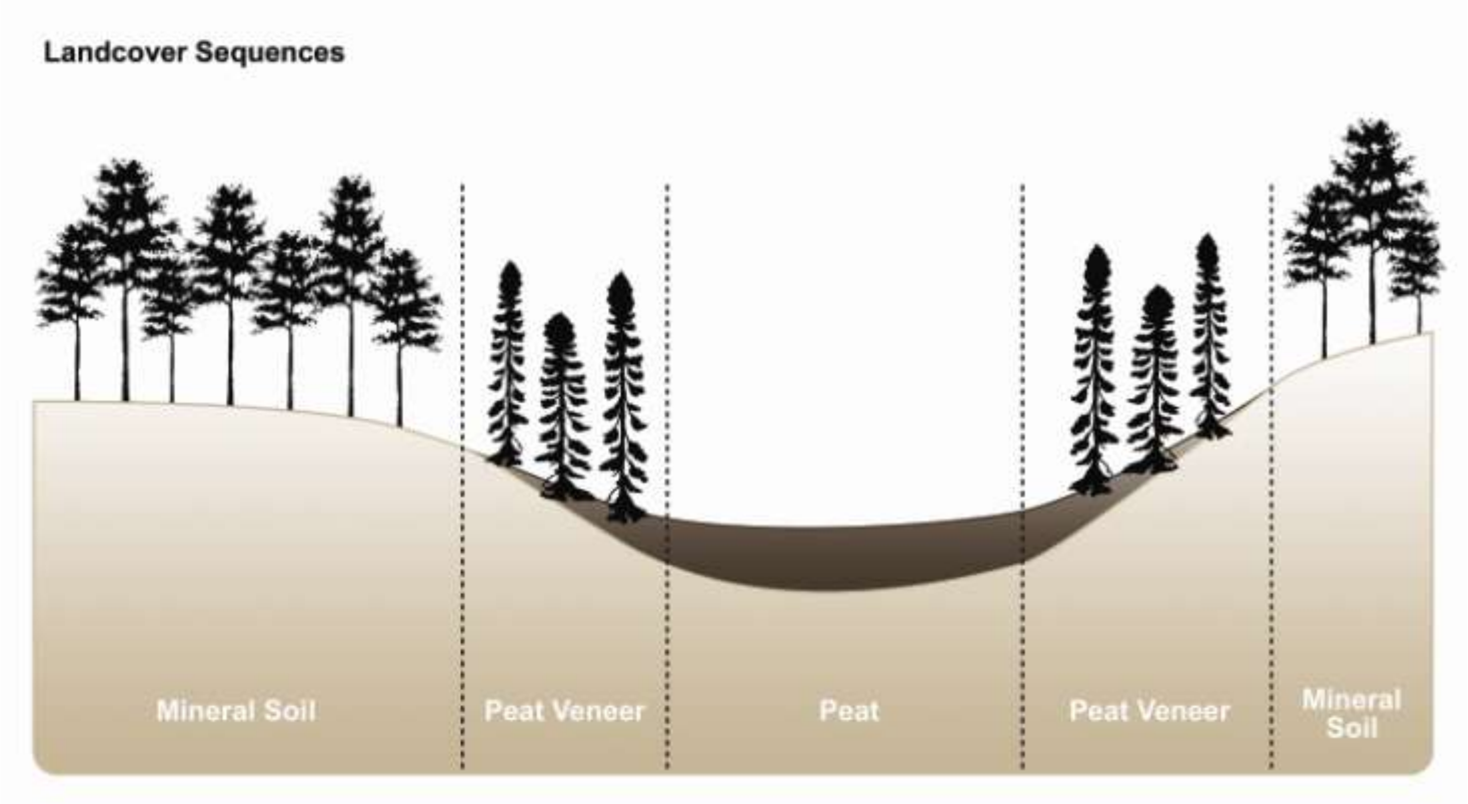


Figure 3B-1: **Schematic diagram of topographic sequences of forest and soil types present in the study area where a low relief and gently undulating topography is present. Peat veneer is also regarded as thin peat. Not to scale**

