A Refined Approach for Examining *Phytophthora ramorum* Risk in the Eastern U.S.

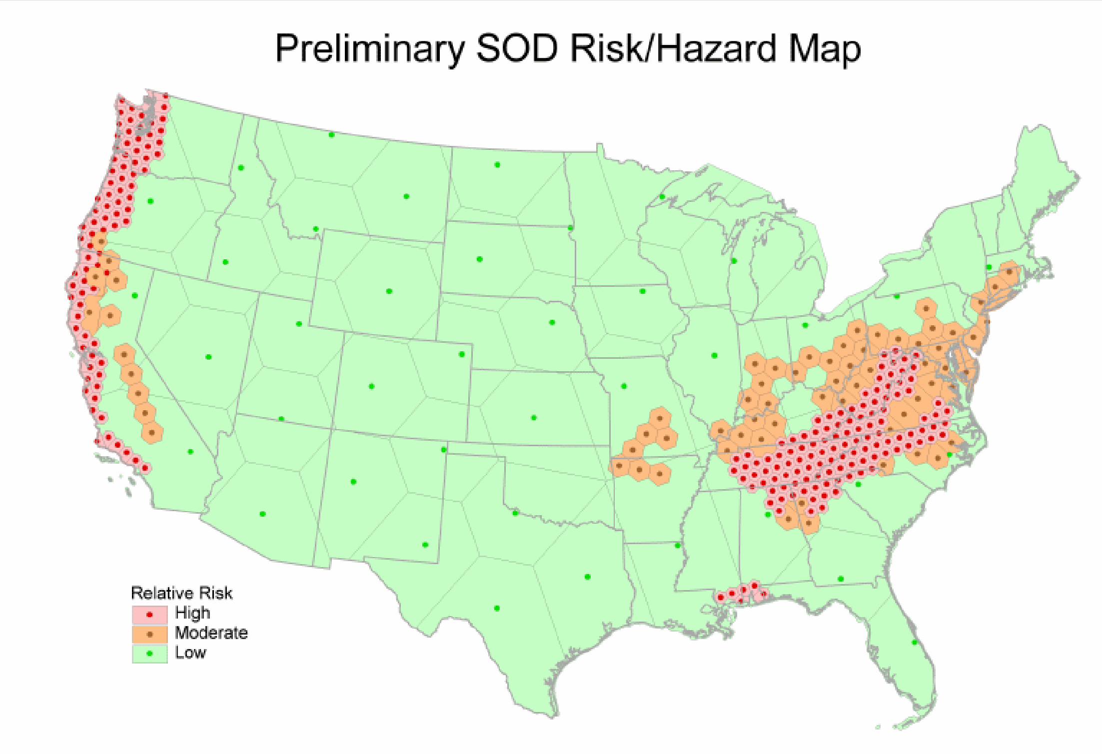
Frank Koch

NC State University

USDA-FS FHM, RTP

In Support of the National SOD Survey Protocol…

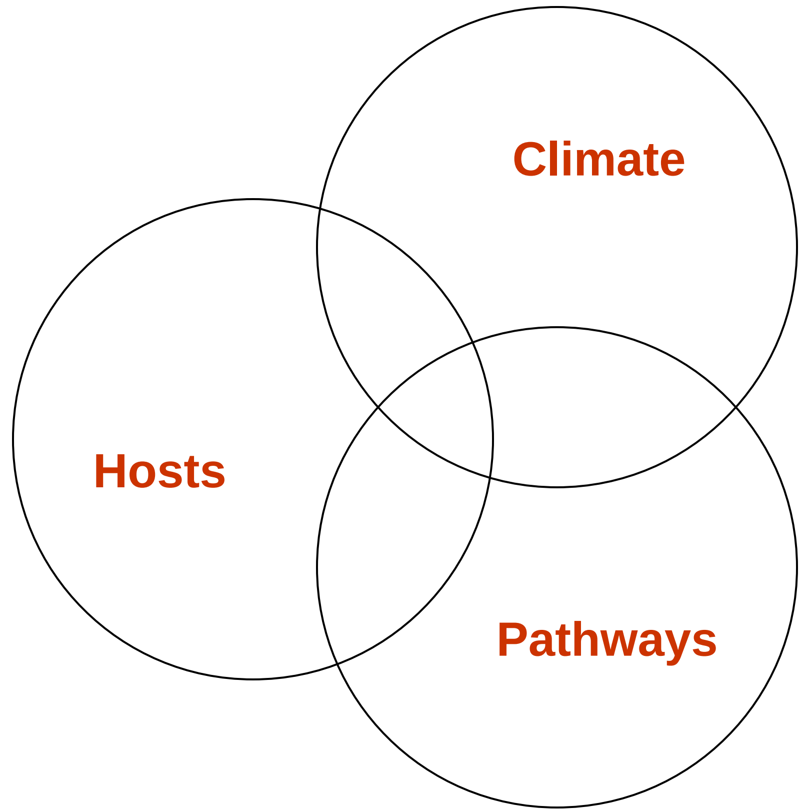
- Two tasks:
  - Risk-based spatial database for SOD
    - Any relevant data layers
  - Use the database in a modeling context



Current Survey Grid

I. Risk-based Spatial Database for SOD

- Basic strategy: Develop data for the three categories of the initial overlay analysis
- Still in progress, but have a number of examples…



Climate

- Previously using PRISM data for climate analyses
  - – 2 km resolution, but expensive
  - – Not updated to current
- Can we generate more up-to-date products?
  - – NOAA daily and monthly station data
- Spatial Interpolation
    - Gradient plus inverse distance squared (Nalder & Wein 1998)
    - Compared favorably to several other methods, including ordinary kriging, detrended kriging, co-kriging, inverse distance squared

From Nalder and Wein (1998)

- GIDS Model (based on the 30 nearest neighbors)

$$Z_k = \frac{\left[\sum_{i=1}^{30} \frac{Z_i + C_x(X_k - X_i) + C_y(Y_k - Y_i) + C_e(E_k - E_i)}{d_i^2} \right]}{\sum_{i=1}^{30} \frac{1}{d_i^2}}$$

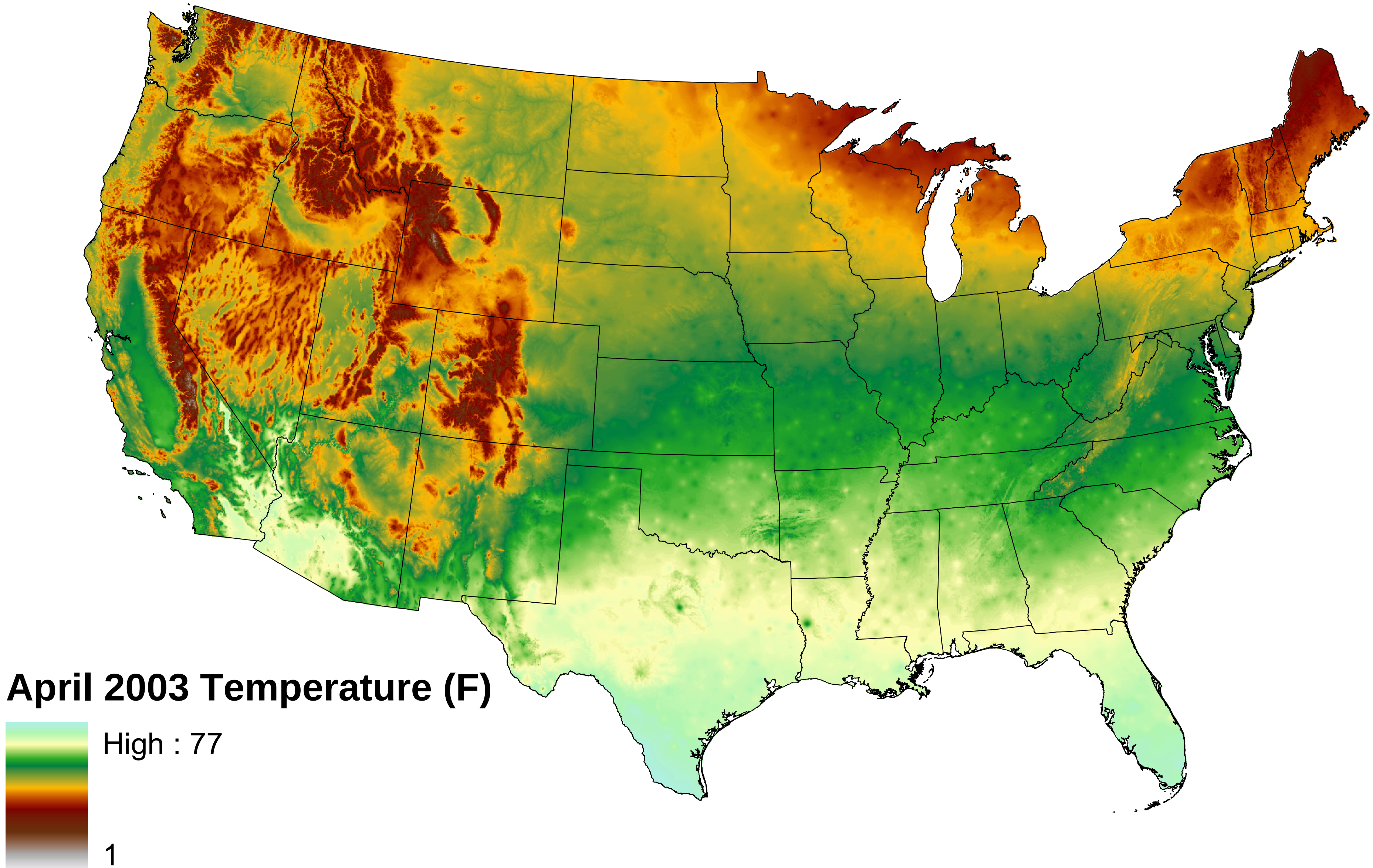
Where,
Z_k = predicted value at unmeasured location *k*
Z_i = measured value at location *i*
X= x-coordinate
Y= y-coordinate
E= elevation
D=distance from measured location *i* to Z
C_x, C_y, and C_e are based on the ordinary least square solution of the following regression model using 30 nearest neighbors to location *k*.
$Z = a + C_x X + C_y Y + C_e E + \varepsilon$
Where, *a* is the intercept and ε is error.

