United States Department of Agriculture

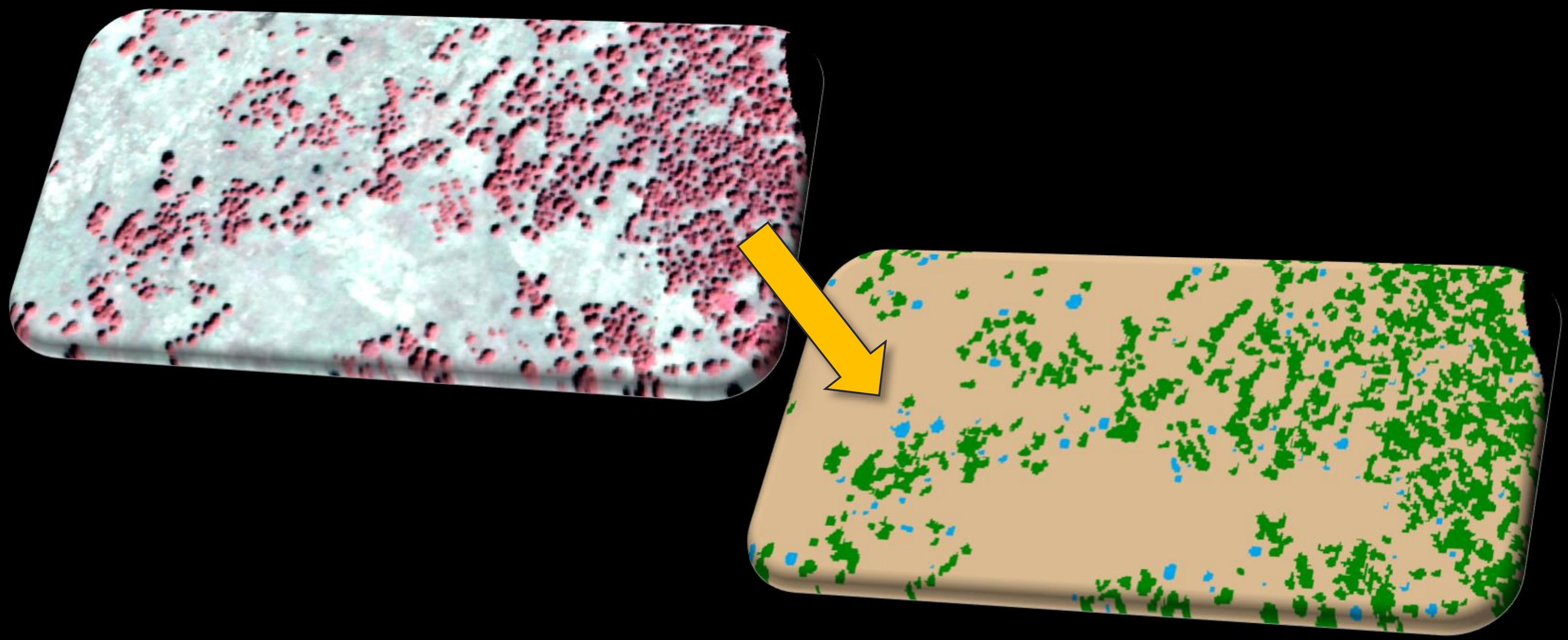
Land Cover Classification Techniques

supervised classification and random forests

Developed by remote sensing specialists at the USFS Geospatial Technology and Applications Center (GTAC), located in Salt Lake City, Utah

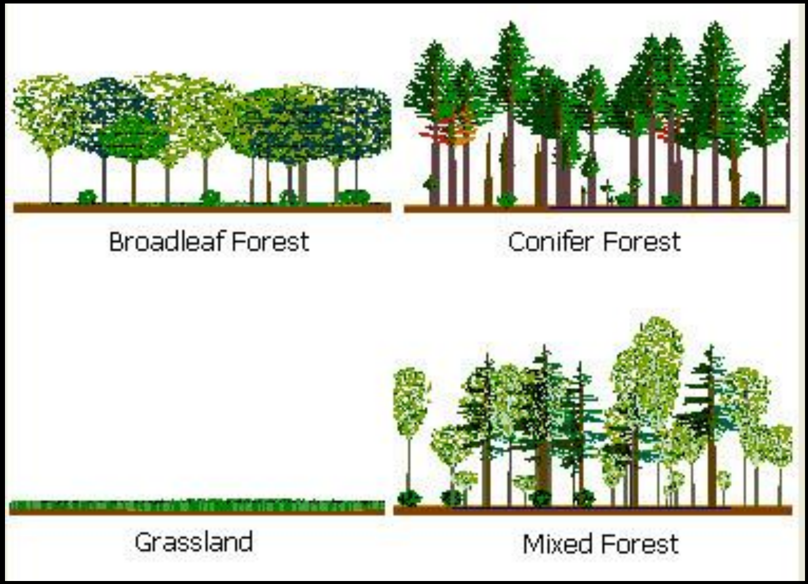
Image Classification

The automated process of categorizing pixels or image objects into a thematic classes (e.g., conifer, deciduous, herbaceous)



Classification Scheme

- Decide on the *scheme*
  - The schema categorizes and labels the land cover *theme* (e.g. vegetation cover)
  - A well-designed classification *scheme* is critical to deriving acceptable and useful information
  - The complexity will affect project accuracy and cost
  - It's not easy—but time spent creating a well designed classification scheme is always well spent!



IPCC greenhouse gas inventory

Estimate changes in 5 carbon pools for six categories of land use.

The six land use categories include:

1. *Forest land*
2. *Cropland*
3. *Grassland*
4. *Wetland*
5. *Settlements*
6. *Other land*

The Classification Scheme – 5 rules

Requirements (for continuous cover classifications)

- Must:
  1. Be <u>exhaustive</u>:
     a. All land cover must be accounted for in the legend
     b. An “Other” category ensures this condition!!
  2. Be <u>mutually exclusive</u>:
     a. Each land cover can be assigned to exactly one class
     b. A land cover feature is considered mutually exclusive if it cannot occur in two classes.
  3. Be composed of <u>Labels and Rules</u> for labeling
  4. <u>Meet the user’s needs</u>
  5. Must be <u>based on what can be interpreted</u> from the imagery (note: this may conflict with the previous requirement in some cases).

The Classification Scheme – Suggestions

- In addition, the scheme should be:
  1. Hierarchical
     a. To be more flexible and better support multiple users.
  2. Based on measurable land cover characteristics
     a. Size class, % canopy cover, dominance, etc.
  3. Avoid subjective, interpretive classes, such as “old growth”.

Classification Scheme -- Bad Example

- Water
- Rock and Soil
- Shrub
- Forest
  - Hardwood
    - Sparse
    - Medium density
    - Dense
  - Softwood
    - Sparse
    - Medium density
    - Dense
- Regeneration

Yes! The data can be collapsed or expanded. → Hierarchical?

No, there is no place for grasslands, mixed forests, etc… an "other" class would satisfy this. → Exhaustive?

None! How much tree cover does it take to separate shrub and forest? No definition for *density* labels. → Labeling Rules?

No! What is "Regeneration"? That could include grasses, shrubs or forest… → Mutually Exclusive?

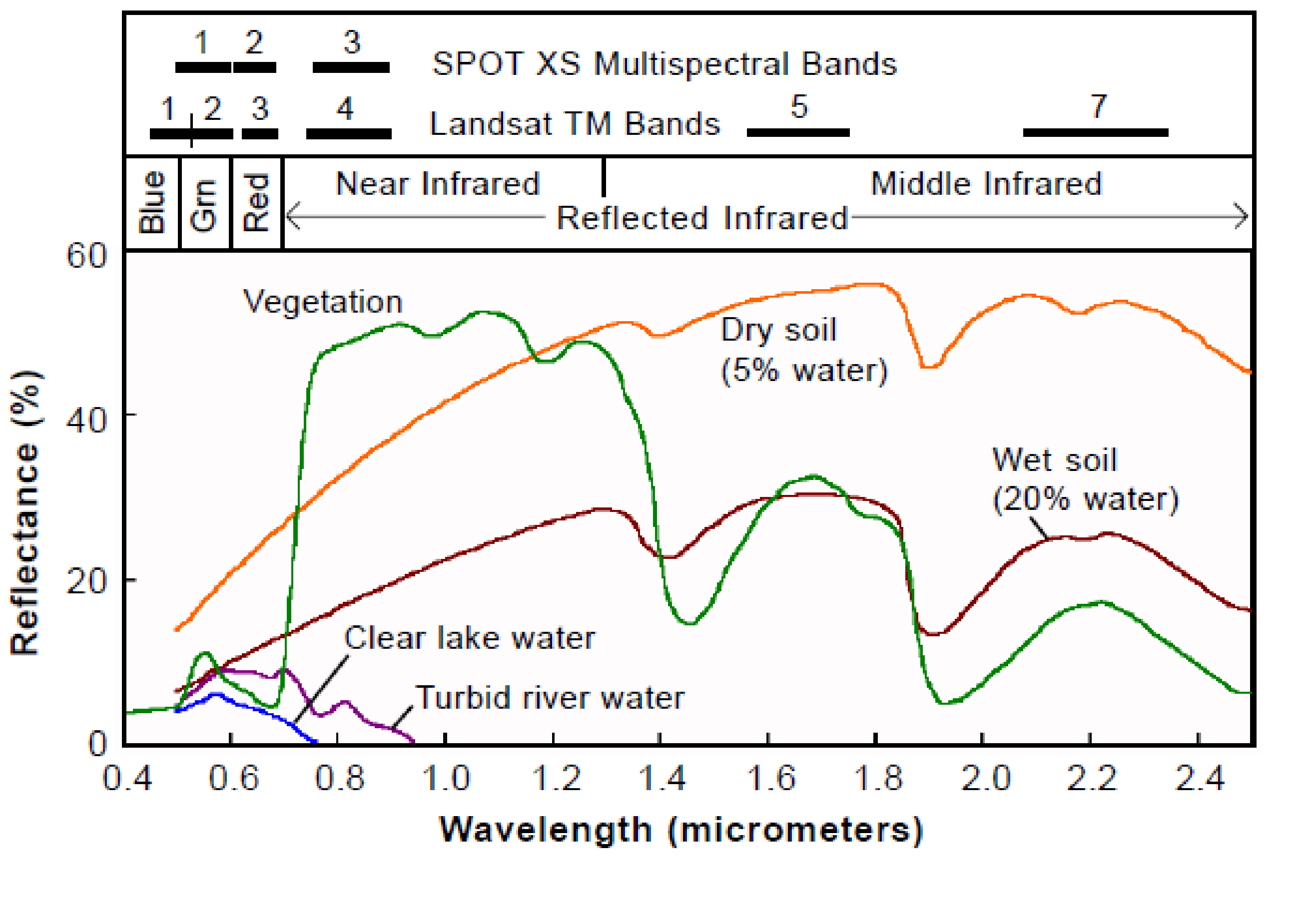
Classification Scheme -- Good Example?

- Water
- Non-Vegetated (< 20% vegetated)
- Rangeland (< 10% tree crown closure)
- Forest (> 10% tree crown closure)
  - Hardwood (65% of trees are hardwood)
    - sparse (10% and < 30% CC)
    - medium density (30% and < 66% CC)
    - dense (66% CC)
  - Softwood (65% of trees are softwood)
    - sparse (10% and < 30% CC)
    - medium density (30% and < 66% CC)
    - dense (66% CC)
  - Other Forest (includes Mixed)
- Other

Classification Scheme -- Good Example?

