GIS and Soil Distribution Modeling

$$\ln\left(\frac{\pi}{1-\pi}\right) = 7.5347277 + (-0.1968505 * miles) + \beta_{GEO} + \beta_{ELEV} + \beta_{VEG}$$

where:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \text{logit link function}$$

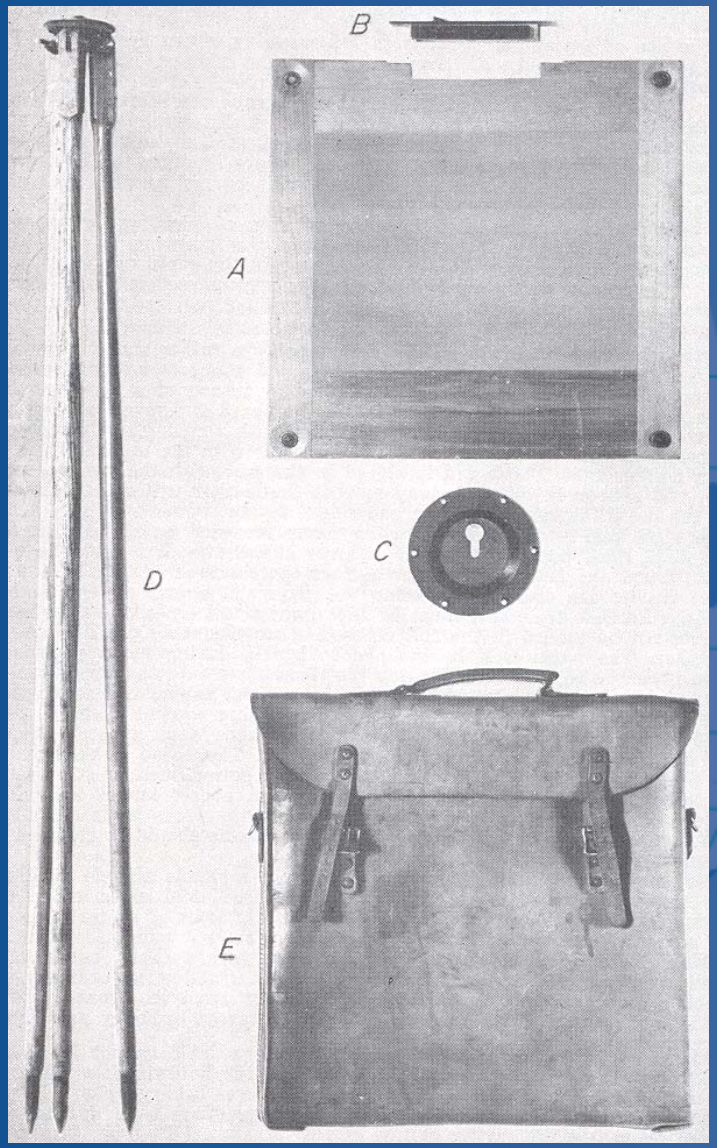
$$\text{miles} = \text{grid of miles from the coast}$$

$$\beta_{\text{GEO}} = \text{grid of coefficients for geologic formations}$$

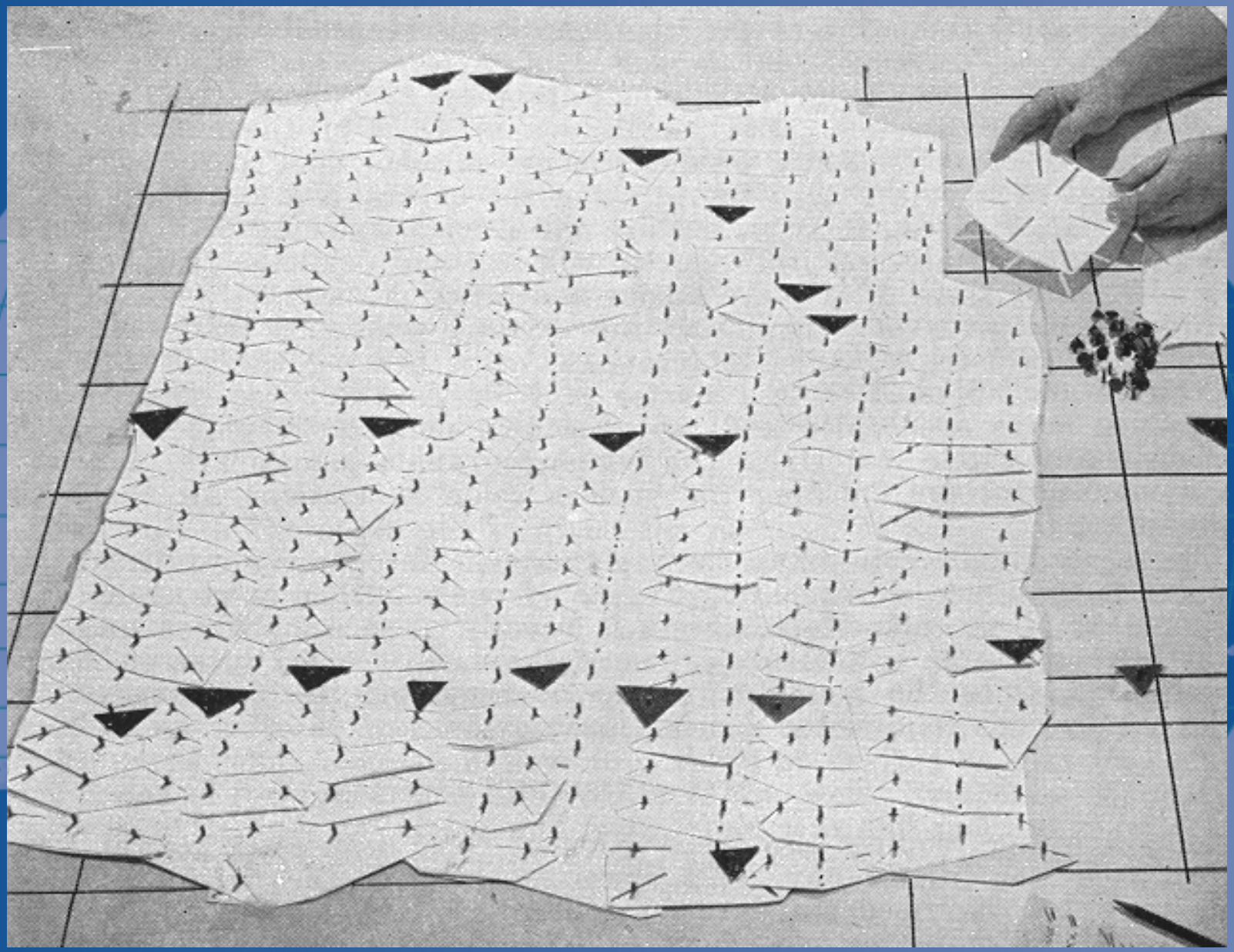
$$\beta_{\text{ELEV}} = \text{grid of coefficients for elevation classes}$$

$$\beta_{\text{VEG}} = \text{grid of coefficients for vegetation classes}$$

David Howell


B
A
C
D
E

Map Mosaic – Slotted Templates



Brief History

“And the time is not far off when the new science, with its great significance for the destiny of mankind, will justly occupy an independent and honorable position,”

Brief History

"With its own strict and well defined problems and methods, in no danger of being confused with existing branches of the natural sciences, and thoroughly permeated by geography."

(Dokuchaev, 1899)

California Status

- Two Soil Scientist / GIS Specialists supporting ongoing soil survey work
- Beginning to Implement GIS in Soil Survey Offices
  - Long-term capacity development
  - On-screen digitizing and editing
  - Data visualization and testing
  - Preliminary soil maps for customers

California Status

- Developing soil distribution modeling methods
- Collaboration with
  - Humboldt State University
  - Bureau of Land Management
  - National Park Service

Soil Distribution Modeling Definition

- The use of relevant data and computer software to ESTIMATE the geographic distribution of soil features or properties.

How is it **different** from traditional field soil mapping?

- It is JUST a model !
- It is a crude and rudimentary tool.
- It is limited by the available digital input data.

How is it **similar** to traditional field soil mapping?

- It is JUST a model.
- It must be based on field data.
- You can never sample 100% of the population.
- It must be confirmed by field work.

What is the purpose?

- Change or augment the type of soils data presented to the users.
- Reduce the time and cost of soil surveys.

How is it done?

- Models are based on data about variables which are easier to obtain (or exist in a different condition) than soils data developed by field investigations.

How is it done?

- Software estimates the relationships between the explanatory variable data and the resulting soil properties or features.

Mathematics/Statistics

How is it done?

- GIS software processes the data about the soil-forming variables, using the mathematical relationships to produce a continuous map for the project area.

How is it done?

- Most models use functional relationships such as:

$$s = f\,(cl, o, r, p, t, \ldots)$$

(Jenny, 1941)

How is it done?

- Jenny intended for the functional relationships to be quantitatively solved in as many places as possible.

- GIS and statistical models provide the tools to do this.

How is it done?

- Jenny stated the functional relationship was between soil-forming factors and soil PROPERTIES. Each soil property may have a unique relationship.

Brief History

- “Soil maps display the areal arrangement of soil properties and types but give no insight into ‘causal’ relationships.”

Brief History

- “The curve, on the other hand, reveals the dependency of soil properties on soil-forming factors, but the conversion of such fundamental knowledge to specific field conditions is impossible unless the areal distribution of the soil formers is known.”

Brief History

- "Clearly, it is the union of the geographic and the functional method that provides the most effective means of pedological research."

(Jenny, 1941)

What Are We Modeling?

- Soil properties are a result of the interaction of soil-forming processes and can be modeled continuously
- Soil Taxonomy applies to individuals, not populations

What Are We Modeling?

- "The explicit prediction of individual properties is advantageous because soil properties usually display contrasting scales of variation."

(McKenzie et al., 2000)

Who is carrying it out?

- U.S.
- England
- The Netherlands
- France
- Australia

Crisp vs. Fuzzy Attributes Polygons vs. Continuous

- Crisp soil attributes are the traditional soil survey classes.
- Polygon soil maps relate soil attributes to each polygon area.
- Fuzzy attributes allow partial membership in several classes.
- Continuous maps attempt to portray continuously varying soil information limited by pixel size.