Regression Model Selection

- Using three possible gradients, there are seven possible models
  1. $Z = a + CxX + CyY + CeE + \varepsilon$
  2. $Z = a + CyY + CeE + \varepsilon$
  3. $Z = a + CxX + CeE + \varepsilon$
  4. $Z = a + CxX + CyY + \varepsilon$
  5. $Z = a + CxX + \varepsilon$
  6. $Z = a + CyY + \varepsilon$
  7. $Z = a + CeE + \varepsilon$
- Test each model to examine if all the independent variables are significant
- If more than one model has all significant independent variables then the model with greatest R^2 is selected
- If no model has all significant independent variables then simple inverse distance square weighting is used (i.e. all coefficients are set to zero)

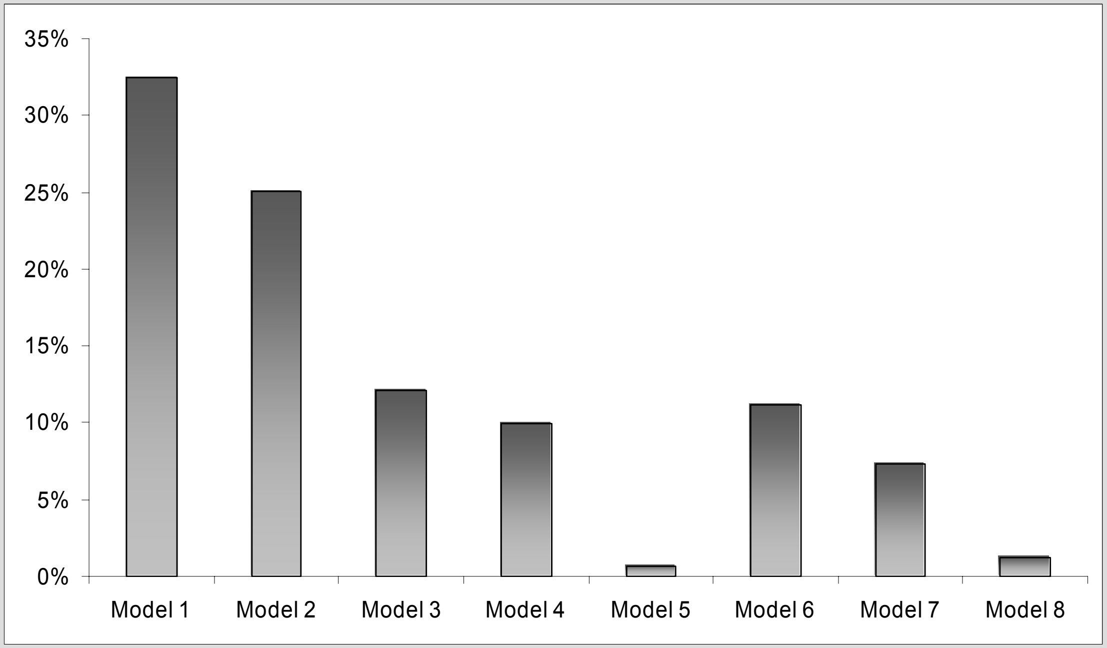
GIDS Surface of Temperature from 2003 NOAA Data

($4km^2$ cells)



