Software and Instructions for kNN Applications in Forest Resources Description and Estimation

by

Reija Haapanen and Alan R. Ek

STAFF PAPER SERIES NUMBER 152

Department of Forest Resources
College of Natural Resources
and
Minnesota Agricultural Experiment Station
University of Minnesota
St. Paul, Minnesota

Software and Instructions for kNN Applications in Forest Resources Description and Estimation

by

Reija Haapanen and Alan Ek[1]

June 2001

STAFF PAPER SERIES NUMBER 152

Department of Forest Resources
College of Natural Resources
and
Agricultural Experiment Station
University of Minnesota
St. Paul, MN 55108

[1]Research fellow and professor, Department of Forest Resources, University of Minnesota, St. Paul, MN; Research supported by NASA, the USDA Forest Service, the National Council on Air and Stream Improvement, the University of Minnesota's College of Natural Resources and Minnesota Agricultural Experiment Station.

Contents

1. Background

The k-nearest neighbors (kNN) method has proven to be a very useful technique to classify and propagate forest field plot information through the landscape. The kNN method was first applied to forest inventory in Finland in the late 1980s and it has been used operationally in the Finnish National Forest Inventory (NFI) since 1990. This classification and estimation process reproduces the covariance structure of the observed data and retains the full range of variability inherent in the sample. Tomppo (1991) and Tokola et al. (1996) describe the use of the kNN method to produce localized estimates and maps.

Testing and the development of software and procedures for applications in the U.S. was begun by Hector Franco-Lopez and Alan Ek in the late 1990s. The applications described here are for use with Landsat TM satellite imagery and USDA Forest Service Forest Inventory and Analysis (FIA) data for Minnesota. However, these applications can be readily adapted to other imagery and forest inventory data formats. The software provides a range of statistical and map analysis and output. An example of map output from the software is shown in figure 1.

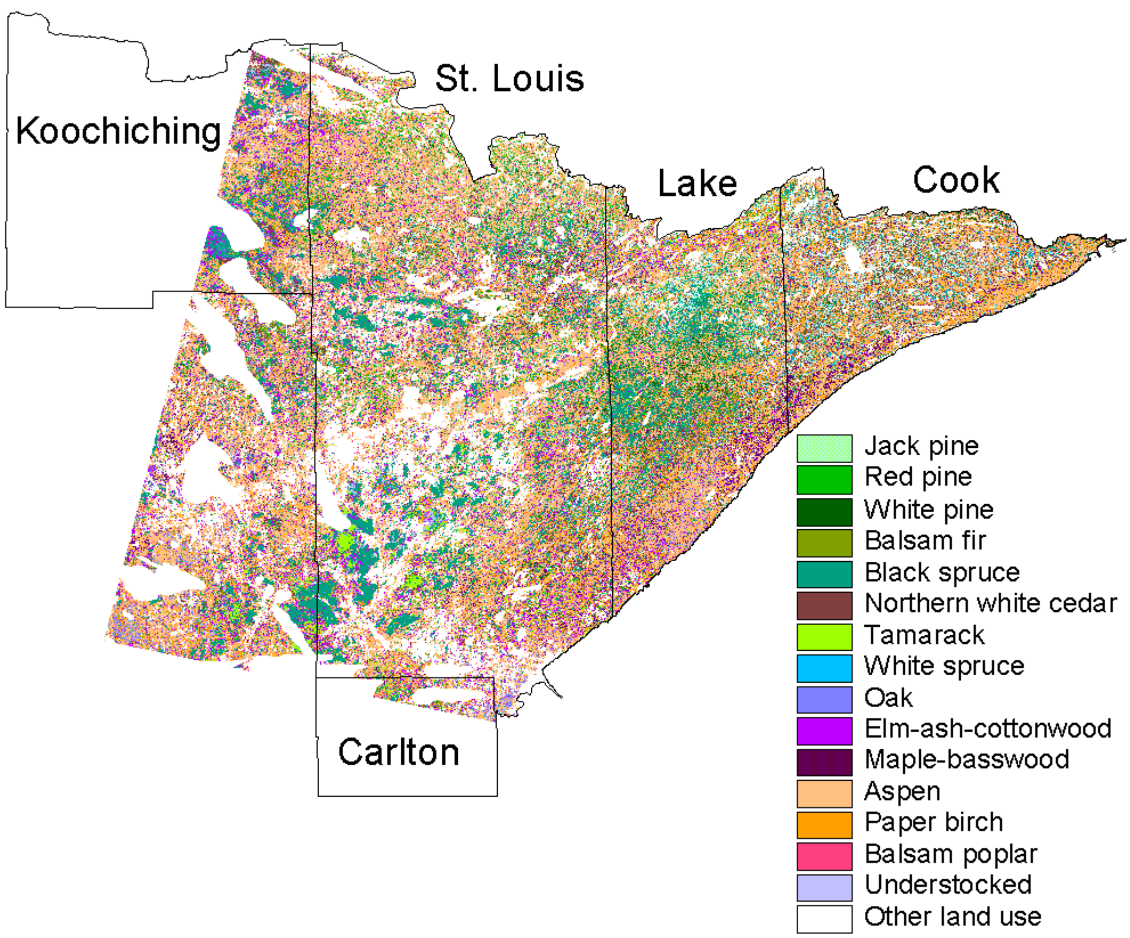

Figure 1. A forest cover type classification map based on the kNN method, Landsat TM, and FIA field plot data for a portion of a five-county area in northeastern Minnesota.