Elevation

Continuous

Raster

Polygons
8 Classes

Vector

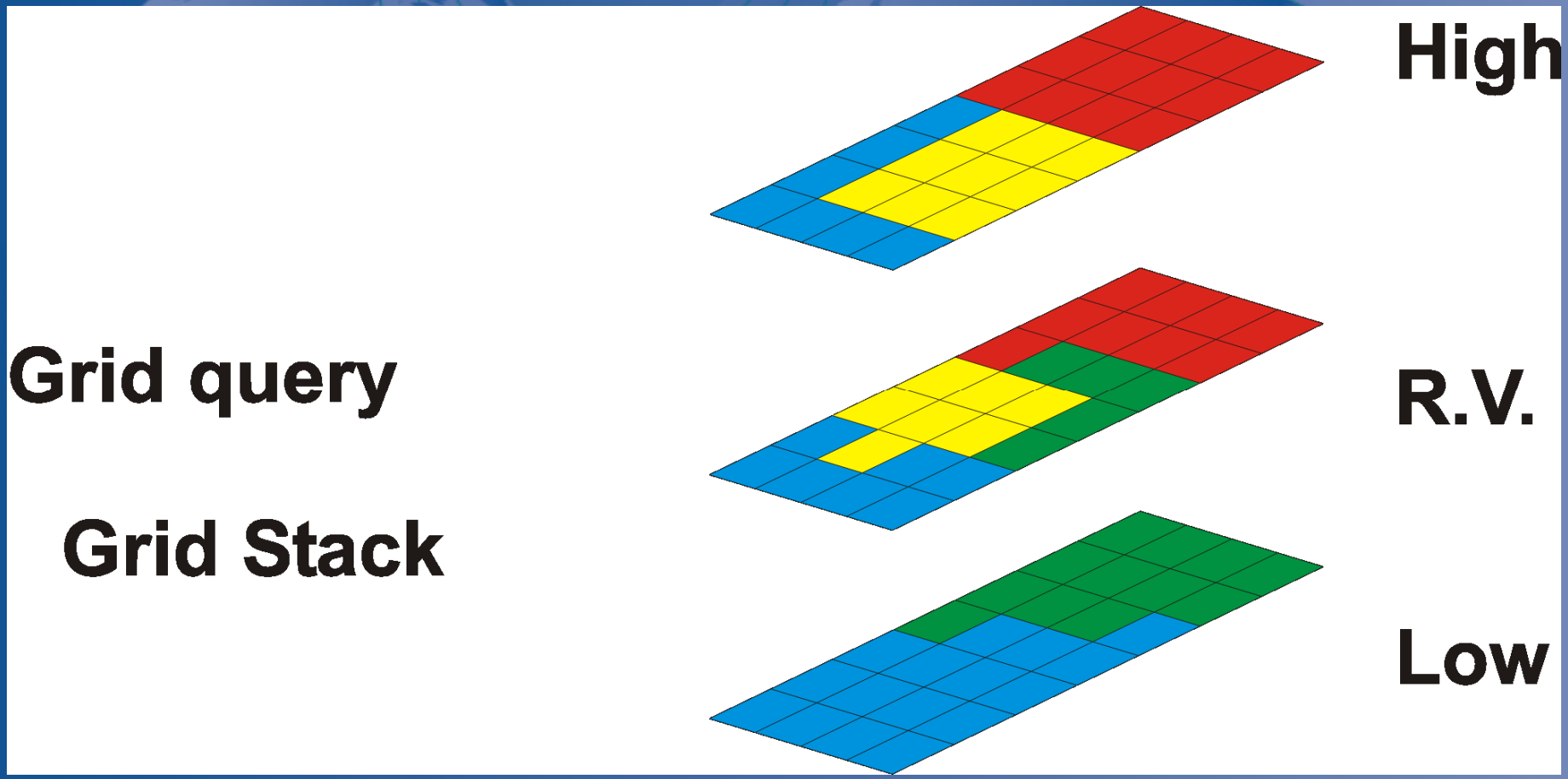
Crisp vs. Fuzzy Attributes Polygons vs. Continuous

- Geographic Modeler Misconceptions
  - Polygons enclose homogenous areas
  - Polygons enclose soil series
  - Polygons show areas that all have the same properties as the modal profile
  - Nearly all soils could be shown if the scale was large enough

Crisp vs. Fuzzy Attributes Polygons vs. Continuous

- Many geographic modelers do not understand Range in Characterstics, similar soils, and minor components, or can not handle them in the spatial context of their GIS model.

Analysis



Current Work

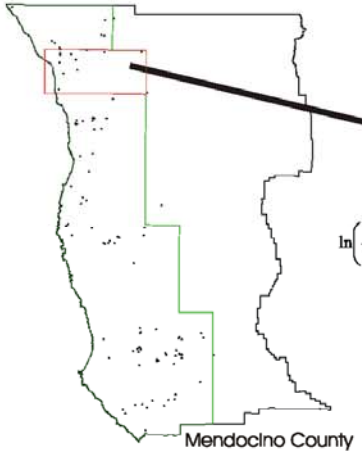
- Geostatistics
- Knowledge Base
- 3-Dimensional Models
- Regression Models
  - Generalized Linear Models
  - Generalized Additive Models
  - Regression Tree Models

Knowledge Base

SoLIM Workshop
Univ. of Wisconsin, Madison
February 16, 2001

Generalized Linear Model example

Calculate models and coefficients for each variable (soil-forming factor) from point data.



$$\ln\left(\frac{\pi}{1-\pi}\right) = (-0.7950694) + \beta_{VEG} + \beta_{GEO} + (6.500472 * miles)$$

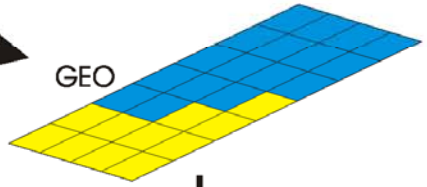
where:

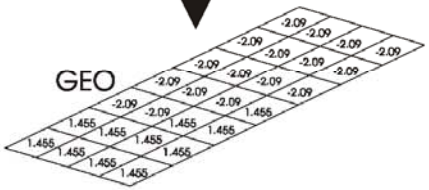
β_{VEG} = grid of coefficients for vegetation classes

β_{GEO} = grid of coefficients for geologic formations

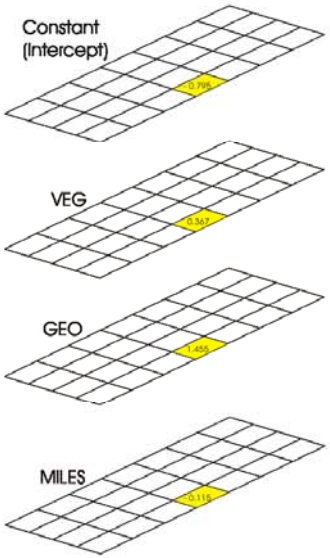
miles = grid of miles from the coast

Assign appropriate coefficient to each grid cell in each digital map of variables (layer)