Temperature Model Selection

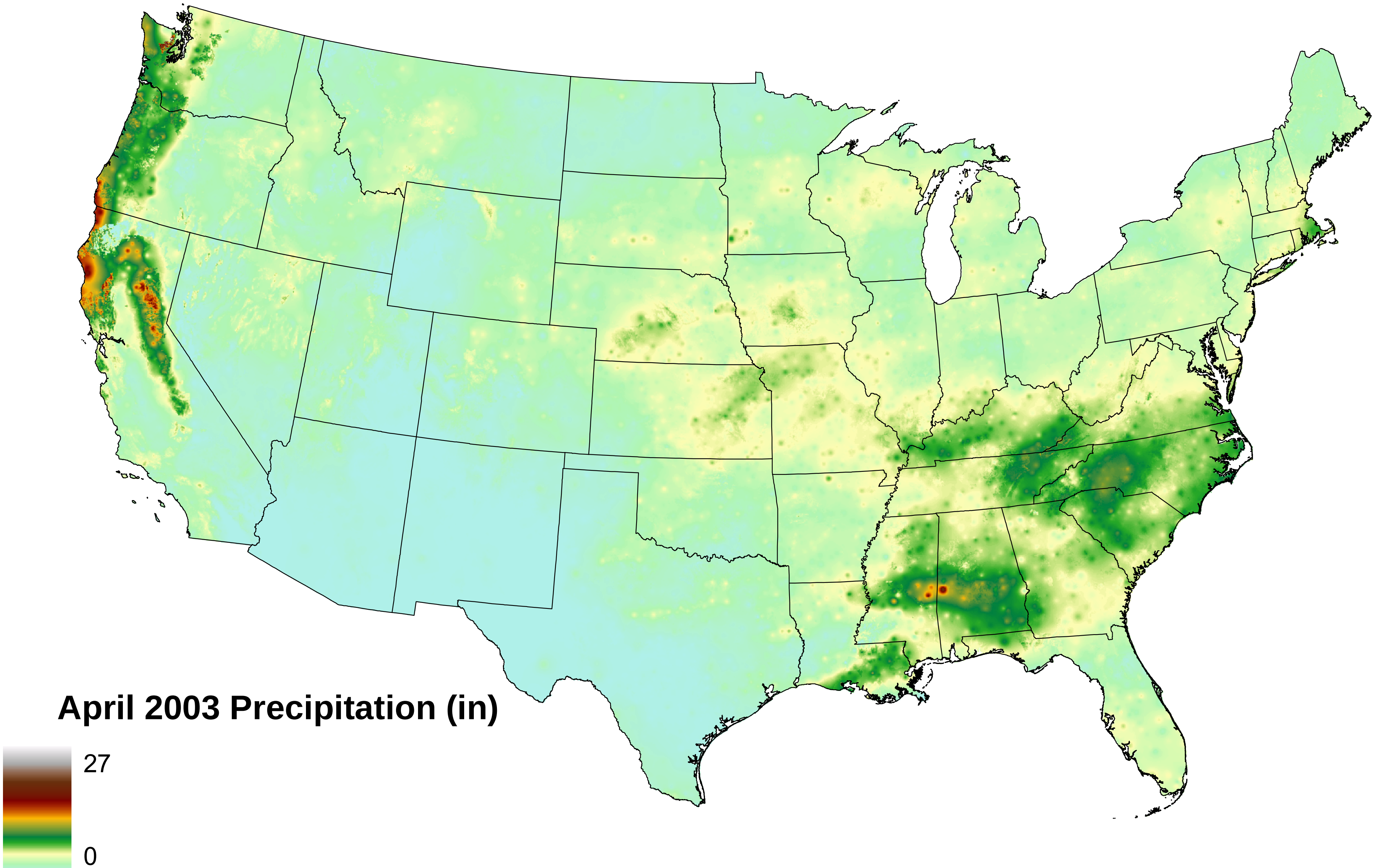
for each prediction point



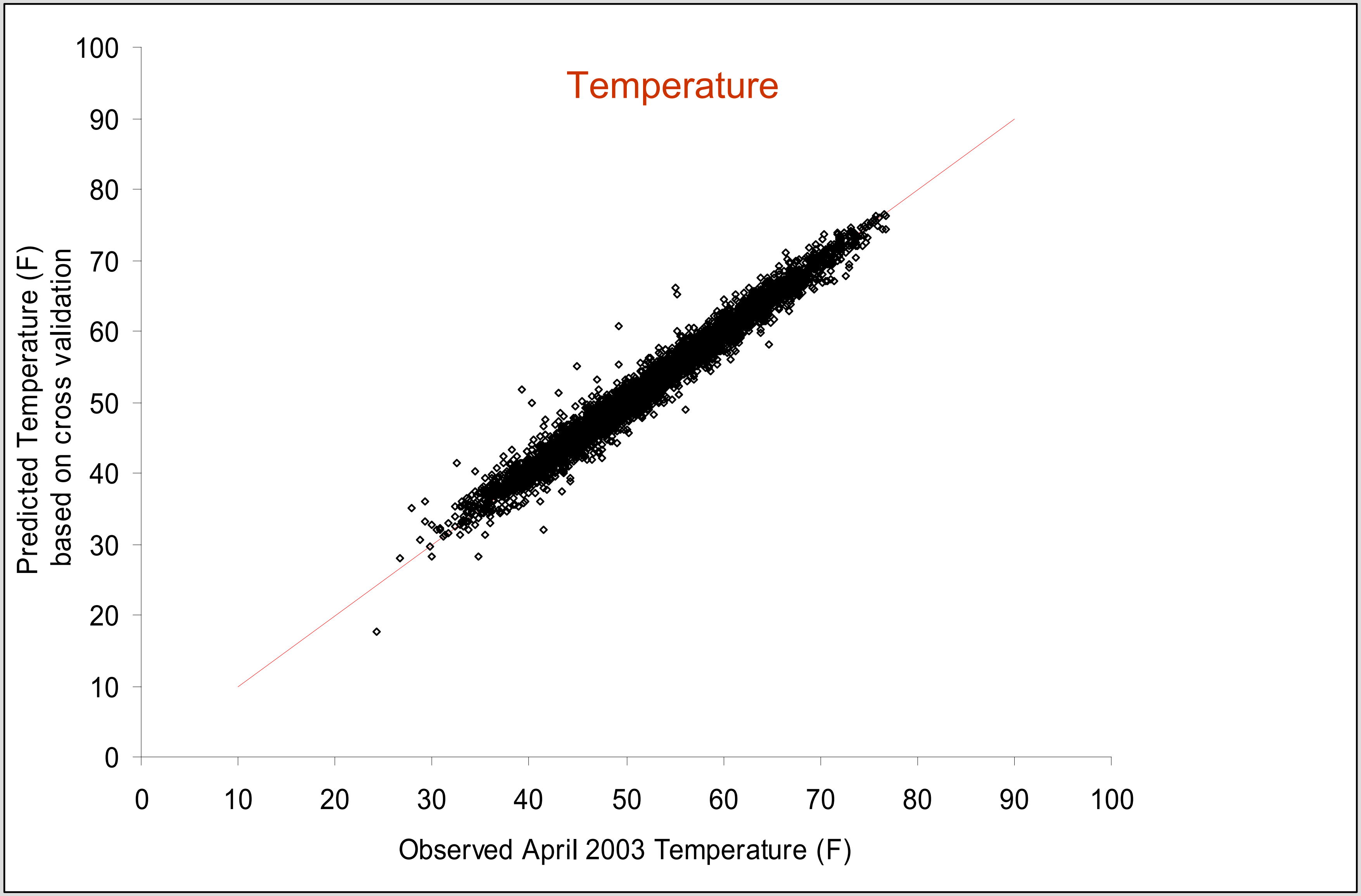
Model

1. $Z = a + CxX + CyY + CeE + \varepsilon$
2. $Z = a + CyY + CeE + \varepsilon$
3. $Z = a + CxX + CeE + \varepsilon$
4. $Z = a + CxX + CyY + \varepsilon$
5. $Z = a + CxX + \varepsilon$
6. $Z = a + CyY + \varepsilon$
7. $Z = a + CeE + \varepsilon$
8. IDW

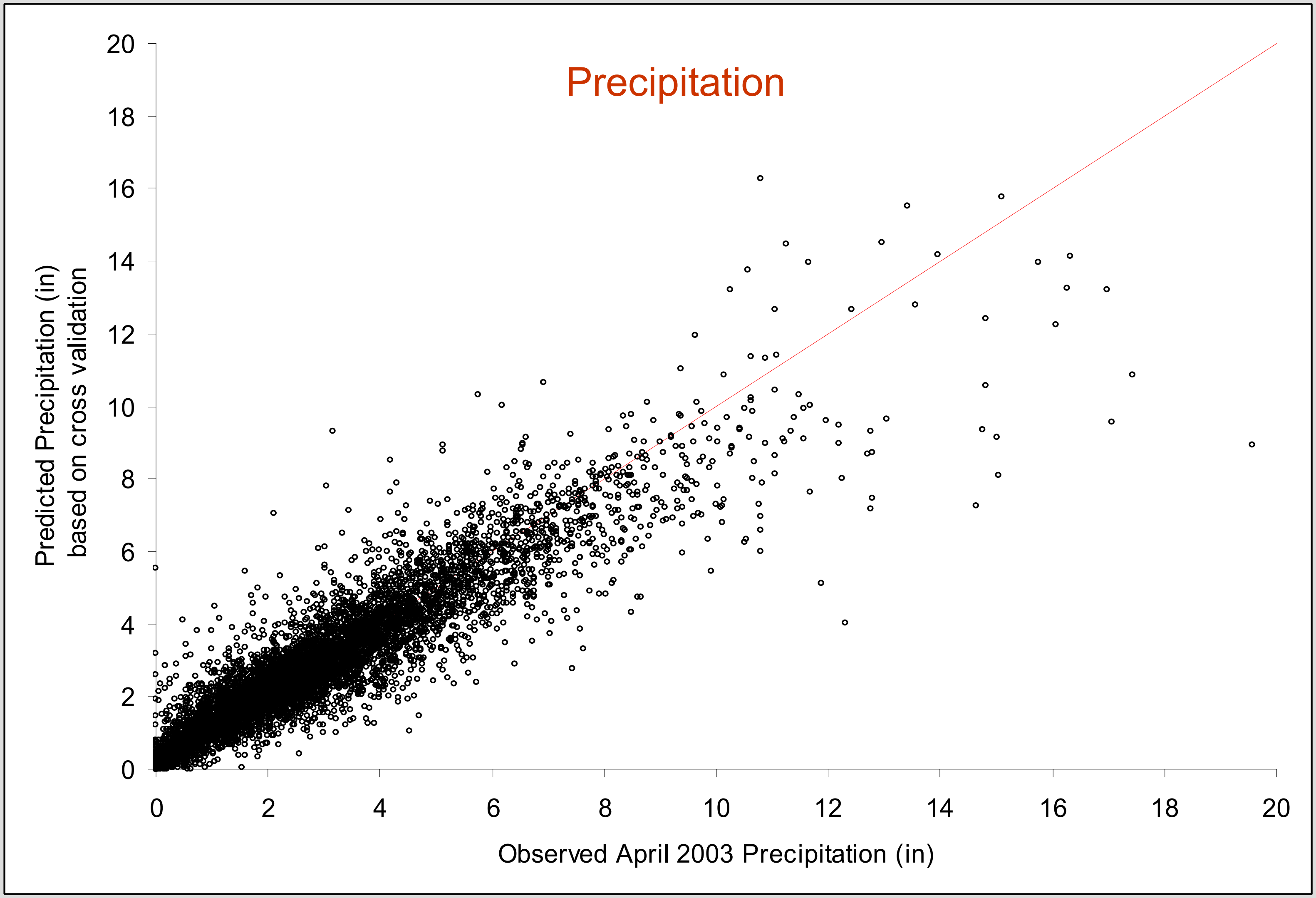
GIDS Surface of Precipitation from 2003 NOAA Data



Cross-Validation Results



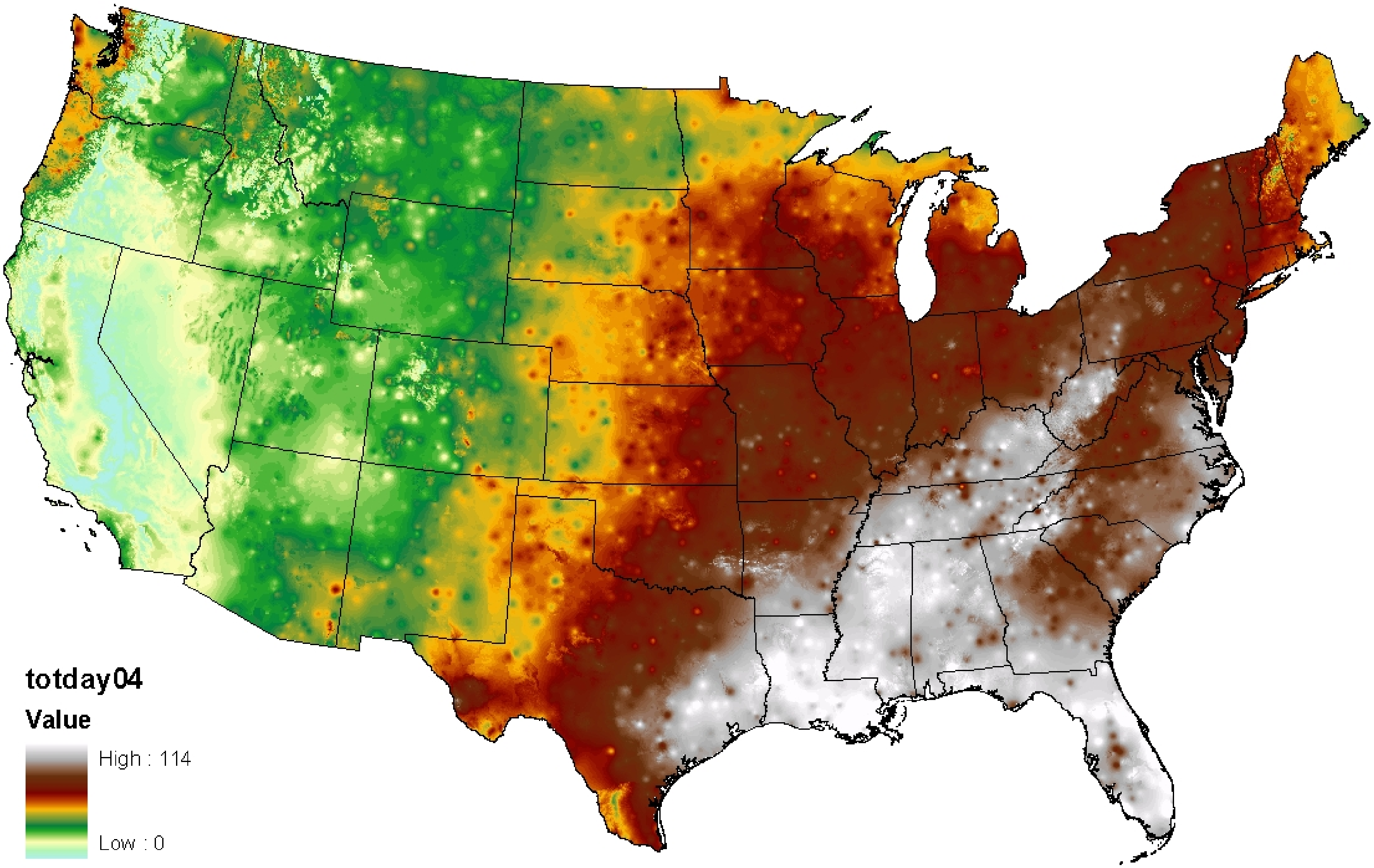
Variable	N	Mean	Mean error	RMSE
April 2003 Temperature (F)	4994	52.5	0.007	1.468
April 2003 Precipitation (in)	6847	2.9	-0.004	0.868