2. Materials

Minimum requirements for the use of the software and procedures describe here are:

- A geo-referenced satellite image in generic binary (BIP, band interleaved by pixel) format;
- Field plot data in text file format and containing the following values:

  Location of plots by x-and y-coordinates, corresponding satellite image spectral values from all bands, timber volume (V) and basal area (BA), or forest type class for the estimation of these variables. Note the coordinates must be in the same projection as the image.

One may also include:

- Background mask for the areas excluded from estimation or classification—such as 0-values outside the satellite image, or for lakes, roads, urban areas, clouds, and other areas outside the area of interest. This step also speeds up the estimation.
- Upland/lowland mask to stratify these spectrally very different land areas. Thus, the field data file must have upland/lowland information in the last column.

At the current stage of software development the image files must be in BIP format. The images can be easily exported to this format with Erdas Imagine's import/export tool. The image files must also be of the same size and location (origin, rows, columns), since the program uses them pixel by pixel and not geographically. Users must also write down the top-left origin of the image and the number of rows and columns, e.g., from Imagine's image info view, to use the horizontal search radius for the reference data. This is also needed to use the stand-alone program that picks the image values from the BIP image.

To illustrate steps described so far, figure 2 shows the resulting layout of field plots on the imagery.

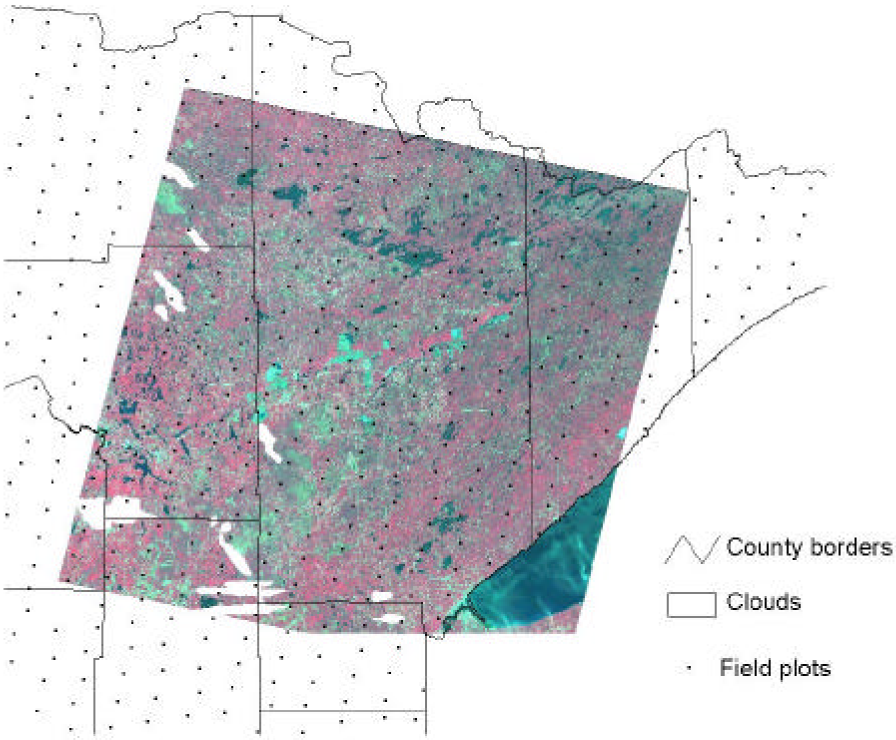

Figure 2. The 1st year (1999) panel of new FIA 4-subplot clusters overlayed on a Landsat TM image.

3. General Information

3.1. The program package and computing environment

The programs are coded in a DOS-environment so they can be compiled in either Unix, Linux or Windows. They are written for the most part in C, but utilize some features of C++.

The kNN software package includes the following programs:

- "**get_values**" for picking the image values from satellite image
- "**4_subplots**" for generating the subplots 2-4 when the FIA center plot is known (this is in accordance with the new design and assumes that the coordinates are in UTM, units meters)
- "**window**" for generating a 3x3, 5x5 or 7x7 window of coordinates around the field plots
- "**optimize**" for finding the weights of satellite image bands
- "**validate**" for calculating cross-validations, bootstrap-errors, biases and general descriptions of the field data
- "**mapmaker**" for generating the actual kNN map

Additionally, the package also includes some utilities functions, which are in the directory "General". Those are compiled to a library that all the kNN programs will use.

"**Base**" directory includes "Base" class, which inherits all important features from other utilities classes. Some of kNN programs inherit "Base" class and others just use its features.
"**Utilitiesclass**" directory includes "Utilities" class, which was written in C++ language.
"**Utilities**" directory includes various functions in C.
"**Initialize**" directory includes "Initial" class, which is used to initialize variables.
"**List**" directory includes list data structures.

If your platform is Unix or Linux, you must write the symbolic links into the ‘include’ directory before compiling the files (project).

The command is: ln -s ../list/list.h
etc.

list.h
element.h
utilities.h
utilitiesclass.h
initialize.h
base.h
element.h

You also have to make a directory named "template" into the "include" directory, where there is a symbolic link to list.cxx

The command is: ln -s ../../list/list.cxx

The compiling order is:

1. List
2. Utilities
3. Utilitiesclass
4. Initialize
5. Base

If the platform is other than Unix or Linux, these or other tasks will depend on the specific compiler.

It is assumed that the user has access to some image processing and GIS software such as Erdas Imagine and Arcview or Arc/Info.