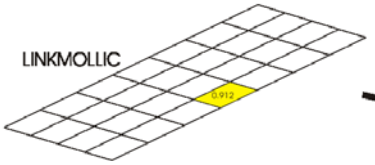




Use map algebra to to add (combine) grids of constant and coefficients



New grid of logit link values



Probability of Mollic Epipedon

= 100 * (exp(LINKMOLLIC) / (1 + exp(LINKMOLLIC)))

or for the example cell

= 100*(2.4893 / 3.4893) = 71.341%

This is calculated cell-by-cell

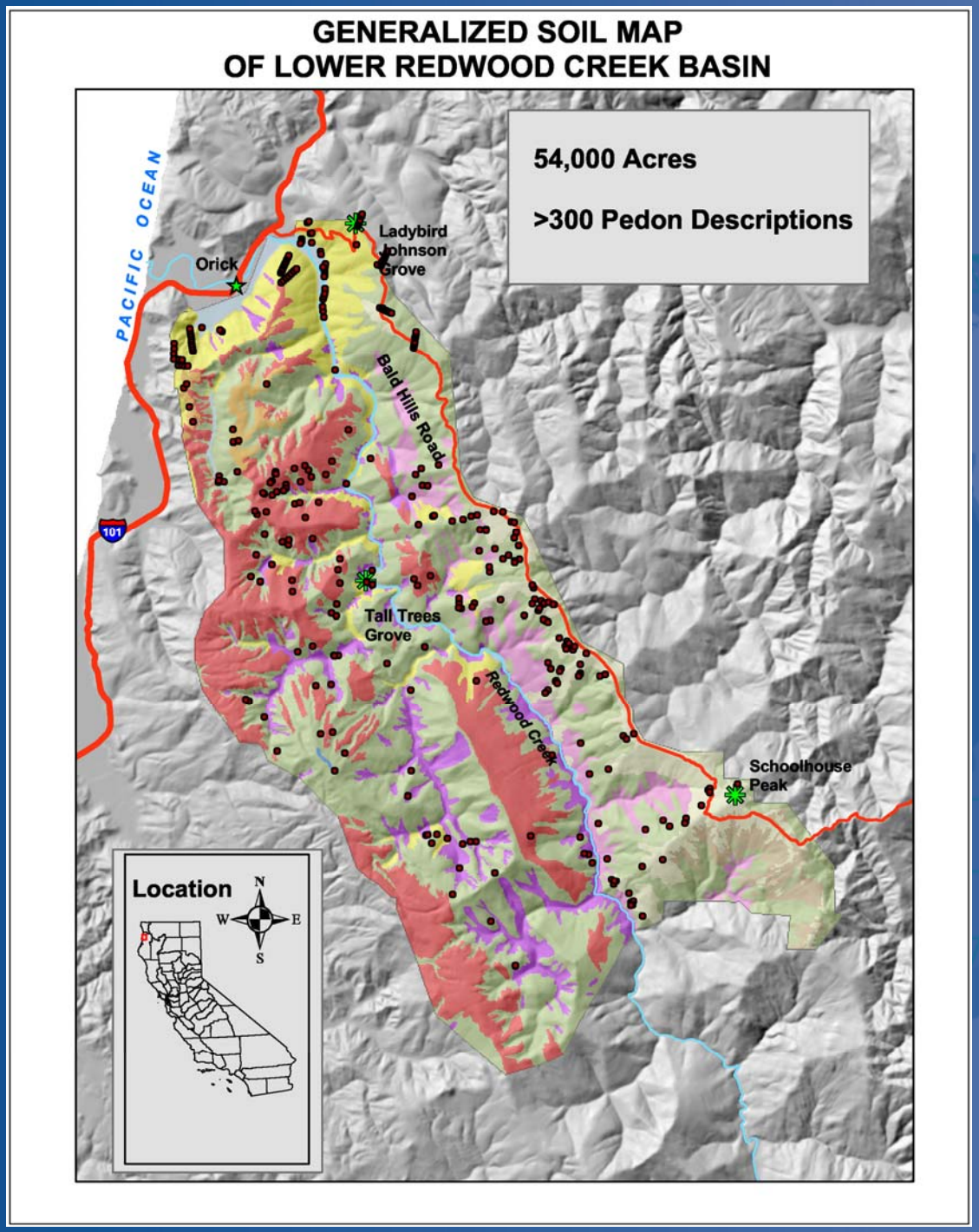
Grid cell size can be selected by the modeler. A common size is 30 meters.

Modeling Process Example – This is an example of a generalized linear model using logistic regression. Other types of models will be evaluated.

(Howell,1999)

Point Data

- Point data are taken from NASIS using “data dump” queries
- Additional class data for each point will be developed

