Geospatial Problem Solving

Spatial Predictive Modeling

Image from NASA: https://commons.wikimedia.org/wiki/File:Earth_Western_Hemisphere_transparent_background.png#filelinks

WV VIEW

AmericaView
Empowering Earth Observation Education
AmericaView.org - Est. 2003

The goal of this module is to provide an introduction to spatial modeling.

We will discuss both liberal and conservative modeling methods.

We will end with a more conceptual discussion of more advanced modeling methods.

Module Topics

1. Modifiable areal unit problem (MAUP)
2. Spatial and temporal autocorrelation
3. Ecological fallacy
4. Conservative models
5. Liberal models
6. Weighted overlay
7. Physical/numeric models
8. Clustering and unsupervised learning
9. Empirical models and supervised learning
10. Machine and deep learning
11. Graphical modeling tools

2

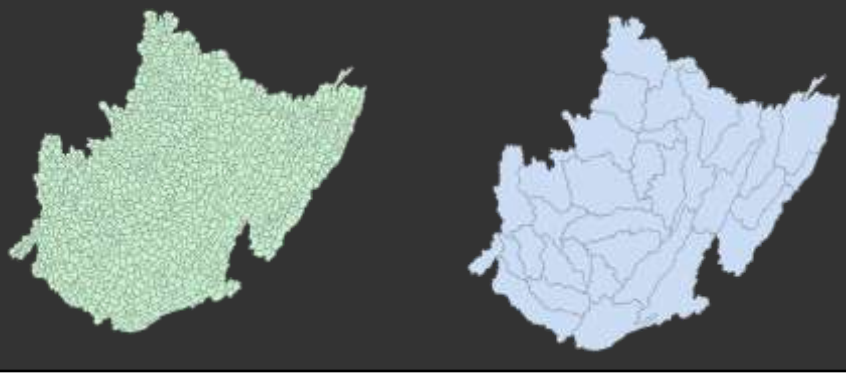
These are the topics covered in this module.

Modifiable Areal Unit Problem (MAUP) and the Ecological Fallacy

3

What is the modifiable areal unit problem (MAUP)?

- A challenge that occurs during the spatial analysis of aggregated data in which the results differ when the same analysis is applied to the same data, but different aggregation schemes are used.
- MAUP takes two forms: the scale effect and the zone effect



4

Before we begin, I will provide a discussion of some issues that may impact the spatial analyses that we perform and the spatial models we create.

We will begin with the modifiable area unit problem or MAUP. This is a challenge that occurs during the spatial analysis of aggregated data in which the results differ when the same analysis is applied to the same data, but different aggregation schemes are used. MAUP takes two forms: the scale effect and the zone effect.

Modifiable Areal Unit Problem (MAUP)

Openshaw (1984)

"The areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating."

Example: Gerrymandering

5

Here is a definition of MAUP after Openshaw (1984):

"The areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating."

A good example is Gerrymandering, where voting districts are altered to increase the likelihood of the desired outcome.

Scale Effect

- The scale effect suggest that analyses will exhibit different results when the same analysis is applied to the same data, but the scale of aggregation is changed.
- For example, analyses using data aggregated by county will differ from analyses using data aggregated by census tract. Often this differences in results are valid: each analysis asks a different question because each evaluates the data from a different perspective (different scale).
- Take home: It is important to consider the scale of your analysis and your aggregating units because you are addressing your question at that specific aggregation scale

6

The scale effect suggest that analyses will exhibit different results when the same analysis is applied to the same data, but the scale of aggregation is changed.

For example, an analysis using data aggregated by county will differ from an analysis using data aggregated by Census tract.

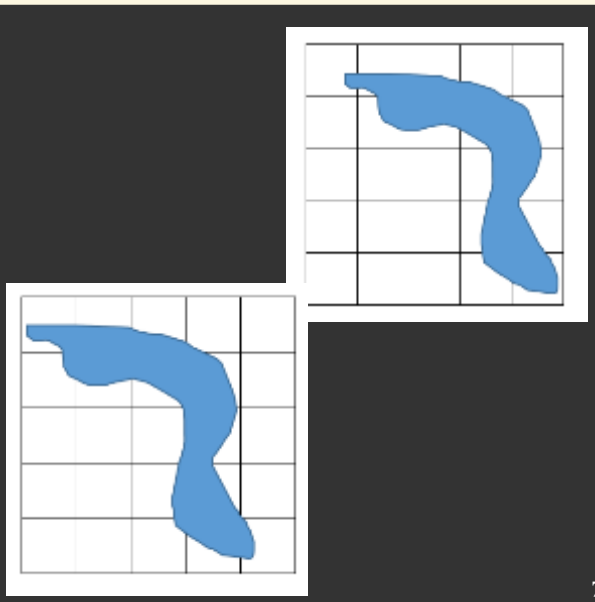
Often these differences in results are valid: each analysis asks a different question because each evaluates the data from a different perspective or scale.

So, scale of aggregation is an important component of any analysis that should be thought through and not ignored.

It is important to consider the scale of your analysis and your aggregating units because you are addressing your question at that specific aggregation scale.

Zone Effect

- The zone effect is observed when the scale of analysis is fixed, but the shape or position of the aggregation units change.
- For example, an analyses using data aggregated into one-mile grid cells will differ from an analysis using one-mile hexagon cells.
- The zone effect is a problem because it is an analysis, at least in part, of the aggregation scheme rather than the data itself.



7

The second component of MAUP is the zone effect.

The zone effect is observed when the scale of analysis is fixed, but the shape or position of the aggregation units is altered.

For example, analysis using data aggregated into one-mile grid cells will differ from analysis using one-mile hexagon cells.

In the example provided, we may see a different outcome simply because the rectangular grid was shifted relative to the blue polygon, even though the sizes are the same.

The zone effect is a problem because it is an analysis, at least in part, of the aggregation scheme rather than the data itself.

Spatial Heterogeneity/Spatial Autocorrelation

- Patterns can vary across space
- Observations or correlations noted at one location may not be true at other locations
- Samples may be correlated based on their proximity to each other in space

8

Patterns can vary over space. This is the concept of spatial heterogeneity.

Observations or correlations noted at one location may not be true at other locations.

So, findings for certain geographic areas may not translate or should not necessarily be applied or assumed at other locations.

Samples may be correlated based on their proximity to each other in space, or spatial autocorrelation may be present. For example, you would expect average annual precipitation to be more similar for two adjacent counties than counties separated by a great distance. It is generally not safe to assume that samples collected over map space are independent of each other. Some methods assume samples are independent, so this may be an issue.

Ecological Fallacy

- The assumption that an individual from a specific group or area will exhibit a trait that is predominant in the group as a whole

9

The ecological fallacy relates to assuming that an individual from a specific group or area will exhibit a trait that is predominant in the group as a whole.

For example, you should not assume that everyone in your neighborhood likes pizza just because the general trend is that your neighbors enjoy pizza.

You should not extrapolate from the population to the individual.

Temporal Heterogeneity/Temporal Autocorrelation

- Observed patterns and correlations may change over time
- Observations near in time tend to be more similar than observations that are further apart in time

10

Patterns can also change through time. Think about seasonal patterns or changes in the climate or landscape over longer time frames. Observations made at a certain time may not hold true in the past or the future.

There can also be a correlation of data through time. For example, you would expect that the temperature five minutes from now is more likely to be similar to the current temperature than the temperature in three weeks.

Many statistical tests assume data points are independent. So, temporal autocorrelation or spatial autocorrelation can be an issue.

Practical Considerations

1. Scale matters
2. Can't assume patterns are consistent over space
3. Do not extrapolate from the population to the individual
4. Do not assume that measures, patterns, or results will be consistent over time
5. Do not assume that samples collected over space are independent

11

How do we practically address or consider these concerns when conducting spatial analysis or making models? Here are some thoughts.

First, scale matters. Think about what scale is most appropriate for your analysis. It might be a good idea to aggregate your data at multiple scales to explore how the patterns observed and the results obtained may vary.

Be careful not to assume that patterns and results observed at one location will hold at other locations. It might be necessary to conduct an analysis in other locations to see if the results hold or are transferable.

Also, do not assume that there is no spatial autocorrelation in your data or that your samples or measures are purely independent. It may be necessary or valuable to consider the impacts of spatial autocorrelation.

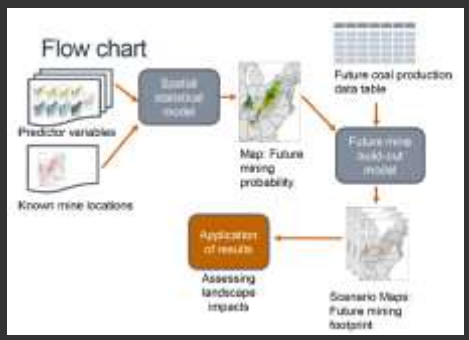
You should also not assume that the general or average characteristics of a population hold for individuals in that population. Do not extrapolate from the population to the individual.

Do not assume that measures, patterns, or results are stationary or consistent over time. If change over time is important, you may need to include a temporal analysis in your study.

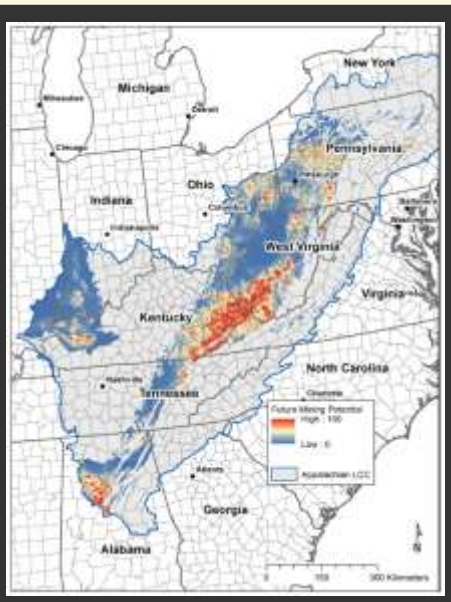
Overview of Modeling

What is a model?

A simplified representation of a phenomenon or a system



Strager, M.P., J.M. Strager, J.S. Evans, J.K. Dunscomb, B.J. Kreps, and A.E. Maxwell, 2015. Combining a spatial model and demand forecasts to map future surface coal mining in Appalachia, *PLoS ONE*.



13

Models are a simplified representation of a phenomenon or a system.

Modeling is different from mapping. With mapping, we are generally recording or documenting what is there.

In contrast, models are generally used to make predictions about occurrence or what will happen in the future.

For example, we can't map the weather in five days because we don't know what it will be.

However, we can predict what the weather may be like in five days. So, this is a model.

The example provided shows a model that predicts the likelihood of coal surface mine expansion in the mid-Atlantic Region of the eastern United States.

What is a model?

"Essentially, all models are wrong, but some are useful."

-Statistician George Box

14

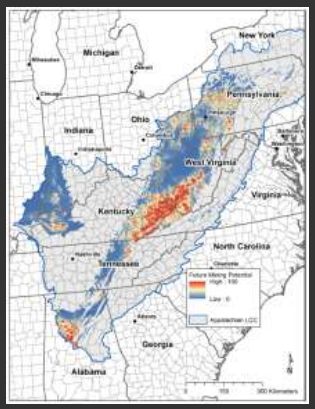
It should be noted that, since models are predictions, they all have inherent error. However, just because a model contains some error does not mean it will not be useful for us to make a prediction or informed decision.

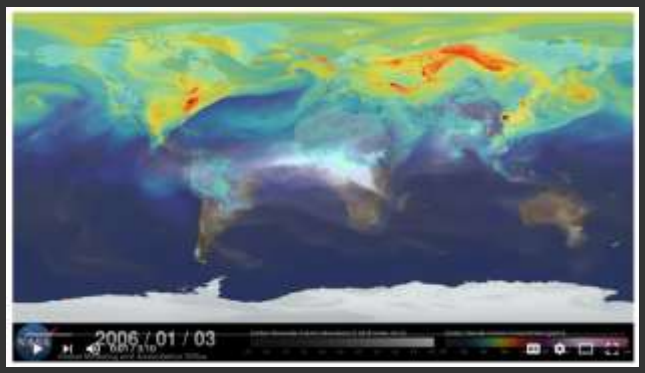
Even though weather predictions have some error, they are still useful to help you make plans, such as how to dress or when to ride your bike to work.

As stated by the statistician George Box, "Essentially, all models are wrong, but some are useful."

Static vs. Dynamic Models

- Static = does not change with time
- Dynamic = changes with time





https://www.youtube.com/watch?v=x1SgmFa0r04

15

Static models are models that do not change with time whereas models that change with time are dynamic.

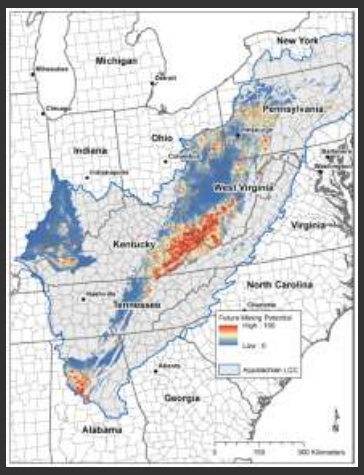
For example, a model of animal habitat could be a static model if only a single or static prediction is made.

Weather models are often dynamic, since they change throughout the time period predicted by the forecast.

This slide provides a link to a model that shows changes in carbon dioxide concentration over time. This would be a dynamic model because it changes with time.

Conservative vs. Liberal Models

- Conservative = binary, 1/0, yes/no
- Liberal = represents gray levels



16

Conservative models generally provide a binary or yes/no output.

For example, a site is deemed to be suitable or not suitable. A site can not be partially suitable.

In contrast, liberal models can represent gray levels.

For example, an index can be calculated to assess the degree to which a site is suitable for a specific purpose.

Deterministic vs. Stochastic

- Deterministic = does not incorporate randomness
- Stochastic = does incorporate randomness

17

Deterministic models do not incorporate randomness whereas stochastic models have a random component.

Deterministic models, if given the same data, will always generate the same results whereas stochastic models may generate different results even given the same data.

An example of a deterministic model is linear regression. If given the same inputs, the same equation will be generated each time.

In contrast, many machine learning methods would be described as stochastic since there is a random component.

Deductive vs. Inductive

- Deductive = conclusion derived from a set of premises (general → specific)
- Inductive = conclusion derived from empirical data and observations (specific → general)

18

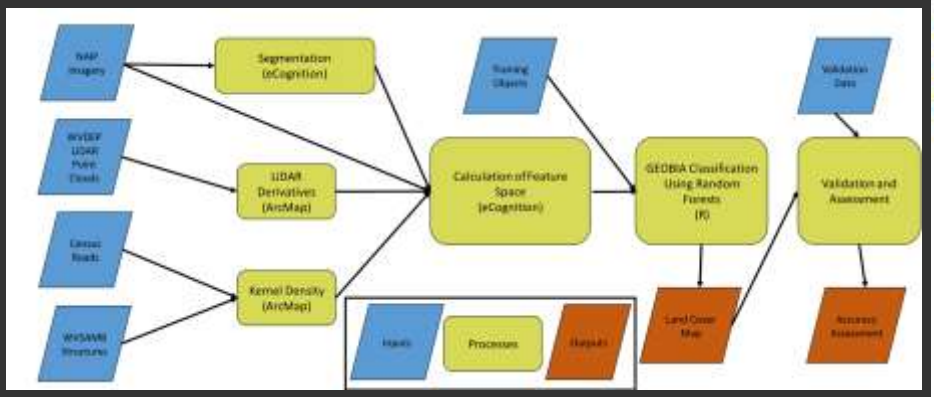
Models can also be deductive or inductive.

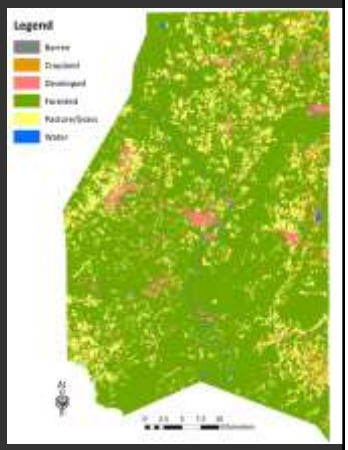
Deductive methods use general premises to derive or come to a specific conclusion whereas inductive models use empirical data and observations to derive a general premise from the specific data.

The scientific method makes use of deductive reasoning. If these general premises are true, then this specific thing should happen.

Sherlock Holmes uses inductive reasoning. Based on specific observations, he concludes who the culprit is.

Modeling Example: Land Cover Classification





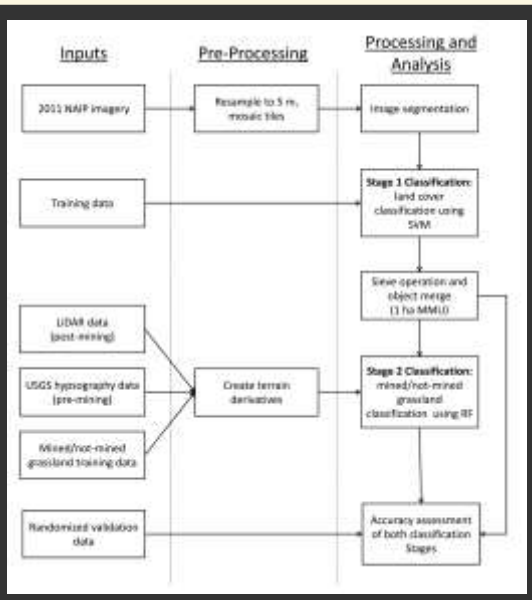
19

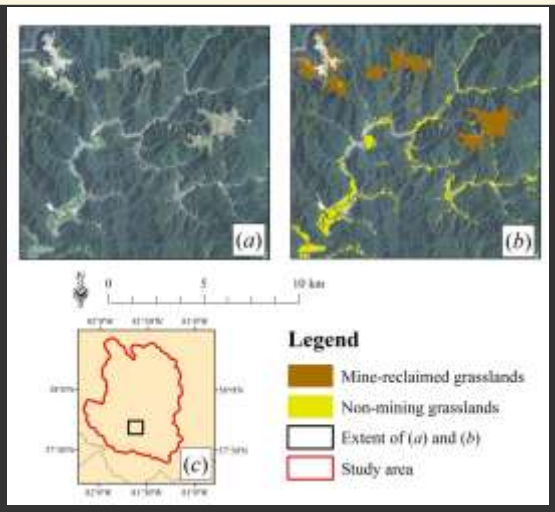
Here, I will provide a few examples of uses of modeling in the geospatial sciences.

One example use of modeling in the geospatial sciences is land cover mapping.

The goal here is to determine the land cover types that occur across a landscape using their signatures within image data.

Modeling Example: Land Cover Classification



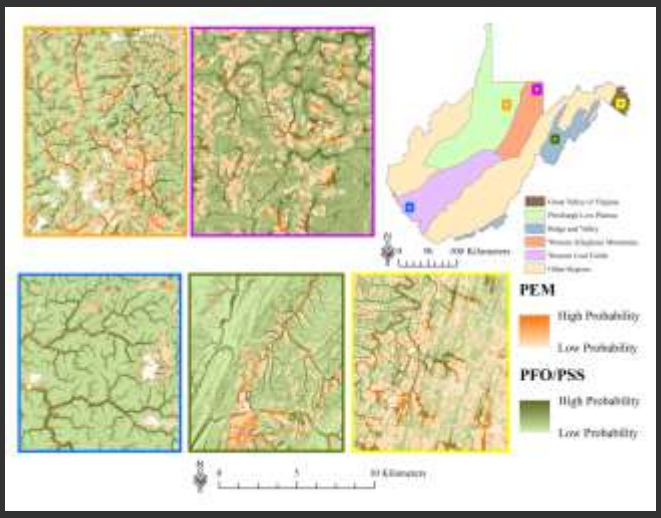


Maxwell, A.E., and T.A. Warner, 2015. Differentiating mine-reclaimed grasslands from spectrally similar land cover using terrain variables and object-based machine learning classification, *International Journal of Remote Sensing*.

20

This slide shows another land cover classification model where mine-reclaimed grasslands are differentiated from other grasslands.

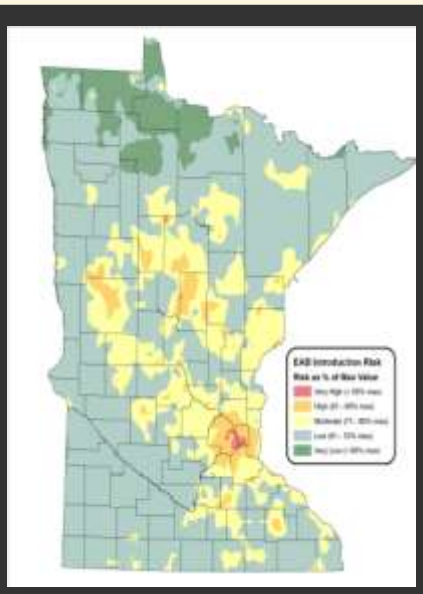
Modeling Example: Wetland Probability

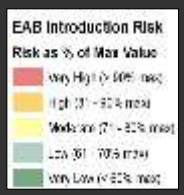


Maxwell, A.E., T.A. Warner, and M.P. Strager, 2016. Predicting palustrine wetland probability using random forest machine learning and digital elevation data-derived terrain variables, *Photogrammetric Engineering & Remote Sensing*.

This slide shows another modeling result in which the probability or likelihood of wetland occurrence is predicted.

Model Example: Pest Risk Mapping (Emerald Ash Borer)





Equation 4 = Same as Equation 2 but Campsites multipled by 25, Seasonal homes multiplied by 25 and Accessibility multiplied by 1 (Figure 13):

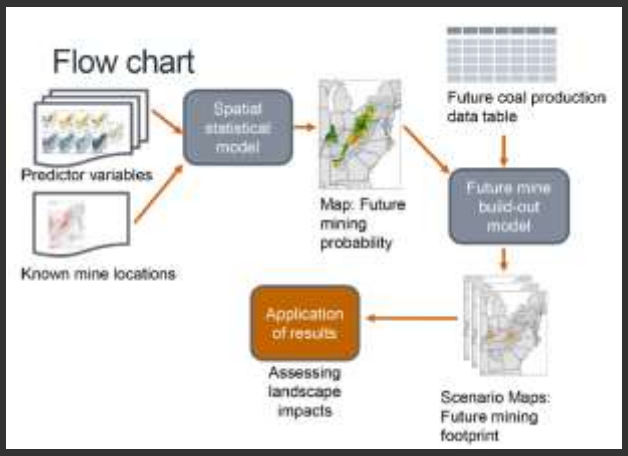
Log((25 * [Reclassified Campsite Density]) +
(25 * [Reclassified Cabin Density]) +
(10 * [Reclassified Urban Centroid Density]) +
(5 * [Reclassified Sawmill Risk Density]) +
(5 * [Reclassified Firewood Dealer Density]) +
(5 * [Reclassified High Risk Nursery Density]) +
(1 * [Reclassified Accessibility]))
= Risk Map Grid Surface

https://www.mda.state.mn.us/

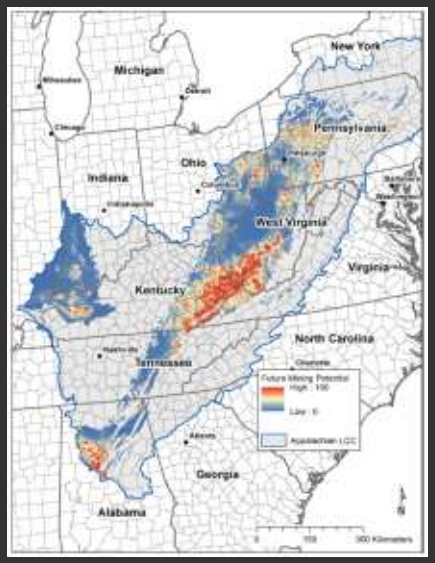
22

As yet another example, this model predicts the risk of an Emerald Ash Borer pest outbreak across Minnesota.

Model Example: Mine Expansion



Strager, M.P., J.M. Strager, J.S. Evans, J.K. Dunscomb, B.J. Kreps, and A.E. Maxwell, 2015. Combining a spatial model and demand forecasts to map future surface coal mining in Appalachia, *PLoS ONE*.

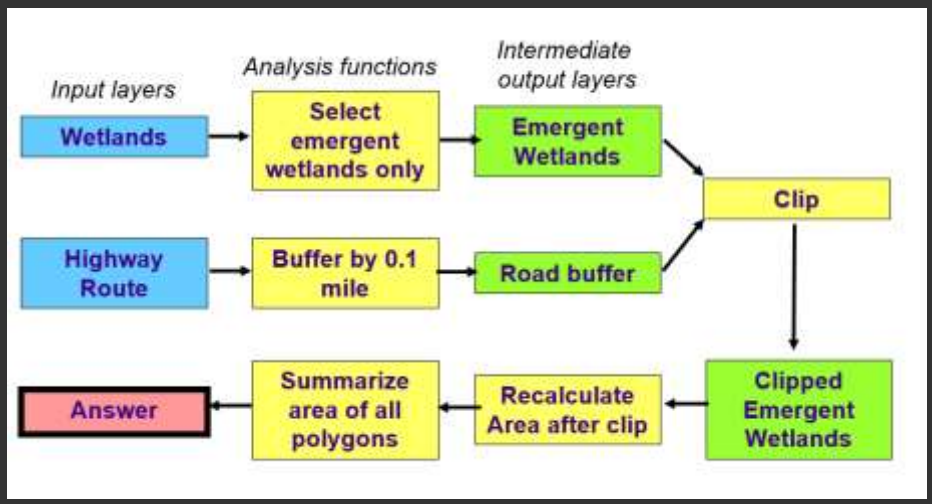


23

Lastly, this slide shows the modeling output for predicting the likelihood of mine expansion discussed earlier.

Conservative Models

Vector Overlay



Example from Mike and Jackie Strager

25

We will now discuss conservative modeling methods. We have already discussed many tools used to produce these models while discussing vector and raster analysis.

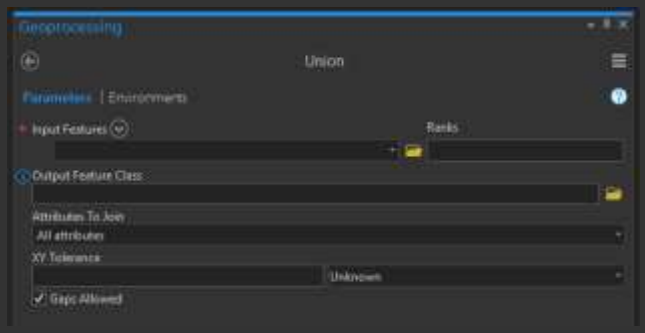
In this example, the goal is to predict areas of emergent wetlands that could be impacted by potential highway construction.

This involves selecting only emergent wetlands from all wetlands, buffering the proposed highway by a defined distance, clipping out the emergent wetlands occurring within this buffer, then summarizing the wetland area.

The power of the spatial analysis techniques that we have discussed is the ability to combine them to solve more complex problems.

Vector Overlay Tools

- Dissolve
- Buffer
- Clip
- Intersect
- Union
- Spatial Query
- Table Query
- Etc.

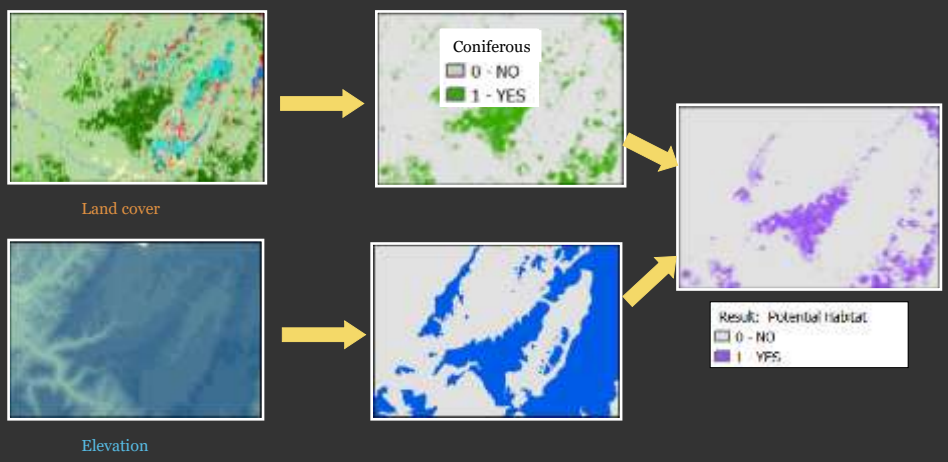


26

This slide lists some tools that are commonly used to perform vector overlay.

Raster Overlay

- Habitat = coniferous forested areas over 1,000 meters in elevation



Example from Mike and Jackie Strager

27

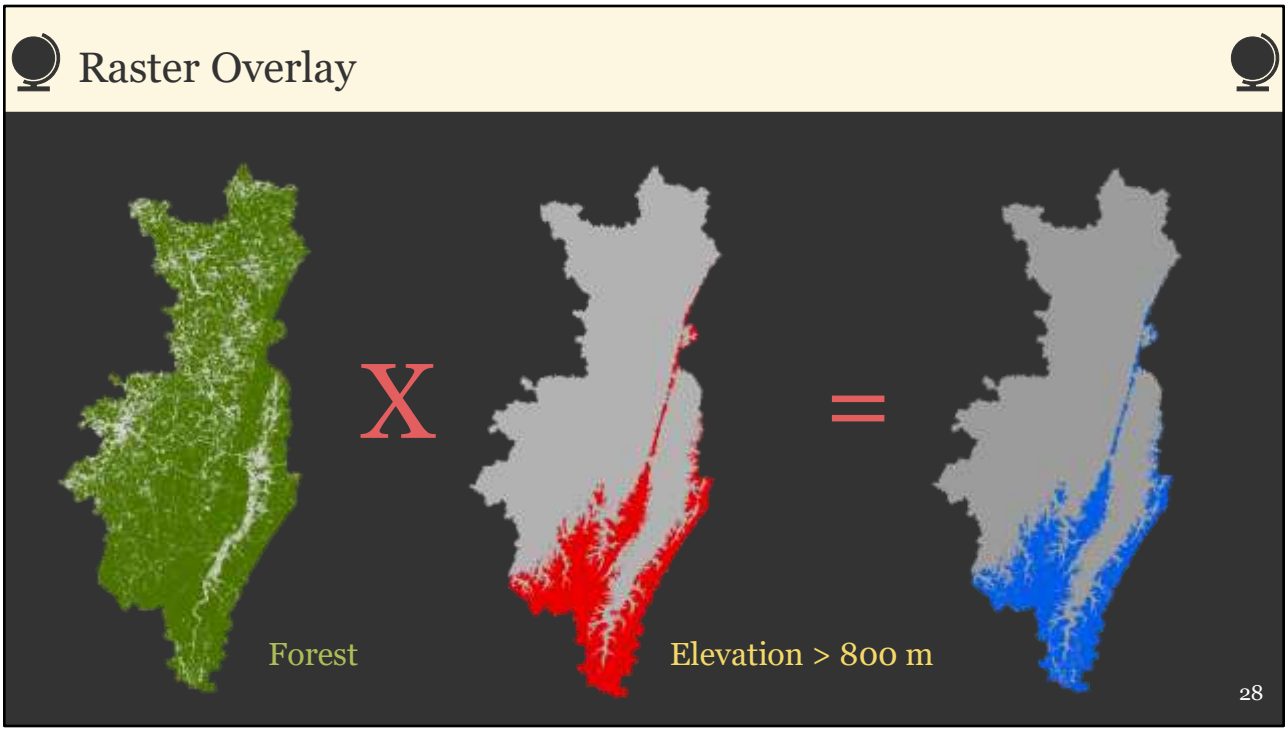
Raster analysis techniques can also be used to created models.

In the example provided here, Northern Flying Squirrel habitat is defined as conifer forests over 1,000 meters in elevation.

Using an elevation dataset and raster math or reclassification, a binary surface can be generated as not conifer forest and conifer forest.

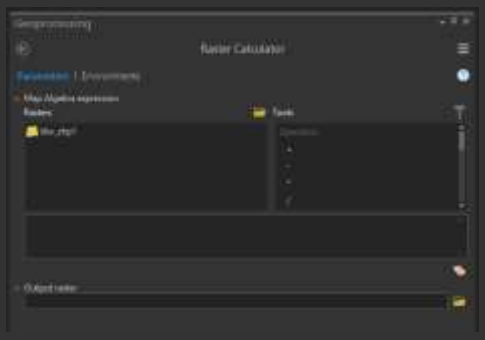
Elevations above 1,000 meters can be determined from a digital elevation model also using raster math.

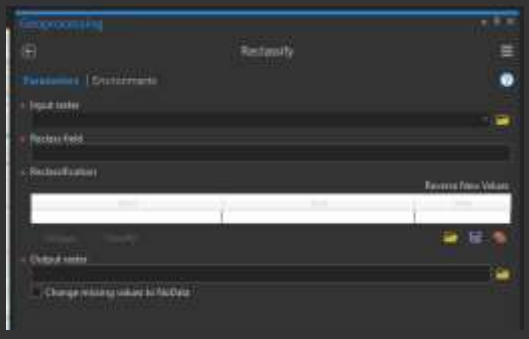
These two surfaces can then be multiplied together to find all areas that are both conifer forests and above 1,000 meters.



Here is another example where potential habitats are defined as forested areas above 800 meters in elevation. By multiplying the two binary surfaces, it is possible to find cells that are both forested and above 800 meters in elevation.

Raster Overlay Tools





29

Raster Calculator and Reclassify are commonly used in raster overlay analysis.

Limitations

- Binary output may not be optimal or appropriate
- Difficult to assign thresholds/cut-offs
- Error can compound

30

All modeling methods have some limitations, and conservative modeling using vector and/or raster overlay is no exception.

First, a binary or yes/no output may not be a desired, optimal, or adequate result.

It is also sometimes difficult to assign thresholds. For example, how do you know what distances from roads or elevations are appropriate?

Also, errors in the input data, such as misclassifications or outdated data, can compound in complex and nonlinear ways in the output, and this can be difficult to quantify.

Weighted Overlay

Overview of Weighted Overlay

- Used to create liberal models

Overall Score = [Criteria Score 1 X Weight 1] + [Criteria Score 2 X Weight 2] + [Criteria Score 3 X Weight 3] + [Criteria Score 4 X Weight 4] + [Criteria Score 5 X Weight 5] + ...

32

A common method used to perform liberal modeling is weighted overlay.

Using this method, weighted scored criteria are combined to produce a suitability or ranked index.

Steps in Weighted Overlay

1. Define the overall problem or goal
   - What are you ranking?
   - What is the required result?
2. Determine evaluation criteria
3. Find spatial data that represent your criteria
4. Determine scoring within criteria
5. Determine weights for individual criteria
6. Calculate result
7. Evaluate and present results

33

These are the general steps in performing weighted overlay.

First, you must define the overall problem or goal. This is key, as a well-defined goal will yield a more meaningful and useful result. For example, if you are modeling animal habitat, what do you mean by habitat? If you are mapping suitable locations to build a new industrial park, what do you mean by suitable? Clearly defining the problem will allow you to produce a result that is usable and informative.

You will then need to define the evaluation criteria. What factors are important for modeling the problem? For example, if you are trying to map animal habitat, what factors impact whether the specific animal will occupy an area? Does elevation matter? How about land cover?

Once you have decided what criteria are important, you will need to find spatial data that represent the criteria. Since your goal is to make a prediction over a defined spatial extent, you must find data that cover this extent or are spatially complete.

You will then need to define the scores. What values indicate suitability? For example, does this species occupy high elevations or low elevations?

Since not all criteria will be of the same importance, you will then need to define weights to approximate the relative influence.

You are then ready to produce the model.

Once the model is produced, you need to evaluate and present the results.

This process is often iterative. So, after evaluating the product and getting feedback, you may need to update or refine the model.

It is important to consider:

1. What is the goal?
2. How will the model be used?
3. Who will use the model?
4. Does your model meet the need?

34

Again, when performing a weighted overlay, it is important that you clearly define the problem and what you are trying to estimate.

What is the goal?

How will the model be used?

Who will use the model?

Does the model meet the need?

Defining Criteria

- Land cover/land use
- Landscape patterns
- Land ownership
- Human Infrastructure
  - Structures
  - Roads
  - Utilities
  - Airports
  - Etc.
- Distance from X
- Density of X
- Soil characteristics
- Terrain characteristics
  - Elevation
  - Slope
  - Aspect
  - Slope position
  - Etc.
  - Climate considerations
  - Temperature
  - Rainfall
  - Etc.
  - Surface hydrology
  - Etc.

Based on:

- Expert knowledge
- Stakeholder preferences
- Consult the literature

35

A wide variety of criteria can be considered, so this can be a bit daunting. This slide provides a list of some common factors considered.

How do you decide on appropriate criteria?

This is generally accomplished by consulting with experts and stakeholders. You can also reference professional and scientific literature.

You should be able to justify what criteria you used and why.

Data Considerations

1. Is the dataset available?
2. Is the dataset of adequate quality?
3. Is the dataset spatial or can it be made spatial? Can criteria be interpreted into spatial and quantifiable variables?
4. What is the spatial resolution?
5. Is it up to date?

36

When defining your criteria it is important to do so critically.

First, a variable will be of little value if there is no dataset available to describe it. For example, if you are interested in including forest types in your model, but no forest classification is available, then this variable cannot be included. It may take some digging to find suitable data.

If you find data to represent a variable, is it of adequate quality? Including data that are suspect, uncertain, inaccurate, or imprecise can potentially reduce the quality of the model. So, it might be best to simply not include it.

Since your goal is to make a prediction over a map space, all variables must be spatial, or you must be able to convert them into spatial data. For example, if you are trying to predict the habitat of a tree species and the literature suggests that it grows "on ridges," you must be able to create a variable that differentiates ridges or provides some measure of ridge vs. valley.

The spatial resolution of the data can be of concern. Since you are likely to use a variety of inputs, it is a common issue that the spatial resolution of the inputs will vary. For example, you may want to include an estimate of mean annual precipitation. However, all the available data may be too coarse compared to your other datasets or to meet your needs.

Lastly, the timeliness of the data may be important, especially for factors that can change rapidly, such as demographics or residential development. This may be less of an issue for factors that tend to be fairly constant or change very slowly, such as geologic conditions.

Defining Scores

- What values are most suitable? Least suitable?
- Methods
- Discrete = Yes/No, 1/0, binary
- Ranges = Break into ranges
- Continuous = Linear increasing, linear decreasing, something more complex
- It is important that all criteria are scored on the same scale

Distance to Towns	
Distance	Score
≤ 0.5 km	10
0.5 to 2 km	7
2 to 4 km	5
> 4 km	3

$$\frac{X_i - X_{min}}{X_{max} - X_{min}}$$

37

Once you have developed your criteria, you will need to apply scores.

The most suitable values will be scored high and the least suitable values will be scored low.

All variables should be scored using the same scale, such as 0 to 1, 0 to 10, or 0 to 100.

There are different methods available for apply scores.

Discrete scores are used when there are only two options: the values are suitable, or they are not suitable, similar to conservative modeling.

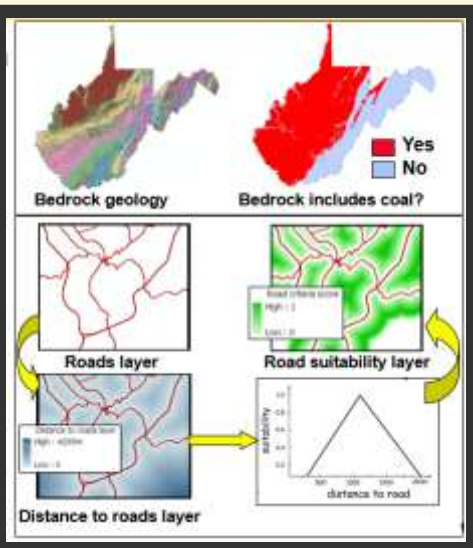
A specific score can be assigned to a range of values, as demonstrated in the provided table. Different ranges of distance from towns are scored using four discrete bins.

Lastly, data can also be scored continuously. For example, as demonstrated in the provided equation, the minimum value can be subtracted from the current value then divided by the range. This will rescale the data from 0 to 1, where 0 indicates least suitable and 1 indicates most suitable. If you would like to invert this, so that low values are more suitable and high values are less suitable, you

can simply subtract the result from 1. There are other more complex functions that can be applied to rescale data.

Defining Scores

- What values are most suitable? Least suitable?
- Methods
- Discrete = Yes/No, 1/0, binary
- Ranges = Break into ranges
- Continuous = Linear increasing, linear decreasing, something more complex
- It is important that all criteria are scored on the same scale



Example from Mike and Jackie Strager

38

This slide provides an example of discrete and continuous scoring.

The bedrock geology is scaled discretely: the unit either has coal or it does not.

In the second example, distance from roads is scored continuously. First, a Euclidean distance grid is generated. Next a function is applied in which moderate values are scored high, or are most suitable, and low or high values are scored low, or are least suitable. The data are rescaled from 0 to 1.

Again, it is important that all criteria are scored using the same scale.

Defining Weights

What values are most important?

Based on:

1. Best judgement
2. Expert knowledge
3. Literature
4. Stakeholder preference (surveys)

Note: There are statistical methods to determine if multiple stakeholders value the same criteria similarly

39

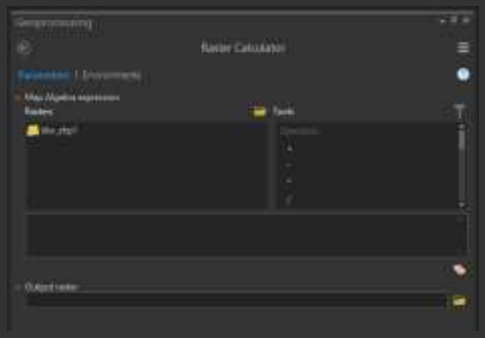
Once the data have been scored, you then need to define weights.

Weights are generally determined based on best judgement, expert knowledge, professional literature, academic literature, or stakeholder preferences.

Stakeholder preferences can be gauged using surveys. Note that there are statistical methods to determine if multiple stakeholders value the same criteria similarly; however, this is outside the scope of our discussion.

Calculating the Model: Raster Calculator

Overall Score = [Criteria Score 1 X Weight 1] + [Criteria Score 2 X Weight 2] + [Criteria Score 3 X Weight 3] + [Criteria Score 4 X Weight 4] + [Criteria Score 5 X Weight 5] + ...



40

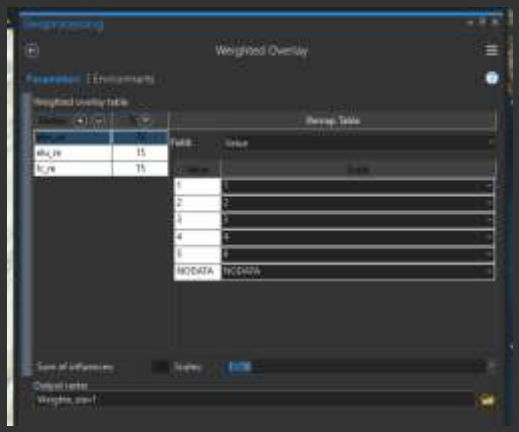
Once you have fully defined your goal, determined your criteria, created scored grids, and defined weights, you are ready to create your model. This can be accomplished using raster math. Each criterion will need to be multiplied by its weight then the results will be added.

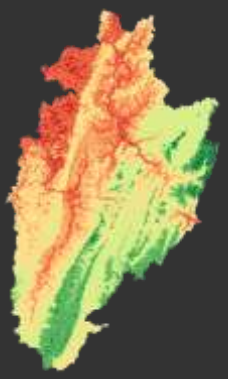
This will create an index score where high values indicate the best or most suitable locations and low values indicate the least suitable locations.

Since this is simply an index, it is unitless.

Calculating the Model: Weighted Overlay Tool

- Provide scored grids
- Define % of influence for each criterion
- Sum of influence must by 100%
- Can define the scale of the output





41

ArcGIS Pro provides a weighted overlay tool. This tool requires scored grids and defined percent of influence for each criterion to specify the weight.

The percent of influence must sum to 100%.

The output can also be rescaled using different ranges.

Calculating the Model

Criteria (C)	Scoring for criterion (C)		Weight of criterion (W)	Score for Site "X"
1. Distance to nearest interstate	10 5 1	within 1 mile within 5 miles within 10 miles	10	5 X 10 = 50
2. Distance to nearest airport	10 5 1	within 10 mile within 50 miles within 100 miles	15	10 X 15 = 150
3. Population in community	10 5 1	> 10,000 5,000 to 9,999 < 5,000	5	1 X 5 = 5

Overall score for site = sum((C1)(W1) + (C2)(W2) + (C3)(W3))= (50 + 15+5) = 205

Example from Mike and Jackie Strager

42

Here is an example of weighted overlay. Three criteria are defined: Distance to nearest interstate, distance to nearest airport, and population in community.

Each are scored using ranges on a scale of 1 to 10. Being nearer to interstates and airports is preferred. Higher population in the community is preferred.

Distance to airports is weighted highest followed by distance to interstates. The population in the community is least important.

The values highlighted in red are the results for a certain location. By multiplying each score by the associated weight then adding them, we can obtain a unitless index value.

Model Validation

How do you know if your model makes sense?

- Not always easy to do in a quantitative/statistical manner
- Visual assessment
- Present to:
  - Experts
  - Stakeholders
  - Public
- Solicit feedback
- Sources of uncertainty and error should be made known

43

Once a model is produced, it will need to be evaluated.

This is generally done based on both qualitative and quantitative methods.

Models can be assessed visually. Does the map output make sense?

It can sometimes be presented to experts, stakeholders, and the public to solicit feedback. Again, this will generally be an iterative process where feedback is integrated to improve the results.

The methods should be well documented. Any sources of uncertainty or error should be made known.

One limitation of weighted overlay is that it is often difficult to statistically or quantitatively assess the output since validation data may not be available.

Sensitivity Analysis

- How does your model change when varying scores/weights?
- How does your model change with changes in criteria?
- How does your model change with data source choice?

44

Another component of model validation is sensitivity analysis.

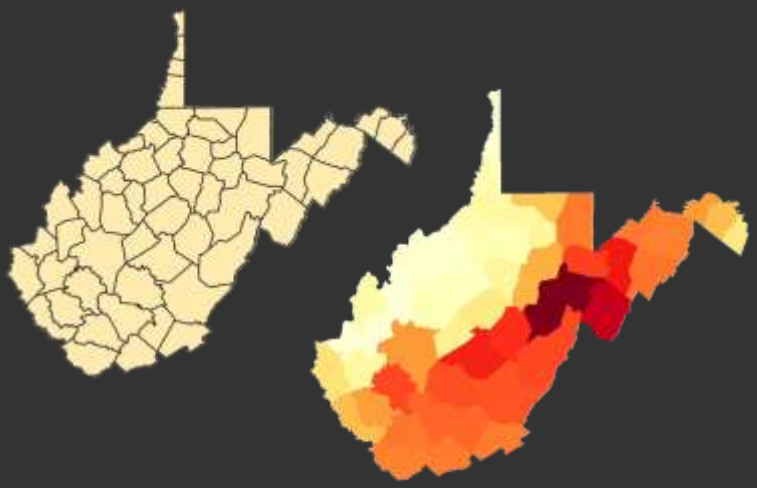
The goal of sensitivity analysis is to determine how sensitive a model is to inputs and parameters, such as scores, weights, input criteria, and data sources.

A model is considered to be sensitive to a specific input or parameter if changing it has a large impact on the output.

For example, if changing the weights only slightly produces a very different output, then the model is sensitive to the weights.

Summarize Model: Zonal Statistics

- Used to obtain statistics for a continuous raster grid within an area or zone as defined by another raster or a polygon vector layer
- Can obtain results as a raster surface for a single statistic



Elevation range by county

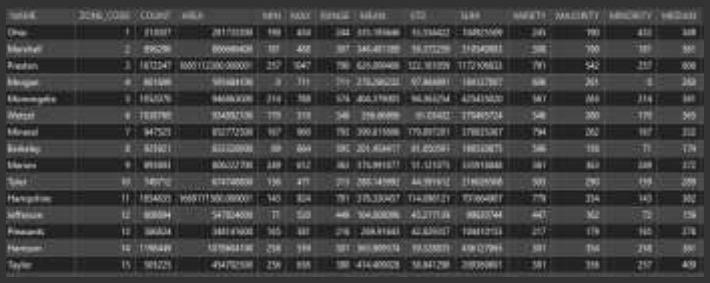
45

You may also need to summarize or generalize your data relative to geographic extents or zones.

For example, you may want to know what county had the highest average score, as calculated from all the pixels in the county. This could be accomplished with zonal statistics.

Summarize Model: Zonal Statistics as Table

- Used to obtain statistics for a continuous raster within an area or zone as defined by another raster or a polygon vector layer
- Can obtain multiple statistics as a table (mean, majority, maximum, median, minimum, minority, range, standard deviation, sum, and variety)
- Can use table join to join the results back to the zone data



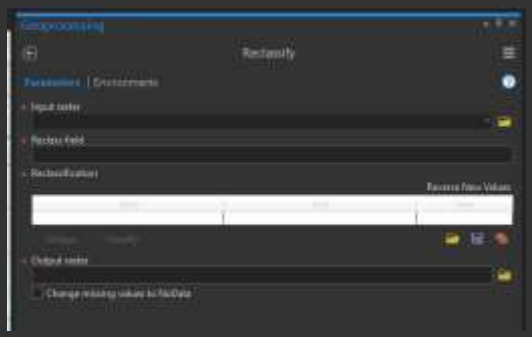
Elevation statistics by county

46

You could also obtain the summary statistics as a table.

Model Generalization

- You can bin your output
  - Not suitable
  - Moderately suitable
  - Highly suitable



47

Instead of using the raw index scores, you may want to rescale them or group them.

It is common to define ranges of index scores to represent different levels of suitability.

This can be accomplished using raster reclassification.

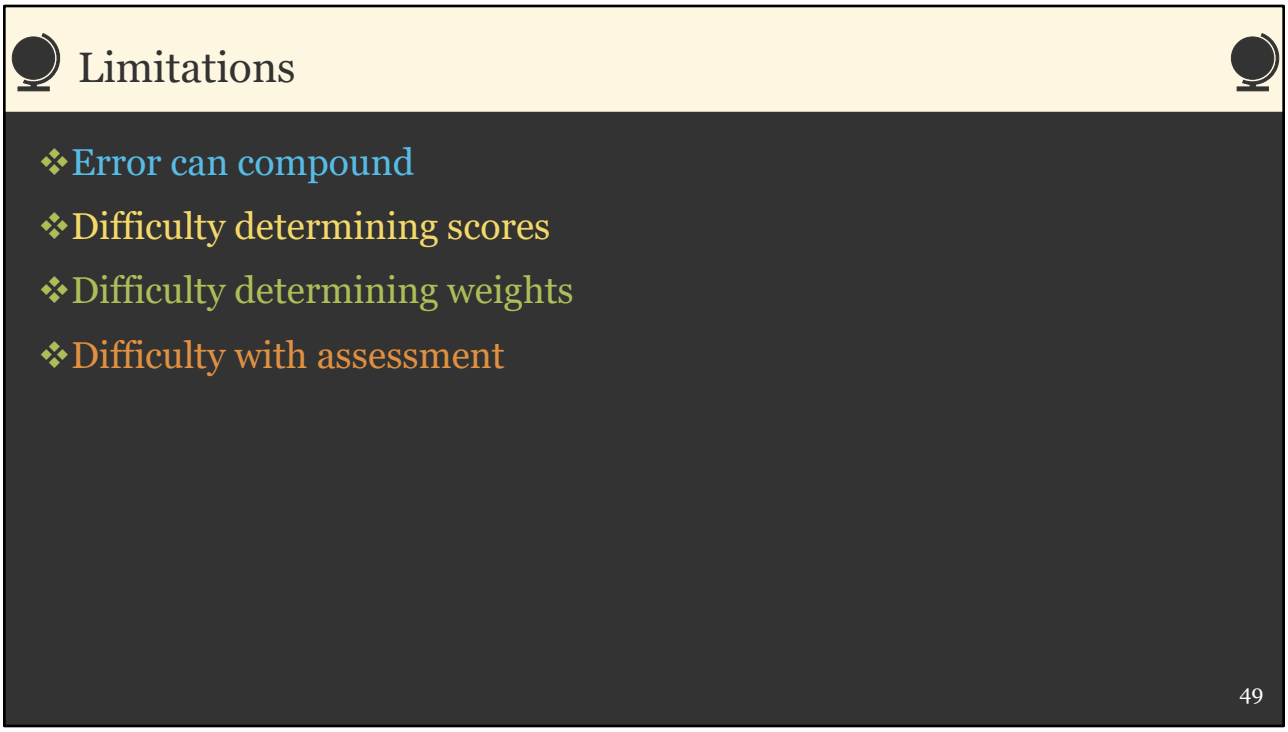
Summarize Model: Tabulate Area

- Used to obtain areas for categories in a categorical raster within an area or zone as defined by another raster or polygon vector layer
- Areas will be in square map units
- Can use table join to join the results back to the zone data

Area of land cover by county

48

If the suitability indexes are grouped or classified, they can be summarized using Tabulate Area.



As with any modeling technique, there are limitations to weighted overlay.

For example, input datasets will have some uncertainty or error, and this uncertainty or error can compound in complex and nonlinear ways. However, this is not an issue unique to weighted overlay.

It can be difficult to define and justify weights and scores.

Lastly, these models can be difficult to assess, especially using statistical or quantitative methods.

Although there are limitations, weighted overlay is still a very powerful and useful technique that has a wide variety of applications.

Other Techniques

Technique	How it works	Strengths	Limitations	Complexity	Typical uses
Boolean overlay (Conservative Models)	Applies binary include/exclude criteria via AND/OR logic on raster or vector layers	Simple; transparent; easy to communicate	No gradation between suitable/unsuitable; ignores trade-offs	Low	Exclusion zoning, conservation buffers
Weighted Overlay (Liberal Models)	Standardizes criteria to a common scale, assigns analyst-defined weights, sums weighted layers	Flexible; intuitive; allows full trade-off between criteria	Subjective weights; full compensation can mask poor scores on key factors	Low–Med	Land-use planning, facility siting
Fuzzy membership/logic	Maps criterion values to [0,1] suitability via membership functions (e.g., sigmoidal, linear); combines with fuzzy operators (AND, OR, algebraic product)	Handles vagueness and continuous gradients; reduces crisp cut-off artifacts	Choice of membership function shape is subjective; results can be hard to interpret	Medium	Habitat modeling, agricultural zoning
Analytic Hierarchy Process (AHP)	Derives criteria weights through pairwise comparisons; checks consistency ratio; weights fed into WLC	Structured, defensible weight derivation; widely accepted	Assumes transitivity; consistency ratio can fail with many criteria; still fully compensatory	Medium	Site selection for infrastructure, disaster risk mapping
Spatial Compromise Programming (SCP)	Identifies locations closest to an ideal solution and farthest from a nadir; balances group utility vs. individual regret using a parameter	Explicit ideal-point reference; handles conflicting criteria; v controls majority/minority rule	Requires defining ideal/nadir; v choice is subjective; computational overhead	High	Multi-stakeholder site selection, climate adaptation planning
MaxEnt	Estimates habitat suitability by maximizing entropy of predicted distribution relative to background, using presence-only data and environmental predictors	Works with presence-only data; strong ecological grounding; handles interactions	Requires careful background selection; overfitting risk; not purely MCDM	High	Biodiversity suitability, invasive species risk

50

This is just an introduction to suitability modeling methods. The table on this slide describes the conservative and liberal methods discussed here along with some more advanced techniques. These advance techniques are often covered in more advanced spatial modeling, analysis, or statistics courses.

Physical Models

Modeling Physical Processes

- A model of a physical process
- Example: Revised Universal Soil Loss Equation (RUSLE)

A = RKLSCP

A = Average soil loss

R = Rainfall-runoff erosivity factor

K = Soil erodibility factor

L = Slope length factor

S = Slope steepness factor

C = Crop management factor

P = Support practice factor

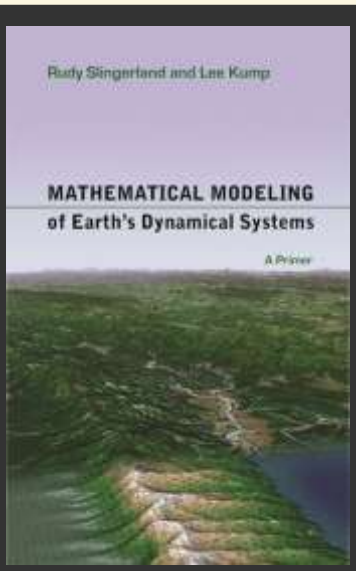
https://www.ars.usda.gov/southeast-area/oxford-ms/national-sedimentation-laboratory/watershed-physical-processes-research/docs/revised-universal-soil-loss-equation-rusle-welcome-to-rusle-1-and-rusle-2/

52

It is also possible to model a physical process using our understanding of the processes and the factors that influence it.

For example, the Revised Universal Soil Loss Equation attempts to predict soil loss based on factors that impact it.

Book Recommendation



- *Mathematical Modeling of Earth's Dynamical Systems: A Primer*
- Rudy Slingerland and Lee Kump

https://press.princeton.edu/books/paperback/9780691145143/mathematical-modeling-of-earths-dynamical-systems

53

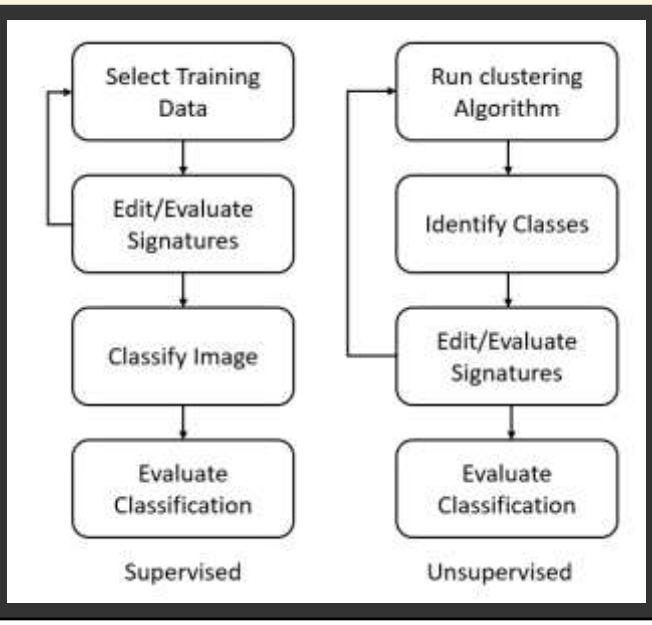
Physical or numerical modeling attempts to implement physical equations that describe landscape, climatological, geochemical, and/or geochemical processes. An example would be erosion and deposition to understand how terrain surfaces change over time.

The inputs to the equations can be represented using geospatial data layers, such as digital elevation models or precipitation raster grids, and the equation can be solved over the landscape surface.

The text referenced on this slide is a good text focused on numeric modeling in the Earth sciences.

Clustering: Unsupervised Models

Unsupervised vs. Supervised Learning



55

In unsupervised classification, the computer uses an algorithm to cluster the data into a set of categories. The analyst must then label and group the clusters into meaningful classes. Cluster is a form of unsupervised learning. Clustering algorithms attempt to find natural clusters in data. Specifically, features that have similar combinations of predictor variable values will be labeled to the same group. For image data, pixels can be placed into a finite set of groups based on the image channel values. It is then up to the analyst to determine what each group represents.

In supervised classification, the analyst must provide examples as training points or polygons. The computer then uses these examples and predictor variables to classify the entire landscape into the provided categories. This is a form of empirical modeling.

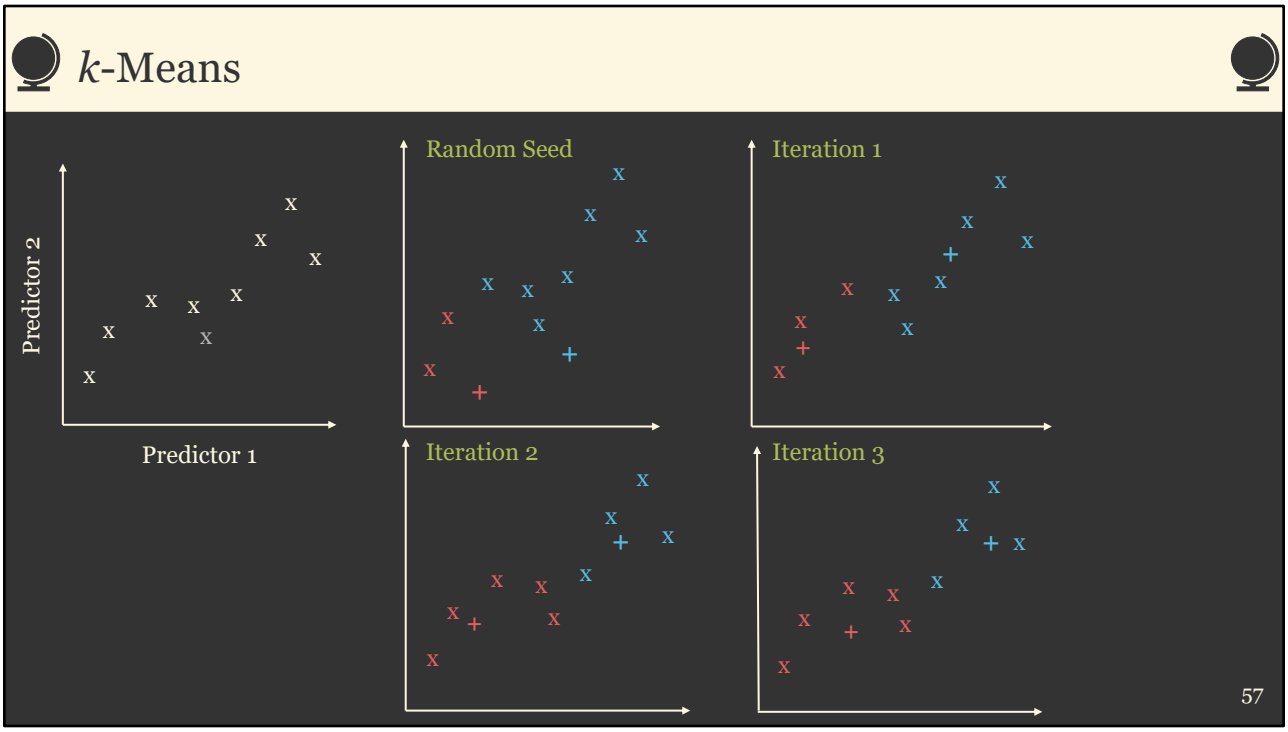
Both methods require user input. The difference is when user input is provided. In supervised classification, the analyst provides input as training samples up front. In contrast, the user will label the clusters at the end of the unsupervised classification process.

k-Means

- Analyst defines the number of clusters
- Algorithm arbitrarily "seeds" or locates the cluster centers in the multidimensional measurement space
- Pixels are assigned to clusters
- Mean vectors for the clusters are determined
- Process is re-ran
- Process continues until mean vectors don't change significantly

56

One commonly used algorithm for data clustering and unsupervised classification is *k*-means. To implement this clustering method, the analyst must define the number of desired classes to differentiate. Generally, more classes should be defined then the number of desired informational classes since the user will group generated classes into a smaller set of meaningful classes. The algorithm then randomly seeds or locates cluster centers in the multidimensional measurement space. Using image data as an example, pixels are then assigned to these cluster centers based on minimal distance. This process continues through multiple iterations to move the cluster centers, or mean vectors, and reduce the overall distance between pixels and these centers. The iteration process will stop once the centers no longer change significantly. Once the final mean vectors are obtained, pixels are assigned to clusters based on which mean vector they are closest to in the multidimensional space.



This slide graphically explains the *k*-means clustering process. To begin, the graph on the left shows the location of data points in a two-dimensional feature space defined by two predictor variables. To include more predictor variables, you can map to more dimensions. For example, with three predictors, you would map a location relative to three values to a location within the volume of a cube. It is difficult to visualize more than three spatial dimensions. However, locations can be mathematically described as a multidimensional vector. This allows for clustering to be performed in many dimensions. We will stick with two dimensions here for simplicity.

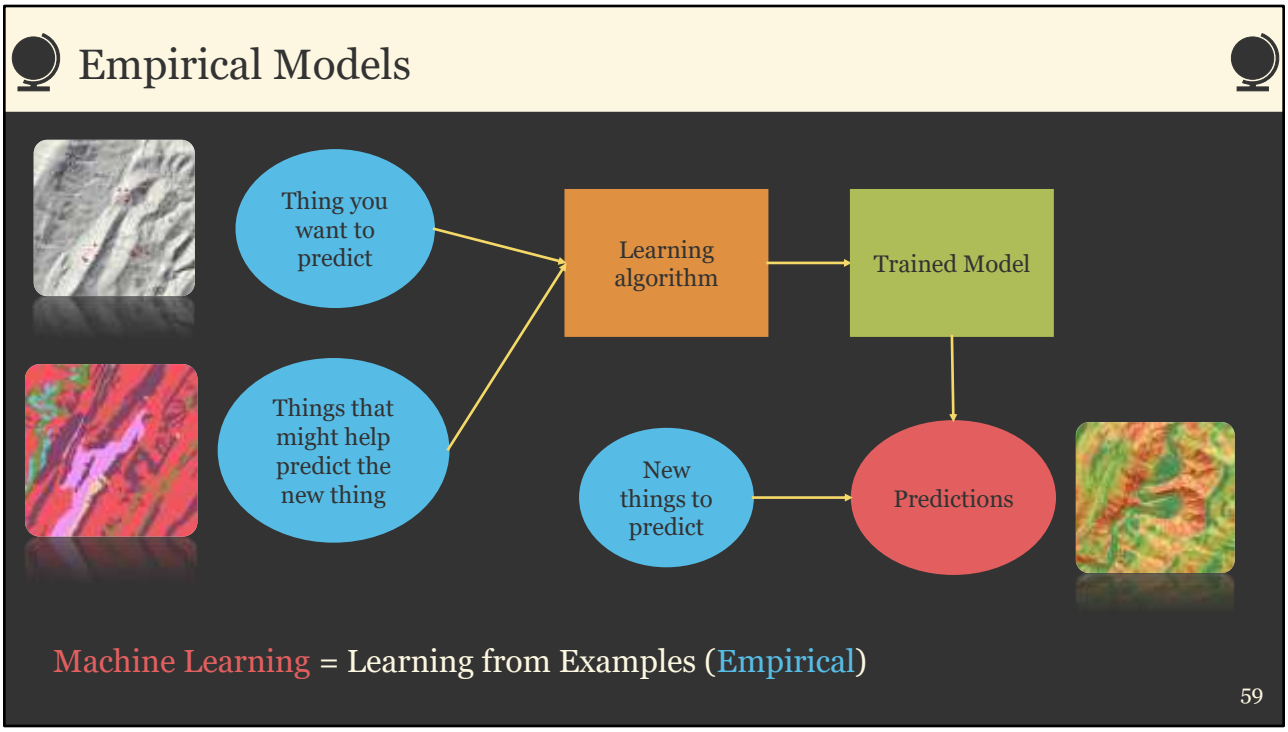
First, cluster seeds are randomly placed in the feature space. Samples are then assigned to these seeds based on proximity. In this example, the red Xs represent data points that have been assigned to the red plus sign mean vector while all the blue Xs have been assigned to the blue plus sign mean vector. In order to assess how well the mean vectors are positioned, the total distance of all the data points to their assigned mean vector are summed.

In subsequent iterations, the mean vectors are moved such that the distance is minimized between each mean vector and all points that were assigned to it in the prior iteration. The data points are then reassigned to cluster centers based on proximity. This process continues until a maximum number of iterations is reached or the mean vectors no longer need to be moved or data points need to

be reassigned to lower the sum of the distances of each data point to its associated mean vector.

There are augmentations of *k*-means in which the number of clusters or mean vectors can change as the algorithm iterates over the data. Based on a minimum number of data points that can be mapped to a cluster, a maximum within-class variance to split clusters, and a minimum pairwise distance to merge clusters, mean vectors are updated, and clusters are merged and spilt.

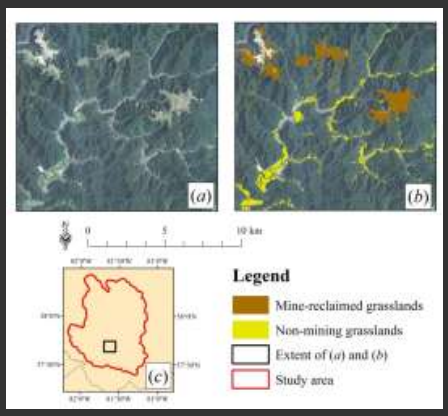
Empirical: Supervised Models

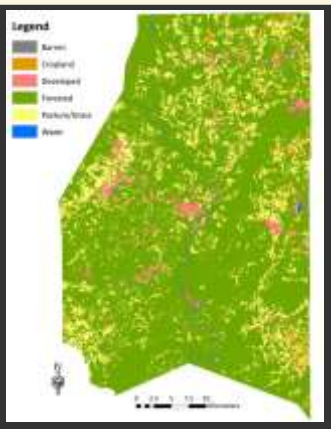


Supervised learning means learning from labeled data. The goal is to provide the computer with examples and predictor variables. An algorithm or method will then be used to model the phenomenon using the examples and predictor variables. This model can then be used to predict a geographic area or an entire population. The model can be assessed using withheld validation samples.

Predicting a Categorical or Binary Outcome

- Logistic regression
- Maximum likelihood
- Parallelepiped
- Minimum-Distance-to-Means
- Machine learning



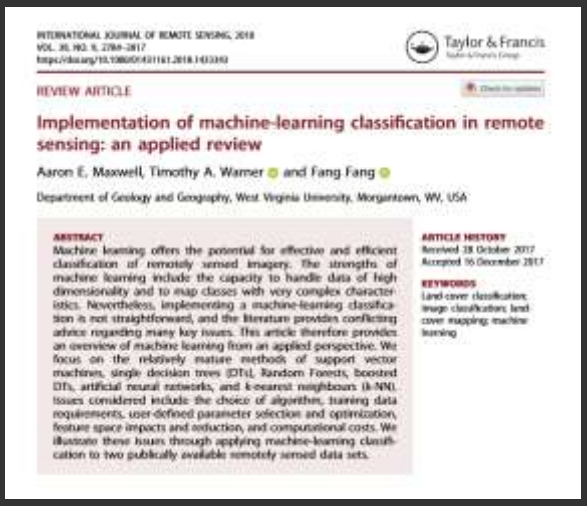


60

Choosing a supervised learning method to use partially depends on the type of data being predicted. For example, certain methods, as listed here, are appropriate for categorical data.

Examples of categorical predications include land cover mapping, forest type mapping, wetland mapping, vegetation mapping, and soil type mapping. Here we are predicting nominal data.

Example: Land Cover Classification



INTERNATIONAL JOURNAL OF REMOTE SENSING, 2018
VOL. 39, NO. 9, 2784–2817
https://doi.org/10.1080/01431161.2018.1433343

Taylor & Francis

REVIEW ARTICLE

Implementation of machine-learning classification in remote sensing: an applied review

Aaron E. Maxwell, Timothy A. Warner and Fang Fang

Department of Geology and Geography, West Virginia University, Morgantown, WV, USA

ABSTRACT

Machine learning offers the potential for effective and efficient classification of remotely sensed imagery. The strengths of machine learning include the capacity to handle data of high dimensionality and to map classes with very complex characteristics. Nevertheless, implementing a machine-learning classification is not straightforward, and the literature provides conflicting advice regarding many key issues. This article therefore provides an overview of machine learning from an applied perspective. We focus on the relatively mature methods of support vector machines, single decision trees (DTs), Random Forests, boosted DTs, artificial neural networks, and k-nearest neighbours (k-NN). Issues considered include the choice of algorithm, training data requirements, user-defined parameter selection and optimization, feature space impacts and reduction, and computational costs. We illustrate these issues through applying machine-learning classification to two publically available remotely sensed data sets.

ARTICLE HISTORY
Received 28 October 2017
Accepted 16 December 2017

KEYWORDS
Land-cover classification; image classification; land cover mapping; machine learning

https://www.tandfonline.com/doi/full/10.1080/01431161.2018.1433343

61

Machine learning methods have been shown to be very applicable to classification problems, such as land cover mapping. Here is a link to an open-source review article on the use of machine learning in land cover classification.

Predicting a Numeric or Continuous Outcome

- Linear regression
- Multiple linear regression
- Polynomial regression
- Machine learning

Percent Canopy Cover NLCD 2011

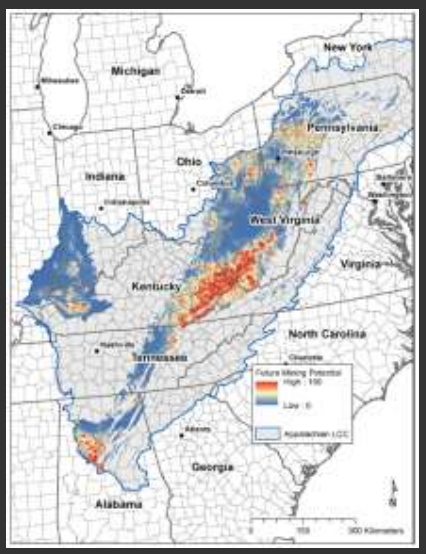
https://www.mrlc.gov/index.php

62

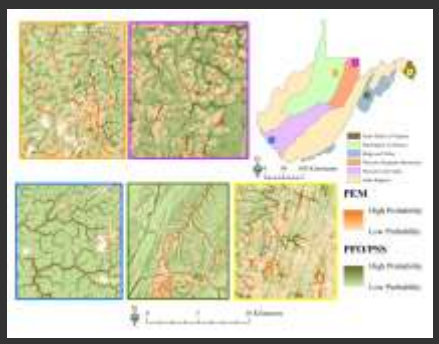
It is also possible to predict continuous variables, such as interval or ratio data. As an example, as part of the National Land Cover Database, percent canopy cover is estimated using machine learning methods.

A wide variety of methods are available to predict continuous variables, such as linear regression and machine learning.

Predicting a Probabilistic Outcome



- Logistic regression
- Maxent
- Machine learning



63

You can also predict probabilities or likelihood. This is often confused with categorical predictions.

Here, the goal is not to predict categories but to predict the likelihood that a location is in a category or could be in that category in the future.

As examples, the maps shown here represent the likelihood of palustrine wetland occurrence and the likelihood of surface mine expansion.

Predicting Count Outcomes

- Poisson Regression
- Machine Learning

64

You can also predict counts. For example, an epidemiologist might be interested in predicting the number of cases of a disease that is likely to occur within a geographic extent.

Developing Training Data

- Spatially explicit
- Representative of the population
- Adequate number of samples
- Adequate number of samples per category
- Accurate

65

Supervised learning requires training samples. These samples will represent examples of the classes or values you are trying to predict.

The algorithm will then use these examples and the predictor variables at these locations to make a model that can be used to predict new locations.

I have found that the quality of the training samples has a large impact on the quality of the prediction.

When creating training samples, they need to be spatially explicit, or you need to know where they exist in the map space.

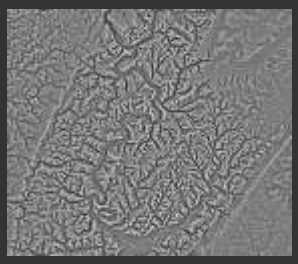
They also must be representative of the population. For example, if you are trying to predict where a certain tree might grow and it is known to grow on both ridges and floodplains, then you need to give the algorithms examples of known locations where the tree has been found to grow in both floodplains and along ridges.

You need to provide an adequate number of samples and an adequate number of samples per class. The number of samples required will vary based on the complexity of the problem and the methods used. I generally try to provide as many quality training samples as possible given limitations and constraints.

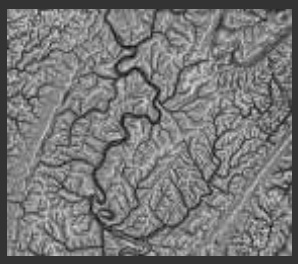
Collecting a large number of training samples can be difficult due to time, cost, and access constraints.

Lastly, the data should be accurate. If the algorithm is given mislabeled or poor examples, it will have difficulty creating a useful model. Garbage in, garbage out.

Developing Predictor Variables



Slope Position



Dissection

Roughness

- Spatially explicit
- Accurately georeferenced
- Accurate/precise
- Timely
- Adequate spatial resolution
- Available

66

You will also need predictor variables that the algorithm can use to predict the features or values of interest.

These will be variables that are thought or known to be correlated with the phenomenon being mapped.

Predictor variables should be spatially explicit and accurately georeferenced. Again, your goal is generally to make a prediction across an entire geographic extent. So, you will need this information across the entire extent.

The data should be accurate and precise. Again, garbage in, garbage out.

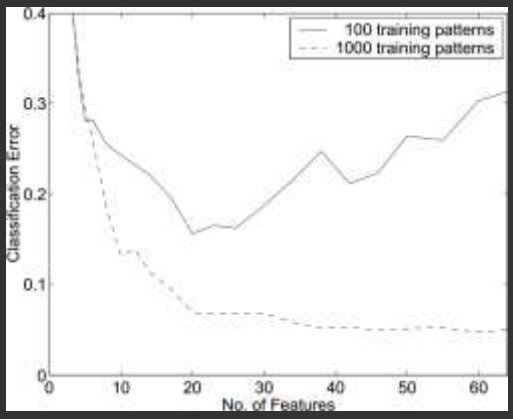
The timeliness of the data may be of concern, especially for variables that can change rapidly.

The data should be of adequate spatial resolution to obtain the desired result and in comparison to the other variables used. You may have to exclude a variable because it is too coarse.

Lastly, the data must be available. You may find that a desired input is simply not mapped, not completely mapped, or not mapped with adequate quality. If you would prefer not to exclude it, then you may need to generate it on your

own, which could be time consuming, expensive, or impractical. There are always some data and modeling limitations.

Developing Predictor Variables



Jain, Anil K., Robert P. W. Duin, and Jianchang Mao. "Statistical pattern recognition: A review." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.1 (2000): 4-37.

- Hughes Phenomenon
- "Curse of dimensionality"

Hughes, G.F. 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory* 14 (1): 55–63. doi:10.1109/TIT.1968.1054102.

Modified by a slide created by Tim Warner

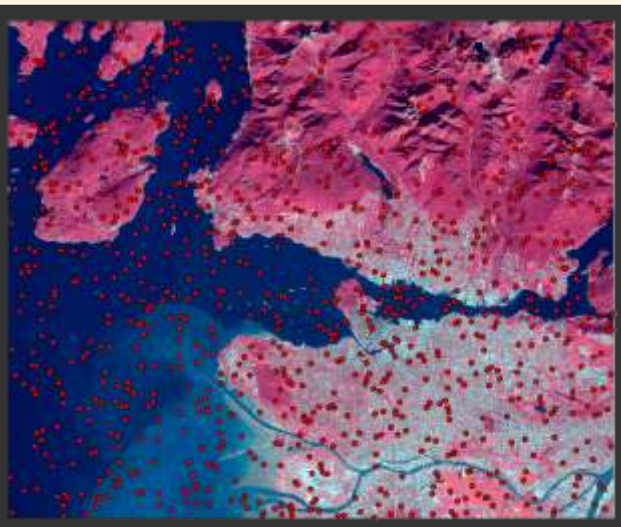
67

It would make sense that including as much information or as many predictor variables as possible would be preferred. For example, if you are buying a car, more information will make you a more informed consumer. However, this has generally been found to not be the case in predictive modeling. This is known as the Hughes Phenomenon or Curse of Dimensionality. This suggests that adding more predictor variables can actually decrease the performance of the model. This is because even though more information is being provided, the complexity of the problem increases.

This problem is generally more pronounced when a small training set is used, as highlighted in the provided figure from Jain et al. This is because more samples will need to be provided to deal with the complex dimensionality of the problem.

Fortunately, some algorithms are fairly robust to this problem. Also, methods are available to reduce the number of variables or select the most useful variables.

Developing Validation Data



- Unbiased
- Randomized
- Non-overlapping with your training data
- Correctly proportioned
- Accurate

68

Validation data must also be provided to assess the model performance and map output.

In order to produce an unbiased estimate of the performance, the validation data must be randomized in some way.

Also, the training and validation data should not overlap, or the same samples should not be included for both training and validation. This is because algorithms tend to do a better job at predicting the training samples as opposed to new locations, a phenomenon known as overfitting. So, including training samples as validation samples could inflate the reported performance.

It is also recommended that training samples be correctly proportioned relative to the map. For example, if you are mapping land cover and 70% of the mapped area is forest, then 70% of the validation samples should also be forest.

Lastly, validation data should be accurate. The goal here is to compare the map product to reference data of higher quality. It is generally assumed that no data are perfect. So, even the validation data will have some error. We try to avoid using the terms “ground truth” or “ground truthing” for this reason.

We do not discuss assessment metrics and methods in this course. This is covered in our *Methods in Open Science*, *Remote Sensing*, *Geospatial Supervised Learning*, and *Geospatial Deep Learning* courses.

Machine Learning

- Supervised learning
- Learn from training samples
- Generalize to unknown samples
- Generally robust
- Can model complex class signatures
- Can use a variety of predictor variables

69

We will now discuss some machine learning method used for supervised learning.

This slide highlights the key characteristics of machine learning methods. These methods tend to be more robust to complex patterns and class signatures than traditional methods. It is also possible to incorporate a variety of predictor variables to potentially improve the classification, such as digital elevation data or ancillary variables.

In the following slides, I provide a conceptualization of three commonly used machine learning methods in remote sensing. I then generally discuss artificial neural networks (ANNs) and deep learning.

Machine Learning

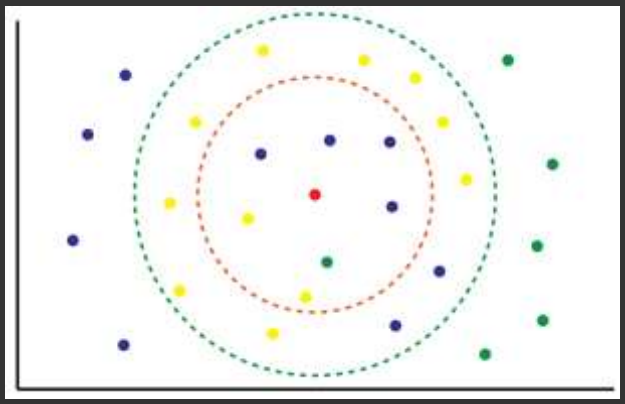
- Artificial Neural Networks (ANN)
- *k*-Nearest Neighbor (*k*NN)
- Classification and Regression Trees (CART)
- Boosted CART
- Random Forests (RF)
- Support Vector Machines (SVM)

70

This slide lists some commonly used machine learning methods. Here, I will provide a brief introduction to *k*-Nearest Neighbor (*k*-NN), random forests (RF), and support vector machines (SVM). I then discuss artificial neural networks (ANNs) and deep learning (DL), which expands upon ANNs.

k-Nearest Neighbor (*k*-NN)

- Compare new sample to nearest training sample(s) in the feature space
- *k* = number of neighbors compared to
- Assign new sample to majority of the neighbors (classification)
- Use average or distance weighted average of neighbors (regression)
- Proportion of neighbors by class (probability)



71

k-NN is conceptually pretty simple. The basic idea here is that a new observation is assigned to the class of the nearest observations in the feature space. In the provided image, the feature space is two dimensional (e.g., two image bands or predictor variables), but distance measures can be extended into multiple dimensions mathematically. The number of neighbors to consider is based on the *k* parameter.

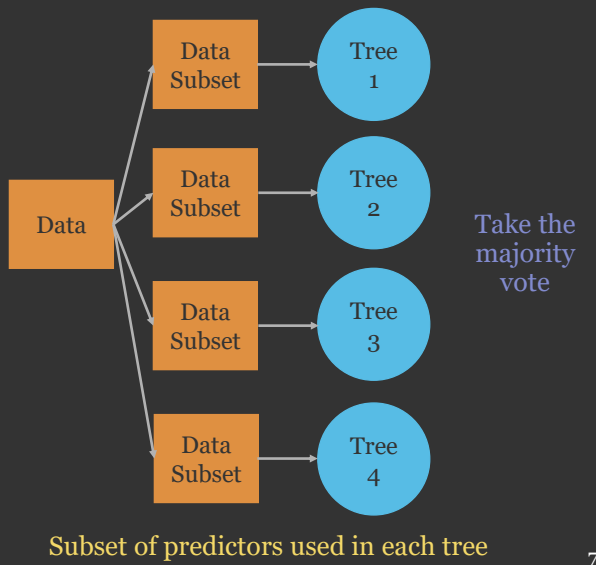
In the example, the red dot represents a new sample or pixel to be classified. The red ring represents 7 nearest neighbors. Of the 7, 4 are blue, 2 are yellow, and 1 is green. So, the new sample would get assigned to the blue class.

The green ring represents 18 nearest neighbors. From these neighbors, the yellow class is most abundant, so the new sample would get mapped to that class.

k-NN can be used to predict classes, as described here. It can also be used to predict continuous variables (regression) and class probabilities.

Random Forests (RF)

- Uses decision trees
- Uses the Gini Index of Impurity
- Ensemble decision tree method
- Uses random subset of predictor variables for splitting at each node
- Uses random subset of training data in each tree
- Attempts to reduce correlation between trees
- Ensemble of weak classifiers



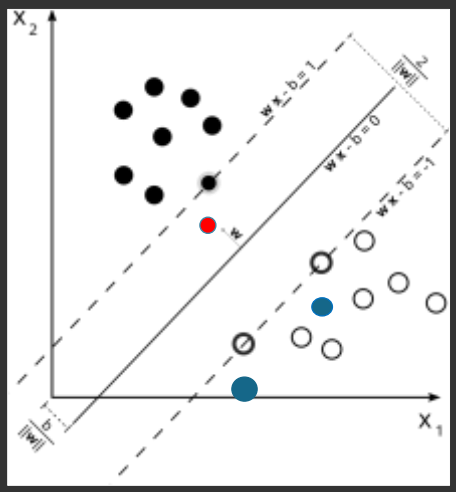
72

Random forest is based on decision trees. Decision trees use a series of binary decision rules to label samples. The decision rules are selected to homogenize the classes through recursive partitioning and are learned from the training examples.

Random forests expands this framework to include multiple decision trees that each use a subset of the training samples and input predictor variables. The goal here is to reduce the correlation between the trees and create a strong ensemble of weak classifiers. It is common to generate hundreds of trees. These trees are combined as an ensemble, and a final classification is determined based on majority voting.

Support Vector Machines (SVM)

- Works with boundary conditions
- Find the linear boundary that gives the best separation between classes (hyperplane)
- Boundary defined by support vectors, or training samples that are nearest to the boundary
- Project the data to a higher dimension
- Two class problem
  - Combine multiple classifications to separate more than three classes
- Non-separable data
  - Positive slack variable (Cortes and Vapnik (1995))
  - Cost parameter (*C*)



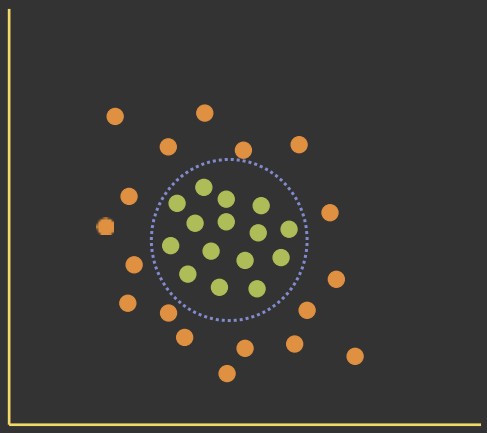
http://en.wikipedia.org/wiki/Support_vector_machine

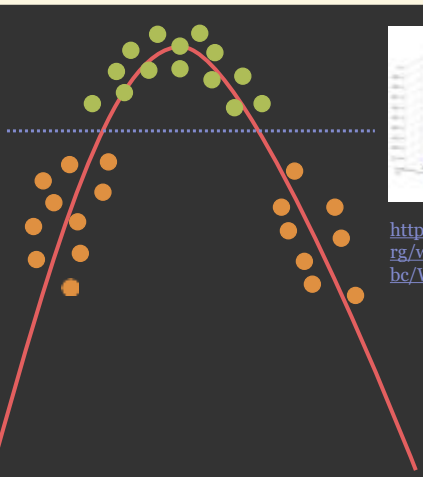
73

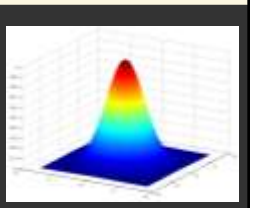
Support vector machines is a kernel-based method. The goal is to find the boundary or hyperplane that separates classes with the maximum margin. In other words, the boundary should maximize the distance between classes as defined by the training samples. Only training samples near the boundary are used to define the hyperplane, and are termed support vectors.

One issue is that classes may not be linearly separable in the feature space; a line cannot be drawn in the space to separate the classes. In order to improve separability, a kernel is used to transform or map the samples to a higher dimensional space where classes are more separable.

The Kernel Trick







https://upload.wikimedia.org/wikipedia/commons/b/bc/Wiki_gauss.png

- Use the kernel trick to transform the feature space into a higher dimensional space where the features are more linearly separable

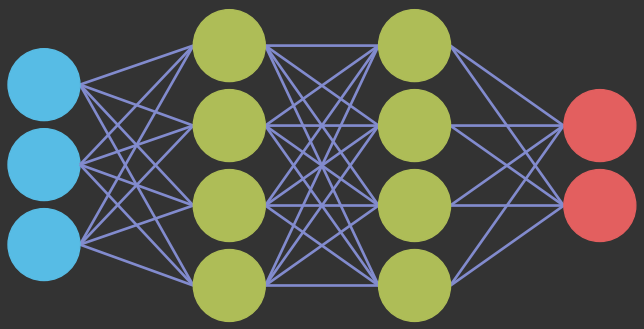
74

This slide further conceptualizes the kernel trick. Again, the goal is to project the input predictor variable space into a higher dimension in which a nonlinear boundary may become more linear and defined as a hyperplane.

As an example, imagine that the separating boundary between two classes in a two-dimensional space would best be defined as a circle, as pictured in the image on the left. The samples could be projected onto the surface of a three-dimensional curve where the separating boundary could be described as a plane. Thus, a nonlinear boundary was made more linear by projecting the data into a higher dimensional feature space.

ANN Structure

- Consist of interconnected neurons or nodes
- Input layer represents predictor variables
- Output layer represents what is being predicted
- Model generated by learning a weight associated with each connection
- Referred to as fully connect architectures or densely connected architectures



75

ANNs are often described as being modeled after the human brain, as they consist of interconnect neurons or nodes. However, how these learning algorithms function is not necessarily comparable to the human brain, so I try to avoid using this analogy.

An ANN consists of neurons organized into layers as demonstrated on the slide. The Input layer represents input predictor variables, such as image bands, and has one neuron for each input, while the output layer represents the desired output, such as land cover categories, and has one neuron for each output class. For binary classification, one or two output neurons may be used. For regression, there will be one output neuron.

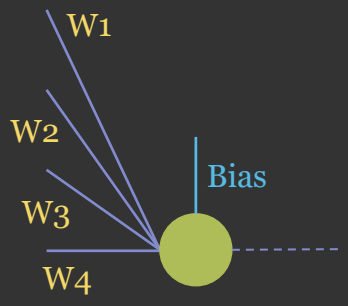
Between the input and output layers are one or more hidden layers containing multiple nodes. Within the network, all neurons in a layer are connected to the neurons or nodes in adjacent layers, and these connections have weights.

On the slide, the conceptualization contains three inputs (e.g., three image bands or predictor variables) and two hidden layers containing 4 nodes each. There are two classes being classified or predicted. Note that all nodes or neurons between adjacent layers are connected. These connections represent the weights.

Through the process of supervised learning, the weights in the network are adjusted in an attempt to improve the prediction of the output.

Weights

- Weights are adjusted during training process
- Similar to slope or coefficient in regression
- Inputs from prior layer multiplied by weights
- Multiple Input X weight combinations are summed
- Weights relate to strength of activation or signal



$$\text{Activation} = b + \sum_{i=1}^{n} x_i w_i$$

76

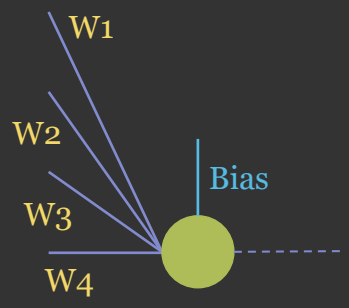
This slide further describes the concept of weights and how a signal or activation is produced at each node.

This is actually very similar to multiple linear regression. The weights are comparable to the coefficients for each predictor variable while the bias (b) is similar to the y-intercept.

So, at this point, the activation or output of the node is a linear combination of the inputs to the node, associated weights, and bias.

Bias

- A constant to shift the activation
- Similar to a y-intercept in linear regression



$$\text{Activation} = b + \sum_{i=1}^{n} x_i w_i$$

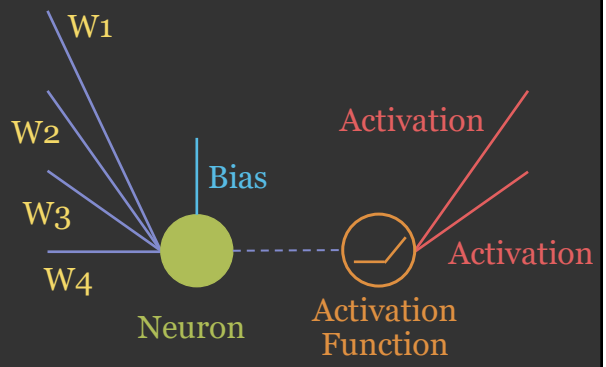
77

Again, the bias at each node is similar to the y-intercept in a linear regression equation. It is not multiplied by any input.

The bias allows the activation to be shifted to larger or smaller values, like an offset. The addition of a bias term further increases the flexibility of the model.

Activation Functions

- Allow for modeling of nonlinear relationships
- A mathematical gate between inputs and outputs



$$\text{Activation} = f\left(b + \sum_{i=1}^{n} x_i w_i\right)$$

78

I left out an important component above in the discussion associated with weights, bias, and activation at each neuron or node. As I mentioned, this is similar to a multiple regression equation where the bias acts as a y-intercept, the weights act as coefficients, and the inputs act as the predictor variable values. In reality, this is more than similar to a multiple regression equation: it is a multiple regression equation.

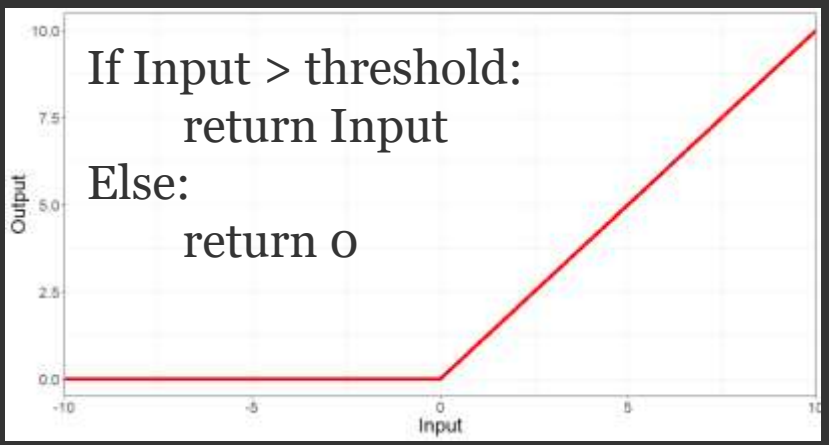
In order to be able to model complex patterns and relationships between variables and make predictions from a complex and noisy signal, it would be useful to be able to model or describe more complex patterns in the data.

This is accomplished by (1) having many interconnected nodes, associated weights, and hidden layers and (2) applying an activation function to transform the activation or signal (the output from the neuron).

Note that just having multiple neurons and multiple hidden layers is not enough to model non-linear relationships. A series of linear relationships is still linear. Thus, activation functions are required to model non-linear patterns and relationships.

Rectified Linear Unit (ReLU)

- Nonlinear
- More computationally efficient than sigmoid or tanh
- Widely used
- Not zero-centered
- Not all neurons updated during backpropagation due to 0 gradient (dying ReLU)

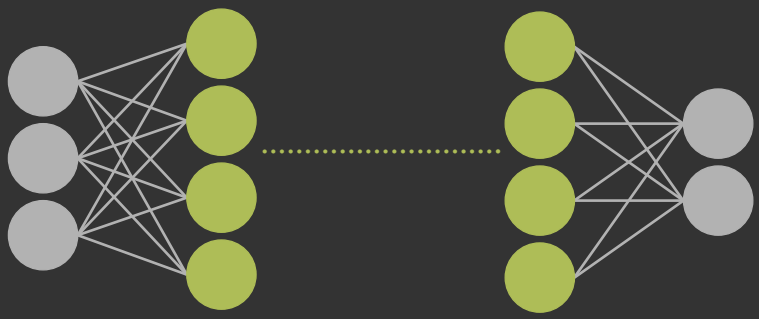


79

The rectified linear unit (ReLU) activation function is currently commonly used. This allows for an activation of 0 below a certain threshold and a linear activation above a certain threshold. Generally, negative values are converted to 0 while positive values are maintained.

ANNs vs. Deep ANNs

- More hidden layers
- Model more complex relationships



80

How is deep learning (DL) different from a traditional ANN? The key difference is the number of hidden layers, or layers between the input layer and output layer.

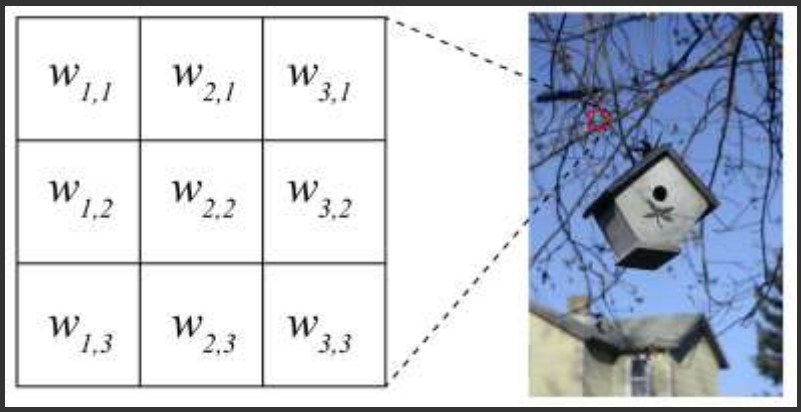
Deep learning algorithms generally have many hidden layers and associated nodes, as opposed to just one or two hidden layers.

So, you can think of "deep" as referring to the depth of the model or number of hidden layers.

Interestingly, it has generally been shown that incorporating more layers is more effective than having fewer layers with a large number of associated neurons. One reason why this might be the case it that a larger number of layers allows for more data abstraction than a small number of layers with a lot of nodes or neurons.

What is Convolution?

- CNNs or ConvNets
- Learn spatial patterns and context
- Based on moving windows
- Apply moving windows to create feature maps



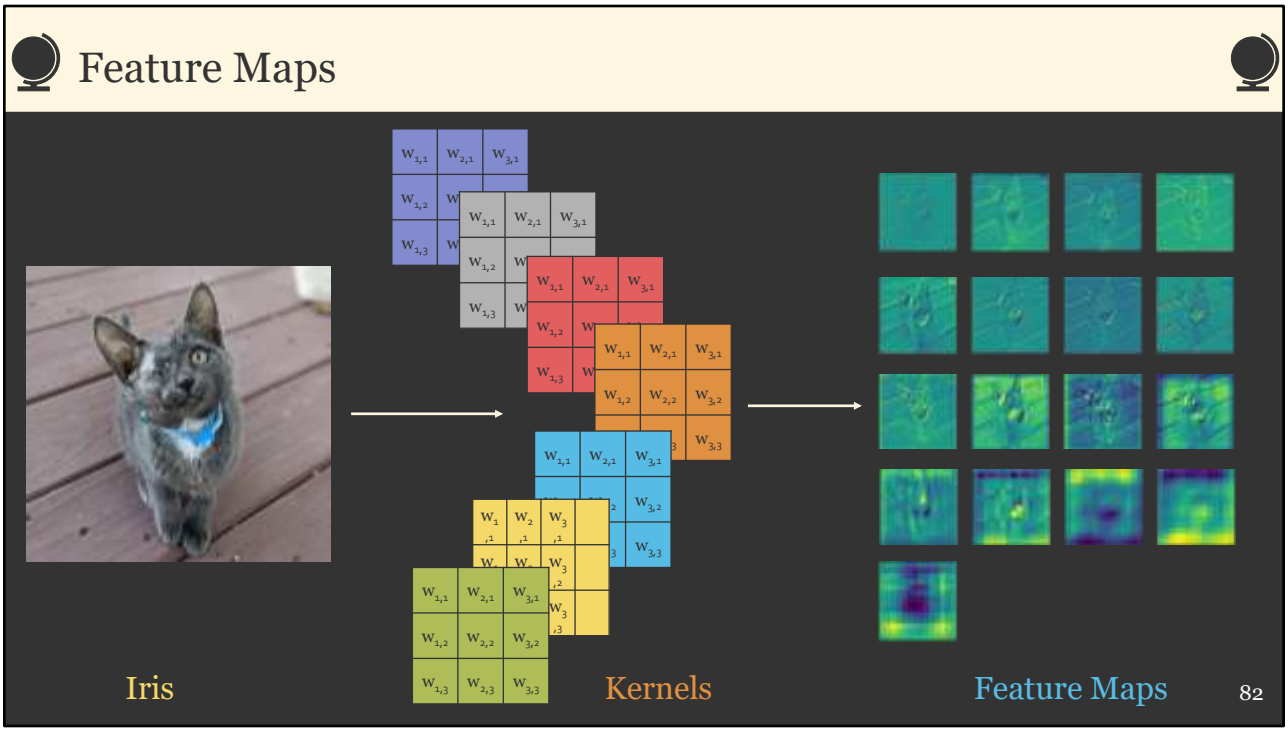
81

Convolutional neural networks, ConvNets, or CNNs are a sub-type of DL ANNs that allow for the incorporation of spatial patterns into the learning process.

Instead of learning weights/parameters associated with links between neurons, as is the case for fully connected layers, convolutional layers learn weights to build a moving window or kernel that is passed over the image to manipulate the data and learn spatial patterns and abstractions.

Once the filter or moving window is learned, it is used to alter the values in the array and produce a feature map.

The basic idea behind CNNs is to learn spatial patterns at different scales.



This slide visualizes some feature maps derived by multiplying the input image or prior feature maps produced by prior layers in the model by the trainable kernel weights and adding a trainable bias term. Again, the goal here is to model useful spatial information that can aid in the classification.

This is the central idea behind CNNs. Learning weights associated with kernels allows for creating spatial abstractions of the data that can be useful for differentiating classes.

UNet

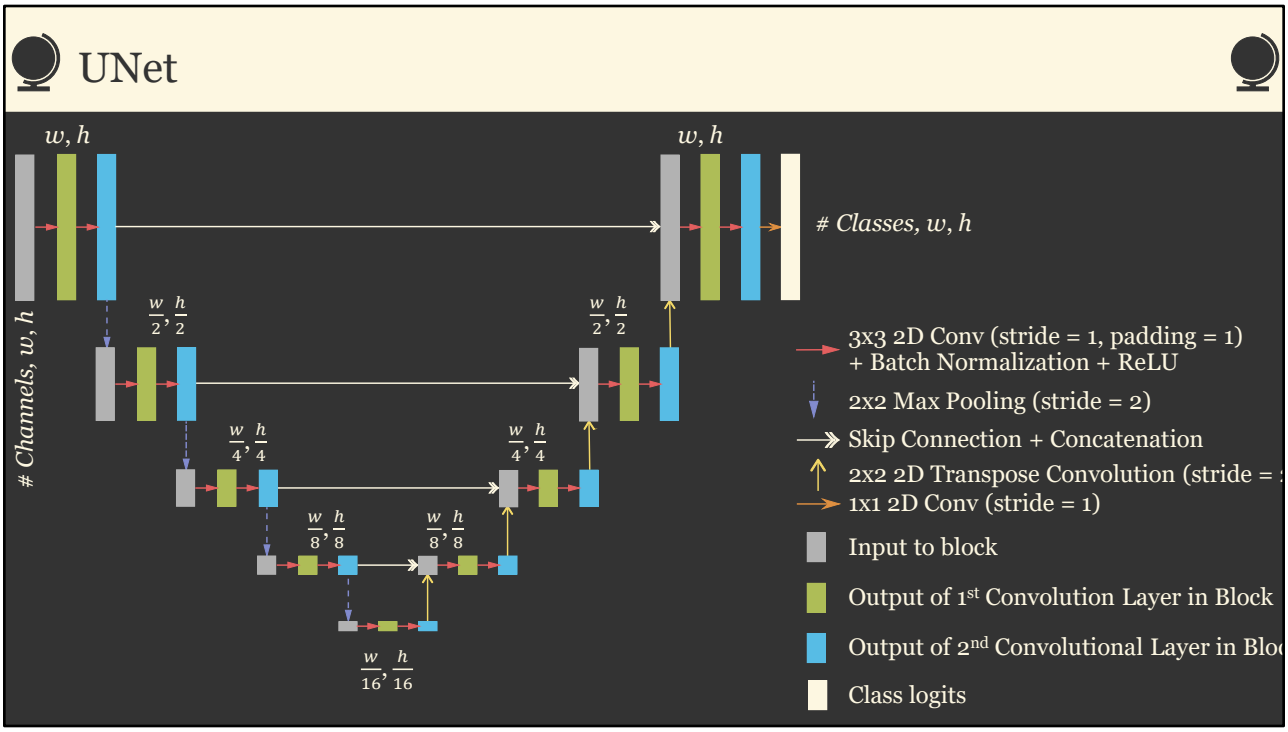

U-Net: Convolutional Networks for Biomedical Image Segmentation

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

https://arxiv.org/abs/1505.04597

The UNet method was first introduced in 2015 and builds upon CNNs to support pixel-level labeling or semantic segmentation. This slide provides a reference to the paper that introduced UNet. Although originally developed for medical image segmentation, it has been shown to have many applications.

There are a variety of CNN architectures for pixel-level labeling. We are introducing UNet here as an example.

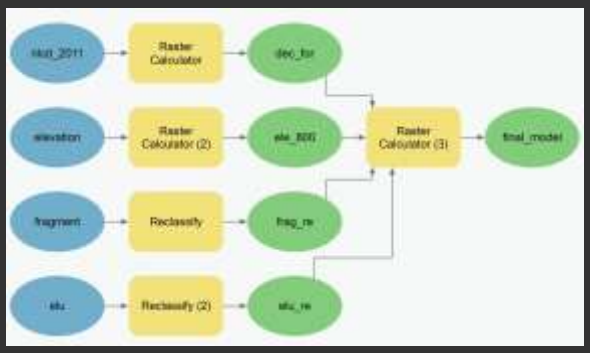


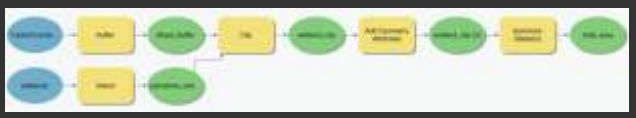
UNet consists of an encoder that generates spatial abstractions of the data at progressively lower resolutions. The bottleneck marks the point at which the feature maps reach their lowest resolution while capturing the greatest concentration of semantic information. The decoder then reconstructs spatial resolution through successive upsampling steps, using convolution operations to refine the learned representations. A classification head generates predictions from the final feature maps. A defining characteristic of UNet is its skip connections, which bridge corresponding encoder and decoder stages to preserve and reuse spatial information that would otherwise be lost during downsampling.
This architecture is designed for pixel-level prediction tasks and can be configured for multiclass classification, binary classification, probabilistic prediction, or regression.

ModelBuilder in ArcGIS Pro

Uses of ModelBuilder

- Allows you to combine analysis tasks and tools into a single process chain
- Graphical interface
- No coding required





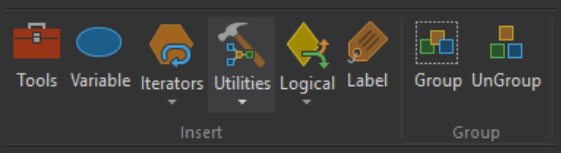
86

ModelBuilder is a valuable tool for performing vector overlay, raster overlay, and spatial analysis in general, especially if the analysis requires many steps.

This tool allows you to combine analysis tasks and tools into a single process chain using a graphical interface. No coding is required.

Components of ModelBuilder

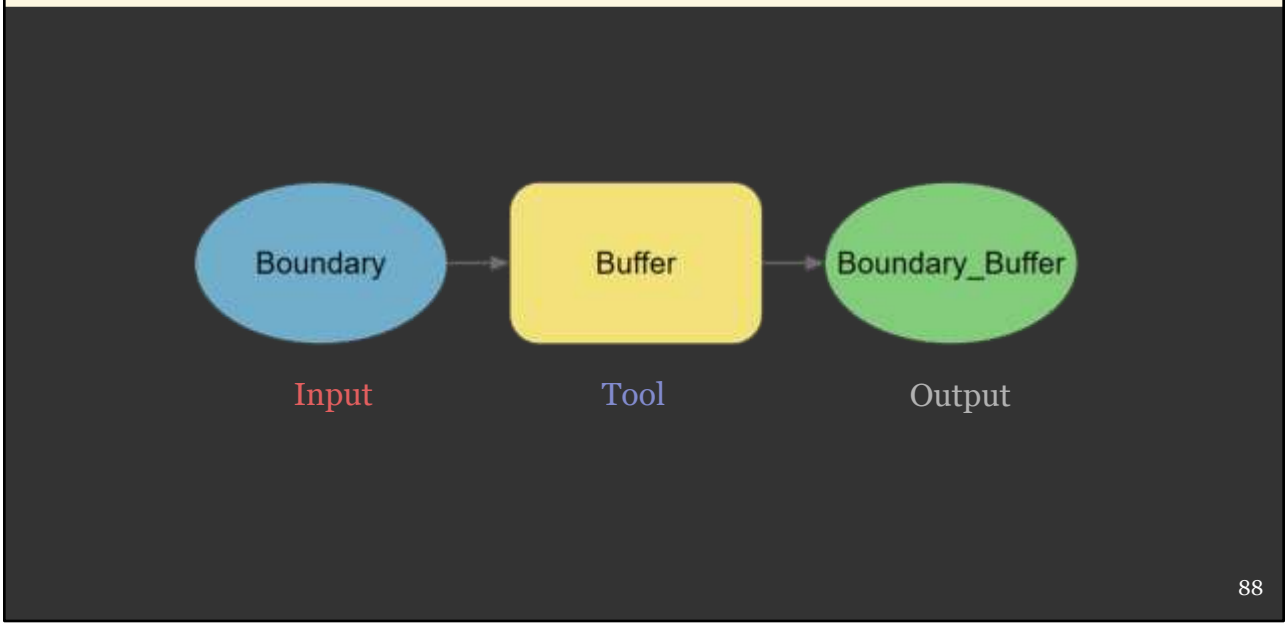
- Input variables and data
- Input parameters
- Tools
- Intermediate outputs
- Final outputs
- Python scripts
- Other models
- Tables
- Iterators
- Etc.



87

Many components can be included in a model. We will discuss some of those components here.

Components of ModelBuilder



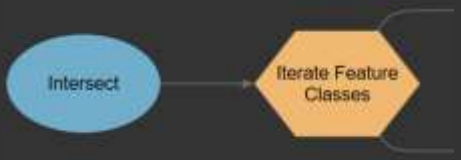
Blue ellipses indicate input data or parameters.

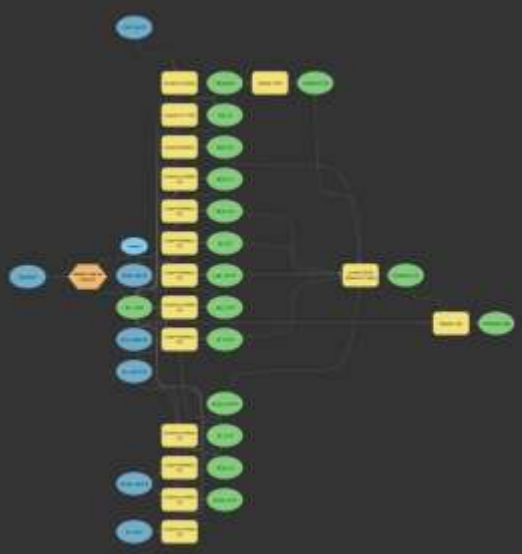
Yellow rounded rectangles show processes.

Green ellipses indicate intermediate and final outputs.

Iterators

- For, While, Iterate Feature Selection, Iterate Row Selection, Iterate Field Values, Iterate Multivalues, Iterate Datasets, Iterate Feature Classes, Iterate Files, Iterate Rasters, Iterate Tables, Iterate Workspaces





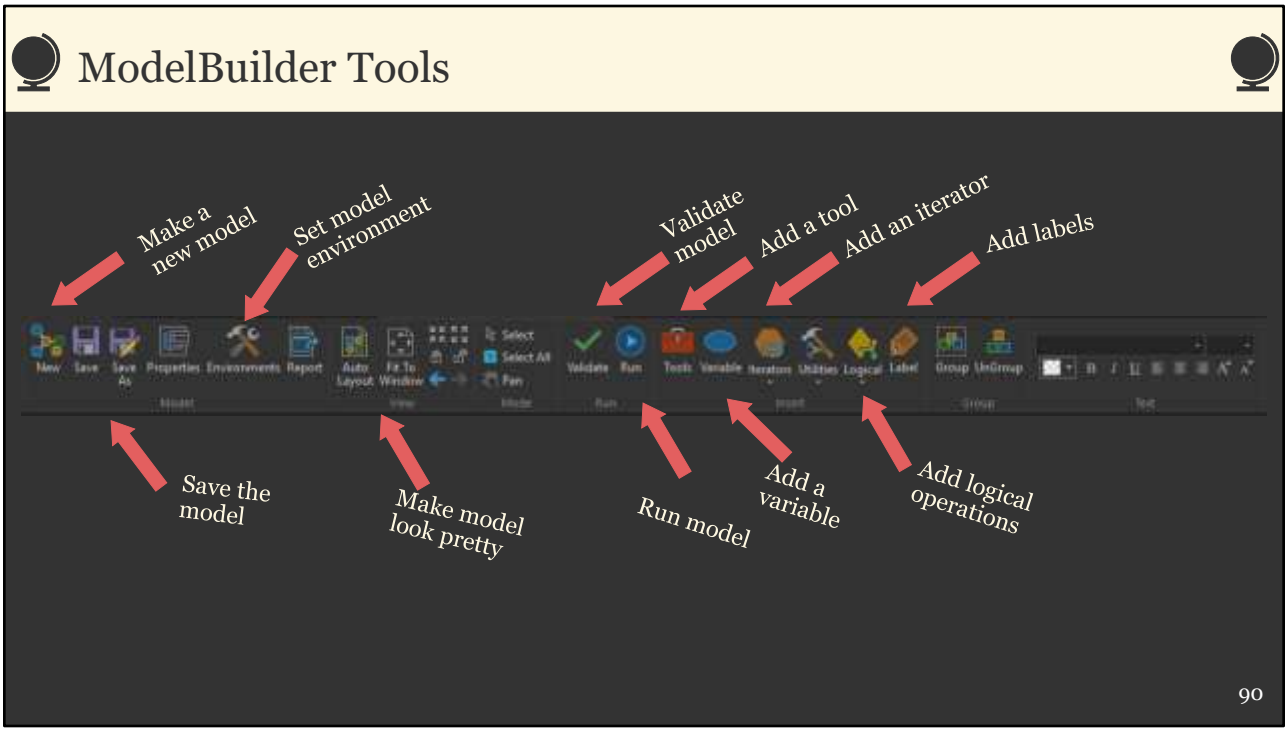
http://pro.arcgis.com/en/pro-app/help/analysis/geoprocessing/modelbuilder/iterators for-looping.htm

89

Iterators can be used to iterate through rows in a table, datasets, feature classes, files, raster datasets, and tables, among other features.

This functionality can be used to perform the same task over many data layers. For example, you could complete a series of steps for all shapefiles in a folder or feature classes in a geodatabase.

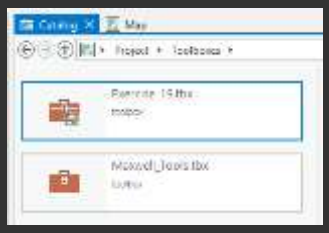
This can be a very easy and efficient way to process large volumes of data.

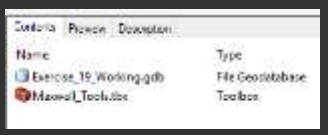


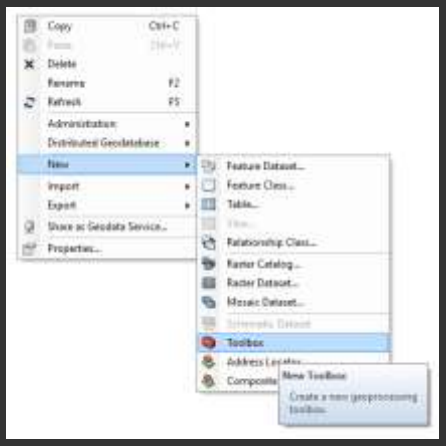
The ModelBuilder Tab in ArcGIS Pro provides access to a variety of tools and options for working with models as described here.

Saving a Model

- Models are saved inside of a Toolbox



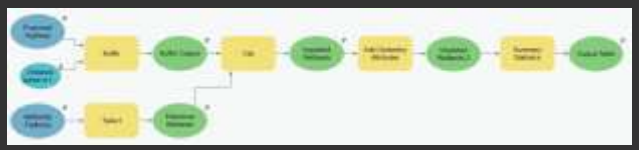


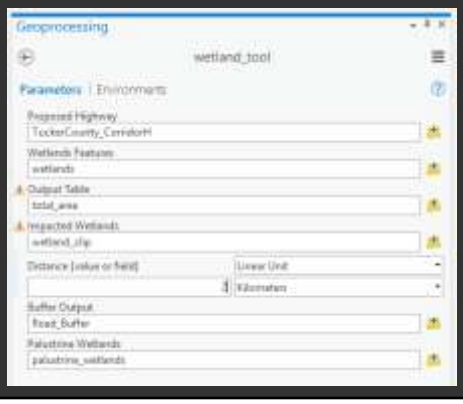


Models must be saved inside of a Toolbox. They cannot be saved in a file folder.

Creating a Tool from a Model

- Must specify user-defined parameters
- May want to rename inputs and outputs





92

If you would like to share your models with others, it would be nice if they could run the model as a standard tool as opposed to editing the model. This will not require any knowledge of ModelBuilder.

This can be accomplished by defining user-defined parameters, providing meaningful parameter names, and producing help and properties for the model. The Toolbox containing the model can then be shared with others.

Graphical Modeler In QGIS

Graphical Modeler

- Processing → Graphical Modeler
- Similar to ModelBuilder in ArcGIS Pro
- Graphical "coding" environment
- Sequence tools to automate more complex analyses
- Can name tool
- Save to .model3 file

94

QGIS also has a graphic modeling framework that does not require coding. This is called the Graphical Modeler.

Inputs

- Database Table
- Raster Layer
- Raster Band
- Vector Layer
- Point Cloud Layer
- Mesh Layer

- Can name
- Can define as mandatory
- Settings/options vary by type

95

A variety of input types can be accepted including database tables, raster grids, raster bands, vector layers (points, lines, or polygons), point clouds (such as LiDAR data), and mesh layers (3D features).

Values and Constants

- Boolean
- Color
- CRS
- Datetime
- Distance
- Duration
- Expression
- Extent
- File/Folder
- Geometry
- Matrix
- Number
- Point
- Range
- Scale
- String
- Vector Field

96

You can also specify a variety of values or constants needed as input or parameters for tools.

Tools

- All QGIS, GDAL, GRASS, and SAGA tools are available
- Can also used tools from plugins
- Can search

97

Tools available within QGIS can be added to the model. This includes tools from QGIS, GDAL, GRASS, SAGA, and installed plugins.

Parameterizing Tools

- Can name tool in model workflow
- Inputs can be:
  - Values
  - Model inputs
  - Prior tool outputs
- Settings vary by tool
- Only final outputs need to be named

98

Tools can be configured using hard coded parameters or can be connected to values or constants, model inputs, or outputs from prior tools in the model workflow. It is not necessary to name or save intermediate output.

Adding Variables

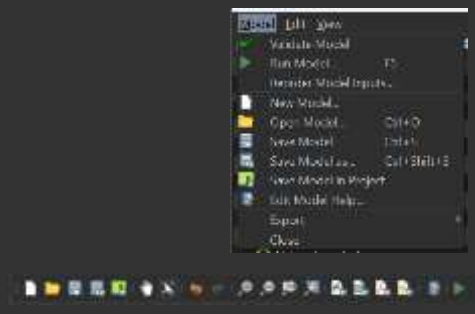
- Can use pre-calculated values provided as a variables
- Allow for user-defined input

99

You can configure the tool to accept user-defined inputs for flexibility.

Saving and Using Models

- Save/Save As
- Save in Project
- Edit Model Help
- Export to Python Script
- Export to Image
- Export to PDF
- Export to SVG

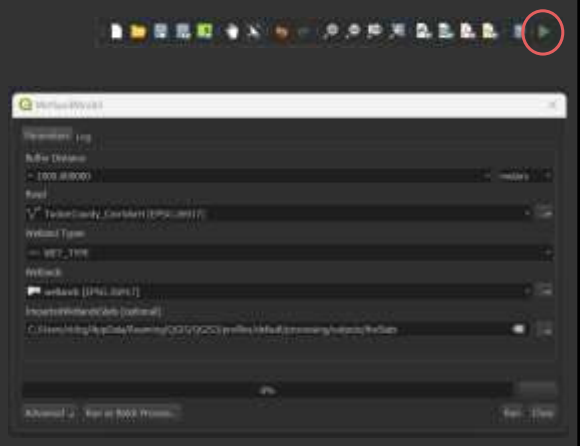


100

You can save the model, create help documentation, export it to a Python script, or export an image or flow chart of the model workflow.

Running Model

- Provide user input
- Name output



101

Models can be executed similar to other QGIS tools. Users can provide input using a wizard interface and can save the output to disk.

QGIS Graphical Modeler Tool

- Use the utah project.
- Use the graphical modeler to develop a workflow to answer the following question:
  - What is the average length of trails in Utah per county?

102

Build a tool to answer the stated question.

This is the end of this lecture module.

Please return to the West Virginia View Webpage for additional content.

Thanks! Hope you found this useful.