QUANTIFYING RURAL REPORTING BIASES IN THE SPC TORNADO DATABASE

Corey K. Potvin[1,2], Chris Broyles[3], Patrick S. Skinner[1,2], and Harold E. Brooks[2]

[1]*Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma*; [2]*NOAA/OAR National Severe Storms Lab*
[3]*NOAA/NWS Storm Prediction Center, Norman, OK*

MOTIVATION

- Major reporting biases exist in SPC tornado database
- Attempts to correct these errors have been limited in scope
- We are developing a much more comprehensive approach
- Resulting improvements in tornado climatology will improve:
  - Tornado risk and economic loss models
  - Detection of climate change
  - Detection of mesoscale tornado maxima ("mini alleys")
  - Diagnosis of effects of surface characteristics (e.g., elevation roughness, land-use type) on tornadoes
  - Analysis of relationships between tornado length/width/rating
- Current focus: rural (under-) reporting bias, **1975-2014**

GENERAL APPROACH

- Calculate tornado counts within 10-km grid cells, then apply Gaussian filter (σ = 10 km) to reduce sampling error
- Estimate fraction of reported tornadoes within each grid cell by assuming the actual count is that within a 10-km cell centered on the nearest 100K+ (population) city or WFO
- Model the reporting fractions using polynomial regression with subsets of the following explanatory variables (**Fig. 1**):
  - Distance from nearest 100K+ city or WFO (**C**)
  - Distance from nearest 5K+ city or interstate (**c**)
  - 10-km population density (**D**)
- Perform 20 regressions (bootstrap fraction estimates), then bootstrap resulting distributions (*n*=10,000) → CIs
- Use regressed (predicted) fractions to correct tornado counts

PRELIMINARY CONCLUSIONS

- Regressing on **C**, **c** or **C**, **D** works well; captures much of effect of 3rd variable (**Fig. 2**)
- Variance dominated by sampling error ($R^2 < 0.25$)
- Estimated 44 % of 1975-2014 tornadoes unreported
- Reporting bias has decreased but remains significant (**Fig. 3**)
- Long-track tornadoes much less likely to be missed (**Fig. 4**)
- Removing bias reduces correlation between tornado frequency and population centers (**Fig. 5**); however, mesoscale maxima ("mini alleys"; Broyles and Crosbie 2004) remain, and early tests indicated they're unlikely to arise by chance (not shown)

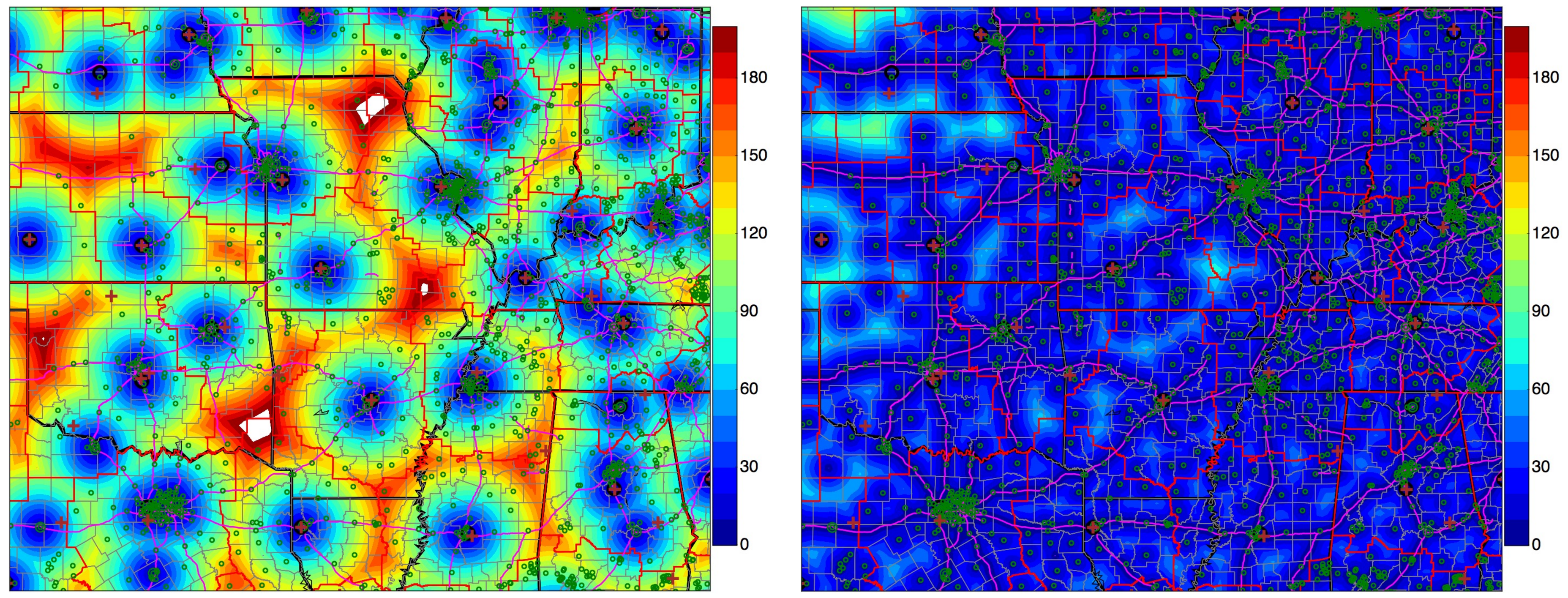

Figure 1. Distance (km) from nearest (left) 100K+ city or WFO and (right) 5K+ city or interstate. Large (small) green circles = 100K+ (5K+) cities, black circles = WFOs, crosses = WSR-88Ds, magenta = interstates, red = NWS County Warning Area boundaries.