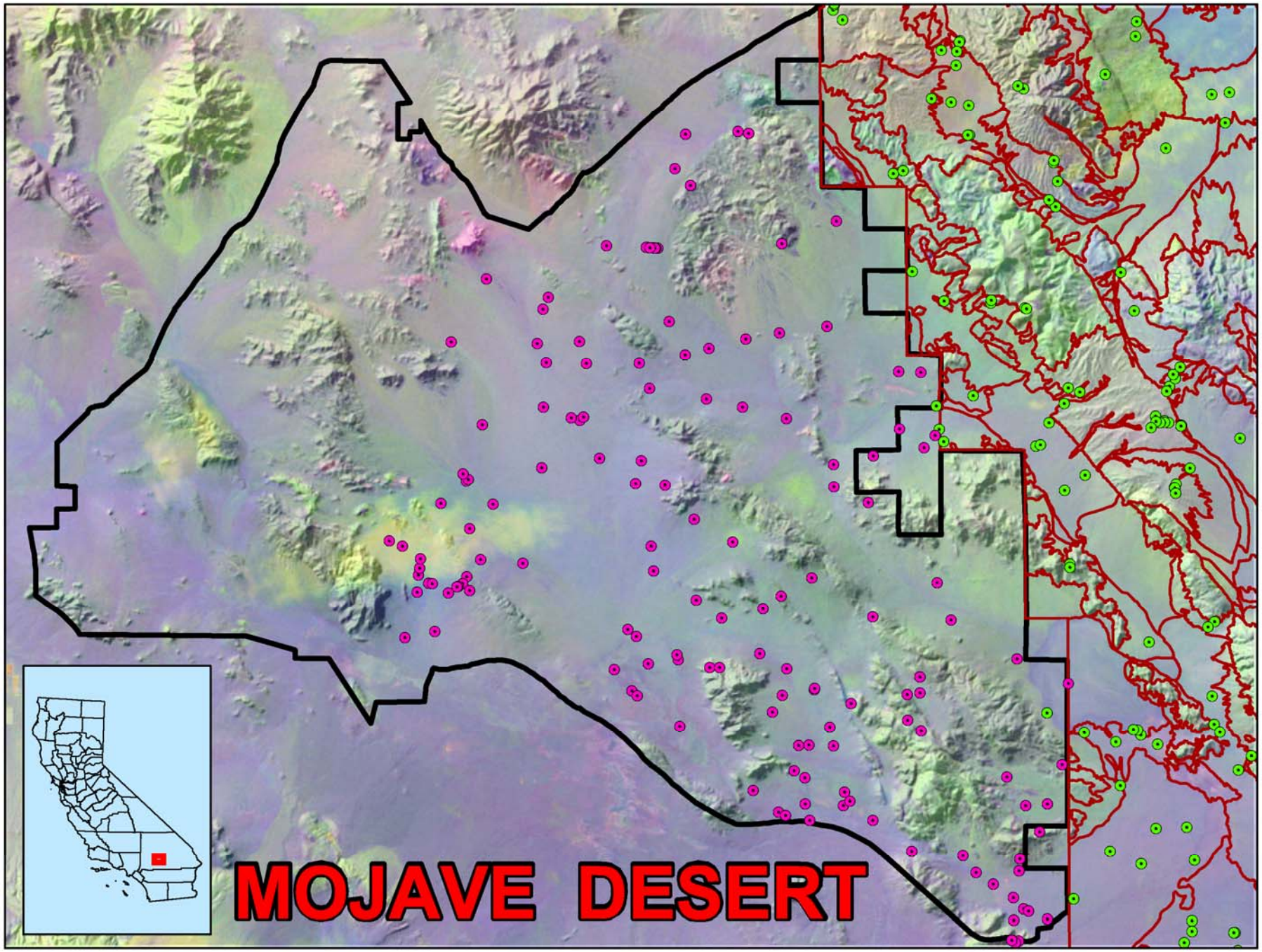


Regression Models

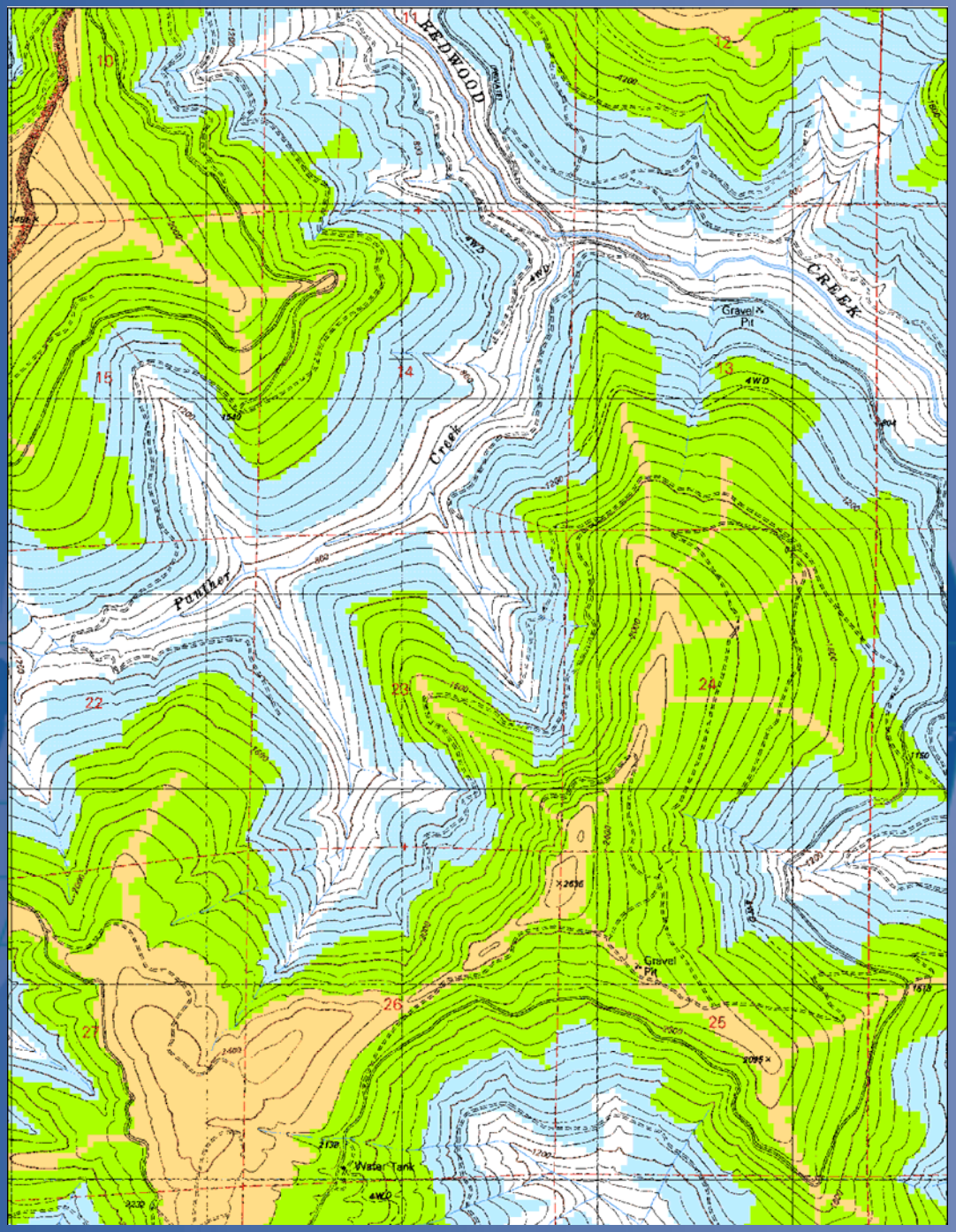
Yoon Kim, Associate Professor, Statistics, Humboldt State Univ.
David Howell, NRCS