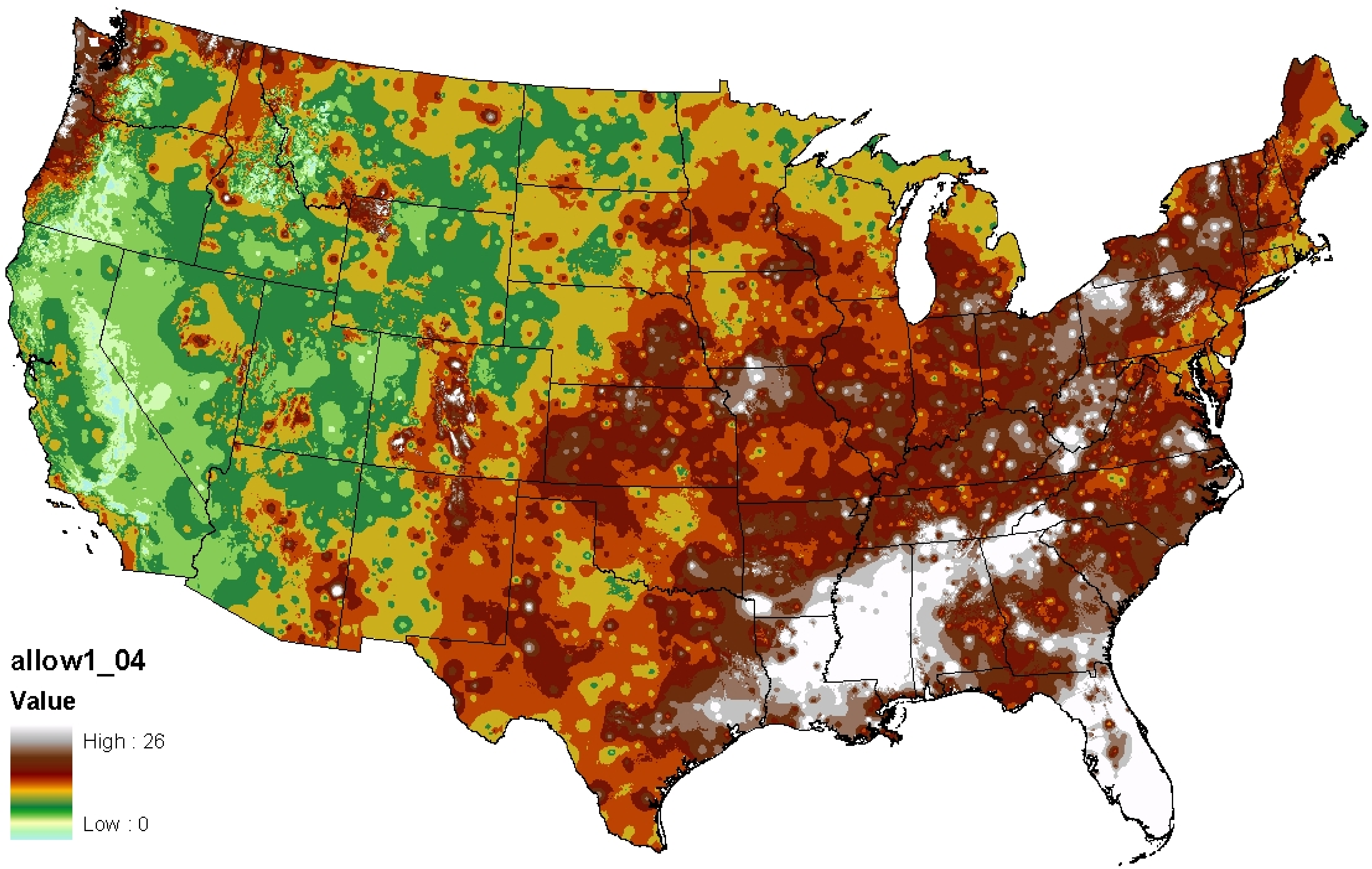
GIDS & Poisson Regression

- GIDS seems to work well, but what about surfaces representing number of days with appropriate conditions?
- Count-based = Poisson distribution
- Like previous GIDS, seven possible models
- Best model chosen based on chi-squared significance (or log likelihood)
- Three-dimensional distance
- Still working out a few things…

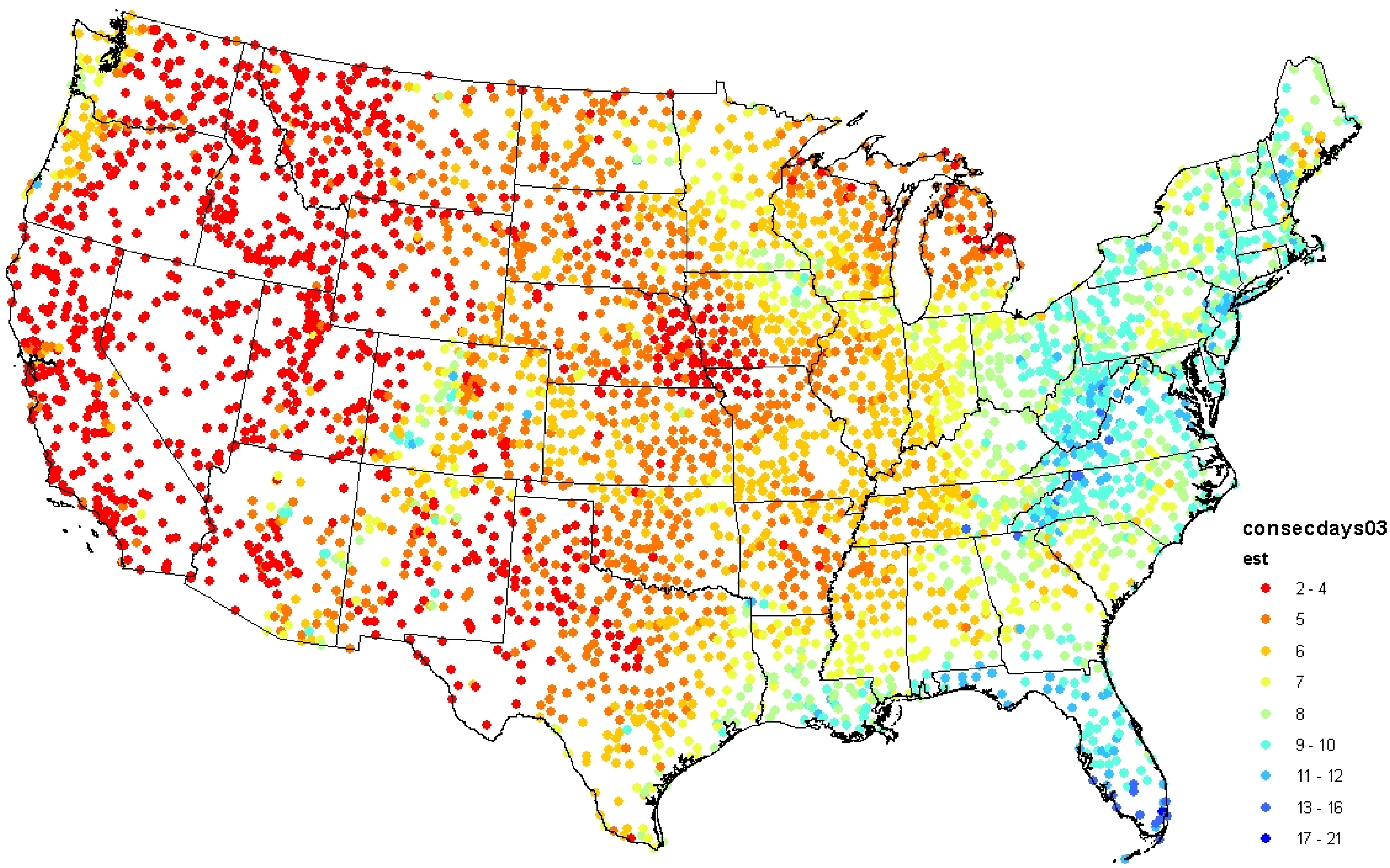
Number of Days in 2004 With Optimal Temperatures (60-80°) and Precipitation > 0.05”



Number of Consecutive Days (Allowing One Day Off) in 2004 with Optimal Temperatures and Precipitation

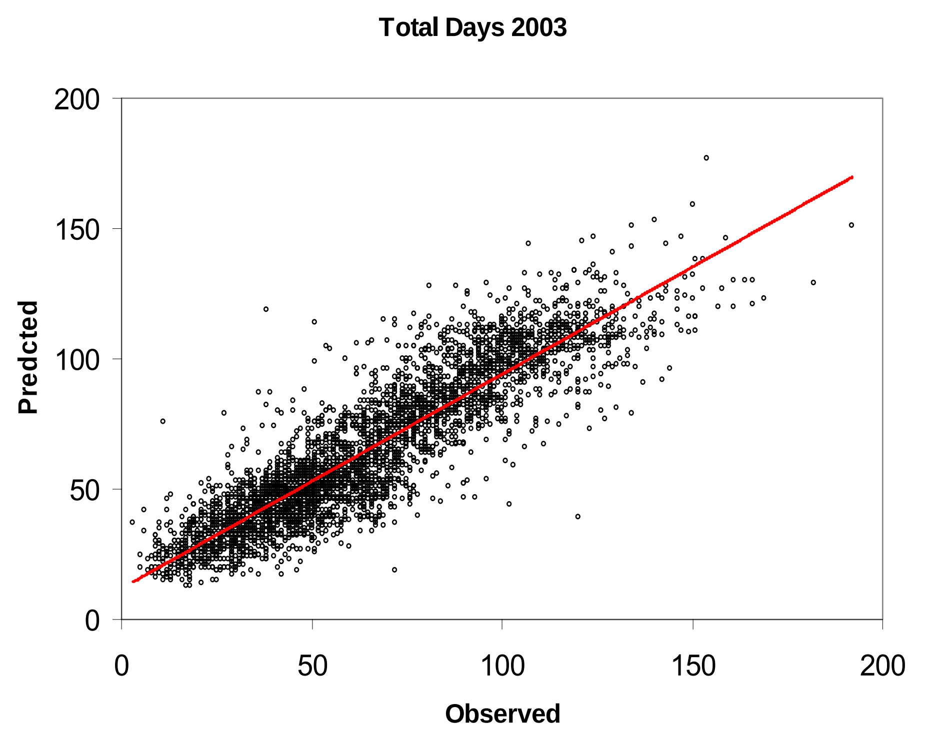


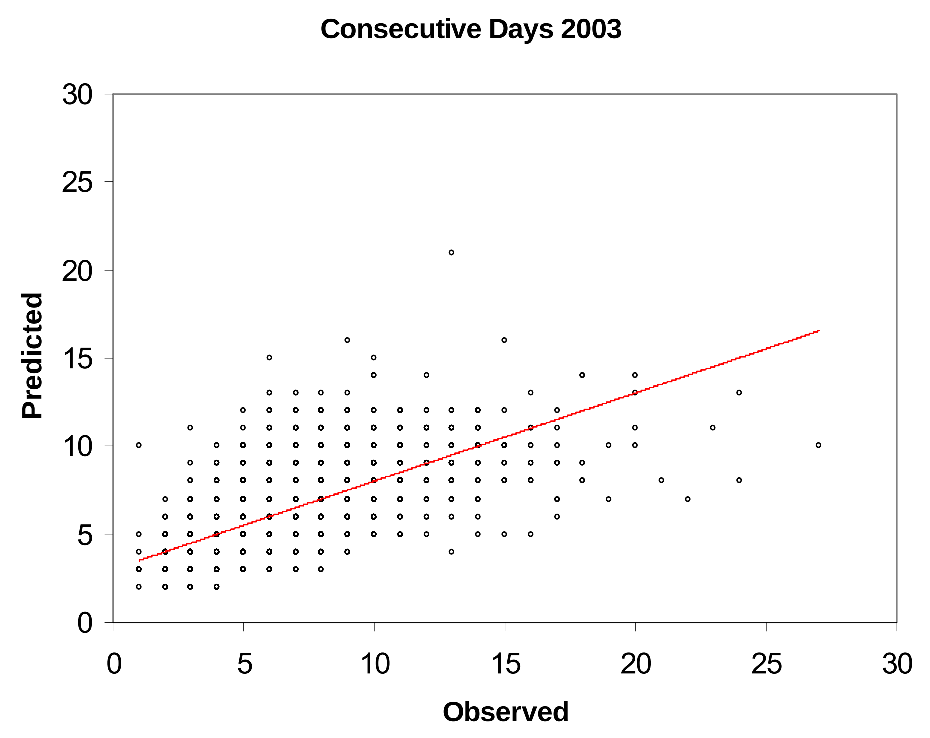
Number of Consecutive Days (No Days Off) in 2003 with Optimal Temperatures And Precipitation – All Reporting Stations (n=4144)



Some Preliminary Results…

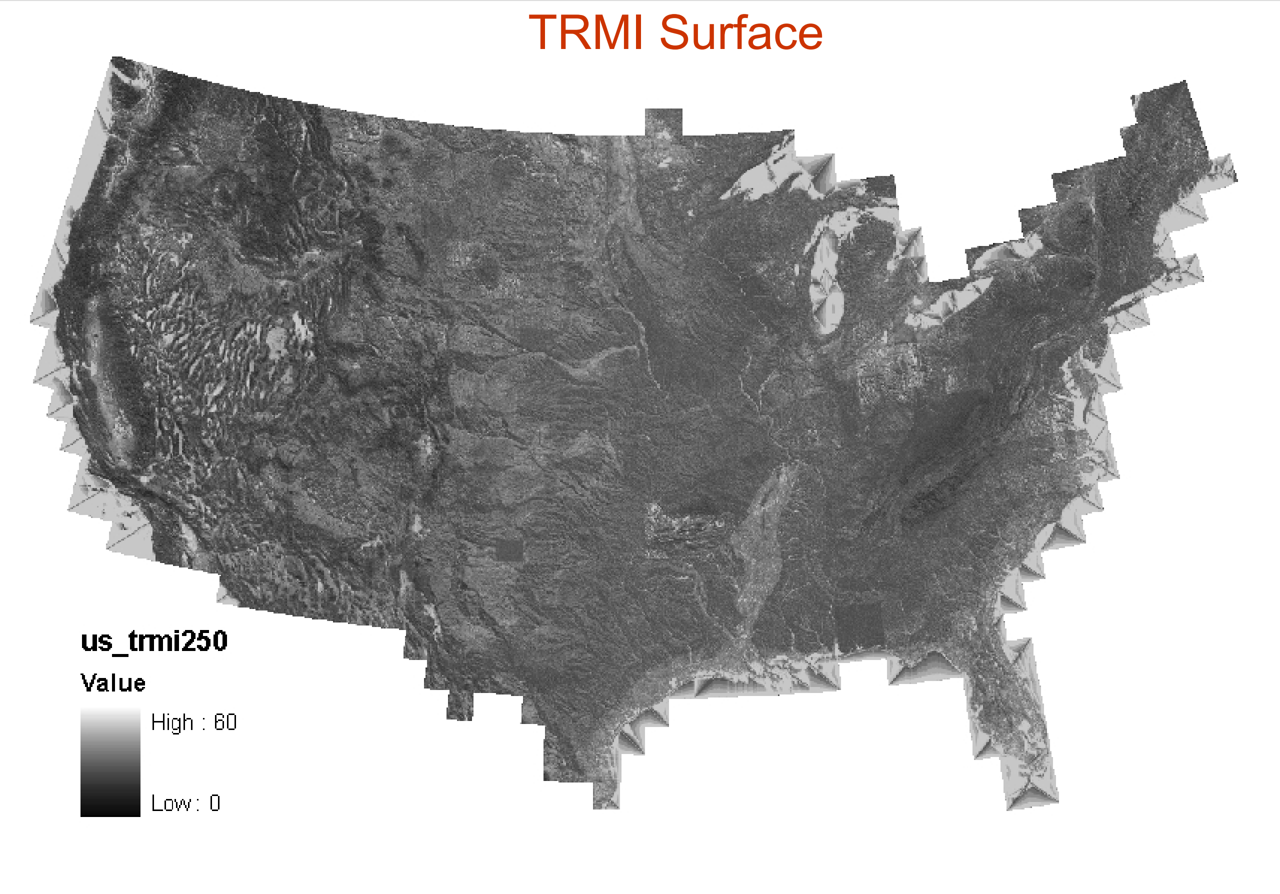
Variable	N	Mean	Max.	RMSE
Total Days 2003	4144	64.84	191	13.36
Consecutive Days 2003	4144	6.03	21	2.15



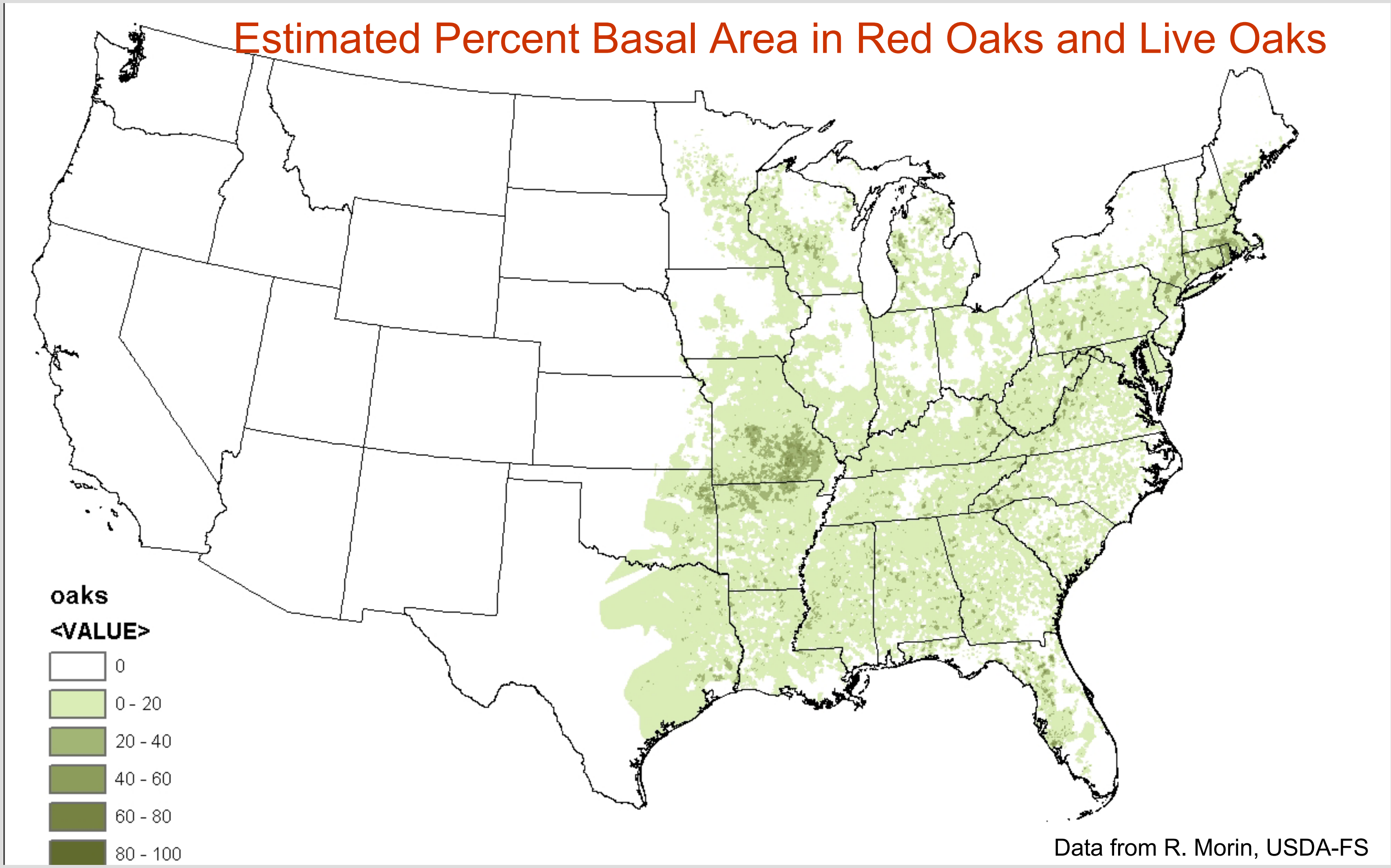


Other Climatic Factors

- Relative Humidity
  - 100% humidity important for *P.ramorum* and other aerials (?)
  - Fewer stations
- Microclimate
  - Topographic Relative Moisture Index (TRMI)
  - represent local conditions (e.g., hollows)