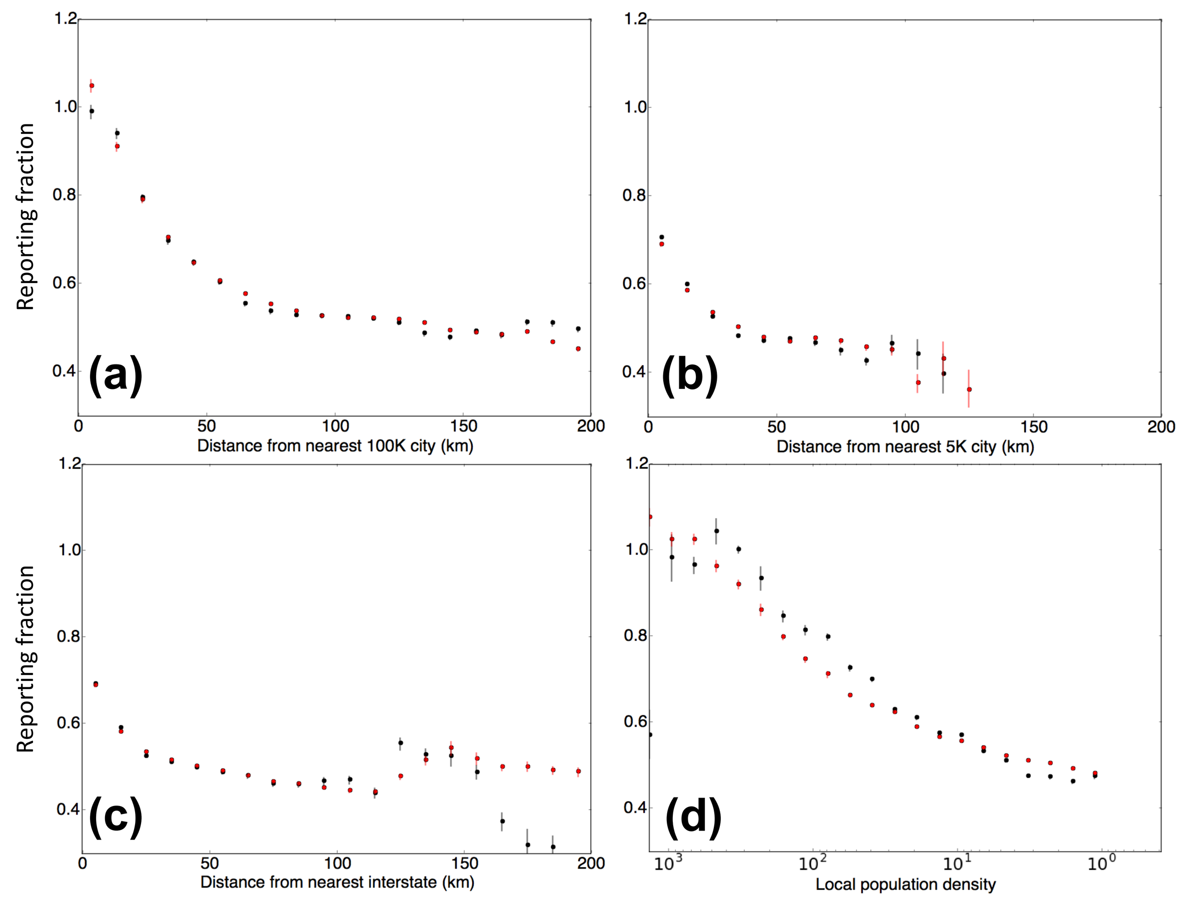

Figure 2. Reporting fractions as functions of distance from nearest 100K+ city (a), 5K+ city (b), and interstate (c), and of population density (d). Black = (median) raw estimates, red = regression on **C** & **c** with 4th-order polynomial. Bars indicate 90 % confidence intervals.