Properties:

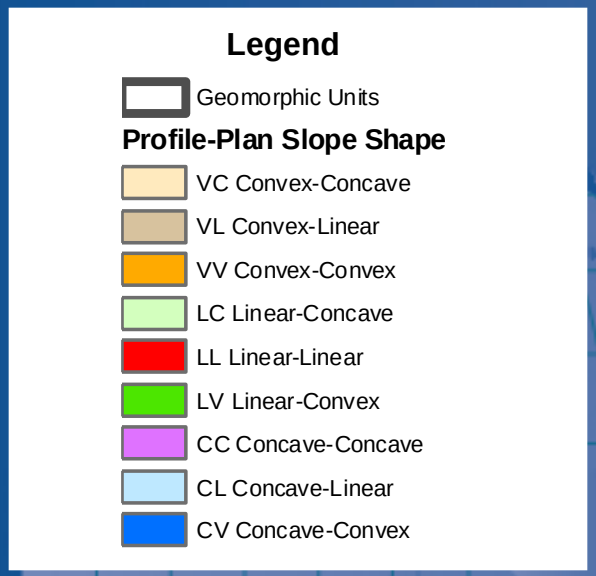
Depth to bedrock
Particle-size family
Rock fragment content
Clay content
Dark epipedon
Epipedon properties
Subsurface horizon
Substratum layer ….

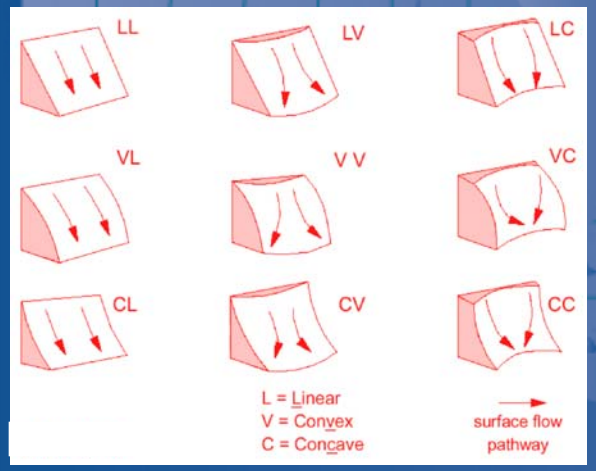

MOJAVE DESERT

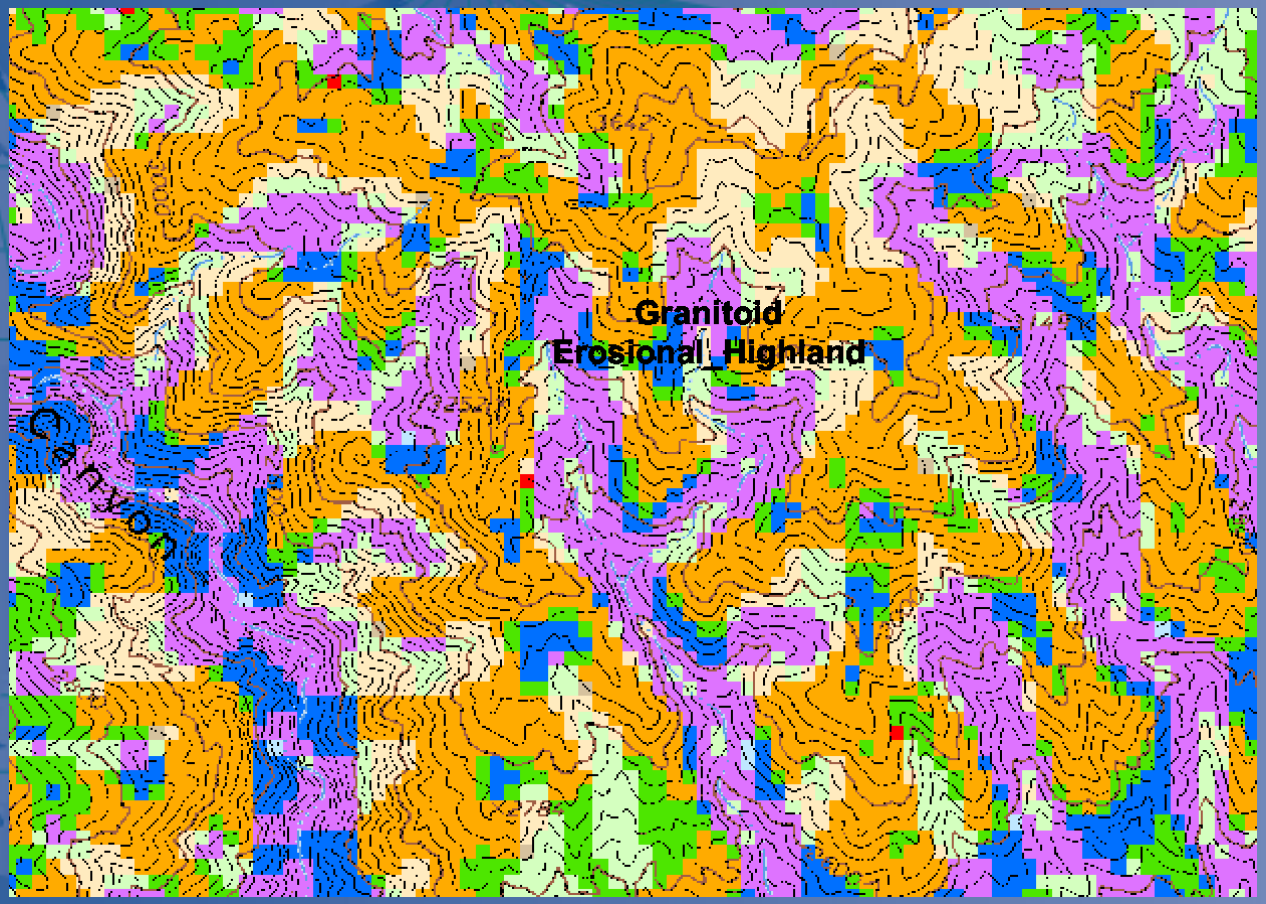
Slope Position



Compound Slope Shape







(Schoeneberger and Wysocki, 1996)

Regression Models

- Generalized Linear Model example
- Limited by digital input data
- Limited in areas where there is a weak or nonlinear relationship between independent and dependent variables

Models

- Limited by digital input data about the soil-forming factors

- Geologic data is the weakest link
- Small scale
- Poor field documentation
- Almost no attribute data
- Geologic formations, not rock types
- Poor correlation and edge matching between datasets
- Not available in digital format in many areas

Models

- Soil short-range variation is too fine for the resolution of current digital input data
- Model outputs must be generalized or fuzzy

Models

- Soil scientists should become familiar with modeling uses of soils data and alternative modeling methods
- Soil scientists should build GIS models so that they reflect field knowledge of soils

Models

- Continuous soil distribution maps are just estimates, they must be confirmed
- Must be based on field data

Models

- GIS modelers should not over-promise or exaggerate their products
- If they do not have field soil survey experience they may be unaware when they are doing this

Models

- Other scientists should not accept their claims without appropriate review and analysis
- Too easy to get lost in the esoteric statistical techniques of a theoretical, emerging, and developmental field

Models

- Soil properties are the real item of interest for environmental modeling
- Model output must allow detailed soil property interpretation for users

Reality Check

- "As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality."

Albert Einstein (as quoted by Black, 1937)

Reality Check

- "In the central part of the United States fairly satisfactory correlations between *s* properties frequently have been reported.

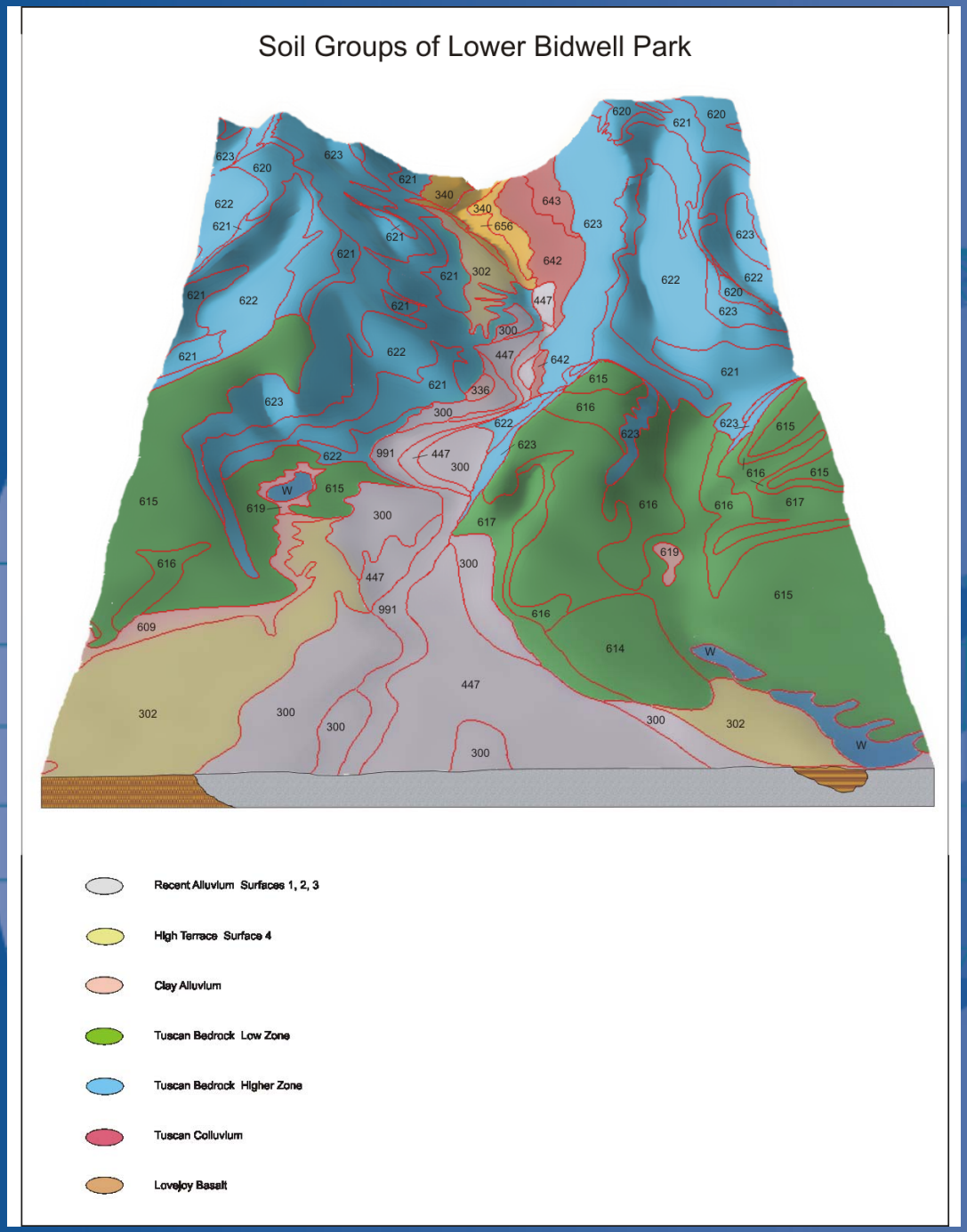
  These functions are valid because these regions are characterized by a relative constancy of soil formers."