3.2. Field data

The volume and basal area programs assume the metric system, thus volume is presented as cubic meters per hectare and basal area as square meters per hectare. The units are not important when you run the optimization program. However, if you use some other units with the cross-validation package component, then you will need to change the class limits for the confusion matrices. Also, the output image for the "mapmaker" program produces an 8 bit-image, thus the value "255" is a limiting factor for volumes.

The programs assume the USDA Forest Service FIA codes for the forest covertype:

1 Jack pine
2 Red pine
3 White pine
6 Exotic
11 Balsam fir
12 Black spruce
13 Northern white-cedar
14 Tamarack
15 White spruce
50 Oak
70 Elm-ash-cottonwood
80 Maple-basswood
91 Aspen
92 Paper birch
94 Balsam poplar
99 Understocked, the species can not be determined

The programs recode these covertype codes into 0,1,2,3,…,15 in order to process the calculations (see coding below). The output from ‘mapmaker’ and cross-validation is again recoded to 1,2,3,4,…,16. If your covertype codes are different or you have a larger set of codes, you will need to change the programs (optimization, cross-validation and mapmaker) and covertype coding accordingly:

```
for (int i=0;i<NPLOTS;i++)
   {
     if (classVar0[i] == 1) classVar[i] = 0;
     if (classVar0[i] == 2) classVar[i] = 1;
     if (classVar0[i] == 3) classVar[i] = 2;
     if (classVar0[i] == 6) classVar[i] = 3;
     if (classVar0[i] == 11) classVar[i] = 4;
     if (classVar0[i] == 12) classVar[i] = 5;
     if (classVar0[i] == 14) classVar[i] = 6;
     if (classVar0[i] == 15) classVar[i] = 7;
     if (classVar0[i] == 16) classVar[i] = 8;
     if (classVar0[i] == 50) classVar[i] = 9;
     if (classVar0[i] == 70) classVar[i] = 10;
     if (classVar0[i] == 80) classVar[i] = 11;
     if (classVar0[i] == 91) classVar[i] = 12;
     if (classVar0[i] == 92) classVar[i] = 13;
     if (classVar0[i] == 94) classVar[i] = 14;
     if (classVar0[i] == 99) classVar[i] = 15;
   }
```

3.3. Satellite images and sample size

For best results, images from several phases of the subject species phenology should be used. The most important image dates for forest type classification are late spring and for volume and basal area late spring and summer. The images should be as cloudless as possible. The most important requirement is a sufficient number of field plots distributed throughout the area of an image(s) to be classified. Thus, for classifying only a small part of a Landsat image, you probably need a much larger area to encompass enough field plots for the classification (Scandinavian studies suggest a minimum of approximately 500 forested field plots per image, e.g., Nilsson 1997).

3.4. kNN algorithm

The kNN algorithm itself is straightforward:

1. For a subject satellite image pixel
2. Calculate Euclidean distances to all pixels carrying field data
3. Sort field plots by Euclidean distances
4. Take k (k = 1, 2, 3,…5 or whatever) nearest field plots and estimate the desired forest variable as a mean, weighted mean or mode (class variable) of the k nearest observations
5. Proceed to the next pixel.

4. Features Specification

The following options can be implemented by appropriate input specifications.

4.1. Stratification by upland and lowland

This procedure gives more reliable results. The spectral response of peatlands differs from that of the mineral soils with the same growing stock. Further, some peatlands cannot be separated from mineral soils. Figure 3 illustrates the effect of stratification on cover type classification.

4.2. Horizontal search radius

This is the maximum distance for inclusion of field plots to be applied in the estimation of subject pixel values. This distance may be different for upland and lowland.

4.3. Minimum distance for field plots to be used in estimation

This is an option implemented for plots composed of a cluster of subplots such as with the new FIA plot design (a four subplot cluster). In the error estimation programs, the nearest neighbors tend to be found in the same cluster. With this minimum distance implemented, the inclusion of subplots from the same cluster can be prohibited. This option is used only in cross-validation analysis.

