University of Victoria | Graduate Studies

Notice of the Final Oral Examination
for the Degree of Master of Science

of

# MORGAN MAGEE

BSc (University of Victoria, 2022)

## "The choice of prediction curve method and its effect on the estimated amount of DNA"

Department of Mathematics and Statistics

Monday, April 22, 2024
12:30 P.M.
David Strong Building
Room C126

Supervisory Committee:
Dr. Mary Lesperance, Department of Mathematics and Statistics, University of Victoria (Supervisor)
Dr. Caren Helbing, Department of Biochemistry and Microbiology, UVic (Outside Member)

External Examiner:
Dr. Julie Zhou, Department of Mathematics and Statistics, UVic

Chair of Oral Examination:
Dr. Michelle Wiebe, Department of Curriculum and Instruction, UVic

Dr. Robin G. Hicks, Dean, Faculty of Graduate Studies

## Abstract

Water samples that contained environmental DNA (eDNA) were taken from multiple rivers where eulachon fish (*Thaleichthys pacificus*) are known to spawn. The samples were split into 8 technical replicates and analyzed using quantitative real-time polymerase chain reaction (qPCR). A qPCR experiment is the real time quantification of DNA amount at the end of a full cycle of heating and cooling. $C_T$ values were determined from a qPCR experiment or a replicate was given NA (not available) if no DNA was detected. Four data sets from two different labs were used, BVL and UVic. Both labs have a gblock and field data set with chemical assays named eTHPA2 and eTHPA6. Gblock data is comprised of gblock samples which were synthetically constructed genes of known concentration (copy number) and measured using qPCR. Field data is comprised of samples taken from river sites in British Columbia where eDNA naturally occurs, and the copy numbers were unknown for the field samples. The field samples were analyzed using qPCR technique to determine the $C_T$ value for each technical replicate. Each data set was split into two subsets named full and partial detect, resulting in eight working data sets. The full data sets were comprised of samples whose technical replicates had (8/8) detects. The partial data set was comprised of samples whose technical replicates had less than (8/8) detects. A binary variable "detect" was defined for the partial detect data sets with the value 0 for $C_T$ = NA and 1 otherwise. This "detect" variable was used to convert the partial detect data to the Poisson mean ($\hat{\lambda}$) scale.

Standard/calibration and prediction curves were built from the gblock data. Standard curves were built using gblock data with known copy number values, and relate $C_T$ and $\hat{\lambda}$ values to copy number values. Standard curves were used to estimate copy numbers given $C_T$ or$\hat{\lambda}$ values for samples with unknown copy numbers. Prediction curves were built by fitting least squares and orthogonal regression using an unweighted and weighted method for each, to the gblock data. Prediction curves were used to estimate eTHPA6 $C_T$ or $\hat{\lambda}$ values given eTHPA2 $C_T$ or $\hat{\lambda}$ values. Plots and model summaries for the four prediction curves for each data set were analyzed. Based off the analysis and recommendation of the literature, weighted orthogonal regression was chosen as the best prediction model for each gblock data set. The prediction curves were applied to the corresponding field data to investigate how well the models predict the values of eTHPA6 given eTHPA2. Three out the four field data sets saw an under prediction in eTHPA6 values given eTHPA2 values, and the other saw over predicted eTHPA6 values given eTHPA2 values, using the prediction curves built from the gblock data set.