Improving Estimates of U.S. Tornado Frequency by Accounting for Unreported and Underrated Tornadoes

Corey K. Potvin,[a,b] Chris Broyles,[c] Patrick S. Skinner,[d,a,b] and Harold E. Brooks[a,b]

[a] *NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma*
[b] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[c] *NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma*
[d] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*

ABSTRACT: Many tornadoes are unreported because of lack of observers or are underrated in intensity, width, or track length because of lack of damage indicators. These reporting biases substantially degrade estimates of tornado frequency and thereby undermine important endeavors such as studies of climate impacts on tornadoes and cost–benefit analyses of tornado damage mitigation. Building on previous studies, we use a Bayesian hierarchical modeling framework to estimate and correct for tornado reporting biases over the central United States during 1975–2018. The reporting biases are treated as a univariate function of population density. We assess how these biases vary with tornado intensity, width, and track length and over the analysis period. We find that the frequencies of tornadoes of all kinds, but especially stronger or wider tornadoes, have been substantially underestimated. Most strikingly, the Bayesian model estimates that there have been approximately 3 times as many tornadoes capable of (E)F2+ damage as have been recorded as (E)F2+ [(E)F indicates a rating on the (enhanced) Fujita scale]. The model estimates that total tornado frequency changed little over the analysis period. Statistically significant trends in frequency are found for tornadoes within certain ranges of intensity, pathlength, and width, but it is unclear what proportion of these trends arise from changes in damage survey practices. Simple analyses of the tornado database corroborate many of the inferences from the Bayesian model.

SIGNIFICANCE STATEMENT: Prior studies have shown that the probabilities of a tornado being reported and of its intensity, track length, and width being accurately estimated are strongly correlated with the local population density. We have developed a sophisticated statistical model that accounts for these population-dependent tornado reporting biases to improve estimates of tornado frequency in the central United States. The bias-corrected tornado frequency estimates differ markedly from the official tornado climatology and have important implications for tornado risk assessment, damage mitigation, and studies of climate change impacts on tornado activity.

KEYWORDS: Tornadoes; Climatology; Bayesian methods

1. Introduction

Some tornadoes go unreported, typically because they occur in a sparsely populated region and therefore escape notice. This underreporting bias introduces sharp spatial gradients in reported tornado counts near cities (e.g., Elsner et al. 2013), and the observed decrease in this bias with time produces a spurious upward trend in tornado frequency that may mask true shifts in tornado activity (e.g., Brooks et al. 2003). The decrease in the tornado underreporting bias results from several factors, including population increases, the installation of the Weather Surveillance Radar-1988 Doppler (WSR-88D) network during the early 1990s, more frequent storm surveys by the National Weather Service (NWS), and increasing numbers of storm spotters and chasers (e.g., McCarthy and Schaefer 2004). Strong tornadoes are more likely than weak tornadoes to be reported, but their intensity is systematically underrated in rural areas, where the maximum winds in a tornado are less likely to be "sampled" by available damage indicators (e.g., Doswell and Burgess 1988; Doswell et al. 2009; Wurman et al. 2021). Tornado path width is also likely systematically

underrated in rural areas because of the lack of damage indicators (Wurman et al. 2021), and it is reasonable to expect that tornado pathlength suffers the same bias. The expansion of the built landscape (Hall and Ashley 2008; Ashley et al. 2014; Ashley and Strader 2016; Strader and Ashley 2015; Strader et al. 2017) is likely reducing both the underreporting and underrating biases with time. Estimating and mitigating the effects of tornado reporting biases and their trends is crucial for a range of endeavors including assessing tornado risk (e.g., Schaefer et al. 1986; Coleman and Dixon 2014; Widen et al. 2013; Strader et al. 2016); developing economic loss models and performing cost–benefit analyses of tornado damage mitigation strategies (e.g., Simmons et al. 2015; Grieser and Terenzi 2016; Romanic et al. 2016); and investigating how tornado activity is modulated by climate (e.g., Lee 2012; Barrett and Gensini 2013; Brooks et al. 2014; Tippett et al. 2015; Allen et al. 2015; Guo et al. 2016; Cook et al. 2017; Strader et al. 2017; Trapp and Hoogewind 2018; Childs et al. 2020; Nouri et al. 2021).

Bayesian inference provides a powerful framework for estimating true tornado frequency in the face of the aforementioned reporting biases and the large sampling errors that arise from the brevity of the tornado record. Bayesian

tornado frequency models are hierarchical, meaning they comprise a series of connected submodels (e.g., Wikle and Anderson 2003). They normally include a reporting bias model that is often a function of population density, a spatial autocorrelation model that accounts for heterogeneity in the true long-term tornado climatology, and a spatial process model that accounts for the random clustering of tornadoes in space and time. This random clustering occurs even with a completely spatially random process, and is amplified in tornado count data by, for example, tornado outbreaks (e.g., Elsner and Widen 2014). The hierarchical model may also incorporate one or more climate covariates to further constrain the analysis (e.g., Wikle and Anderson 2003; Elsner and Widen 2014; Cheng et al. 2016; Nouri et al. 2021), as well as time-dependent terms that account for trends and interannual variability in tornado occurrence (e.g., Elsner et al. 2016). The model is fit to reported county-level or gridded tornado counts to produce probabilistic estimates of reporting bias, true tornado frequency, and/or regression coefficients describing the relationship between tornado frequency and climate covariates.

Potvin et al. (2019, hereinafter P19) used a Bayesian model to estimate the tornado reporting rate (TRR)—the proportion of all occurring tornadoes that were reported—and thereby the true tornado frequency over the central United States. The Bayesian model was based on those presented in Anderson et al. (2007) and Elsner et al. (2016) but included important modifications to mitigate a solution nonuniqueness (i.e., parameter confounding) problem that may have degraded estimates of TRR and true tornado frequency in previous studies. The solution nonuniqueness arises because a given reported tornado frequency can be explained equally well by a high true frequency and low TRR as by a low true frequency and high TRR. Statistical models for estimating the true tornado frequency should therefore include constraints that properly "anchor" the analysis. The P19 model addressed this problem by assuming that TRR = 1 (i.e., all tornadoes are reported) in grid cells where population density exceeds a prescribed threshold. The model estimated that only 45% of tornadoes over the central United States were reported within the period 1975–2016.

The current study uses a statistically more rigorous version of the P19 model to assess how TRR and true tornado frequency vary with tornado damage rating, pathlength, path width, and time over the period 1975–2018. This new analysis allows us to assess how reporting bias differs between weak and potentially destructive tornadoes, shorter- and longer-track tornadoes, and narrower and wider tornadoes. Large, statistically significant trends in frequency are found for certain ranges of tornado intensity, track length, and width, but it is unclear how to distinguish between true changes in tornado characteristics and secular trends arising from changes in tornado damage assessment practices, including the switch to reporting maximum instead of average damage path width in 1995 and the implementation of the enhanced Fujita (EF; Wind Science and Engineering Center 2006) scale in 2007 (e.g., Edwards at al. 2021). The stratified tornado frequency estimates from our analysis provide a comprehensive view of the true tornado frequency over the central United States.

2. