Creation of an n-Dimensional Spatial Database Instead of a Static Map

Andy Lister

USDA Forest Service
Northeastern Research Station
Forest Inventory and Analysis

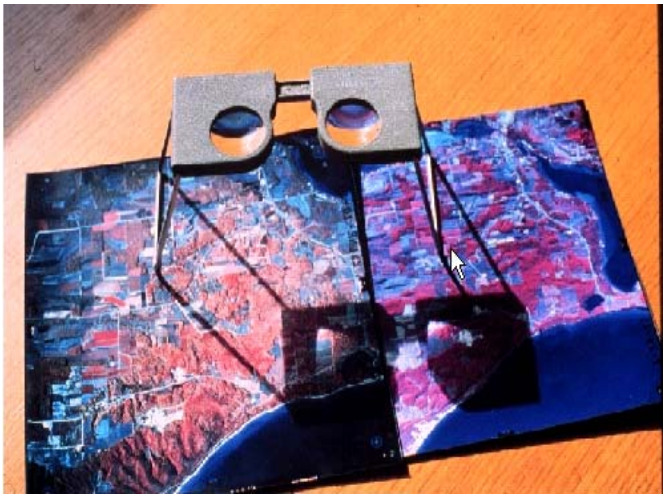
The USDA Forest Service's Forest Inventory and Analysis Unit (FIA) collects information on the trends in and condition of the Nation's forests. In addition to tabular reports, maps of forest attributes are produced. Risk maps are of particular interest, because managers, policymakers and scientists require them to make decisions about investments of personnel and money used to mitigate and respond to risks of things like fire or pest outbreaks. Nearest neighbor estimation methods have been used by FIA for several years. We have developed a methodology to implement nearest neighbor methods to create an n-dimensional map of forestry data that can be summarized at will to create a single static map. The most striking advantage of this spatial database over static maps is its flexibility — the user can populate and summarize the database quickly and efficiently using a simple lookup table to create a static map based on the users' requirements. The current presentation will focus on a brief overview of nearest neighbor methodology, and present the database construction technique. The pros and cons of the technique will be discussed, using a dataset from the Northeastern United States as an example.

Nearest Neighbor Tools in Leica Imagine Software

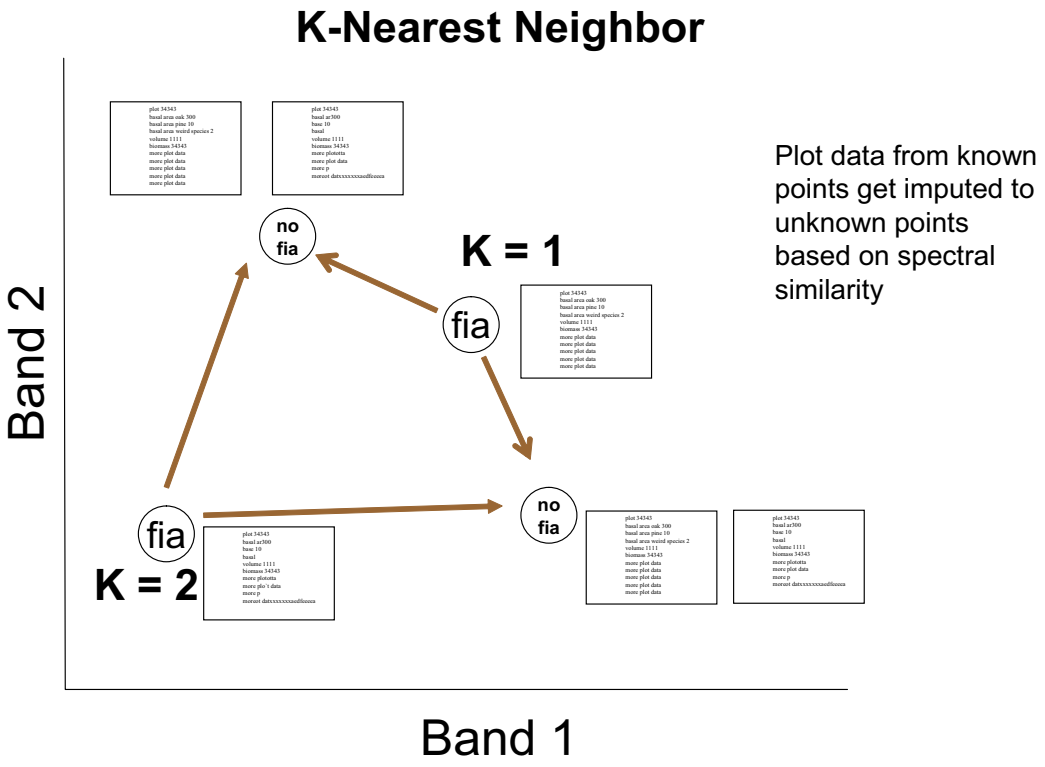
Andrew J. Lister

USDA Forest Service
Northeastern Research Station
Forest Inventory and Analysis

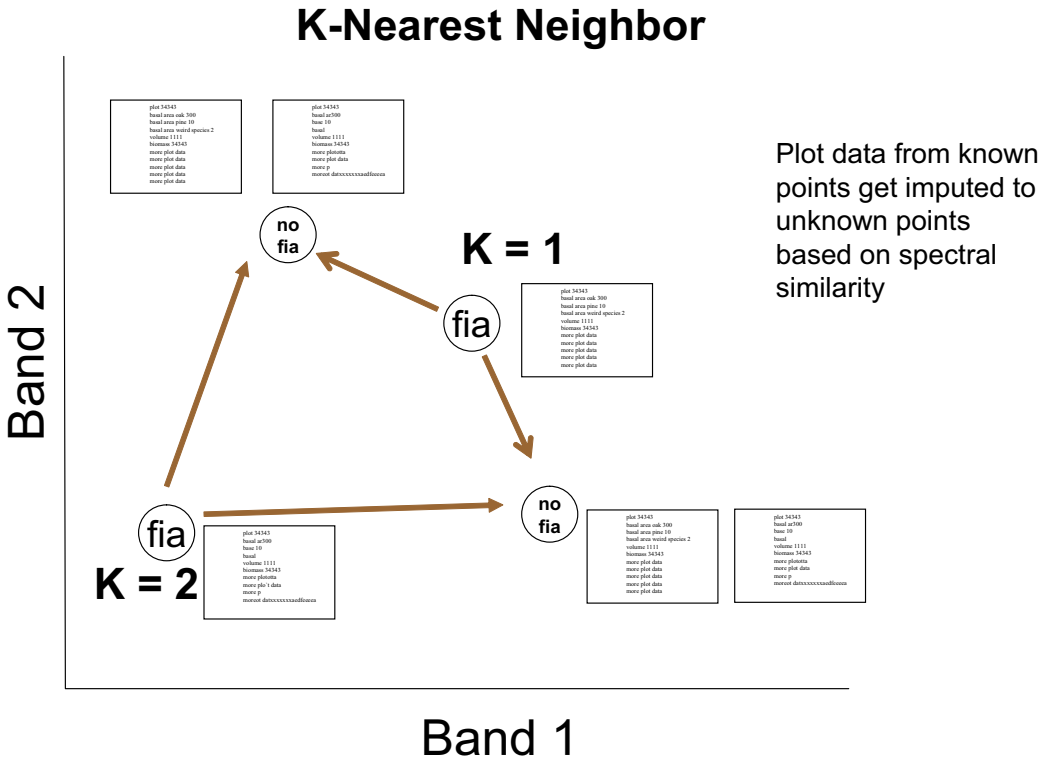
Rationale for needing classified maps: plot stratification for variance reduction



> 400,000 photointerpretation points



Plot data from known points get imputed to unknown points based on spectral similarity



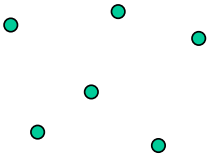
Plot data from known points get imputed to unknown points based on spectral similarity

Overview of Approach to KNN

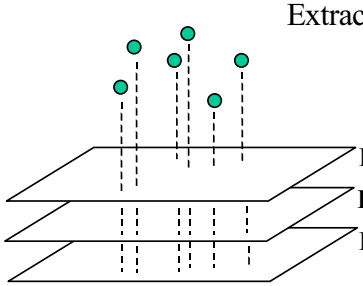
Prepare inventory data

plotid	x coordinate	y coordinate	forest/nonforest	volume
1	435602	883325	1	2000
2	435602	883816	1	185
3	435605	883994	1	333
4	435611	884789	1	1800
5	435616	885009	0	0
6	435617	885809	0	0
7	435622	885846	1	555
...				

Convert to ESRI shapefile



Extract values of predictors with plots to create training data set



plotid	Pred.1	Pred.2	Pred.3
1	39	101	187
2	224	59	239
3	190	128	227
4	186	105	179
5	33	29	242
6	240	47	121
7	194	110	193
...			

Example of Approach to KNN

Generate plot to plot Euclidean distance matrix, then rank plots by distance

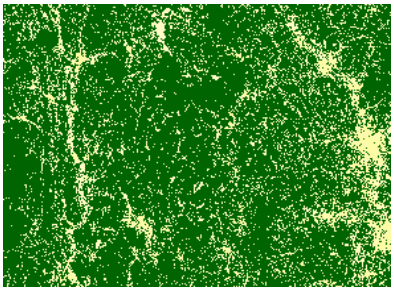
plotid	1	2	3	4	5	6	7
1	0.00	196.71	158.52	147.27	90.80	218.34	155.38
2	196.71	0.00	77.85	84.62	193.36	119.68	74.95
3	158.52	77.85	0.00	53.38	186.21	142.47	38.68
4	147.27	84.62	53.38	0.00	182.08	98.20	16.88
5	90.80	193.36	186.21	182.08	0.00	240.45	186.77
6	218.34	119.68	142.47	98.20	240.45	0.00	106.16
7	155.38	74.95	38.68	16.88	186.77	106.16	0.00

	plotid						
rank	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	5	7	7	7	1	4	4
3	4	3	4	3	4	7	3
4	7	4	2	2	3	2	2
5	3	6	6	6	7	3	6
6	2	5	1	1	2	1	1
7	6	1	5	5	6	5	5

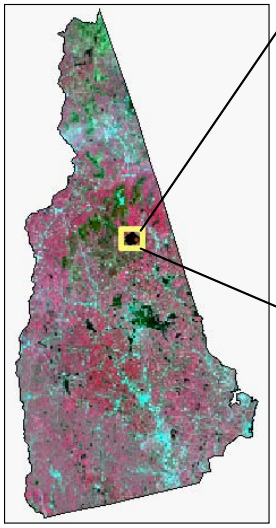
With this information, it is simple to recode the plotid with the level of the attribute of interest (volume, biomass, forest/nonforest, etc.) and conduct summaries for different levels of K

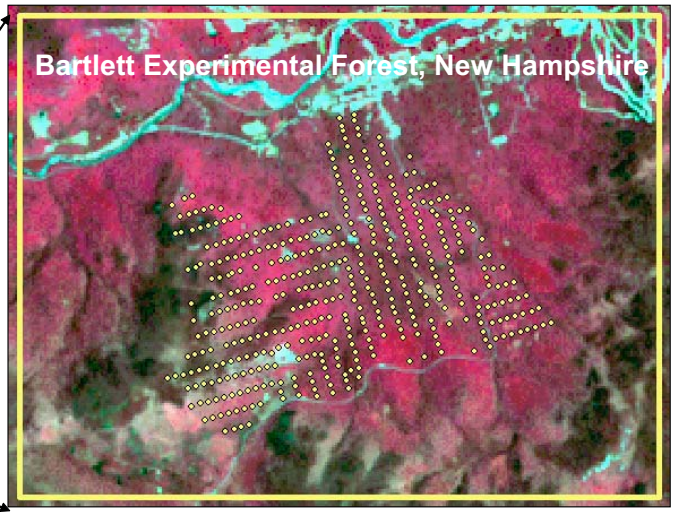
final map = majority landuse (or average or median volume) found at a pixel location across all 7 layers

e.g., for k=7 there are 7 layers in fuzzy output



Real example of KNN



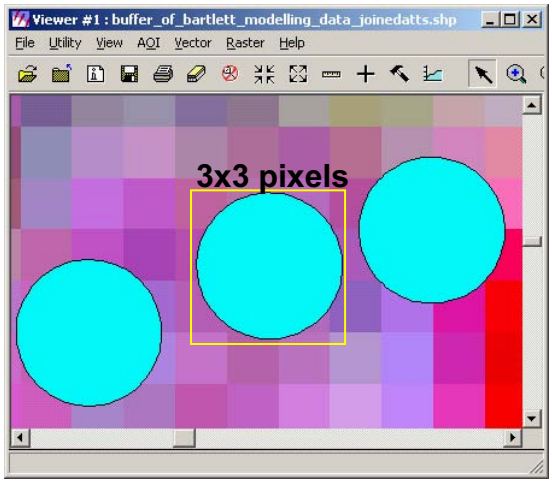


- 437 0.1 ha square plots
- 200 meters apart
- Basal area of all live trees recorded by species

Real example of KNN

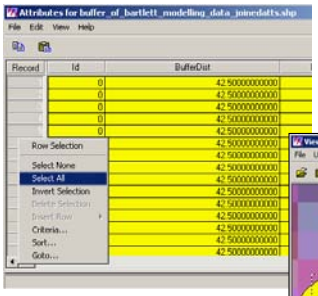
Step1: Buffer the points

Here, I chose 42.5 meter buffer to approximate a 3x3 pixel window. There is some new thought (a la Ray Czaplewski) that it might be sensible to resample the image to new, larger pixels centered on the plots. Pros and cons for discussion.

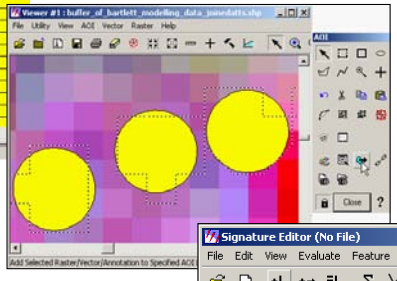


Real example of KNN

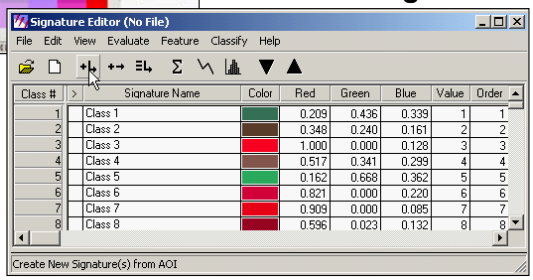
Step2: Select attributes,



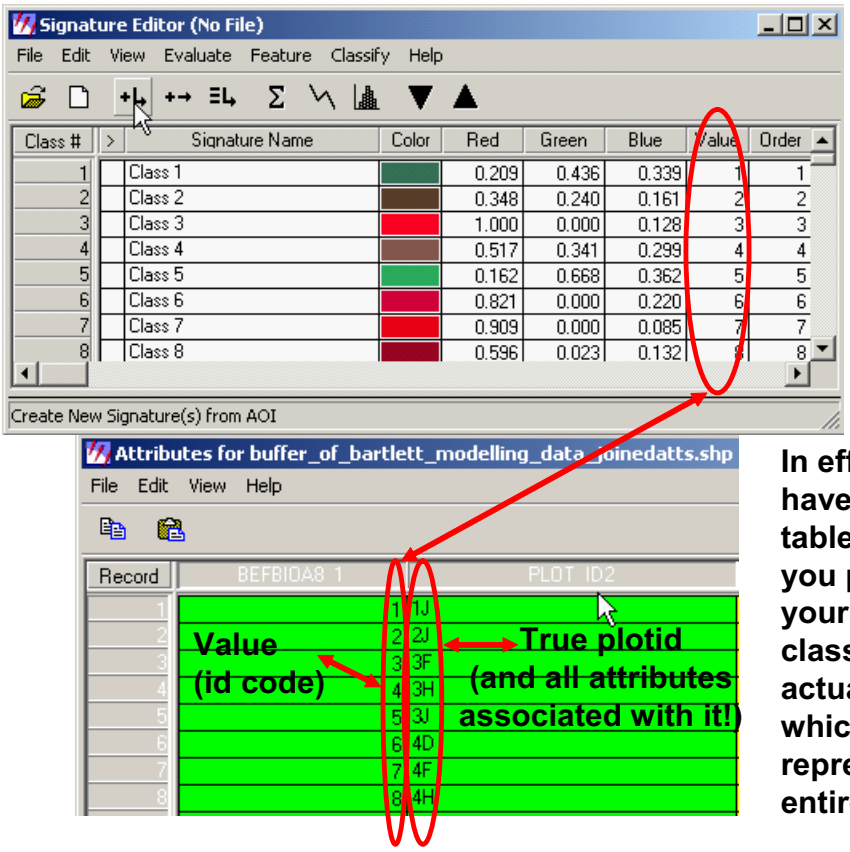
make AOI's,



collect signatures.



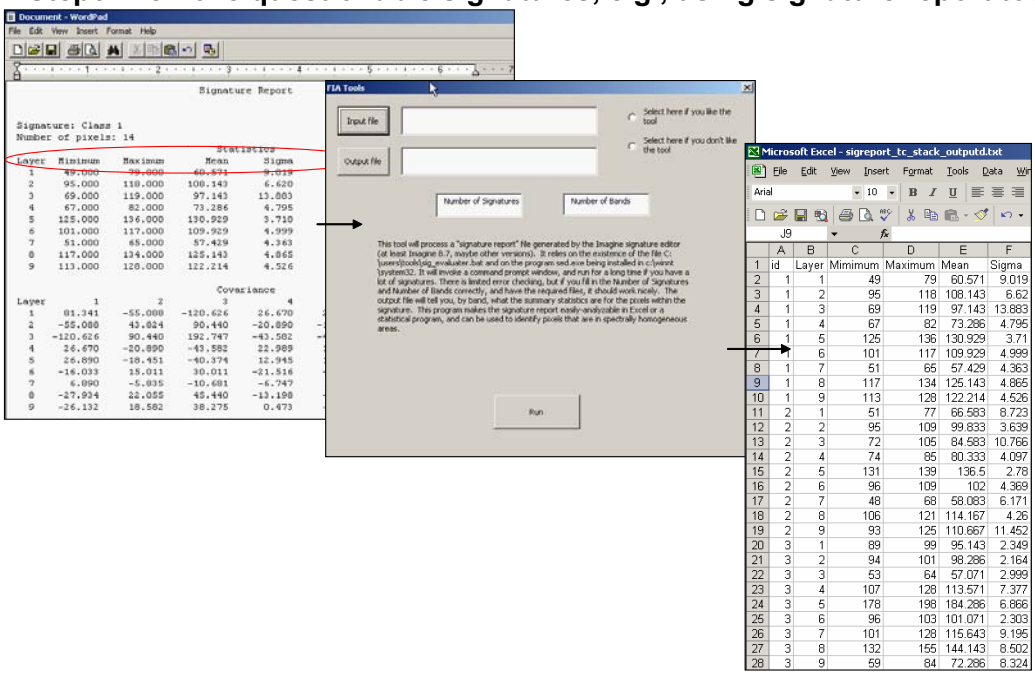
Real example of KNN



In effect, you have a lookup table. The value you predict in your kNN classification is actually plotid, which actually represents the entire plot record!

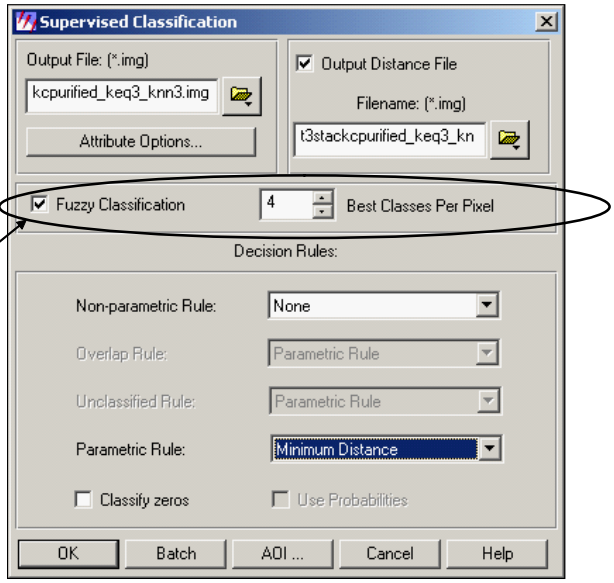
Real example of KNN

Step3: Remove questionable signatures, e.g., using signature report tool



Real example of KNN

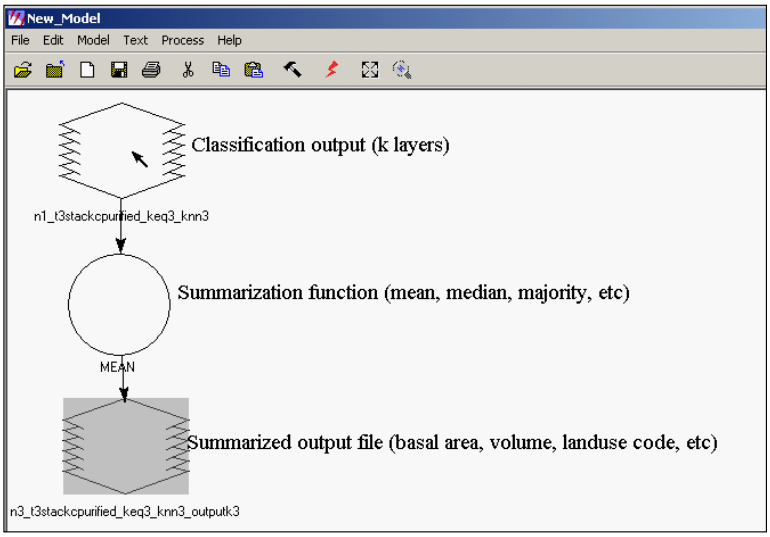
Step4: Once you've removed undesirable signatures, classify your image using "Fuzzy Classification", with the plotid code as the value to be predicted.



Chose 3 neighbors using a heuristic – how well do various levels of k classify the points?

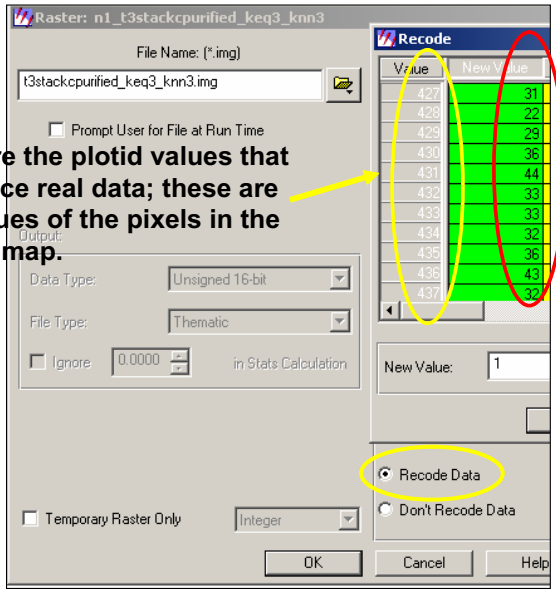
Real example of KNN

Step5: Take the classification output, apply lookup table and produce a map of any attribute in the original database that can be associated with a plotid (total ba, ba of red maple, percent damage, species richness, etc. etc. etc..). This is done by recoding the knn output..



Real example of KNN

Recoding can be done by simply copying and pasting the lookup table value column into the recode dialogue box…

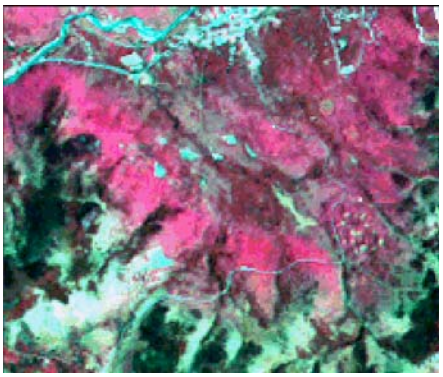


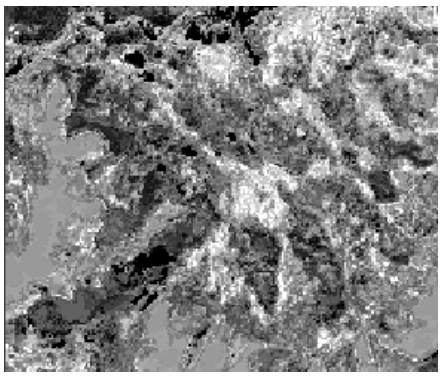
Here are the plotid values that reference real data; these are the values of the pixels in the k layer map.