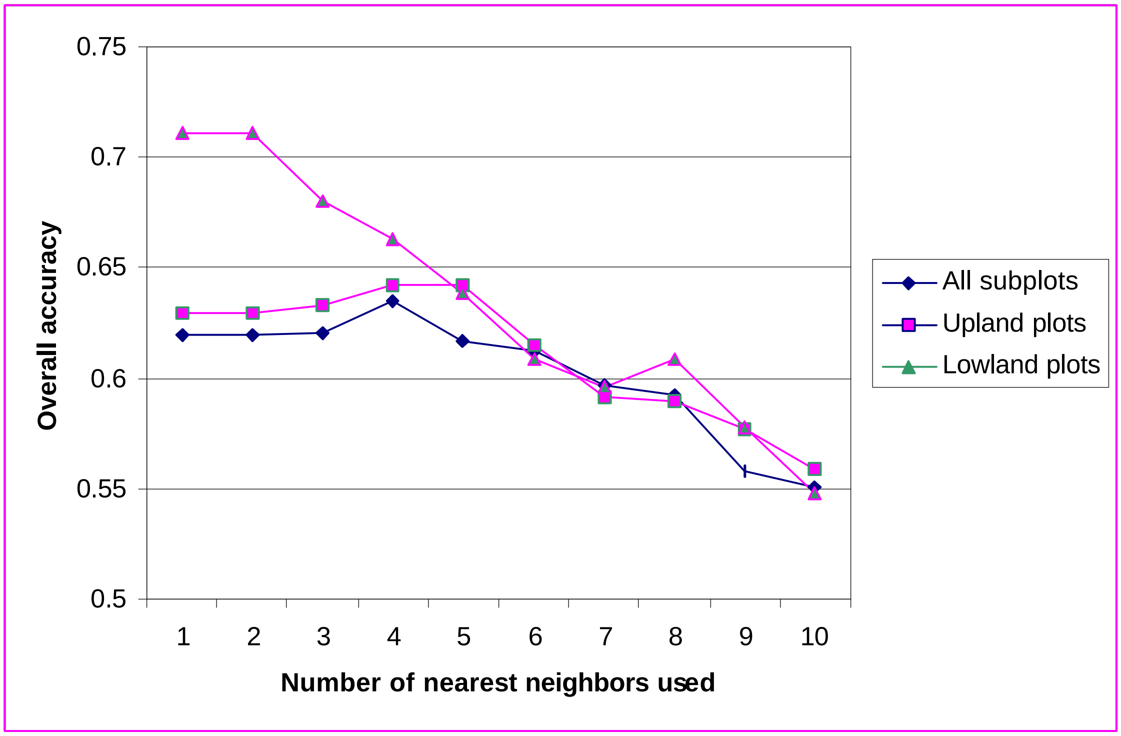

Figure 3. Estimates of accuracy for forest cover type classification, when upland-lowland stratification is used.

4.4. Use of mask image

When a mask image is provided as input to the mapmaker program, the program skips calculations when the "background" is true for a pixel and inserts the value "255" for the result. This speeds up the classification significantly and also improves the map. Given 255 as the background or mask value, the greatest output value for volume or basal area is 254. This mask value "255" was chosen in order to save the value "0" for volumes of 0 m3/ha. The mask image can contain, e.g., state borders, other land use classes (water, urban areas, cultivated land) and clouds.

5. Program Inputs and Programs

5.1 Program inputs

Most of the programs utilize a "header" text file designated "parameterfile.txt," which includes all the parameters used in or to specify computations. All these files are described below, but they are used selectively in the programs—see the listing of parameterfile.txt at the beginning of the respective program descriptions:

<u>Inputfile</u>

This is the user data in a text file. Note there is no header row, data is separated by spaces or tabs, and the following order is used for optimization, cross-validation and mapmaker programs:

x-coordinate	y-coordinate	ch_1	ch_2	ch_3	ch_4	ch_5	ch_6	ch_7	volume	ba	land_use
521970.72	5177875.12	63	28	19	120	85	166	26	159.82	38.93	1
521970.72	5177911.69	65	27	19	121	69	166	20	86.7	17.22	1
522002.4	5177856.83	62	23	17	98	65	166	16	106.32	39.86	1
521939.05	5177856.83	63	24	17	90	69	166	19	53.66	16.4	1
472465.74	5318991.12	64	24	19	78	63	164	18	0	29.46	1
472465.74	5319027.7	60	23	18	76	58	165	20	36.55	9.52	1
472497.41	5318972.83	62	23	20	77	59	164	18	35.21	29.99	1
472434.06	5318972.83	62	24	19	68	61	166	19	16.09	18.45	1
486392.79	5312647.16	69	29	22	121	102	166	28	26.49	5.1	0
486392.79	5312683.74	69	30	23	118	105	167	33	0	0	0
486361.11	5312628.87	67	28	23	114	102	167	29	83.18	14.1	0

Outputfile
The file name for the output.

Nmax
The number of iterations used in Nelder & Mead optimization.

Maxline
This is used in reading the input text files; It has no role in calculations.

Coeffsfile
This is the file carrying coefficients for image bands. The coefficients come from the Nelder & Mead optimization. The next item specifies whether it is to be used or not.

Coefficients
Enter 0 if you do not have coefficients or do not want to include them in calculations. Enter 1 if you have specified a coefficients file is to be used.

N_all_bands
The total number of satellite image bands in a user data set

Nbands
This is used in the error estimation programs (cross-validation). It indicates the number of bands to be used. Note, users are limited to a continuing set of bands, e.g., 0-9, 23-25, 1-17. This option allows for testing the use of separate dates, some bands within a date, etc. In all other programs, this is the total number of image bands in your data set.

Nplots
The number of records in your data set.

Kneig
The number of neighbors to be used in calculating kNN based estimates.

Howmanyclasses
The number of classes used in forest type classification.

Bootstrapsamples
The number of bootstrap samples to be developed. To skip the bootstrap estimate of error, enter 0. The bootstrap error calculations take more time than the cross-validation error estimation.

Weightfunction
This item specifies the weighting function for Euclidean distance:

0 = equal weights
1 = weights inversely proportional to the distance (1/d)
2 = weights inversely proportional to the squared distance ($1/d^2$)

Sortband
This item is used in two situations:

- Error estimation programs when using bootstrapping
- Mapmaker program

The sorting of bands is used to speed the nearest neighbor search and "sortband" specifies which band to use for sorting. Note that when sorting, the result is not always the nearest neighbor. This is the cost of optimizing the processing time.

In error estimation programs, if you choose to test a smaller set of bands within your data set (say, you have 27 bands and you want to use bands from 0 to 8 [1-9 in real life]), be sure to set this "sortband" to a number <= 8. See <u>N_all_bands</u> and <u>Nbands</u> above. However, if you do not use the bootstrap procedures, this selection is ignored.

Radius
This is the horizontal distance limit for the nearest neighbor search with uplands. We recommend using it only with a sufficient amount of data, that is, at least 300-500 observations should be left inside your search radius. Enter the radius in meters. These values can lie between 40-100 kilometers, that is 40,000–120,000 meters. If you do not plan to use this radius, enter 0. A separate radius can be specified for both upland and lowland.

Radius2
This is the horizontal distance limit for nearest neighbor search for lowlands. See the explanations above and below. A rule of thumb is that lowlands can effectively use a greater radius than uplands.

Peat
If you have access to a lowland or peatland mask and have entered values for that into your field data file, enter 1. Otherwise enter 0. The lowland should be recoded as 1 and the upland as 0.

Drop own cluster
The new 4-plot FIA design typically provides precise results for forest cover if all subplots are used. This result is a function of the very close proximity of subplots within

a plot (see figure 4). Use this option to drop the subplots from the same cluster while calculating cross-validation error. If you want to drop the closest subplots, enter 1, otherwise enter 0.

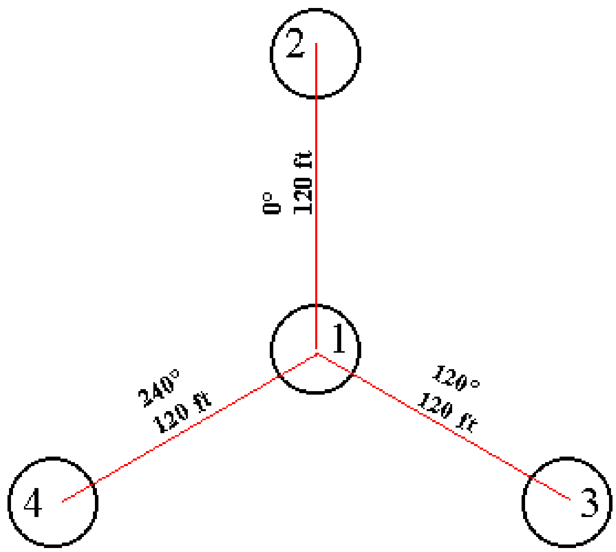

Figure 4. Layout of the new FIA plot with four subplots.

<u>Drop_threshold</u>

This item allows the user to specify a minimum geographical distance between subplots in error calculations. To drop all other subplots in the same cluster, use, e.g., 70 (distance in meters). To drop subplots closer than 40 meters, use 40. Figure 5 illustrates the effect of limiting the inclusion of subplots.

<u>Startb</u>

Default 0. If you specified a different number in <u>N_all_bands</u> and <u>nbands</u>, use this parameter to set your starting band. Note that the indexing begins at 0 (= band 1).

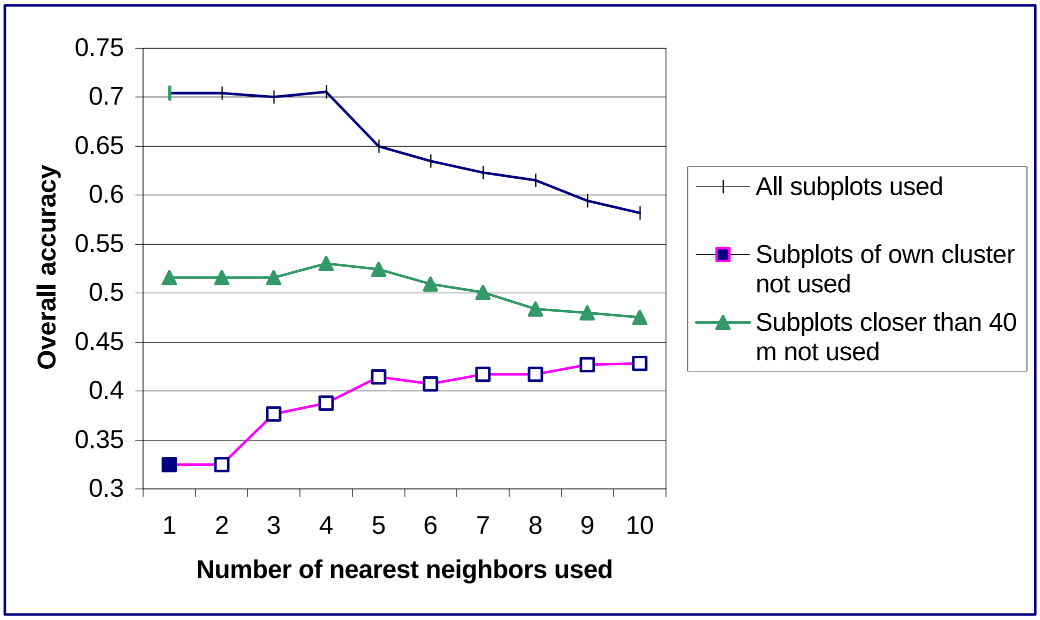

Figure 5. Effect on overall accuracy results from limiting the inclusion of FIA subplots.

5.2. OPTIMIZATION

This program calculates the Nelder & Mead optimization (Nelder and Mead 1965) in order to find the optimal weights for different bands and reduce the residual mean square error (RMSE) of estimates. The algorithm is somewhat sensitive to local optima and thus needs to run from several different starting points to find the overall or best weights, that is, the set closest to the "real" optimal weights. The program requires a long time to run. The suggested different starting points are set to 1,2,3,…,20. The user can also refine the program to run for a reduced set of starting points by modifying the following code and in particular the values for k:

```
    for (int k=1;k<=20;k++)
    {
       lambda=(double)k;
       for (int i=1; i<=NBANDS;i++)p[1][i]=1;
       for (int i=2; i<=NBANDS+1;i++)
       {
             for (int j=1;j<=NBANDS;j++)
             {
                   if (j==i-1) p[i][j]=1+lambda;
             else p[i][j]=1;
             }
       }
```

The parameterfile.txt for the optimization program is written as follows with HOWMANYCLASSES used only in forest type optimization runs:

```
INPUTFILE  p691.txt
OUTPUTFILE output691.txt
NMAX 5000
MAXLINE 600
```

```
NBANDS 27
NPLOTS 691
KNEIG 1
WEIGHTFUNCTION 0
HOWMANYCLASSES 16 (for class variables only)
RADIUS 0
PEAT 1
```

5.3. CROSSVALIDATE

Cross-validation is handled with two programs, one for continuous variables and the other for class variables.

The parameterfile.txt for these programs is

```
INPUTFILE  p696.txt
OUTPUTFILE output.txt
COEFFSFILE coefficients.txt
COEFFICIENTS 0
MAXLINE 600
N_ALL_BANDS 34
NBANDS 34
NPLOTS 696
KNEIG 1
BOOTSTRAPSAMPLES 0
HOWMANYCLASSES 16 (for class variable only)
WEIGHTFUNCTION 0
SORTBAND 0
RADIUS 0
RADIUS2 0
PEAT 1
DROP_OWN_CLUSTER 0
DROP_THRESHOLD 70
STARTB 0
```

Note that currently the use of upland/lowland stratification (PEAT), horizontal distance limit for neighbor search (RADIUS & RADIUS2) and DROP_OWN_CLUSTER and DROP_THRESHOLD work for cross-validation estimation of error only, not for apparent error or bootstrap errors.

The output of volume and basal area cross-validation is illustrated below:

```
INPUT FILE: p696_new.txt
OUTPUT FILE: output.txt
DISTANCE: EUCLIDEAN
WEIGHTING FUNCTION: NO WEIGHTS.
NUMBER OF PLOTS: 696
NUMBER OF BANDS: 34
NUMBER OF NEIGHBORS: 1
NUMBER OF BOOTSTRAP SAMPLES: 0
PEAT:     1
VOLUME APPARENT RMSE = 0
BA APPARENT RMSE = 0
**ALL PLOTS**********
VOLUME MIN AND MAX                      = 0 and 384.71 m3
VOLUME MEAN                             = 64.1413 m3
```

```
VOLUME STANDARD DEVIATION              = 70.0947 m3
BA MIN AND MAX                         = 0 and  67.7 m2
BA MEAN                                = 15.4648 m2
BA STANDARD DEVIATION                  = 12.1298 m2

VOLUME CONFUSION MATRIX
           Reference 0-40     40-80    80-120   120-160    160-  m3
Classified                                                        User's acc.
   0 -   40 m3       217        54        35        12        23    0.636364
  40 -   80 m3        47        35        17        16        10    0.28
  80 -  120 m3        40        14        24        12        15    0.228571
 120 -  160 m3        15        20        13         7         8    0.111111
 160 -      m3        17        10         9        11        15    0.241935
Producer's acc.     0.645833  0.263158  0.244898   0.12069  0.211268

VOLUME CROSS-VALIDATION RMSE         = 81.1354 m3,   126.495 % of mean
VOLUME BIAS                          = -1.40832

BA CONFUSION MATRIX
           Reference 0-10     10-20     20-30     30-  m2
Classified                                              User's acc.
   0 -   10 m2       153        58        36        26    0.56044
  10 -   20 m2        59        66        48        14    0.352941
  20 -   30 m2        39        48        48        25    0.3
  30 -      m2        22        19        21        14    0.184211
Producer's acc.     0.56044   0.34555  0.313725  0.177215

BA CROSS-VALIDATION RMSE             = 14.5889 m2,   94.3365 % of mean
BA BIAS                              = -0.203434
---------------------------------------------------
**LOWLAND***********
NUMBER OF PLOTS                         = 164
VOLUME MIN AND MAX                      = 0 and 254.07 m3
VOLUME MEAN                             = 42.9926 m3
VOLUME STANDARD DEVIATION               = 53.712 m3
BA MIN AND MAX                          = 0 and  58.75 m2
BA MEAN                                 = 13.0046 m2
BA STANDARD DEVIATION                   = 11.4917 m2

VOLUME CONFUSION MATRIX
           Reference 0-40     40-80    80-120   120-160    160-  m3
Classified                                                        User's acc.
   0 -   40 m3        82         8        12         3         5    0.745455
  40 -   80 m3         7         7         3         0         1    0.388889
  80 -  120 m3         9         4         6         3         1    0.26087
 120 -  160 m3         4         0         2         0         0    0
 160 -      m3         4         1         1         0         1    0.142857
Producer's acc.     0.773585      0.35      0.25         0     0.125

VOLUME CROSS-VALIDATION RMSE         = 66.624 m3,   154.966 % of mean
VOLUME BIAS                          = -2.87561

BA CONFUSION MATRIX
           Reference 0-10     10-20     20-30     30-  m2
Classified                                              User's acc.
   0 -   10 m2        50        22         8         7    0.574713
  10 -   20 m2        20        14        10         2    0.304348
  20 -   30 m2         9         6         6         1    0.272727
  30 -      m2         3         5         0         1    0.111111
Producer's acc.    0.609756  0.297872      0.25 0.0909091

BA CROSS-VALIDATION RMSE             = 14.8892 m2,   114.491 % of mean
BA BIAS                              = -0.42872
---------------------------------------------------
**UPLAND**********
NUMBER OF PLOTS                       = 532
VOLUME MIN AND MAX                    = 0 and 384.71 m3
VOLUME MEAN                           = 70.6608 m3
VOLUME STANDARD DEVIATION             = 73.2422 m3
BA MIN AND MAX                        = 0 m2 and  67.7
BA MEAN                               = 16.2232 m2
BA STANDARD DEVIATION                 = 12.2307 m2
```

```
VOLUME CONFUSION MATRIX
           Reference 0-40     40-80    80-120   120-160    160-  m3
Classified                                                          User's acc.
   0 -   40 m3       134        42        23         9        20    0.587719
  40 -   80 m3        42        29        13        16         7    0.271028
  80 -  120 m3        27        14        20         9        13    0.240964
 120 -  160 m3        12        18        11         7         9    0.122807
 160 -      m3        15        10         7        11        14    0.245614
Producer's acc.     0.582609  0.256637   0.27027  0.134615  0.222222

VOLUME CROSS-VALIDATION RMSE         = 86.8592 m3,   122.924 % of mean
VOLUME BIAS                          = -0.0196241

BA CONFUSION MATRIX
           Reference 0-10     10-20     20-30     30-  m2
Classified                                                User's acc.
   0 -   10 m2        97        36        25        20    0.544944
  10 -   20 m2        44        45        40        12    0.319149
  20 -   30 m2        33        45        42        24    0.291667
  30 -      m2        17        18        22        12    0.173913
Producer's acc.    0.507853    0.3125  0.325581  0.176471

BA CROSS-VALIDATION RMSE             = 14.9178 m2,   91.9537 % of mean
BA BIAS                              = 0.167669
```

