MODELING AND CORRECTING REPORTING BIASES IN THE SPC TORNADO DATABASE

Corey K. Potvin\(^1,2,3\), Chris Broyles\(^4\), 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 Laboratory; \(^3\)School of Meteorology, University of Oklahoma; \(^4\)NOAA/NWS Storm Prediction Center

**MOTIVATION**

- Many tornadoes are unreported or mischaracterized, especially in rural areas, leading to serious biases in tornado climatologies
- Goal 1: Develop a comprehensive statistical model to estimate these biases
- Goal 2: Correct the biases to improve tornado climatologies

**METHODOLOGY**

- Use 1975-2014 SPC tornado database over most tornado-prone region of the U.S.
- Calculate tornado counts within 10-km grid cells (Fig. 1a)
- Demonstrate that tornado reporting rate (TRR) = 1 at grid cells within 10 km of 100K+ city or NWS Weather Forecast Office
- Develop & apply method to generate initial TRR estimates over entire analysis grid; very noisy due to sampling error (Fig. 1b)
- Train multivariable polynomial regression model to generate spatially smooth TRR estimates (Fig. 3) that preserve important geographical dependencies (Fig. 2)
- Divide reported tornado counts by modeled TRR to obtain bias-corrected tornado counts, then apply Gaussian smoother to mitigate sampling errors (Fig. 4)

**CONCLUSIONS**

- Only 45% of 1975-2014 tornadoes reported, implying ≈2400 annual U.S. tornadoes
- Under-reporting bias worse earlier in record, but remains large (Fig. 5)
- Bias increases with damage rating due to lack of rural damage indicators (cf. Fig. 3a, c); actual number of F/EF-3+ tornadoes estimated to be nearly three times that reported
- Much smaller bias for long-track tornadoes (cf. Fig. 3a, b), implying path length is better proxy for intensity than damage rating in rural areas
- Bias corrections are critical for identifying climate signals that may be masked by detrending; note differences between red, green curves in Fig. 6
- Mesoscale maxima persist in corrected tornado counts (Fig. 4b); "mini tornado alleys"?
- The basic methodology could be applied even in data-sparse parts of the world

---

**Fig. 1.** (a) Gridded 1975-2014 tornado counts; (b) initial TRR estimates. Circles = 100K+ cities (green), WFOs (black).

**Fig. 2.** Domain-mean initial (unregressed) TRR versus selected geographical variables. Smaller dots = fewer grid points.

**Fig. 3.** Modeled TRR for (a) all tornadoes, (b) 15+ mi path lengths, (c) F/EF-3+.

**Fig. 4.** Smoothed (a) reported and (b) bias-corrected all-tornado counts.

**Fig. 5.** Annual series of 5-year domain-mean TRR.

**Fig. 6.** 5-year reported (black), bias-corrected (red), and de-trended, renormalized reported (green) tornado counts.