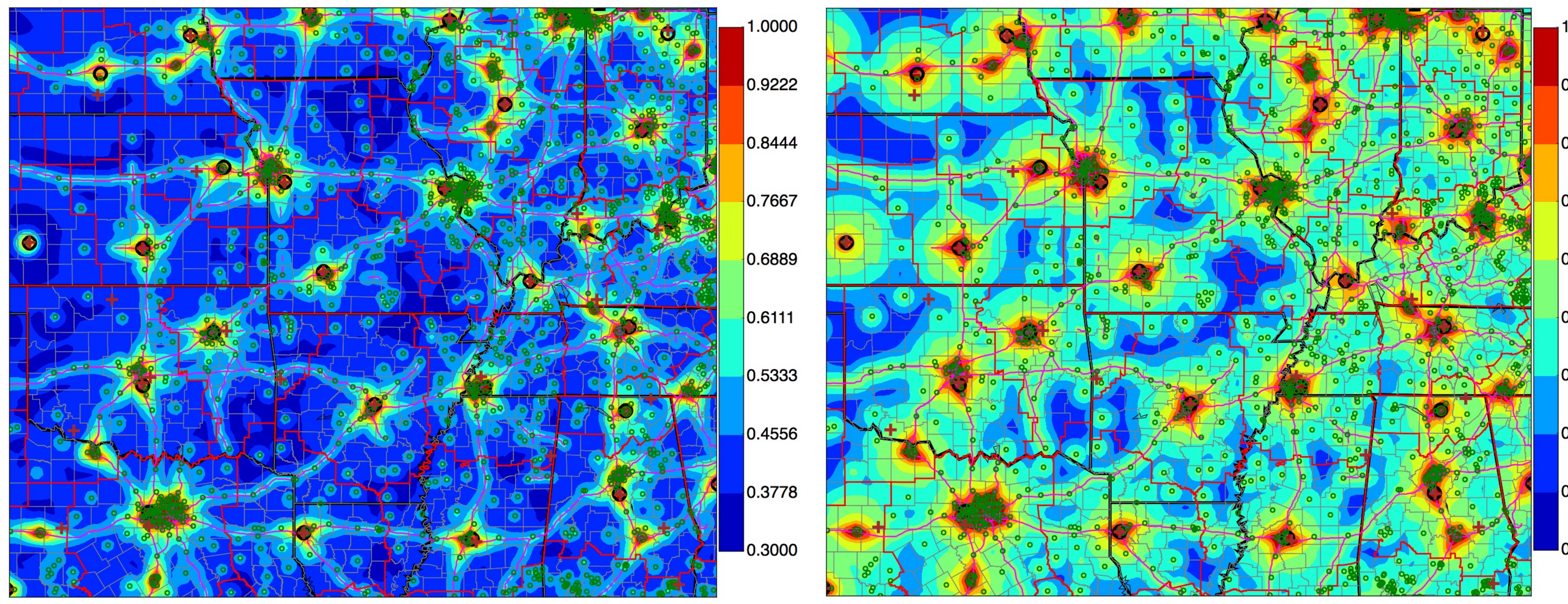

Figure 3. Regressed reporting fractions for (left) 1975-1994 and (right) 1995-2014 tornadoes. Estimated 51% (38%) of 1975-1994 (1995-2014) tornadoes missed.

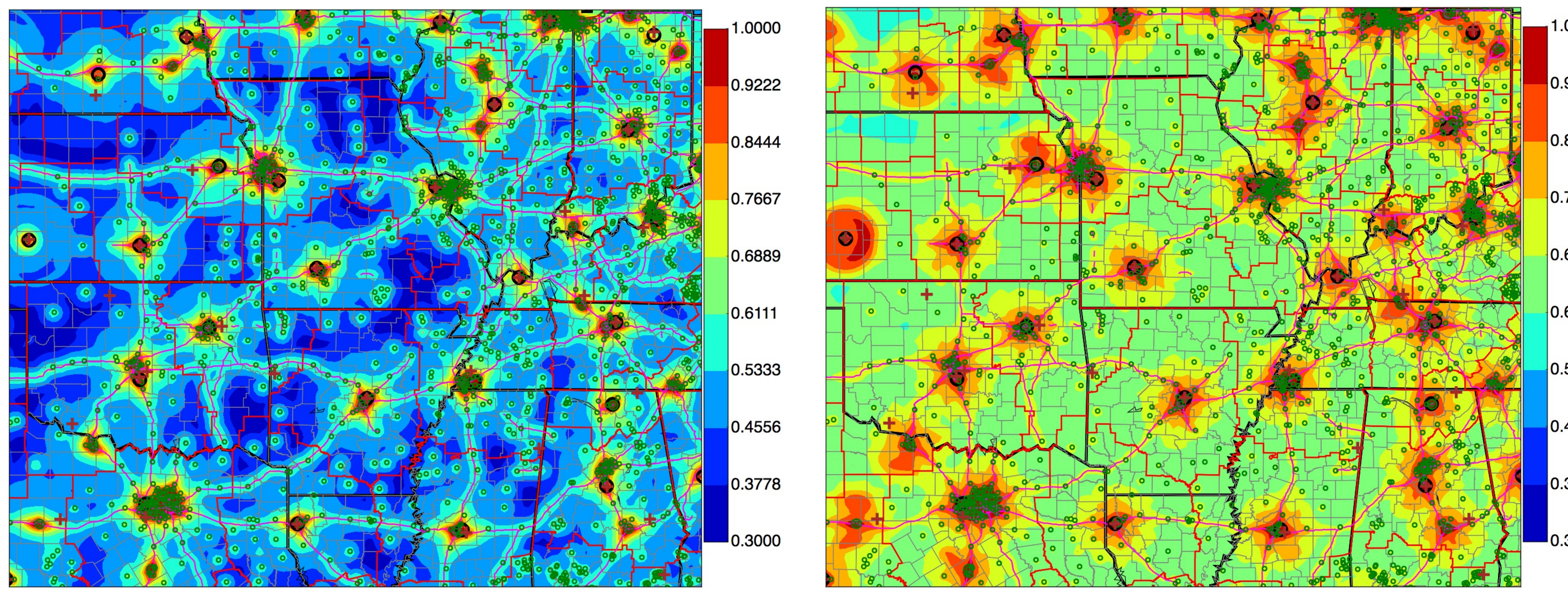

Figure 4. Regressed reporting fractions for (left) 0-5mi and (right) 10mi+ tornadoes. Estimated 47% (28%) of 0-5mi (10mi+) tornadoes missed.