5.4. MAPMAKER

This set of programs processes the final map. Currently it interactively asks for the important parameters from the user. The run time is several hours for an entire Landsat scene, depending on the number of bands. With several image dates, you treat the images separately as in the example, or as one big image. If possible, consider the use of a background/cloud mask (background, lakes, roads, urban areas, clouds, etc.) to speed up the classification. The mask causes the program to proceeds directly to the next pixel when it finds "mask" values. The program creates also a text file named weights.txt, where the weights for all field plots are stored.

```
C:\KNN_2001\2001_mapmaker3>map
Enter your choice:
0 - Estimate volume
1 - Estimate basal area
2 - Forest type classification
2
Enter your choice:
0 - use only sat. image
1 - use forest/non-forest mask
2 - use forest/non-forest mask & peat mask
2
Enter number of satellite images, max 4 (enter -1 to quit the program):
4
Enter image file 1:
mar_9ch.bin
Enter image file 2:
apr_9ch.bin
Enter image file 3:
may_9ch.bin
Enter image file 4:
jul_7ch.bin
Enter mask file:
lu.bin
Enter peatmask file:
```

```
peat.bin
Enter plot file:
data_722.txt
Enter output file:
cover_type_peatmask.bin
Use coefficients?  0 - No, 1 - Yes
0
read_spec 1
Enter number of bands for the first image(enter -1 to quit the
program):
9
Enter number of bands for the second image(enter -1 to quit the
program):
9
Enter number of bands for the third image(enter -1 to quit the
program):
9
Enter number of bands for the fourth image(enter -1 to quit the
program):
7
Enter number of plots (enter -1 to quit the program):
722
Enter number of forest type classes (enter -1 to quit the program):
16
Enter number of nearest neighbors to be used in estimation
(enter -1 to quit the program):
1
Enter a weight function to be used in estimation
(enter -1 to quit the program):
0 - Equal weights
1 - Inversely proportional to the distance
2 - Inversely proportional to the
square distance
0
Enter the most variable band(enter -1 to quit the program):
Count from 0 in the order that bands appear in the input plot file
For example, if the 6th band is the most variable band, enter 5
13
nband : 34
plot_file ok
nplot: 722
sortband :13
(If you do not want to limit the search area give 0 )
Give the search radius for upland in meters:0
Give the search radius for lowland in meters:0
Split data: lowland or upland
```

5.5. 4_SUBPLOTS

This program creates the new FIA subplots 2, 3 and 4 when the cluster center (plot 1) is known. The coordinates must be in a metric system. This program reads the following parameterfile.txt:

INPUTFILE plots.txt
OUTPUTFILE mn_4subplots.txt
X 6
Y 1804

Where X is the number of columns in the input file and Y is the number of observations (plots) in the input file. The input file must have the following data:

```
State unit county plot subplot x-coordinate y-coordinate
```

5.6. WINDOW

This program creates a 3x3, 5x5 or 7x7 window around the existing coordinates. Actually, what is created are just coordinates. The program is run from the command prompt in the following style: *prog input_file window_size.* The input file must have the following data:

```
state unit county plot subplot x-coordinate y-coordinate
```

The output is named "three_window.txt", "five_window.txt" or "seven_window.txt", depending on the selected window size

5.7. GET_VALUES

This program gets image values for the field plots.
The "parameterfile.txt" has the format:

```
OUTPUTFILE pixelval_9072_apr.txt
INPUTFILE c:\knn_2001\2001_mapmaker3\apr_9ch.bin
COORDINATEFILE three_window9072.txt
COLUMNS 9433
ROWS 8175
NBANDS 9
ORIGO_X 419219
ORIGO_Y 5392534
PIXEL_SIZE 30
```

The input file has the following order:

```
state unit county plot subplot x-coordinate y-coordinate
```

6. Special Notes

6.1. Preparing multi-temporal images for kNN estimation

1. Import channels 1-5 and 7
2. Stack layers in imagine → interpretation → utilities → layerstack
3. Find control points.
4. Geo-reference setting the output coordinates so that they are multiplies of 30 meters, e.g. x = 425790, y = 5388780. This way seamless mosaicing of classified output maps is ensured and shifting of the images many times is avoided.
5. Create an image where common image area for all dates has value 1, other pixels 0.
6. Cut all images to same size, defined by common area, see 30-meter rule in step 4.