Here are the values of one attribute from the plot database that I want to display in a map… total basal area

Real example of KNN

Output: The image visually represents spatial heterogeneity of biomass on the landscape.

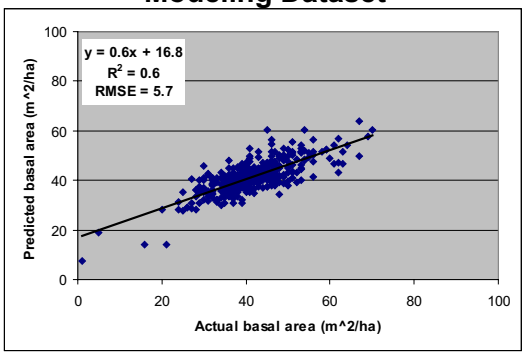




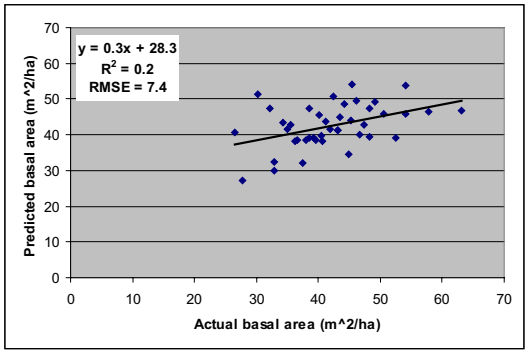
Real example of KNN

Output: Numerically, the values are underpredicted in the upper and overpredicted in the lower tails.

Modeling Dataset



Test Dataset

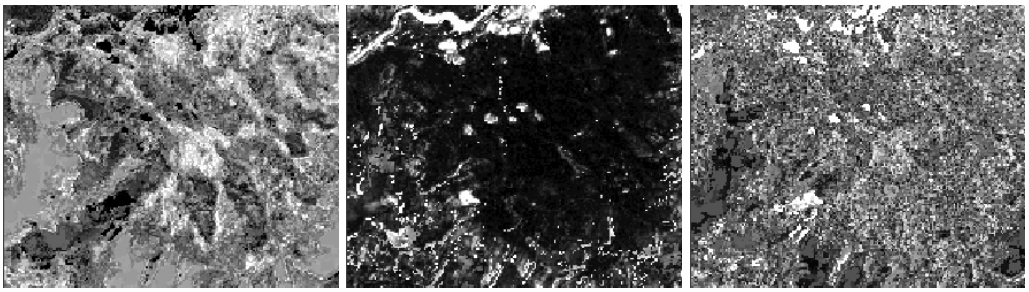


All data are those that remain after the anomalous signature removal procedure

Real example of KNN

Confidence Maps: Can create using a distance image, or a variance calculation on the stack of k layers, all with simple models.

Original Image **Mean Distance** **Variance of Basal Area**



Summary – Why might Imagine's KNN be good for risk mapping?

1. **Intuitively easy**
2. **Flexible**
3. **Already built into Imagine – most operations are point and click, copy/paste; easily customizable**
4. **Allows you to assess model outputs both numerically and visually**
5. **Creates intuitive confidence maps**
6. **Most excitingly, by creating a stack of k maps with pixels that are labeled with plot id code, one can apply a lookup table to the stack of maps and create a map of any attribute associated with plotid.**

Summary – Why might Imagine's KNN be BAD for risk mapping?

1. **Intuitively easy,** *but is it too easy? Oversimplistic?*
2. **Flexible,** *but does it tempt you to make maps of things with no functional relationship to the "predictors"?*
3. **Already built into Imagine,** *but by automating a process, do you run the risk of becoming a robot?*
4. **Allows you to assess model outputs both numerically and visually,** *but do visual assessments blind you to the true utility of a map (ugly maps can be useful, pretty maps can be useless)*

For Additional Information:

Andrew J. Lister
alister@fs.fed.us
610.557.4038

USDA Forest Service
Northeastern Research Station
Forest Inventory and Analysis
11 Campus Blvd., Ste 200
Newtown Square, PA 19073