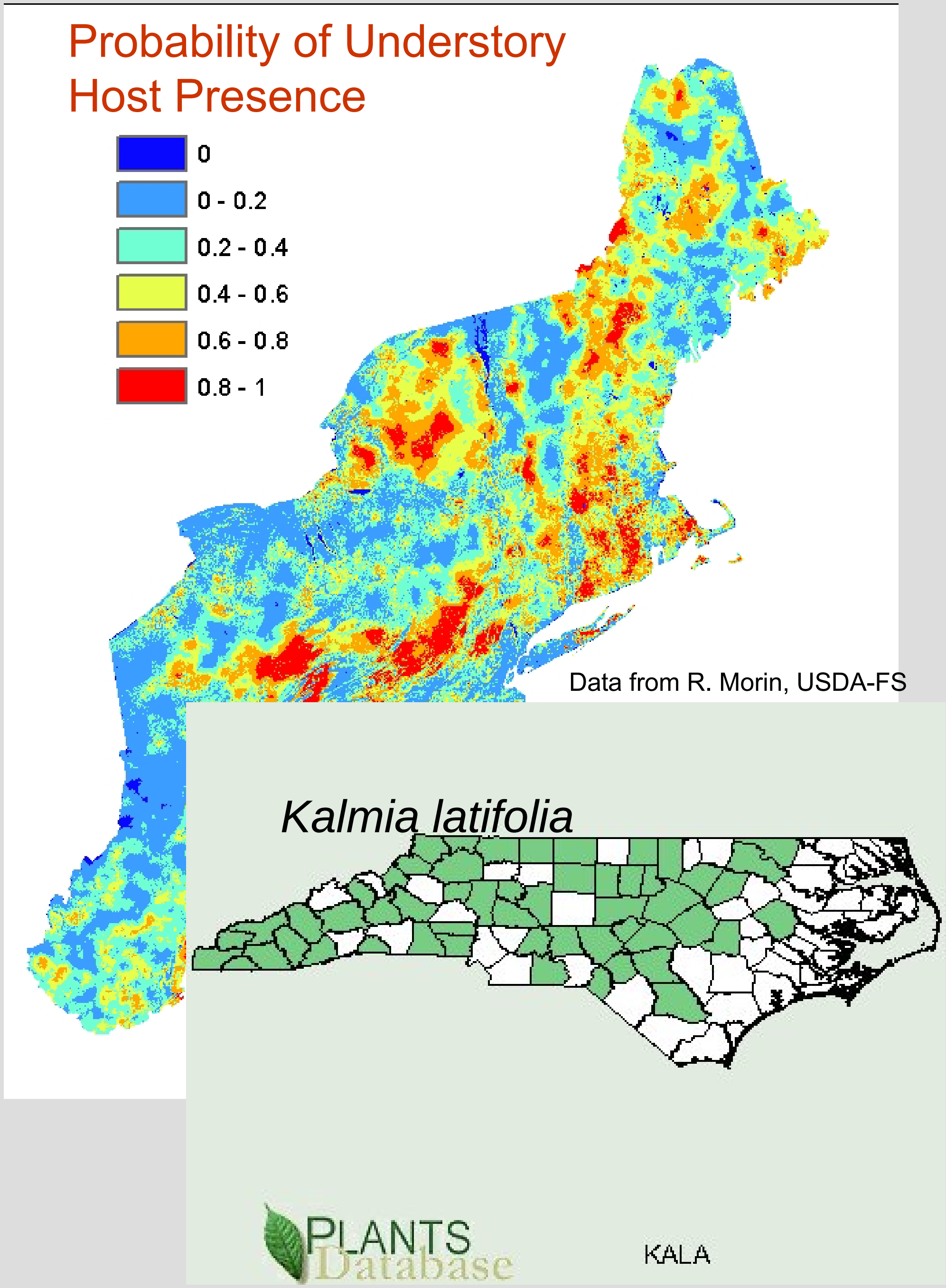
Hosts



- Generated by kriging of FIA plot basal area data - 1 km^2 cells
- High-frequency pattern to FIA data, so not well predicted by smooth interpolators (poor RMSE for validation or cross-validation)

What About Understory Hosts?

- Already have estimate of understory hosts for NE U.S.
  - Again, kriged from plot data…
- Archival understory data also exists for SE U.S.
  - Strange format
  - Don’t know how many plots
- Given limitations of interpolated FIA data…
  - County-level distribution maps from PLANTS national database, other sources