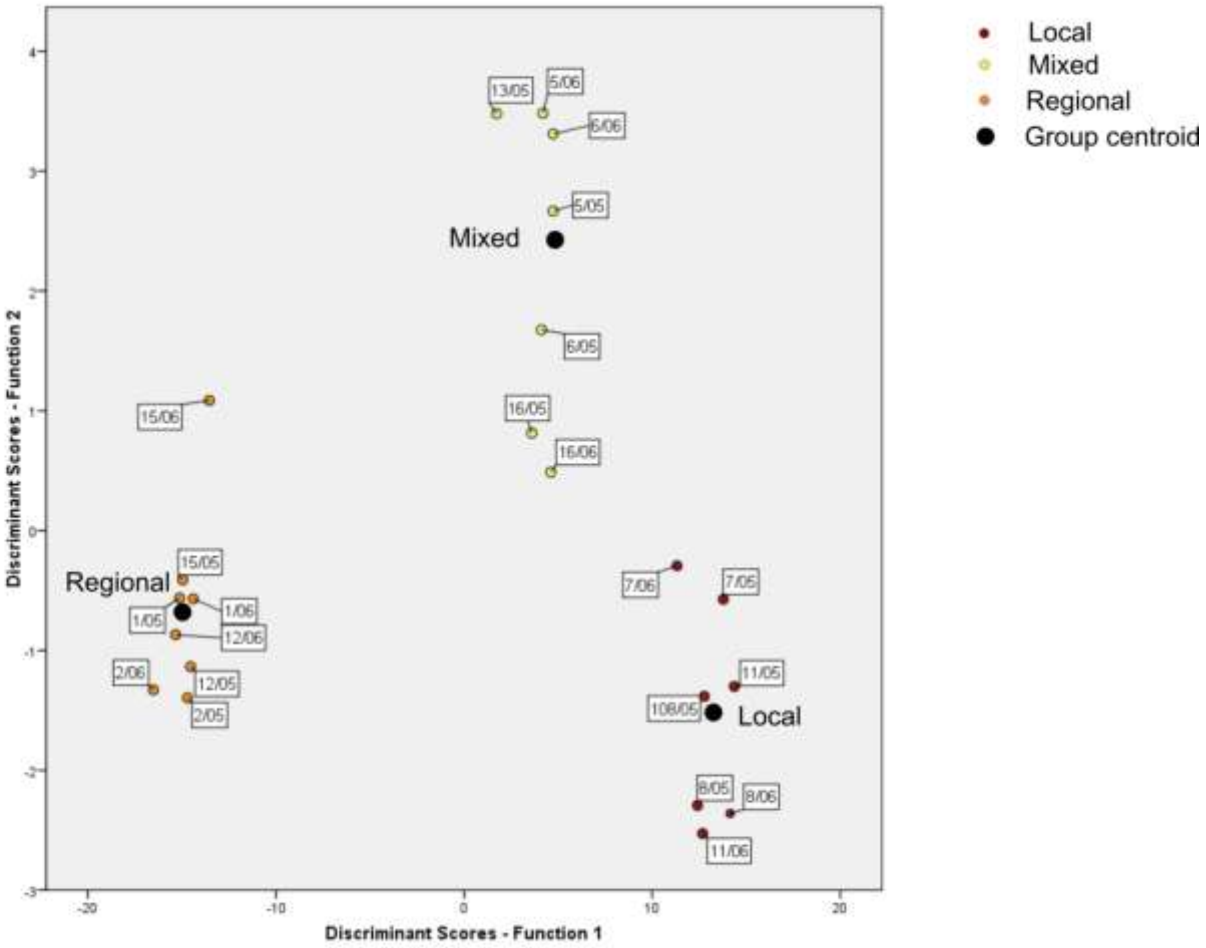


Figure 3B-2: Discriminant Analysis grouping of water quality and light attenuation sites shown in Map 3B-1. Sites are shown (1–15) by year (05/06) by water mass type

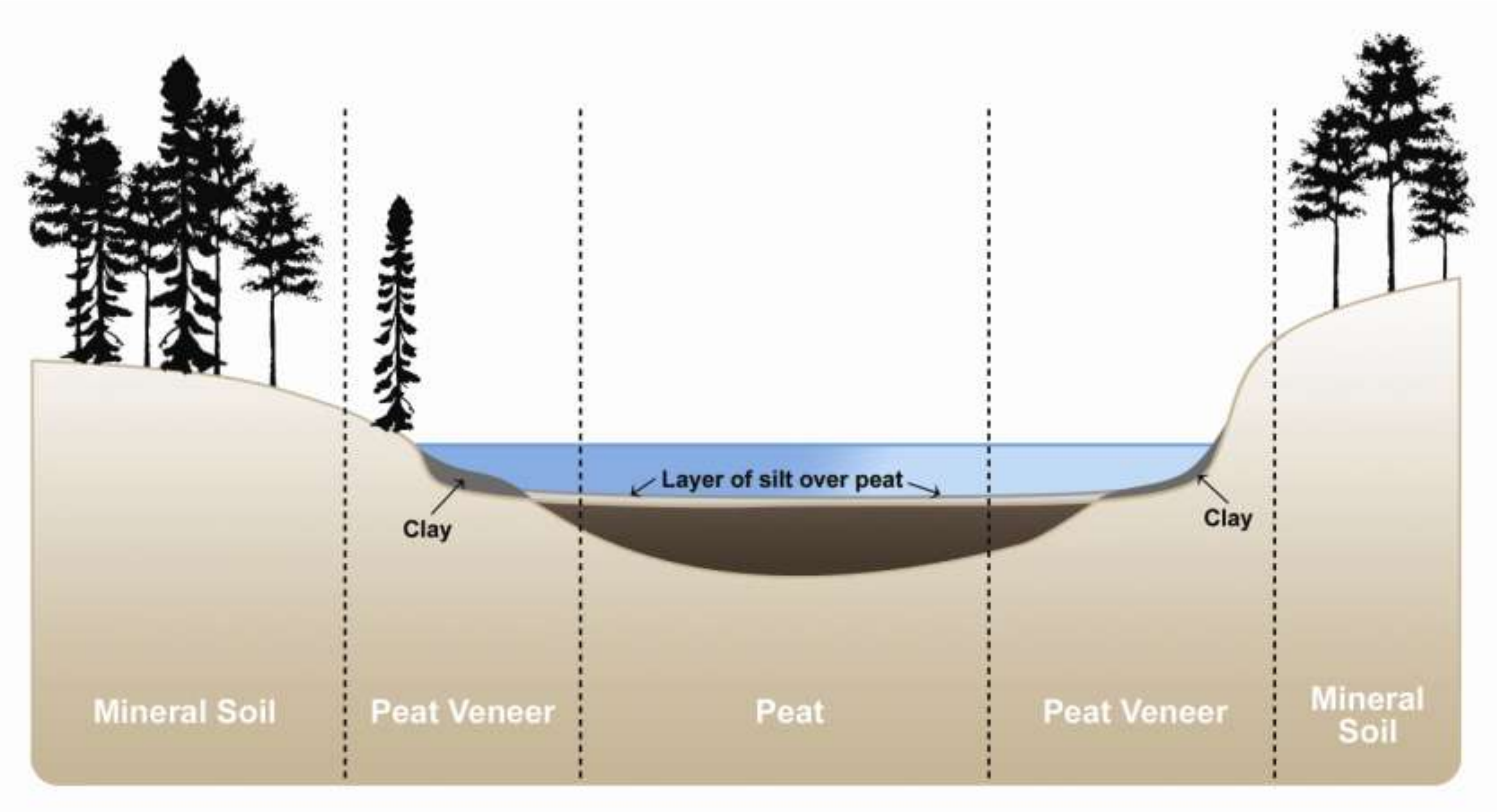


Figure 3B-3: Schematic illustration of post-Project flooded terrestrial habitat showing the development of a clay–aggregate nearshore matrix from a pre-flood peat veneer, and superimposition of silt over the pre-flood peat in deeper areas. Peat veneer is also regarded as thin peat

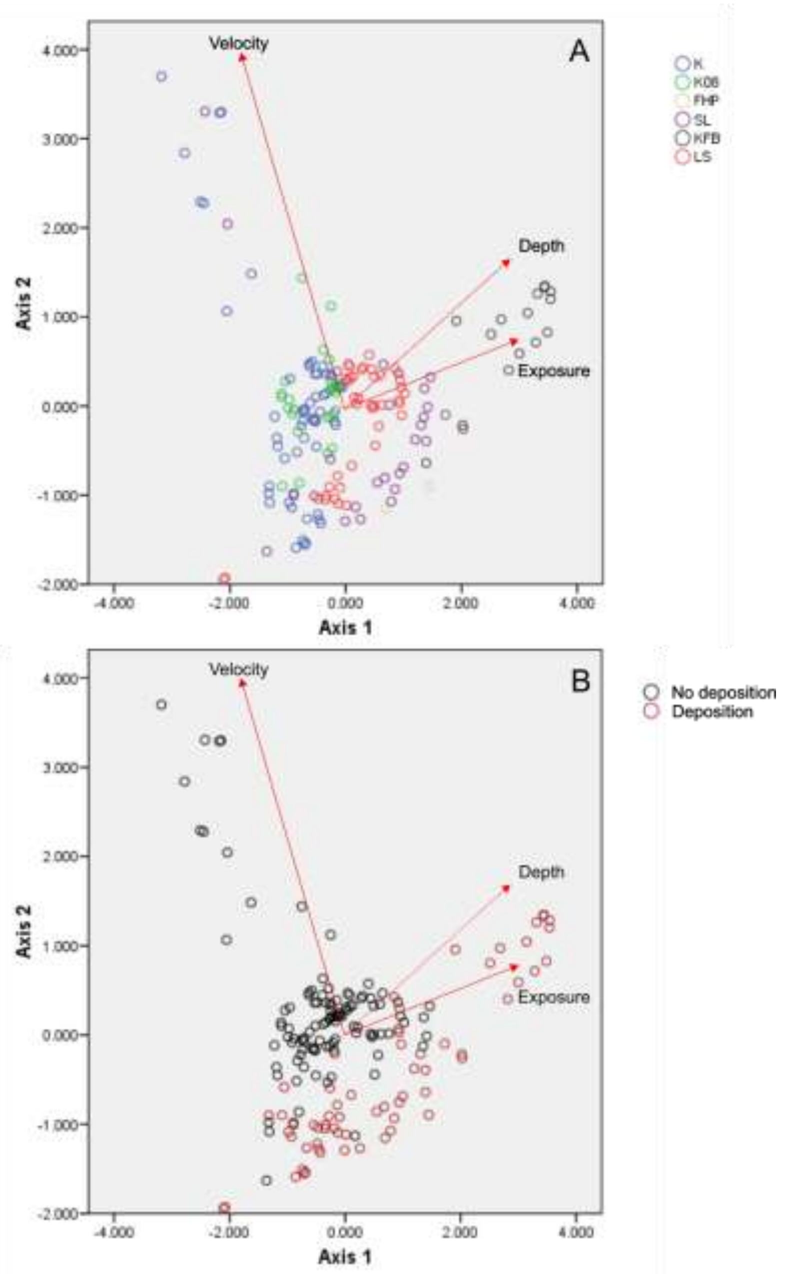


Figure 3B-4: **Principal component analysis correlation biplot for 95th percentile inflow scenario of the lotic deposition model showing: (A) scatter of data used to build the model by study area (K = Keeyask 2006, K08 = Keeyask 2008, FHP = Fish Habitat Preferences in Stephens Lake 2006, Stephens Lake thalweg studies 2007, KFB = Kettle reservoir 2006, LS = Limestone reservoir 2006) with arrows indicating correlation amongst variables and each PCA axis, and (B) the same data but classified according to deposition or no deposition as observed in the field**