- Water
- Non-Vegetated (< 20% vegetated)
- Rangeland (< 10% tree crown closure)
- Forest (> 10% tree crown closure)
  - Hardwood (65% of trees are hardwood)
    - sparse (10% and < 30% CC)
    - medium density (30% and < 66% CC)
    - dense (66% CC)
  - Softwood (65% of trees are softwood)
    - sparse (10% and < 30% CC)
    - medium density (30% and < 66% CC)
    - dense (66% CC)
  - Other Forest (includes Mixed)
- Other

•This is a graph of a spectral response curves for several different types of material.
•It represents the visible to NIR portion of the electromagnetic spectrum
•As you can see materials (e.g., dry soil, wet soil) can be identified and differentiated from each other as long as their signatures are unique from one another.



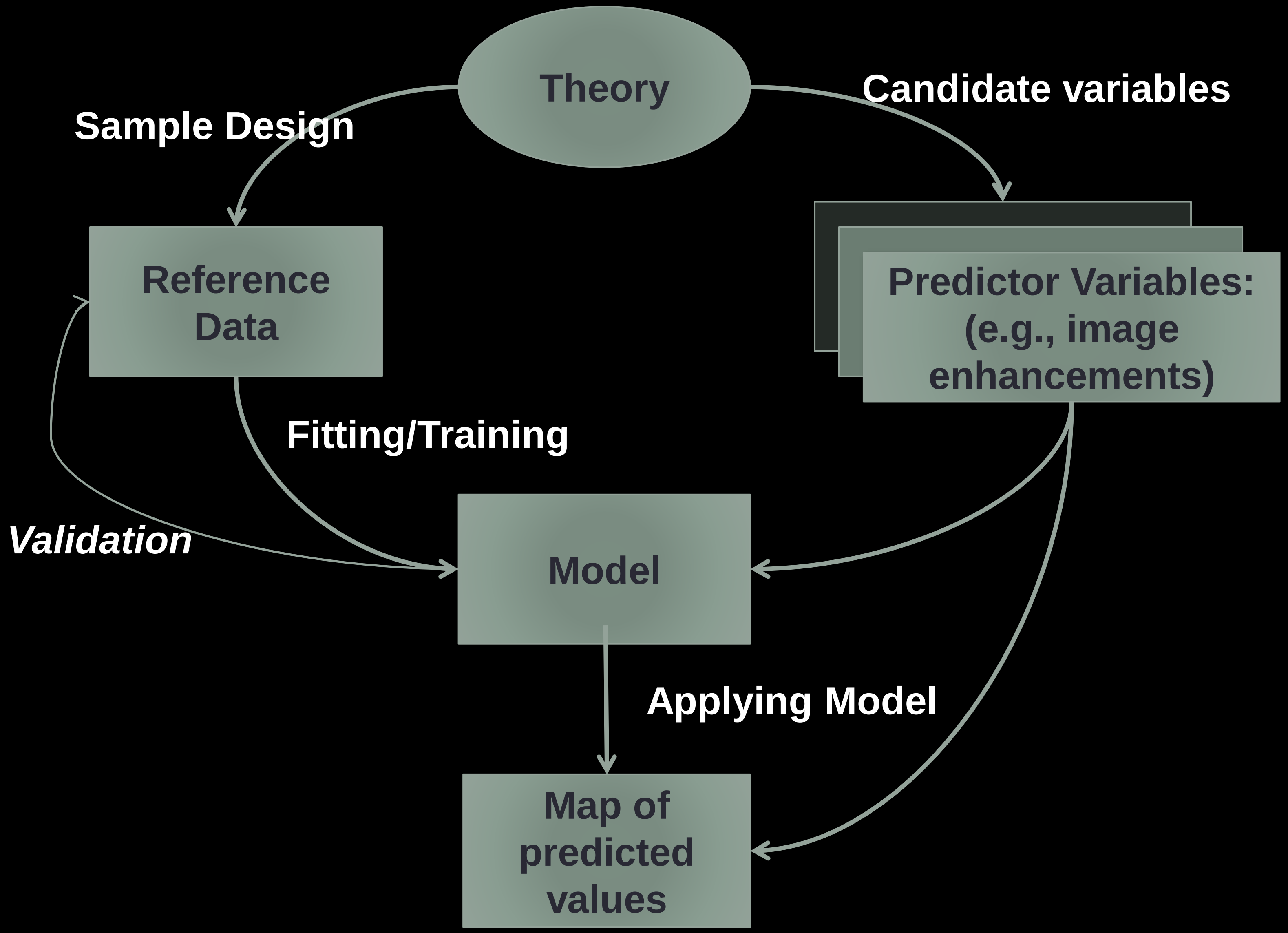
Spatial Models

Linking goals (land cover categories)

with

Land cover characteristics (spectral signatures)

Spatial Modeling Process



Adapted from Guisan and Zimmerman (2000) & Franklin (2009)

Spatial Modeling Process

Theory

Sample Design

Candidate variables

Reference Data

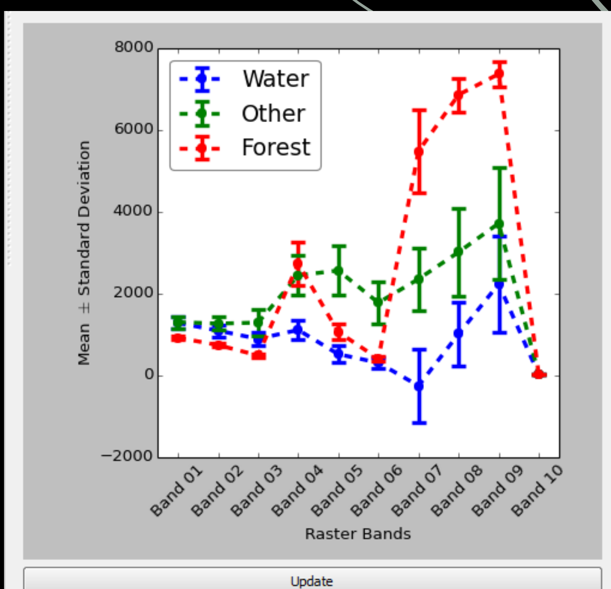
Predictor Variables: (e.g., image enhancements)

Fitting/Training

Model

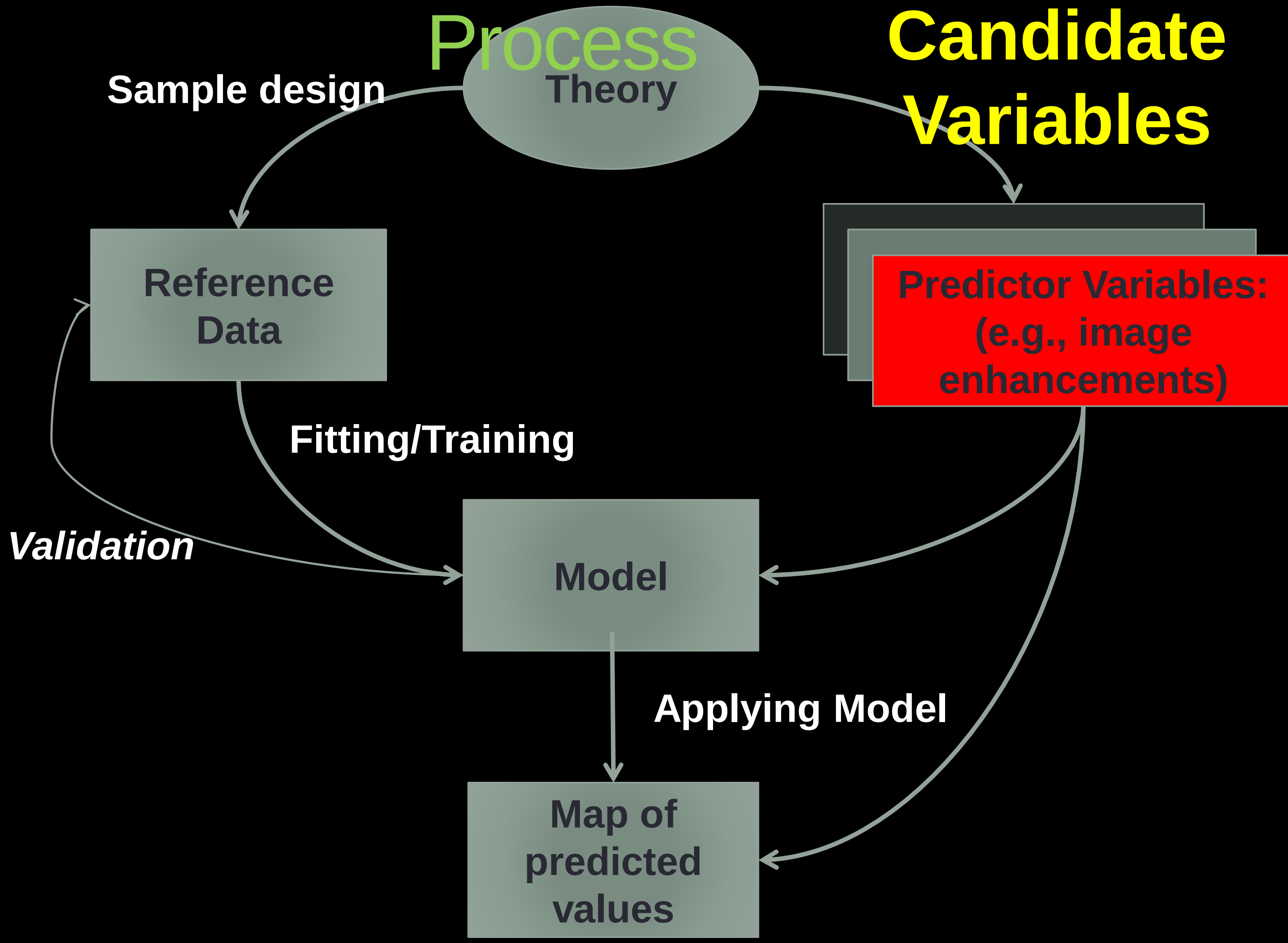
Applying Model

Map of predicted values



Adapted from Guisan and Zimmerman (2000) & Franklin (2009)

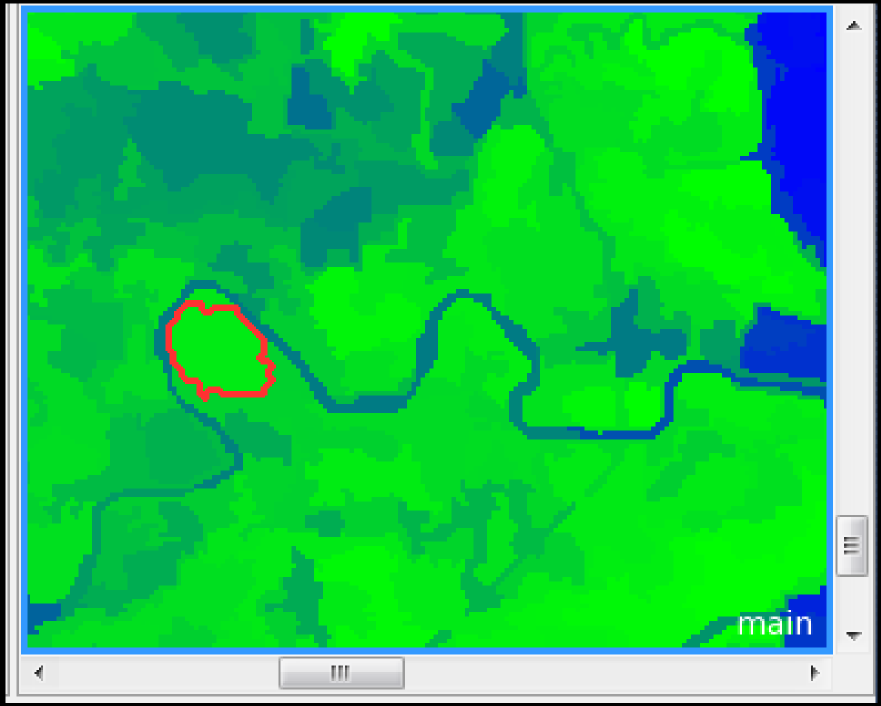
Components of the Spatial Modeling Process



Adapted from Guisan and Zimmerman (2000) & Franklin (2009)

Predictor Variables

- What does water look like?
- What does land look like?
- What information do you think will help differentiate water from land?



Predictor Variable Examples

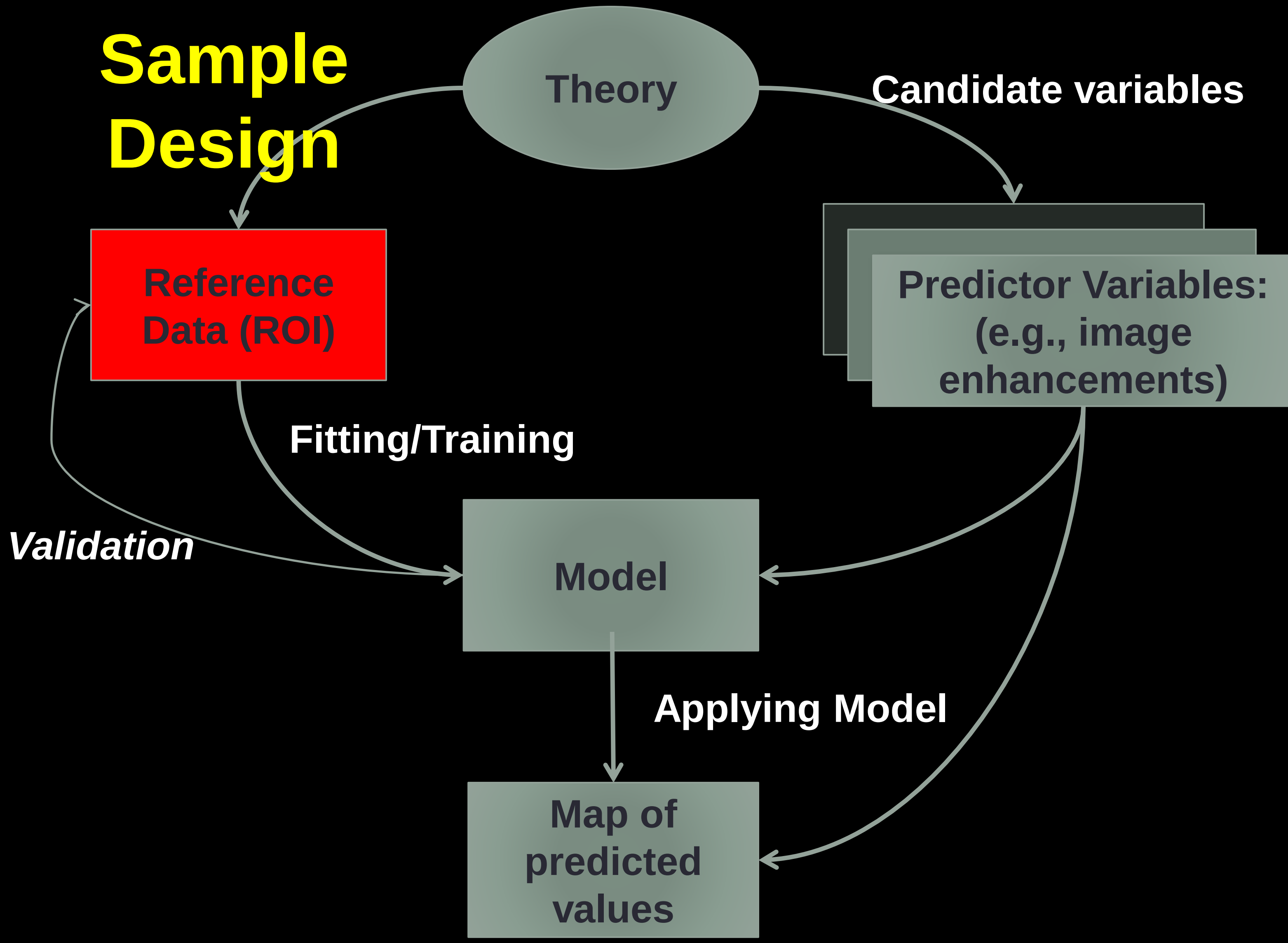
- Impervious surface metric
- Modified normalized difference water index
- Normalized difference vegetation index
- Elevation
- Aspect

Predictor Variables

Theory/Knowledge to get predictor variables:

- Tone/color:
- Shape:
- Size:
- Association:
- Shadow:
- Pattern:
- Texture:

Spatial Modeling Process



Adapted from Guisan and Zimmerman (2000) & Franklin (2009)

Training Data

- Required for the advanced classifiers
  - Interpreted manually
  - Imported from a point or polygon shapefile
    - Polygons can end up overlapping multiple segments
    - Points are more precise and have less potential for error

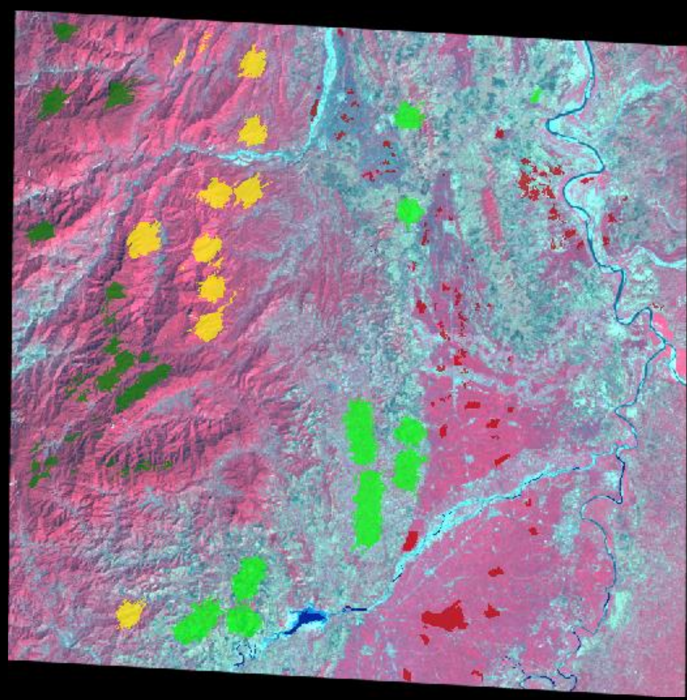
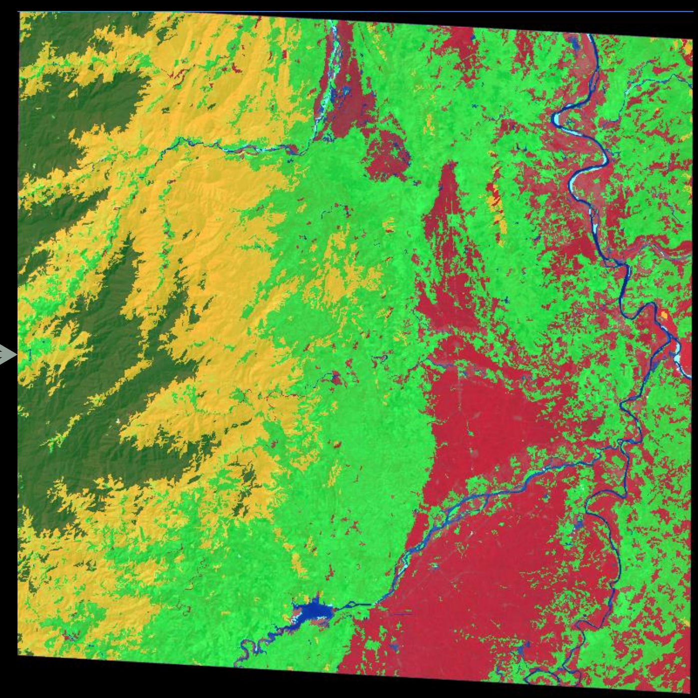


Image with samples collected

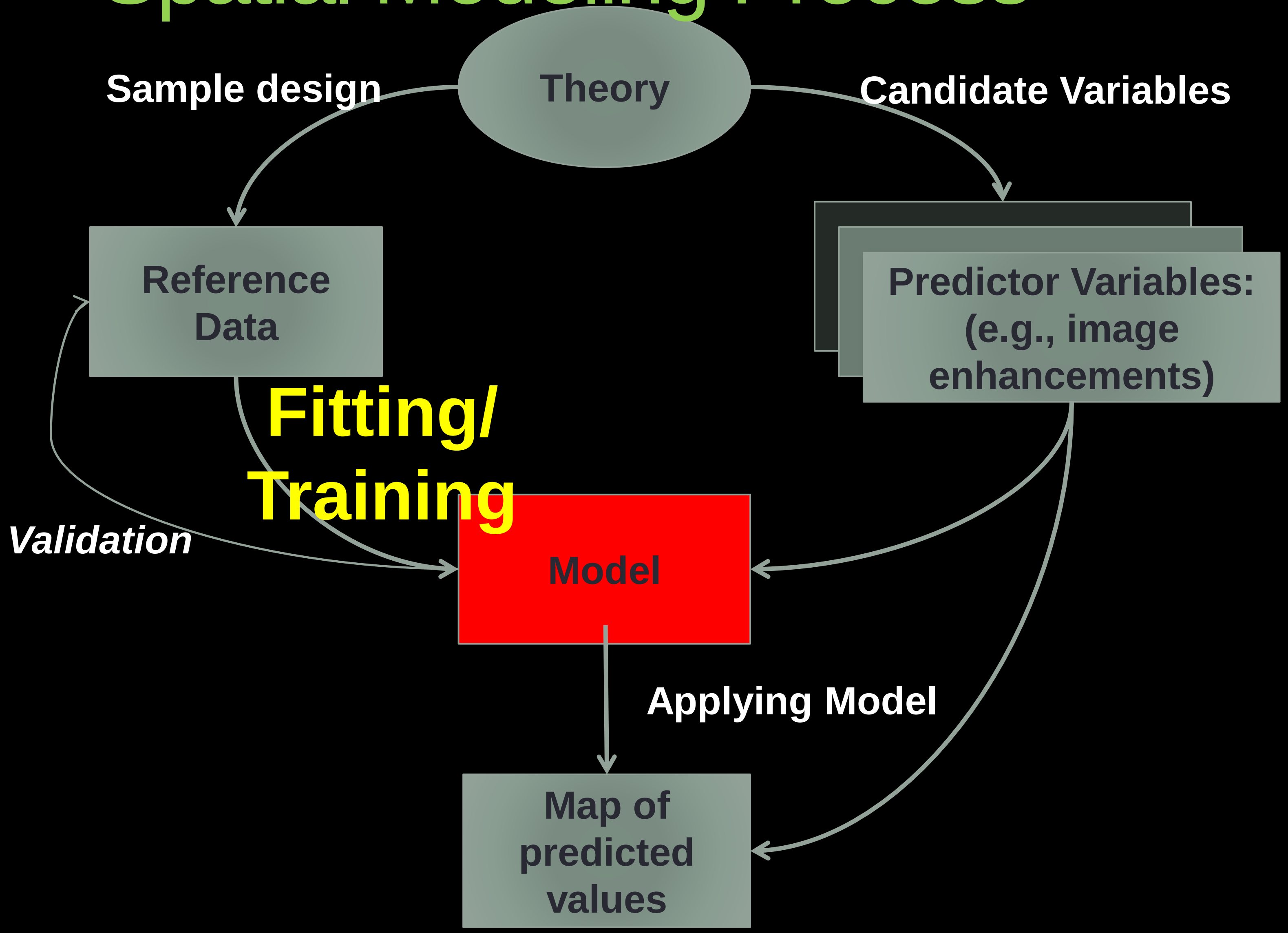


Classified image

Required Elements: training data

- Sample (training) data considerations:
  - Collected in field or digitized from high resolution imagery
  - Covers full region
  - Full range
  - Random, or stratified random is best, often an opportunistic sample is all that is available

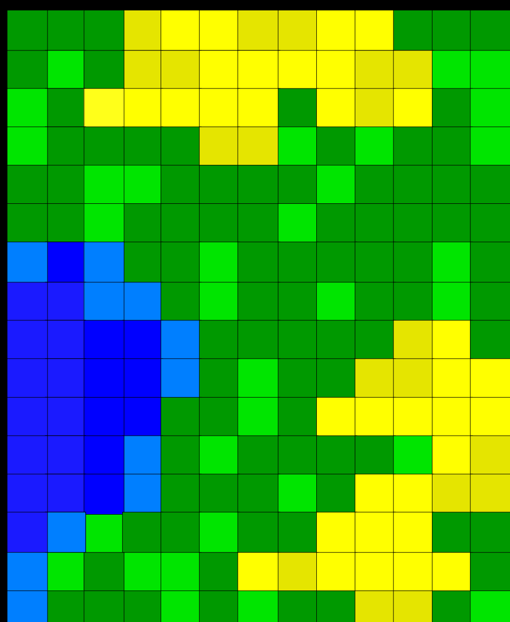
Spatial Modeling Process



Adapted from Guisan and Zimmerman (2000) & Franklin (2009)

Unsupervised Classification

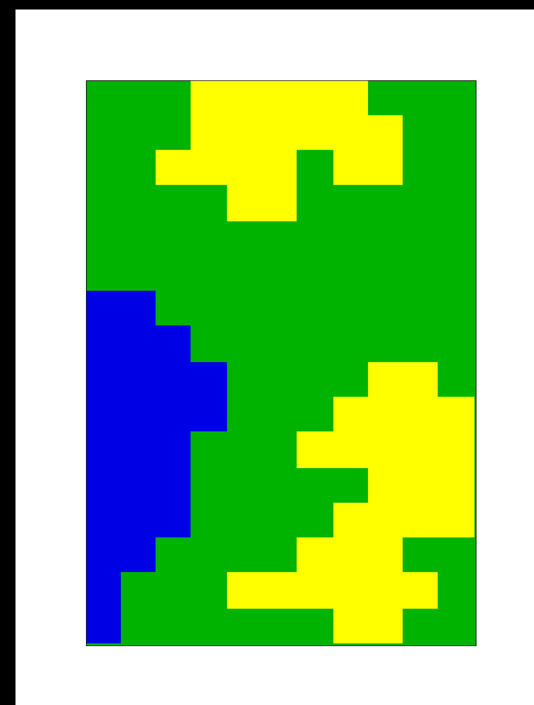
After the computer finishes processing, the analyst then attempts to assign these spectral classes into informational classes. It is a simple process to regroup (recode) the clusters into meaningful information classes (the legend).



Labels

- Water
- Water
- Conifer
- Conifer
- Hardwood
- Hardwood

The result is a Land Cover Map

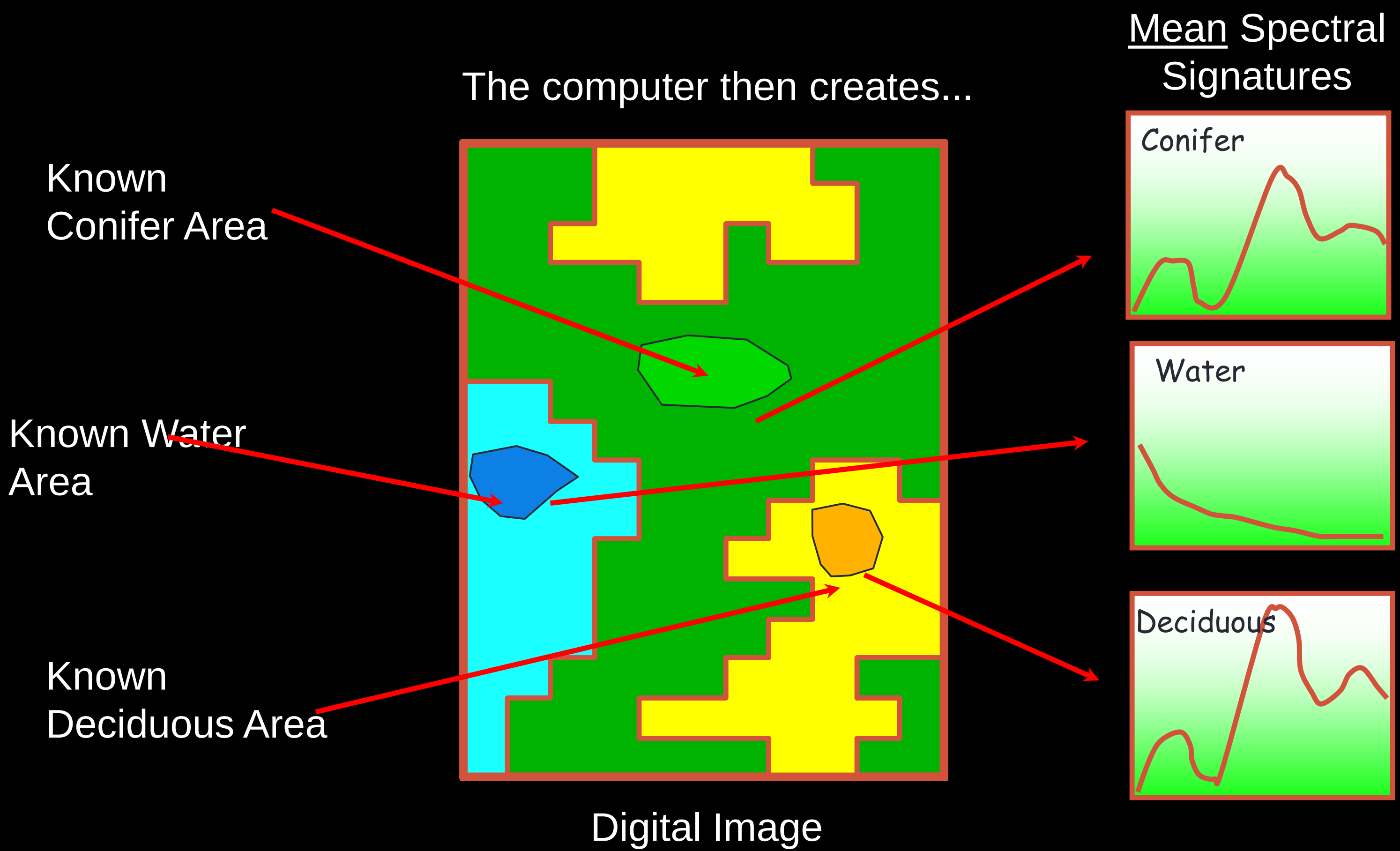


Legend

- Water
- Conif.
- Hardw.

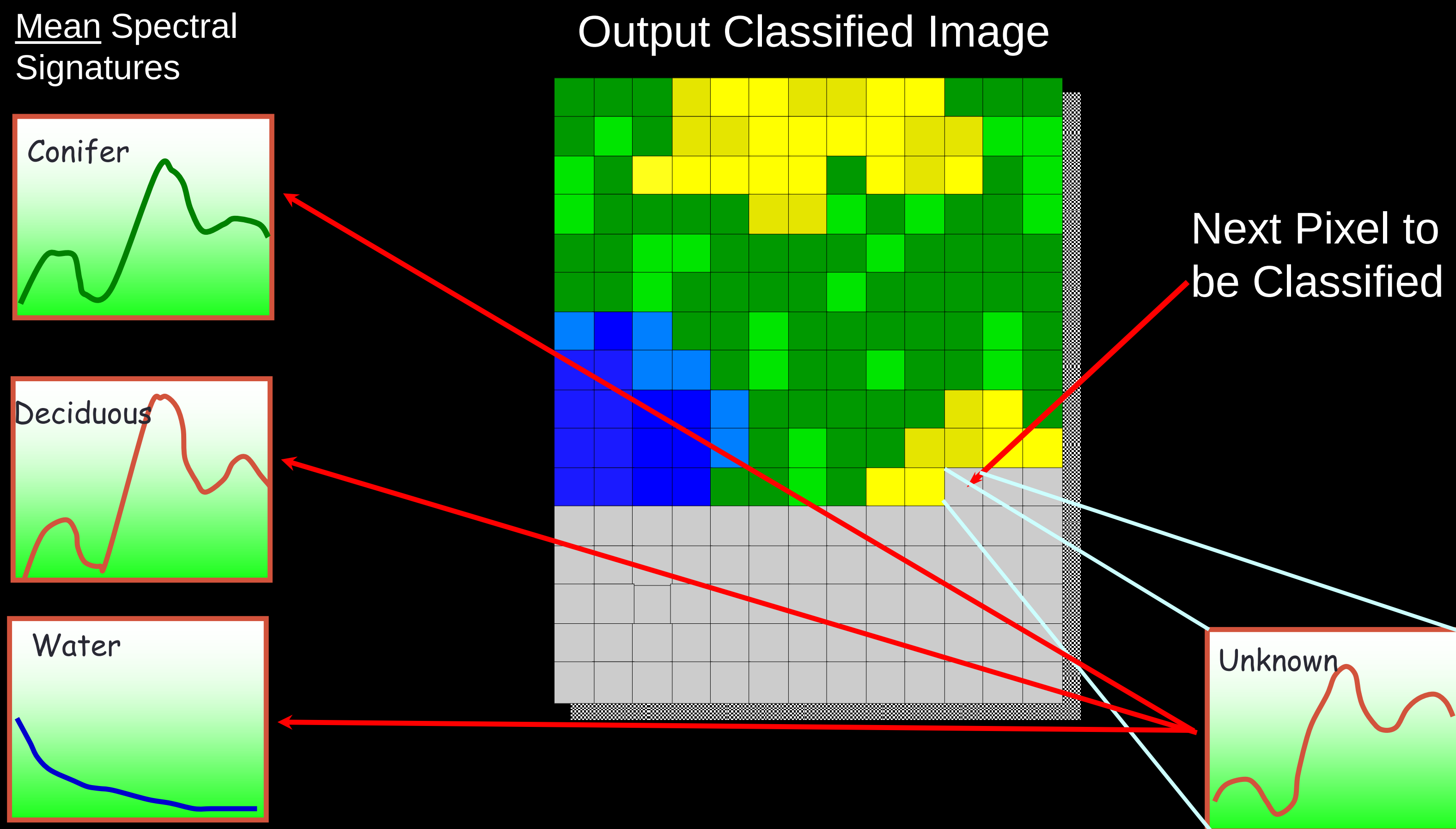
Supervised Classification

Supervised classification requires the analyst to create training data, either in the field or from an image data source.



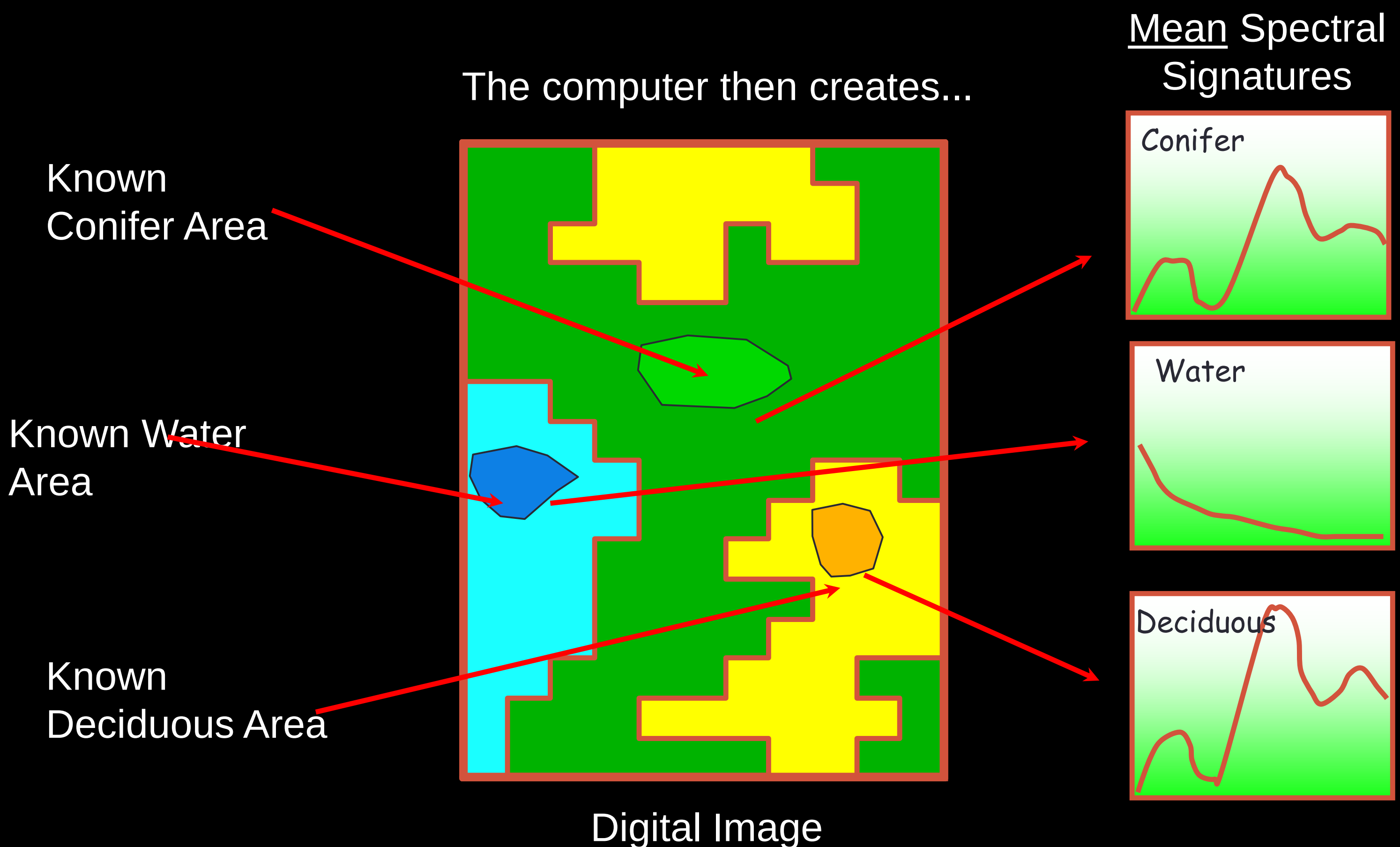
Supervised Classification

Using the statistical information provided by the training regions, the computer attempts to determine all remaining pixels in the image falling into these defined classes.



Supervised Classification

Supervised classification requires the analyst to create training data, either in the field or from an image data source.



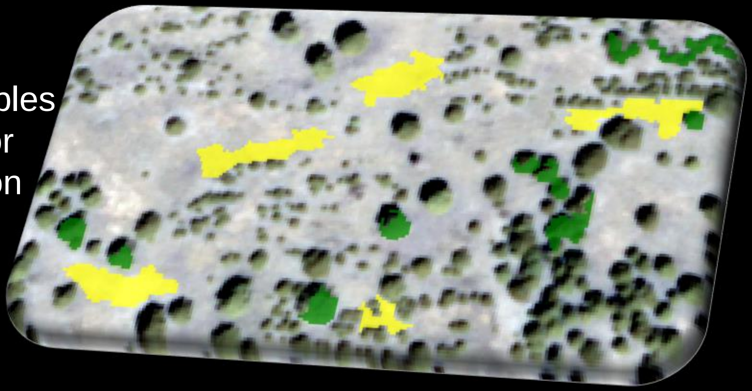
Supervised Classifiers

- **Nearest neighbor**
- Bayes
- Decision Tree
- K-nearest neighbors
- Support Vector Machines
- **Random forests**

Classification Process

- Classifies based on descriptors applied to the classes
- Nearest neighbor
  - Select the features the classification will be based on (e.g., red band, NDVI, ancillary data, etc)
  - Select training data

