### LVI Evaluation of Photo Sampling Intensity for the kNN

Our current understanding of the photo sampling procedure is as follows (Chris Buston, pers. comm. 2013):

LVI starts with segmentation of the Landsat images. Once non-treed areas segments have been accounted for, and once recently disturbed areas (stand ages less than 40 years old) have been removed, what remains are the segments representing the population of interest. Strips are then flown to acquire aerial photography. In the case of Quesnel this involved the use of the Digital Camera System and in the case of Williams Lake digital colour infrared photography. It is assumed that these strips are located systematically starting at a randomly selected point with an orientation that is also selected randomly. However, it is understood that this kind of process may not always be feasible in very mountainous regions. Also the sampling grid may simply be overlaid onto the National Forest Inventory grid to capitalize on that sampling program. Beyond that it is assumed that individual segments are then selected for sampling along the systematic grid at regular intervals, once again starting at a randomly selected location within, and along the centreline of the previously established strips. Where the points intersect with a given segment, that segment is then classified as a whole according to VRI standards.

Complications do arise in this process. One complication is that from a photo interpreter’s perspective the preferred option might be to further subdivide the segment into distinct polygons to recognize forest types that would be declared as being different according to VRI standards; the current approach is based on requiring the interpreter to estimate a single set of descriptors. An alternative would be to proceed with subdividing the segment as necessary and to assign each subdivision an appropriate VRI label. The Enhanced LVI process would have to be developed to allow for this condition; procedures are available to do this.

Another complication is that in some cases the chosen segment may extend outside of the photo sample strip; this case is currently being handled by rejecting the photo sample and moving to the next one. As a result this becomes a problem of missing data … a problem for which the nearest neighbour concept was originally designed to address. This is an important aspect of the process that relates to sampling intensity as a function of the proportion of cases when this condition holds true. It is also important to identify the partial segment and associated metrics as part of the reference dataset, albeit with missing VRI data. Once again the Enhanced LVI process would have to be developed to also allow for this condition. This and the previous condition would then be integrated into the evaluation of sampling intensity as outlined below.

Another complication relates to the inclusion of ancillary attributes, for example those leveraged from a previous VRI inventory and associated with each individual segment. Under conditions where a segment involves intersection with more than one polygon, the current process appears to resolve this problem by simply assigning the one that is most dominant. An alternative would be to include each and every one of the polygon attributes (perhaps excluding some polygons below a minimum size) so as to better represent the original data. All of these decisions and issues affect the final product and relate to the manner in which samples are utilized to derive population estimates.

Finally if consideration is also being given to inclusion of NFI photo plot inventory data, then this would represent a considerably different sample design involving within and between photoplot sources of error. In addition greater consideration would have to be given to spatial autocorrelations as factor influencing sampling error. It is assumed that this sample design is not included as part of the request for proposals.

It is proposed that the above mentioned issues also be considered within the context of the evaluation and development of the enhanced LVI process. The basic evaluation of sampling intensities will then be completed using the reference dataset from Williams Lake as a known population. A Monte Carlo sample simulation model will be used to look at different sampling sizes and intensities. One hundred systematic samples will be drawn using a random start representing three (or more) sampling intensities. The variable selection process, model calibration, variable weighting, and nearest neighbour assignments (kNN = 1, kNN = 5) will be applied in each case. Summary statistics will be developed to represent the quality of the outcomes with respect to a defined Y-variable dataset. These will be agreed upon in advance with the MoFLNRO. For example, for any given Y-variable, the predicted values could be regressed against the actual values using the unsampled portion of the population to produce an overall empirical distribution of the expected errors of estimation. Biases could also be summarized at standard intervals between the minimum and maximum values and presented as box plots. With respect to species distributions the results could be summarized in terms of the proportion of cases (median, 25th and 75th percentile as well as minimum and maximum) where there is agreement in the leading species, as well as leading plus secondary species.

The results of this analysis are