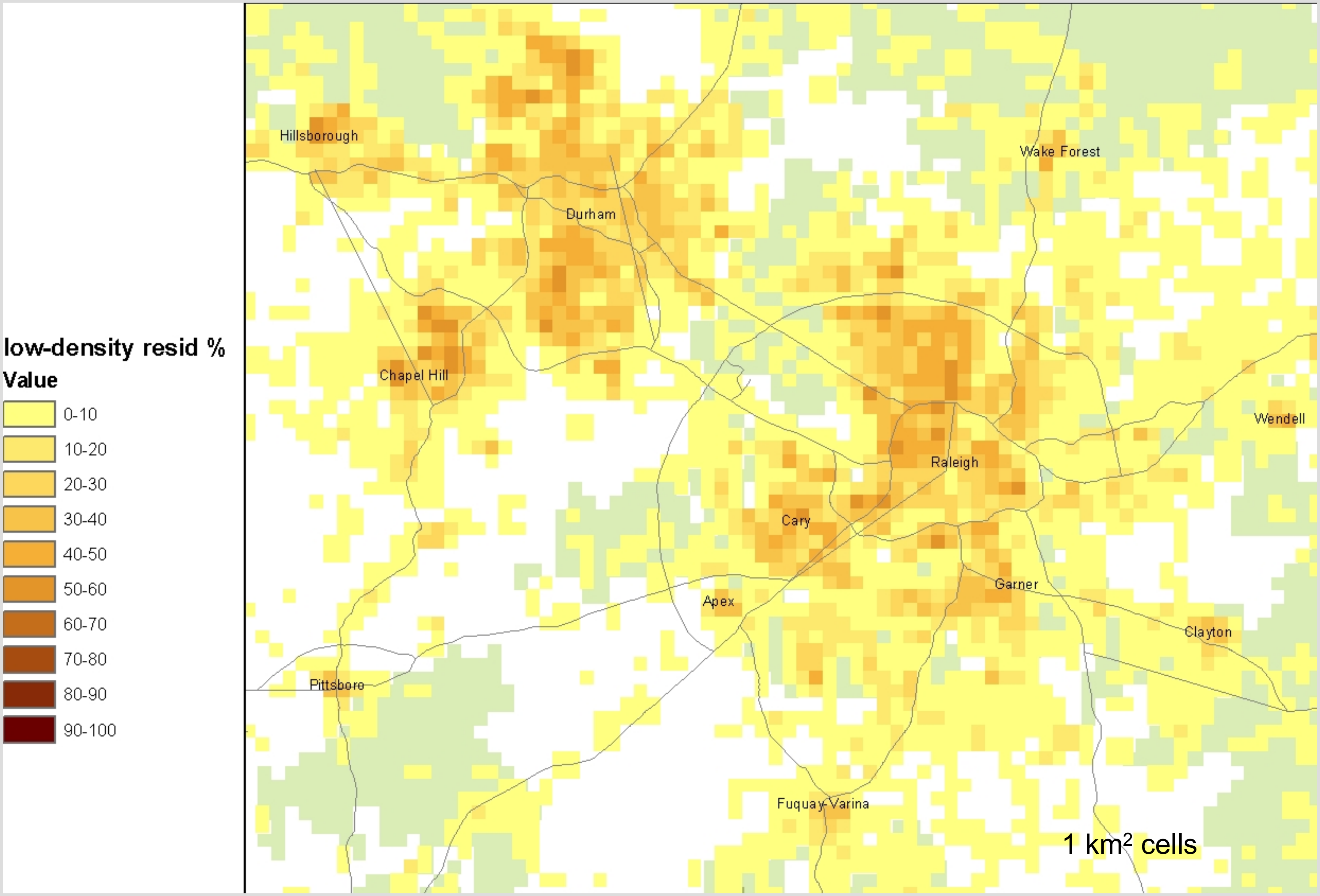


Data from R. Morin, USDA-FS

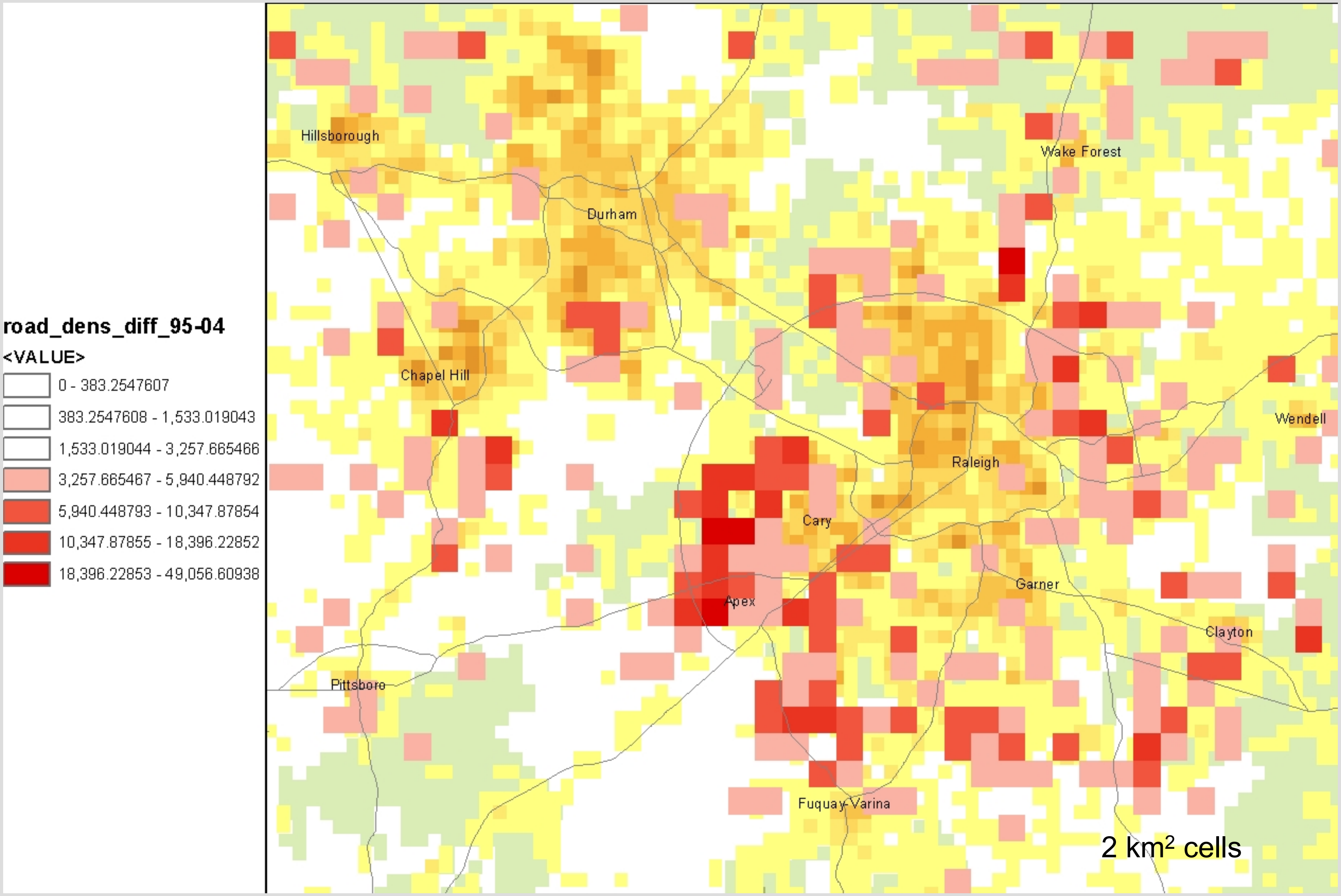
Pathways

- Number of potential areas for refinement
- For example, how to get at landscapes where people plant (potentially infected) nursery stock?
  - Percentage of low density residential from land cover data
    - Existing neighborhoods where people may be planting
  - Changes in road density = areas of new construction
  - Nighttime lights expansion (Imhoff et al. 1997)
      - Broad estimate of suburbanization/expansion

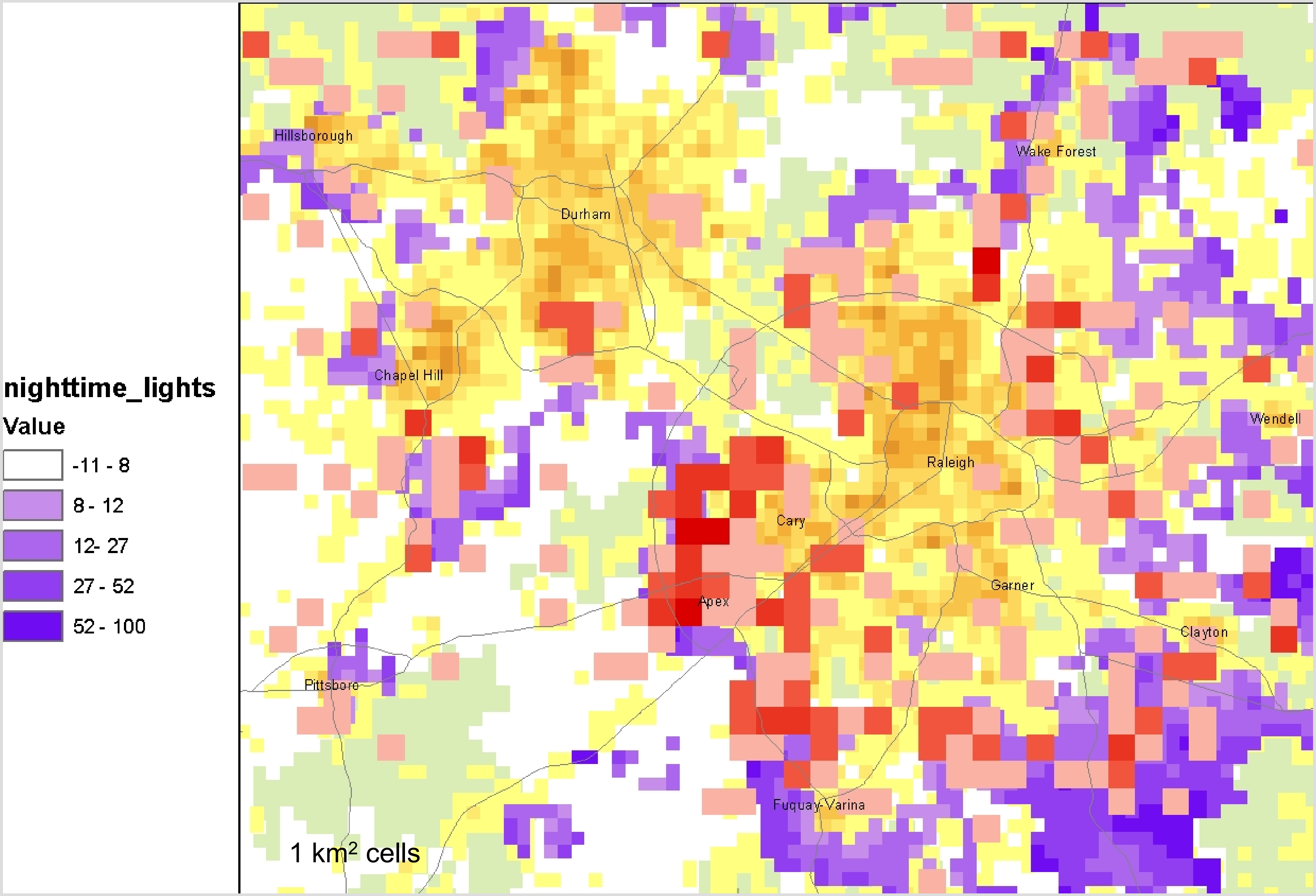
Percentage of Low-Density Residential Land Cover (from NLCD)



Change in Road Density 1996-2005



Nighttime Lights Expansion (Using 1992 and 2000 Imagery)

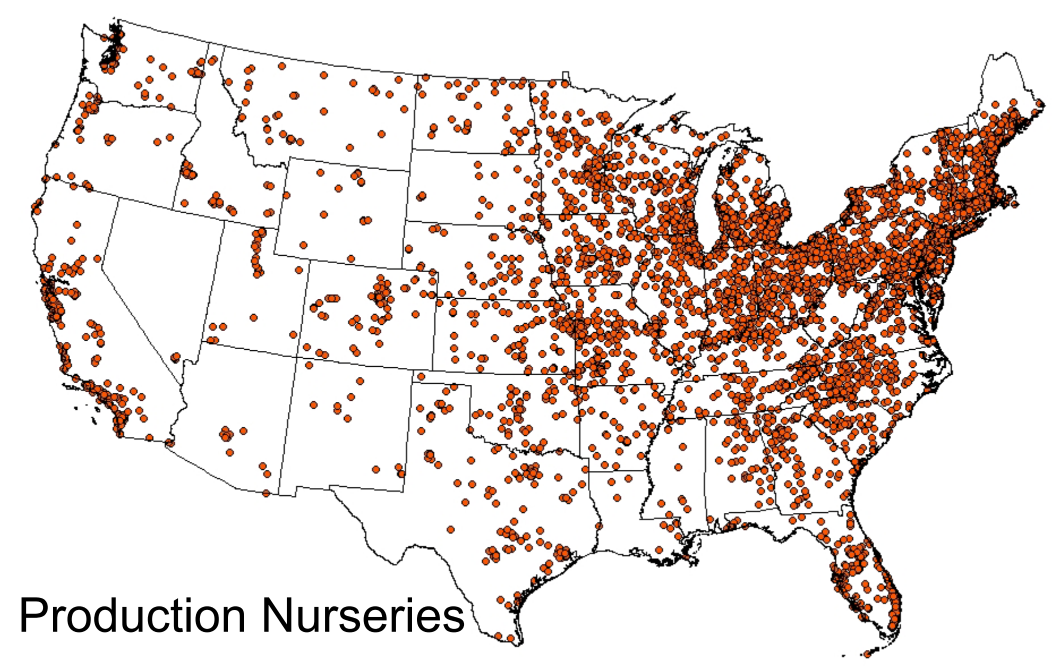


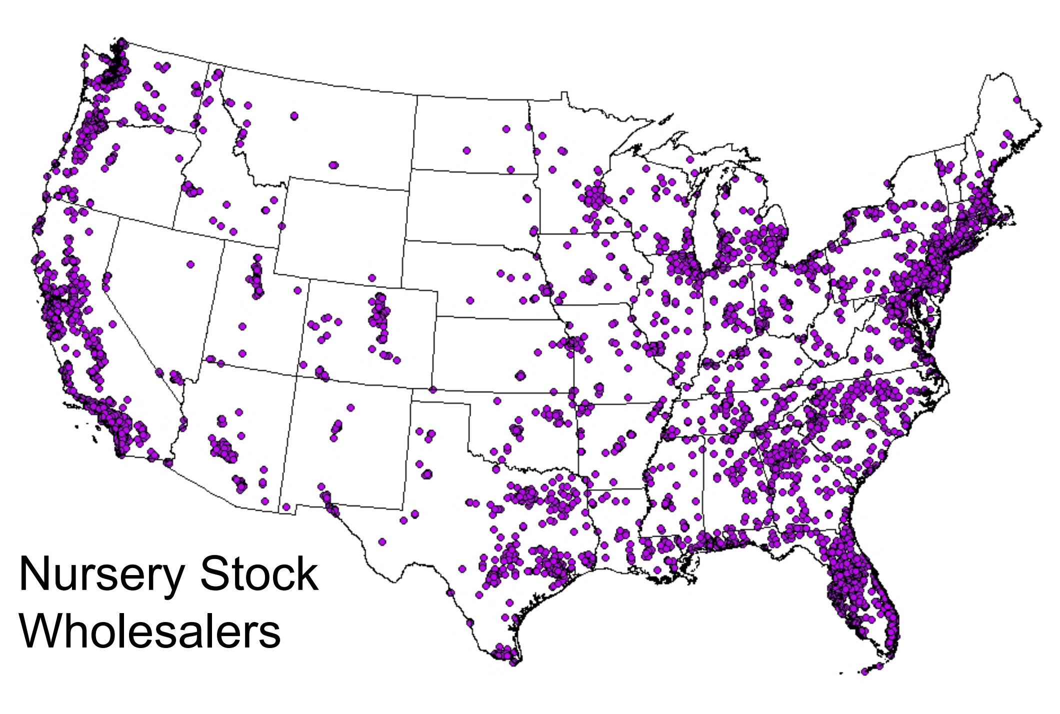
Spread via Roads

- Freight Analysis Framework
  - Volume of cargo moved by trucks
  - Nursery stock movement