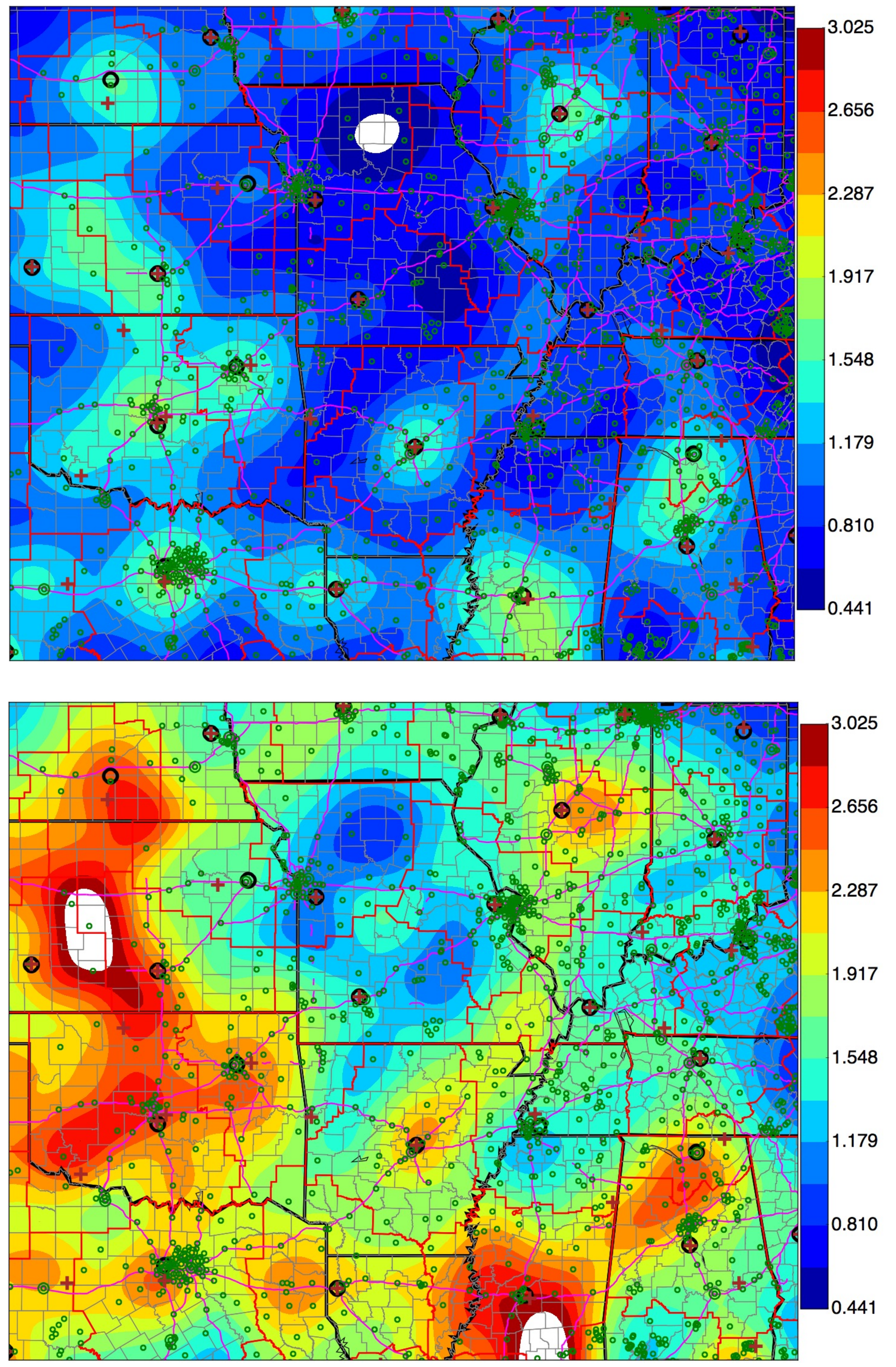

Figure 5. Raw (top) and corrected (bottom) tornado counts (1975-2014) per 10-km grid point, smoothed with a σ = 50 km Gaussian kernel.