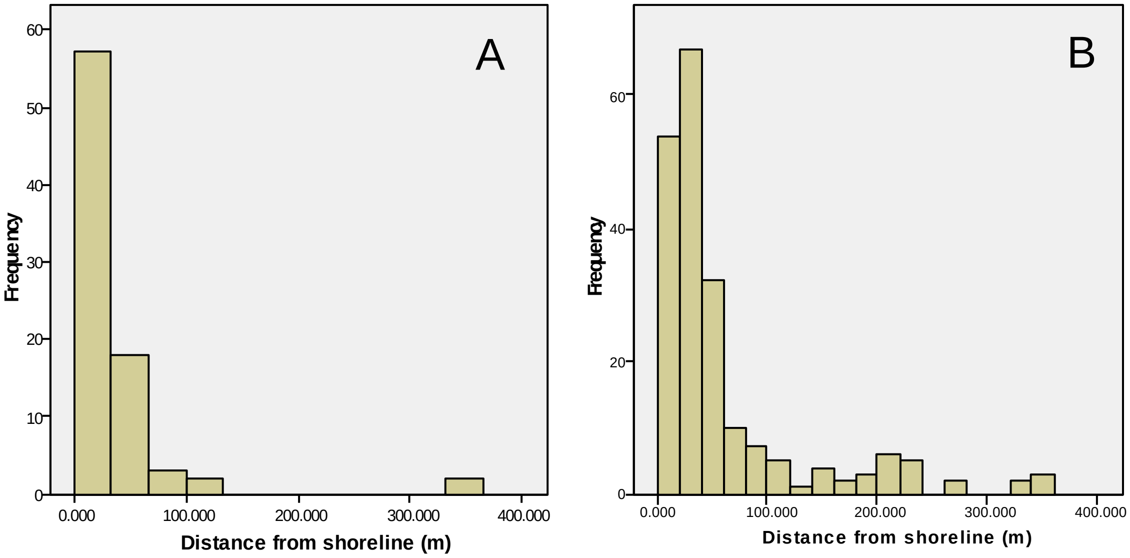


Figure 3B-5: Frequency histograms of the minimum distance of *M. sibiricum* (A) and *P. richardsonii* (B) to the shoreline of Stephens Lake at about the 95th percentile water level

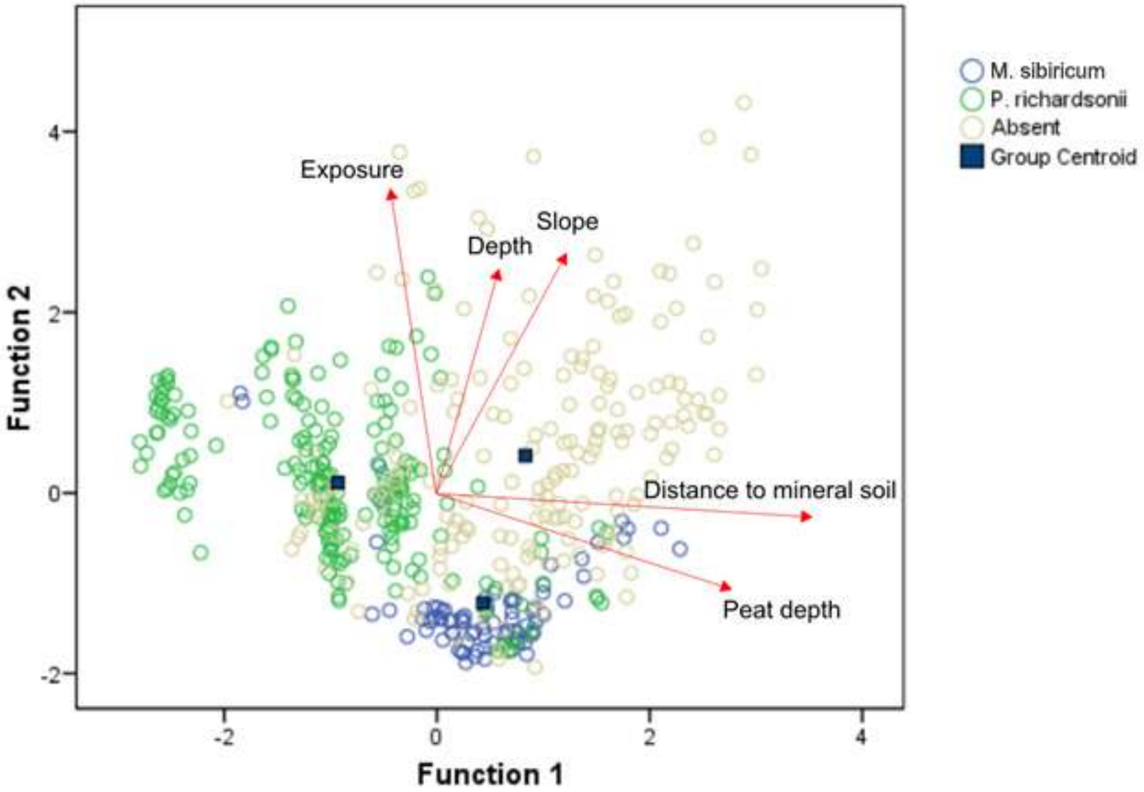


Figure 3B-6: **Discriminant analysis scatter plot showing the predictive macrophyte model built from Stephens Lake and applied to the proposed Keeyask reservoir at FSL. The predictive reservoir model contained three existing environment variables (exposure, depth, slope) and two pre-flood variables (distance to mineral soil, peat depth) that are surrogate variables used when the substrate grain size is unknown in this future scenario**