7. Cut "common area" image (step 7) to the same size
8. Digitize clouds for all dates in imagine (AOI-tools, create polygon area of interest)
9. Open "common area" image, open all cloud aoi’s on image
10. Select Raster tools → box select aoi → create a box over your clouds
11. Set clouds to 0 in raster tools with "fill area"
12. Save edits → save → top layer as
13. Export your cloud and image mask to grid
14. Select field plots in Arcview.

6.2. Processing of field data

6.2.1. Normal data set, one pixel per subplot

1. Place all of the subplots in Arcview
2. Add a grid theme, with image area (value 1) and clouds (value 0)
3. Get grid values for the plots (example scripts can be found from ESRI’s page)
4. Select plots with value 1
5. Create a new shapefile
6. Get the lowland values from land-use grid
7. Export to text file
8. Get satellite image values, date by date
9. Move the data to access
10. Add the field variables you wish to work with

6.2.2. Location optimization, a 3x3 pixel window around subplot

1. Place all your subplots in Arcview
2. Add a grid theme, with image area (value 1) and clouds (value 0)
3. Get grid values for your plots
4. Select plots with value 1
5. Create a new shapefile
6. Export to text file
7. Create a 3x3 window around the subplots with "window"–program
8. Get satellite image values, date by date
9. Move the data back to Arcview
10. Combine all the image dates with "spatial join", be careful with the order of dates
11. Get the lowland values from land-use grid
12. Move the data to access
13. Add the field variables you wish to work with

6.3. Coordinate conversion

The geographical coordinate values (lat/lon) can be converted to UTM in the following way: import the field data to Arcview, convert file to a shapefile, import to Arc/Info, define and change projection, and write out the coordinates as shown below. Remember to check the datum your points are in.

```
Arc: shapearc plots2000 plot_cov
5243 Type 1 (POINT) shape records in PLOTS2000.
```

```
Arc: projectdefine cover plot_cov
Define Projection
Project: projection geographic
Project: units dd
Project: datum nad83
Project: parameters
Arc: project cover plot_cov pl_utm

**************************************************
*      The INPUT projection has been defined.     *
**************************************************
Use OUTPUT to define the output projection and END
to finish.

Project: output
Project: projection utm
Project: units meters
Project: zone 15
Project: datum nad83
Project: parameters
Project: end
Arc: ungenerate point pl_utm crdi_utm.txt
```

7. Literature Cited and References

Franco-Lopez, H., A. R. Ek, and M. E. Bauer. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbor method. *Remote Sensing of Environment*. In press.

Katila, M., and E. Tomppo. 2000. Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment* 76 (1) (2001): 16-32.

Muinonen, E., and T. Tokola. 1990. An application of remote sensing for communal forest inventory. In *The Usability of Remote Sensing for Forest Inventory and Planning. Proceedings from SNS/IUFRO Workshop,* 35-42. Report 4. Umeå, Sweden: Remote Sensing Laboratory. Swedish University of Agricultural Sciences.

Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. C*omputer Journal*. 7: 391-98.

Nilsson, M. 1997. Estimation of Forest Variables Using Satellite Image Data and Airborne Lida*r*. Ph. D. Dissertation. *Acta Universitatis Agriculturae Suecicae*. Silvestria 17. pp. 1-17.

Tokola, T., J. Pitkänen, S. Partinen, and E. Muinonen. 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing* 17(12):2333-2351.

Tokola, T., and P. Kilpeläinen. 1998. The forest stand margin area in the interpretation of growing stock using landsat TM imagery. *Can. J. For. Res.* 29:303-309.

Tomppo, E. 1986. Stand delineation and estimation of stand variates by means of satellite images. In *Remote sensing-aided forest inventory. Seminars organized by SNS and Taksaattoriklubi*, 60-75. Hyytiälä, Finland, December 10-12, 1986.

Tomppo, E. 1991. Satellite image-based national forest inventory of Finland. *International Archives of Photogrammetry and Remote Sensing* 28(7-1): 419-24.

Tomppo, E. 1993. Multi-source national forest inventory of Finland. In *Proceedings of Ilvessalo Symposium on National Forest Inventories*, eds. A. Nyyssönen, S. Poso, and J. Rautala, 53-61. Research Papers 444. Helsinki, Finland: The Finnish Forest Research Institute.

Tomppo, E. 1996. Multi-source national forest inventory of Finland. In *New Thrusts in Forest Inventory. Proceedings of the Subject Group S4.02-00 "Forest Resource Inventory and Monitoring" and Subject Group S4.12-00 "Remote Sensing Technology"*, eds. R. Päivinen, J. K. Vanclay, and S. Miina, 27-41. EFI Proceedings no. 7. Joensuu, Finland: European Forestry Institute.

Tomppo, E. 1990. Designing a satellite image aided National Forest Survey in Finland. In *The Usability of Remote Sensing for Forest Inventory Planning. Proceedings from SNS/IUFRO workshop in Umeå 26/28 February 1990*, 43-46.

USDA Forest Service. 1999. National Forest Inventory and Monitoring Core Field Guide. Volume I: *Field Data Collection Procedures for all Plots.* Version 3/99 – National Core. Version 10/99 North Central. St. Paul, MN.

USDA Forest Service. 2000. Forest Inventory and Analysis National Core Field Guide. Volume I: *Field Data Collection Procedures for Phase 2 Plots*. Version 1.4. 171 p.