Reality Check

- "In other parts of the country, notably on the Pacific Coast, correlations are notoriously poor.

  The combination of high mountain massives and a vast ocean produces such wide variations in climate, organisms, parent material, and topography that any effort to establish general functions among soil properties of the *s* type must lead to dissapointments."

(Jenny, 1941)



Block Diagrams

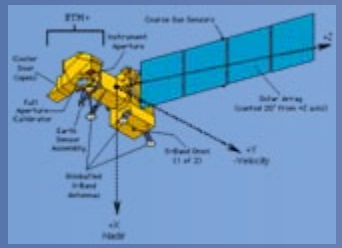
ArcGIS™

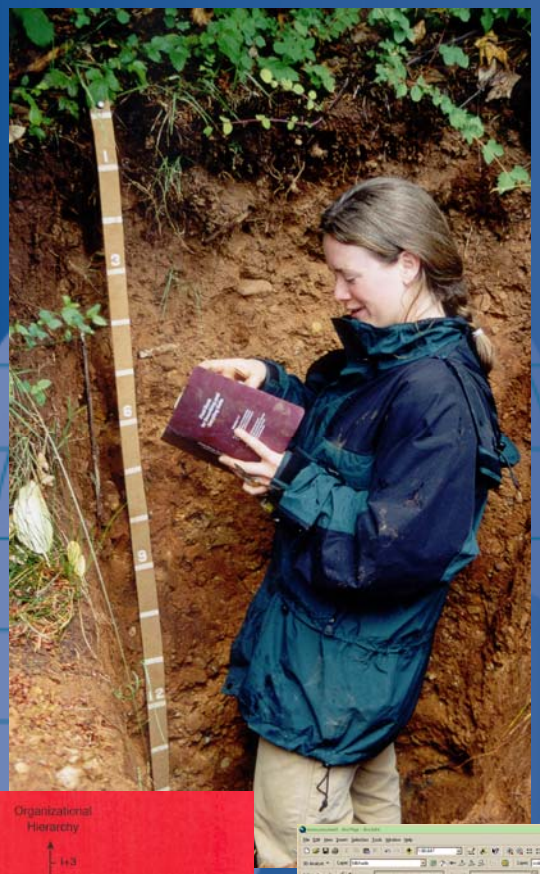
ArcScene™

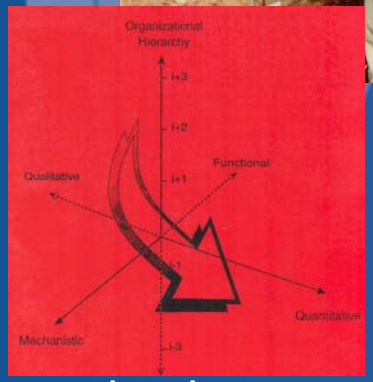
Illustration Software

- **Bodenkunde**
  - Soil knowledge, soil science (Sprengel, 1837)
- **Erdkunde**
  - Knowledge of the world, geography

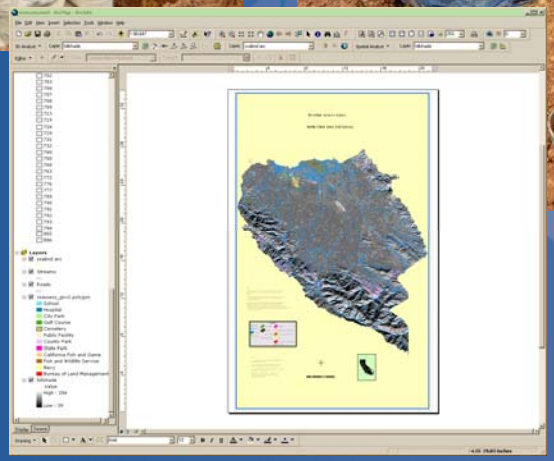
Soil Survey of the Future







(Hoosbeek and Bryant, 1992)



$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 \ldots + \beta_n\chi_n$$

$$\left(\frac{\pi}{1-\pi}\right) = e^{\beta_0+\beta_1\chi_1+\beta_2\chi_2\ldots+\beta_n\chi_n}$$

$$\pi = \frac{e^{\beta_0+\beta_1\chi_1+\beta_2\chi_2\ldots+\beta_n\chi_n}}{1+e^{\beta_0+\beta_1\chi_1+\beta_2\chi_2\ldots+\beta_n\chi_n}}$$

Thank You!

david.howell@ca.usda.gov