Training samples selected for classification

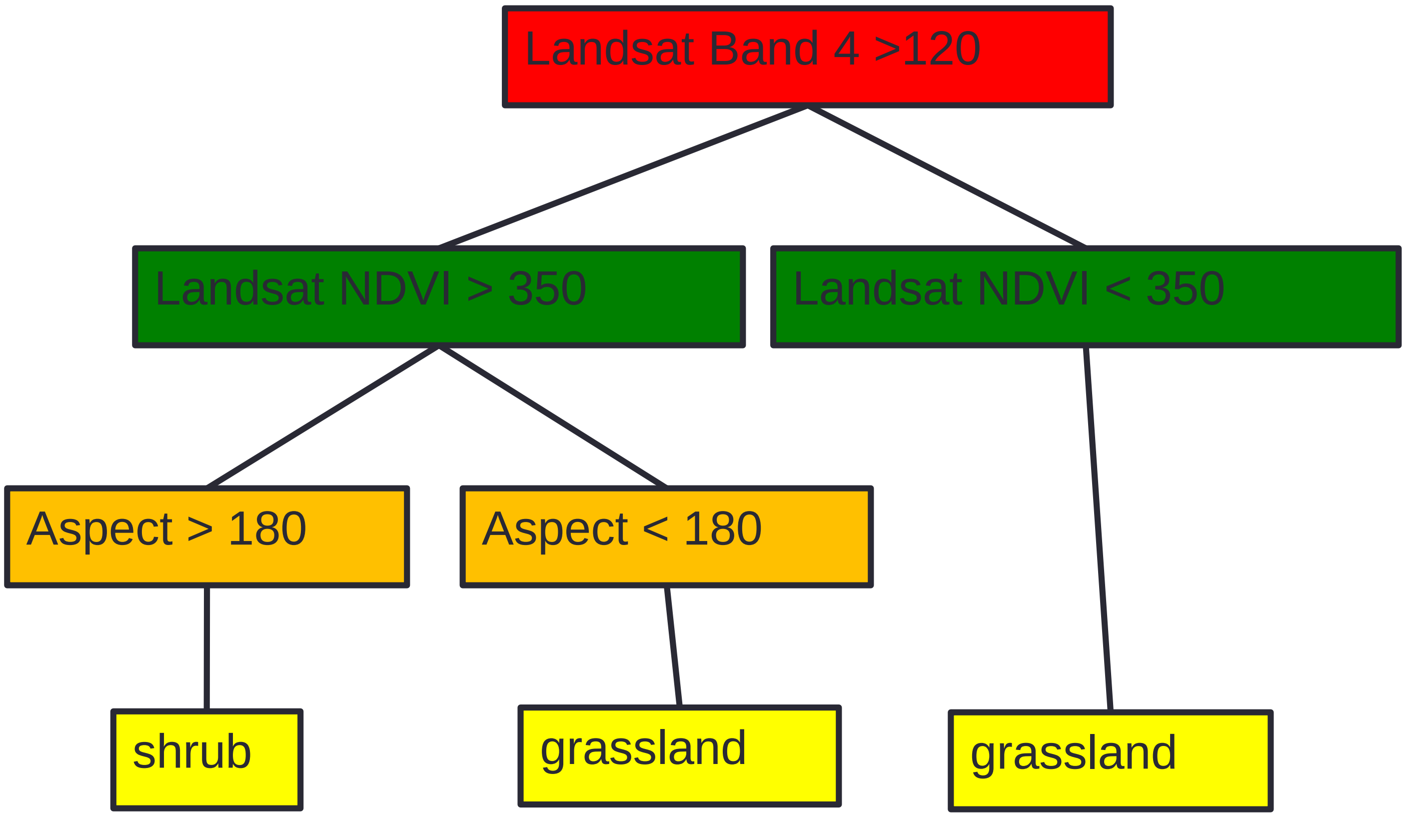


Decision Trees

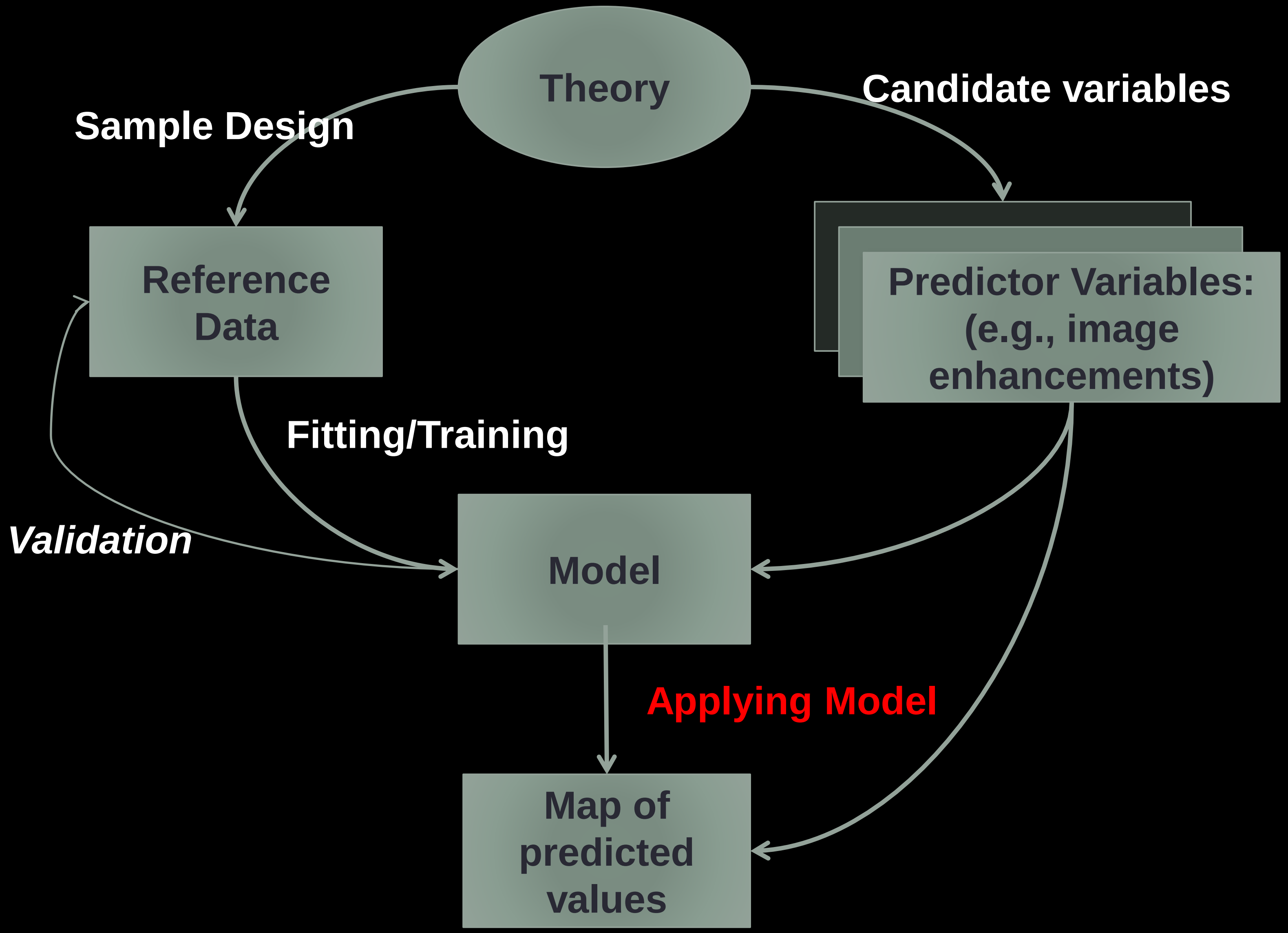
- Also known as CART (Classification and Regression Tree) when they refer to both:
  - **Classification tree** = predicted outcome is the class to which the data belongs.
  - **Regression tree** = predicted outcome can be considered a real number (e.g. NDVI value)
- Data mining technique
- Uses training data to develop a tree-like set of rules to determine the class for certain combinations of input data
- Simple to use and can produce good results

Example of a Decision Trees

For example - separating shrub from grassland



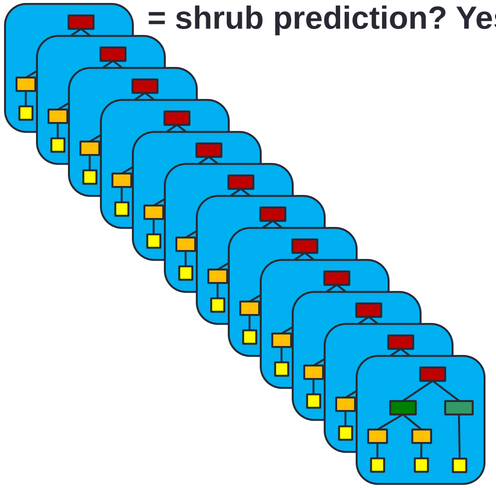
Spatial Modeling Process



Adapted from Guisan and Zimmerman (2000) & Franklin (2009)

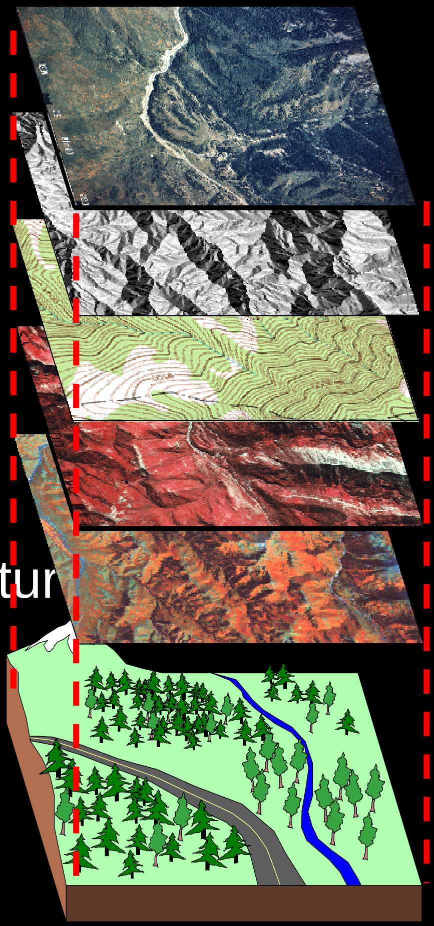
Random Forest

- Ensemble-based machine learning technique
  - Ensemble = “many of”
  - Machine learning technique = Classification and Regression Trees (CART)
  - The “Random” in Random Forest describes that the selection of predictor variables is random for any given CART model (“tree”).



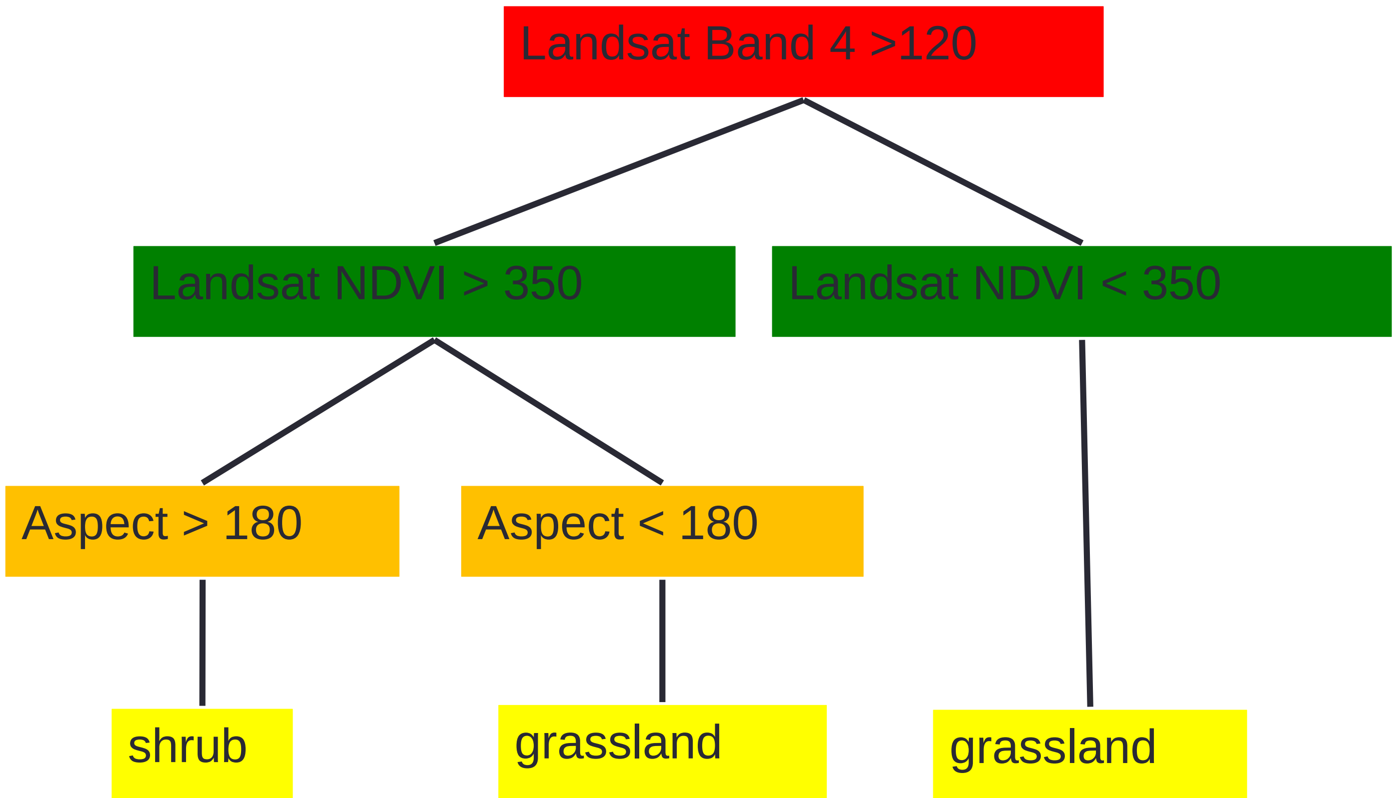
What is input into Random Forest?

- Training or reference data (point data)
  - Examples of each class (e.g., Conifer, Aspen, Grass, Shrub, Road, Sagebrush, Shadow, Water, Soils)
- Predictor variables such as:
  - Multispectral imagery
  - Panchromatic imagery
  - Topographic variables: Elevation, Slope, & Aspect.
  - Bioclimatic variables: Temperature, Precipitation, Moistur Index, Potential Global Radiation, Vapor Pressure, Humidity, Degree Days.
- Derived Predictor variables such as:
  - NDVI
  - Tasseled Cap transformations (soil brightness, greenness, wetness)

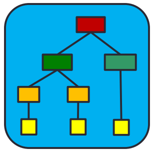


CART

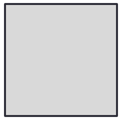
Random input variables are selected and used on your training data to make trees.



This is an example of **1** Classification tree – hundreds are created randomly using your reference/training data and built for your various Classes.

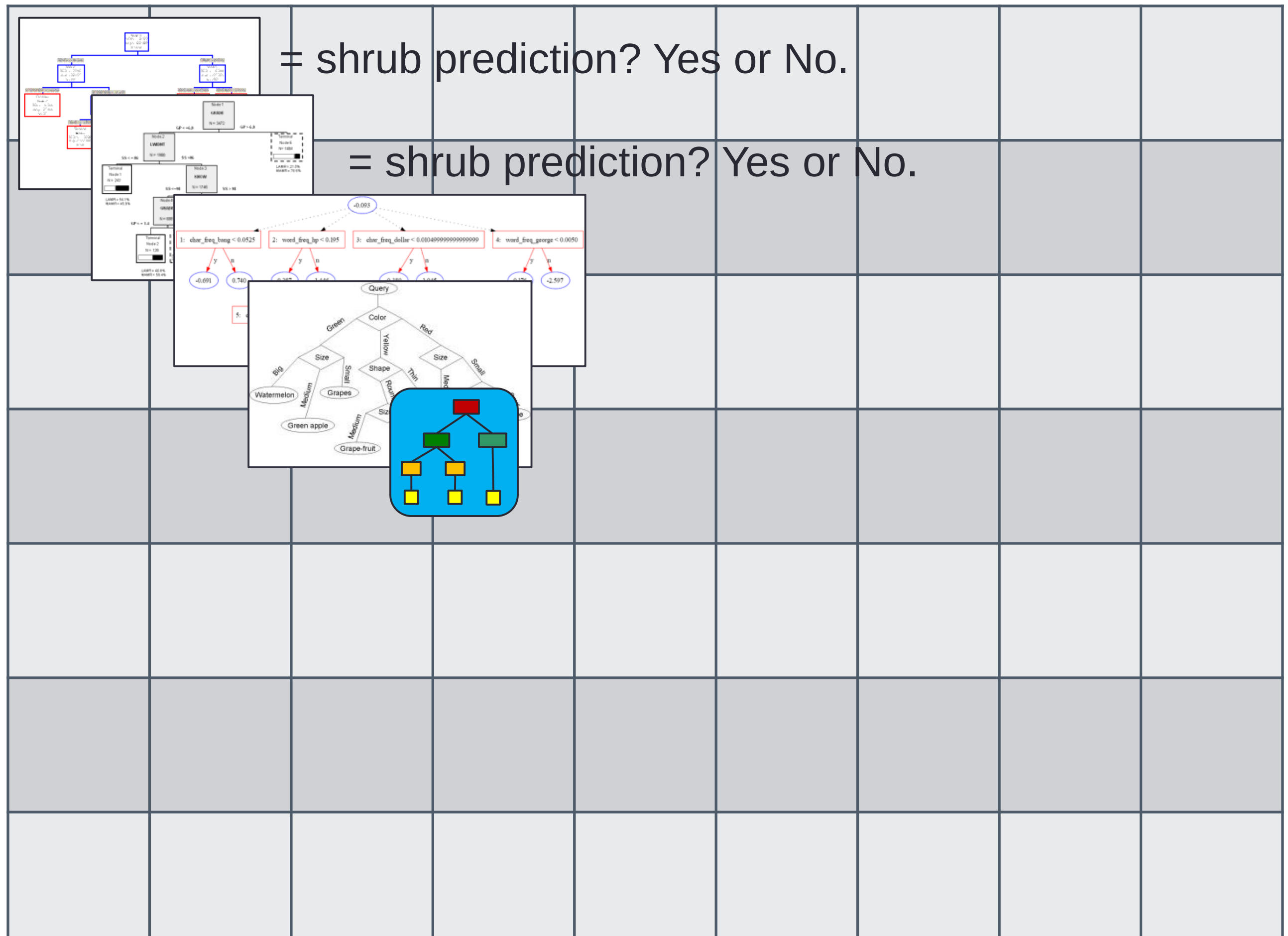


= A single CART model created during the Random Forest process.

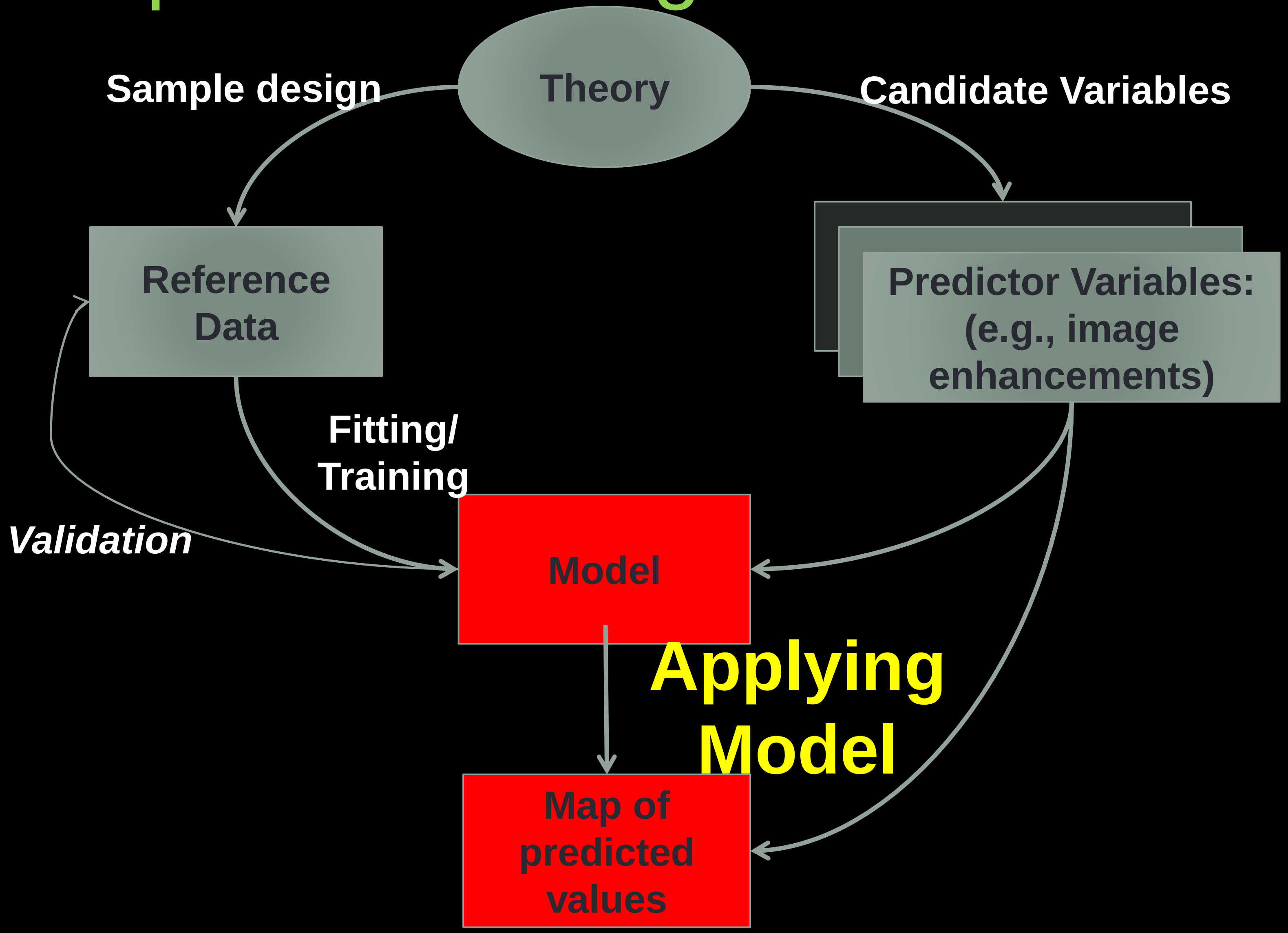


= A single pixel

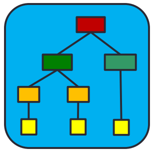
The CART models make a class prediction for each pixel using the predictor variables that were used as inputs to that particular CART model.



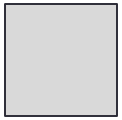
Spatial Modeling Process



Adapted from Guisan and Zimmerman (2000) & Franklin (2009)



A single CART model created during the Random Forest process.



A single pixel

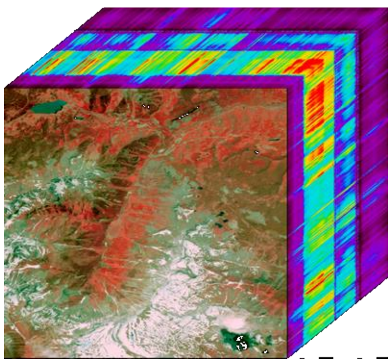
In this example, the most frequently occurring class modeled from the CART models for this pixel is Shrub.

So the final class for that pixel is Shrub.

What comes out of Random Forest?

Inputs

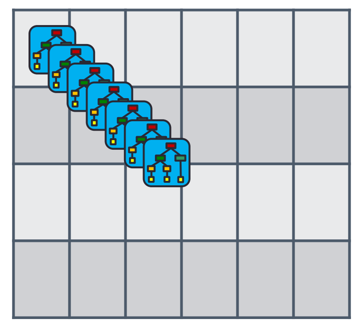
Outputs



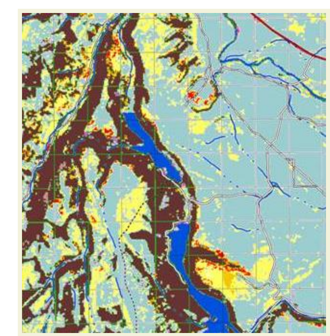
Input image-cube

(58 layers)

Field & PI training samples



Random Forest models are created & applied



Classification & labeled segments

United States Department of Agriculture

Qualitative assessment of land cover map output

Forest Service

Ways to improve classification

- Improving ROI
- Select better predictor variables
- Modify land cover class categories
- Sieve
- Try an object based approach

Unit of Analysis

- Pixel Based Classification
  - Uses spectral signatures
  - Ancillary data (elevation, etc)
- Object Based Classification
  - We can use more than spectral information!
  - Texture
  - Contextual relationships
  - Etc.

Object-based Image Analysis

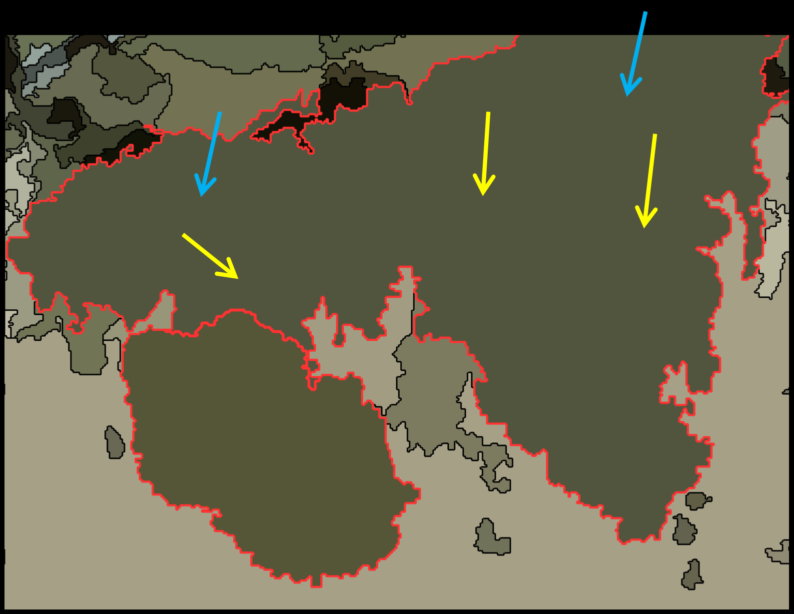
- What is object-based image analysis (OBIA)?
  - Image processing technique that uses image “objects” or “segments” as the basic unit instead of pixels

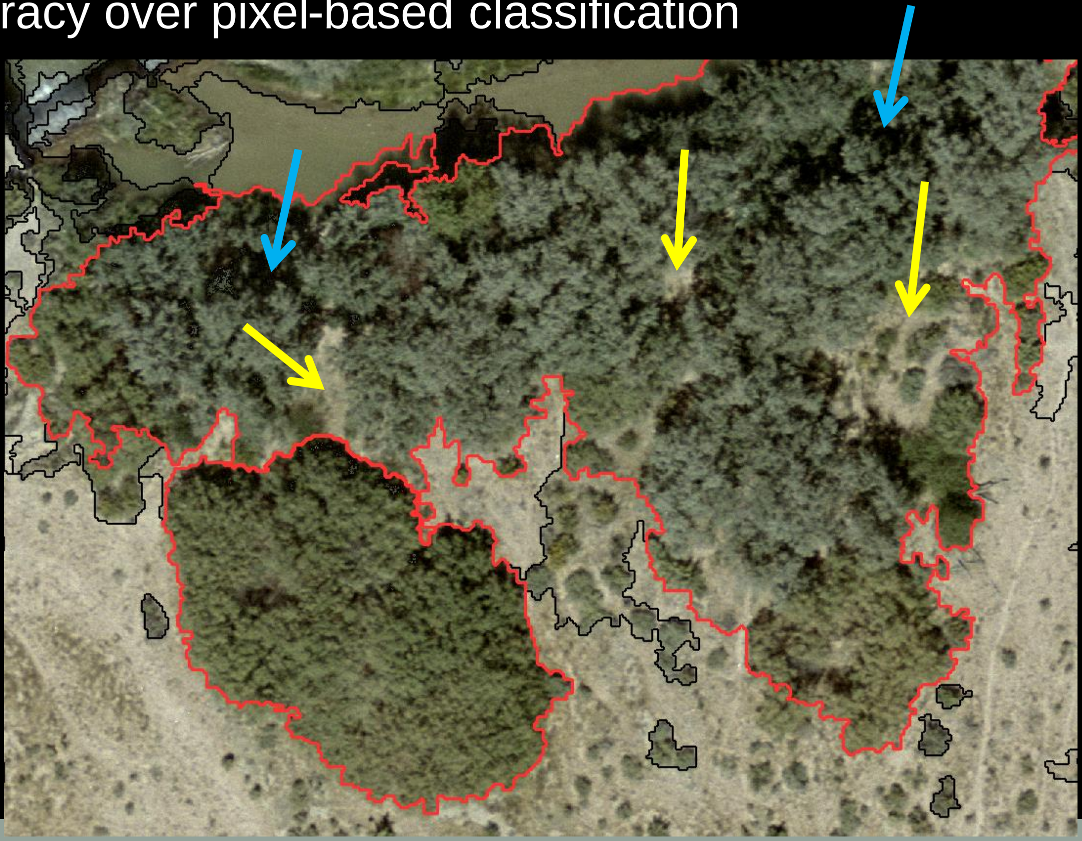
Object-based Image Analysis

- Division of imagery into homogeneous areas
  - Image segmentation
- Analysis performed on object or segments, not pixels
- Especially useful for high-resolution imagery, but beneficial for coarse resolution as well

Why object-based classification?