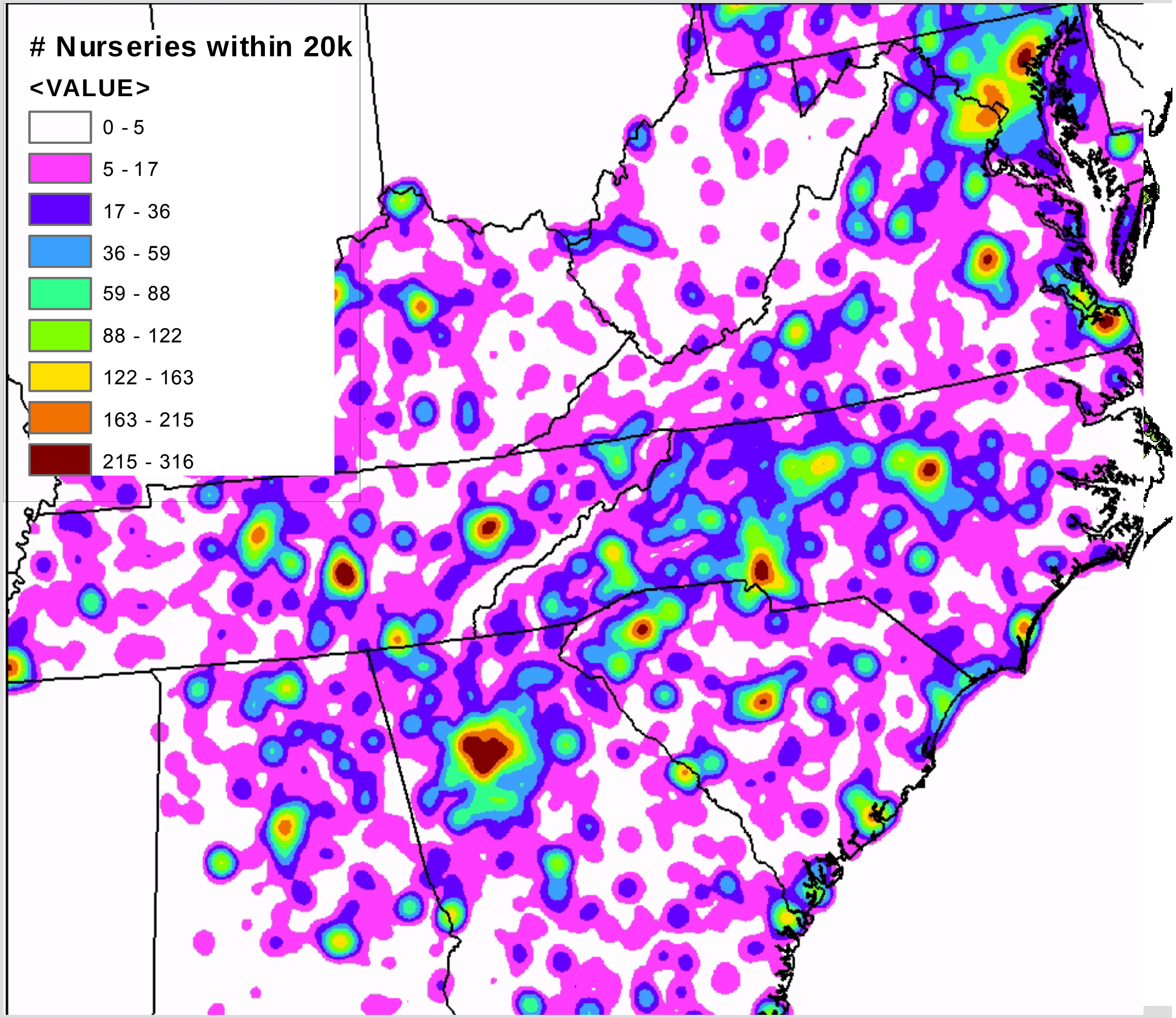
Nurseries

- ReferenceUSA database
  - 13 million U.S. businesses
  - Have geographic coordinates
  - Search for primary descriptions
- Comprehensive lists



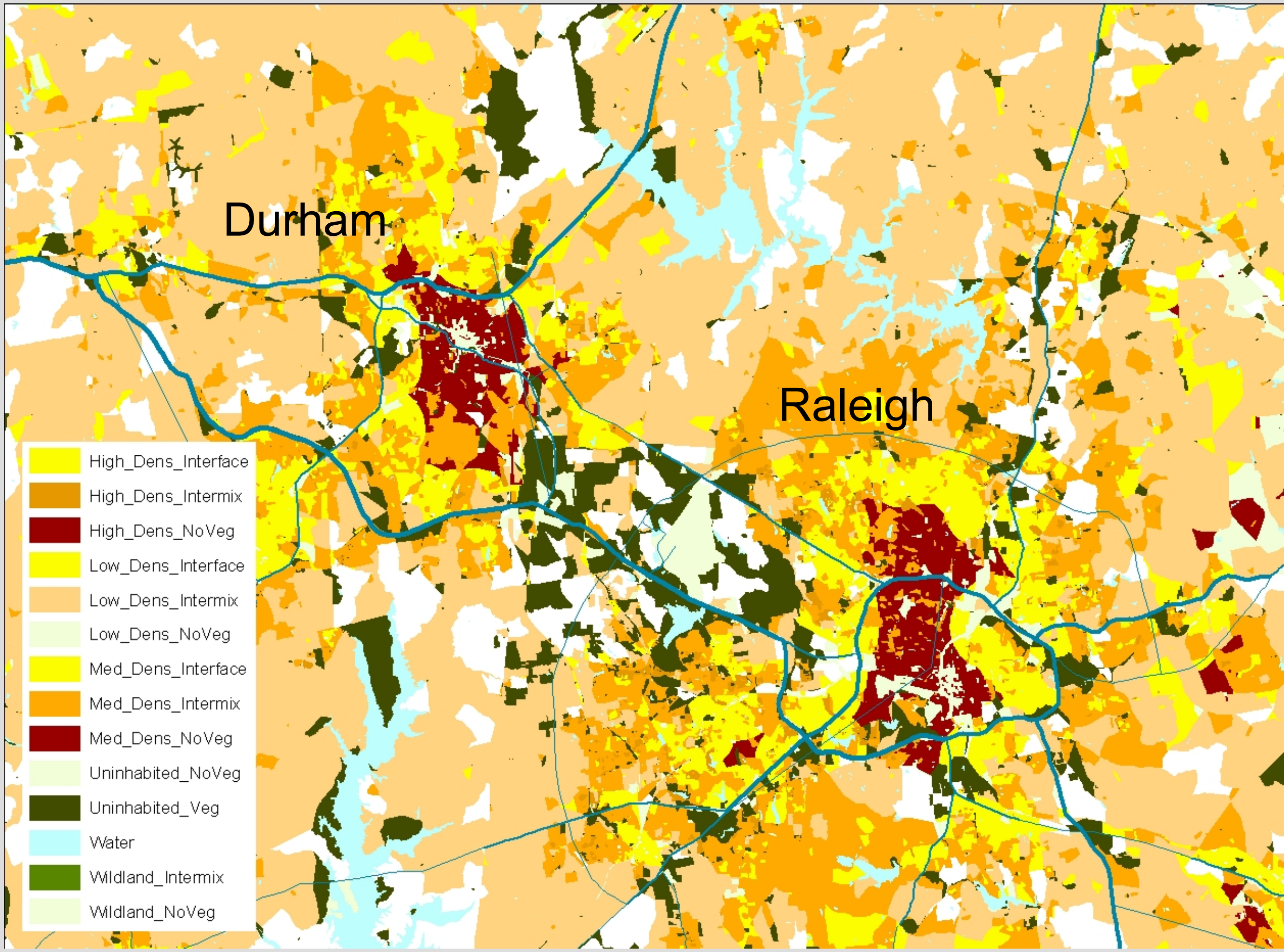


- When add retail locations, including home improvement centers, data layer gets quite large
- Can calculate grids of number of nursery locations within distance of a point
  - # possible exposures to (potentially) positive nurseries



Other Pathway Factors

- Housing / population density
- Road proximities
- Wildland-urban interface
  - Compiled for fire risk, but relevant for *P.ramorum*, other insects and pathogens



II. SOD Risk Modeling

- Build cost surface(s) incorporating all of these spatial data layers
  - Ideal spatial resolution?
- Model SOD movement after hypothetical introduction
  - Cellular automaton approach
  - Transition probabilities = interaction between cost and infection rate

SOD Risk Modeling (cont.)

- Many uncertainties, so try several different scenarios
  - Test range of infection rates (low, moderate, high)
  - Different cost surfaces
- Through repeated runs, can develop per-pixel risk ratings