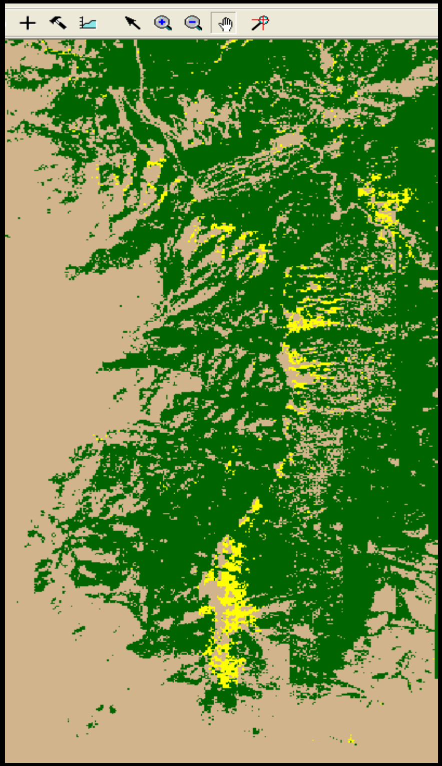
- Spectral data is averaged for each segment
  - Useful when dealing with features that have a range of spectral values (high texture)
- Segments filter out “noise” or meaningless information present in individual pixels and groups
  - Helps to reveal recognizable features
- Increase signal-to-noise ratio
- Can improve classification accuracy over pixel-based classification

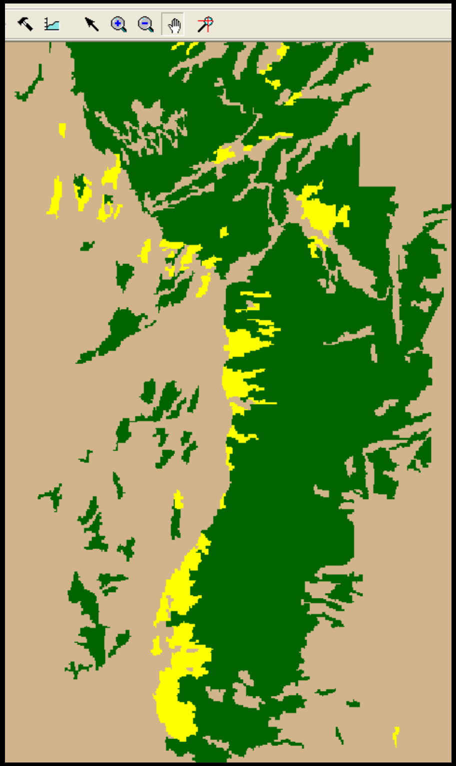




Why object-based classification?

- Object-based image analysis
  - Can use spectral AND other information (size, shape, context, texture, et cetera) for classification
  - Eliminates the pixel-based speckling or “salt and pepper” that we sometimes see



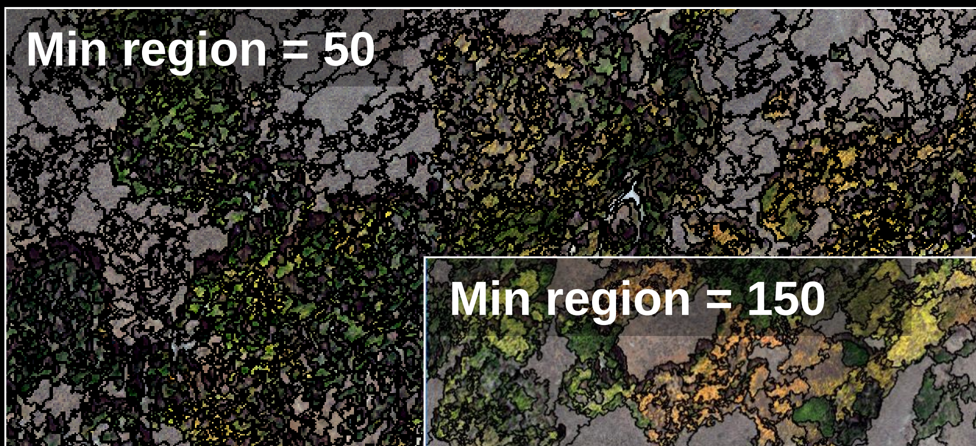


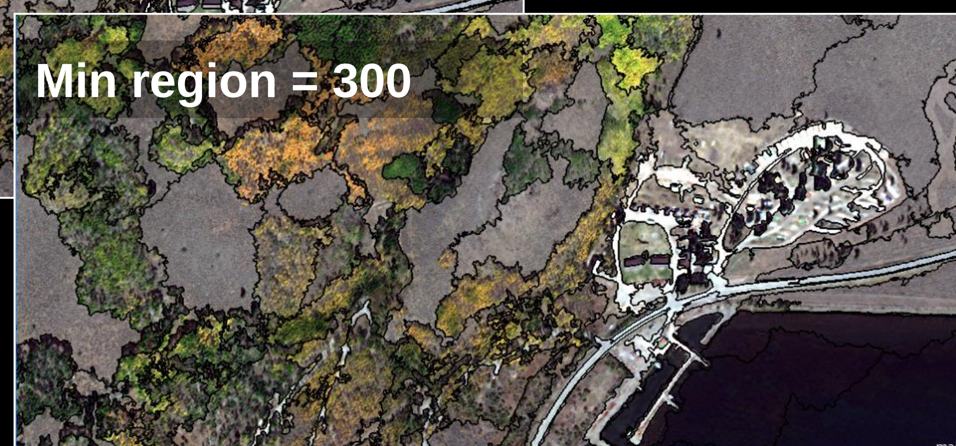
Segmentation

- Mean shift
  - Most commonly used
  - Robust algorithm
  - Delineates homogeneous areas
  - Memory and processor intensive

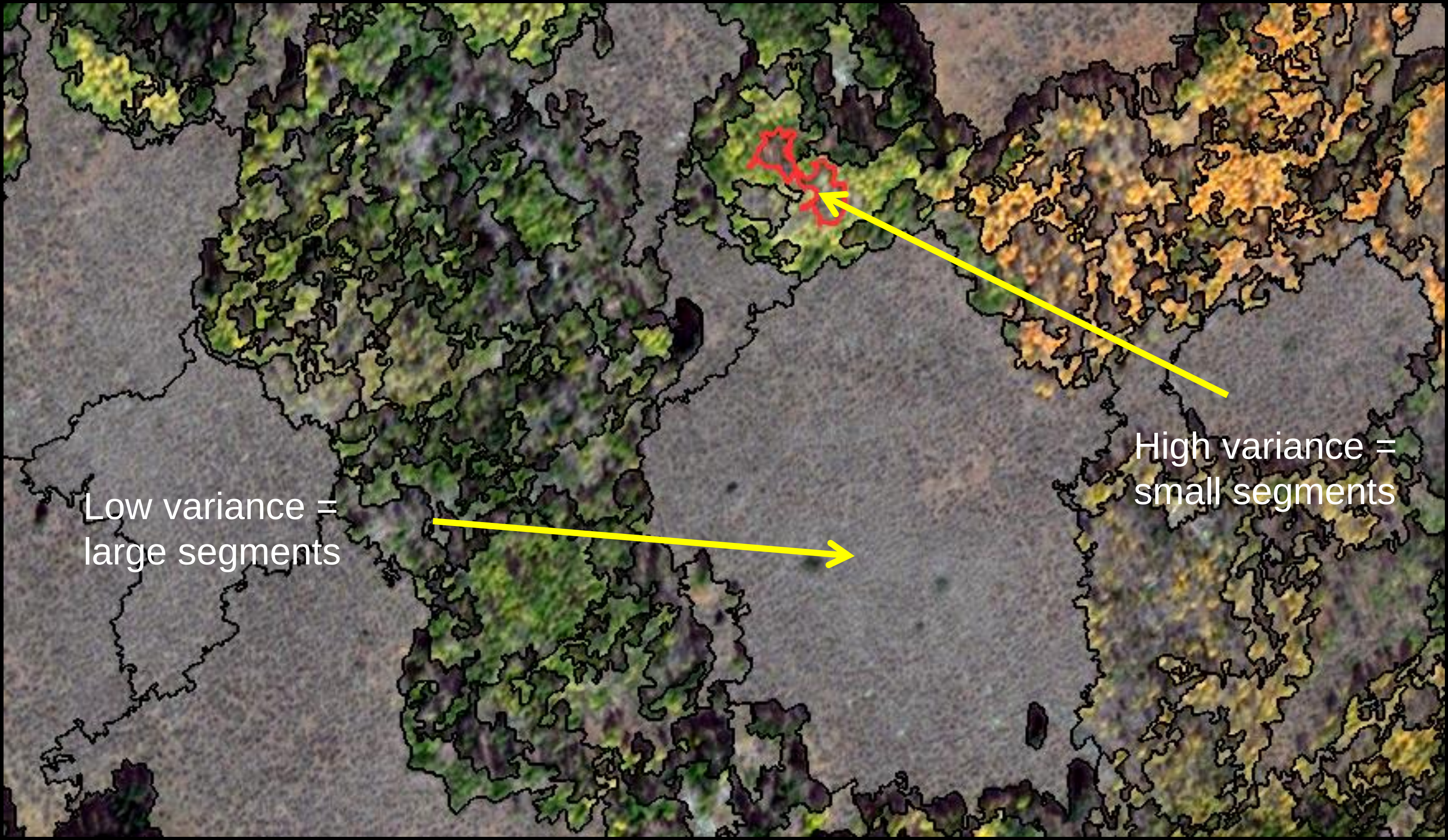
Segmentation

- Minimum region size controls the size of the segments





Segmentation



United States Department of Agriculture

If you have questions please contact
Bill Silva,
at USFS Geospatial Technology and Applications Center (GTAC):
billsilva@fs.fed.us
801-975-3804

Please contact Sarah Marlay,
at USFS International Programs
to learn more about international training opportunities:
sarahemarlay@fs.fed.us

Forest Service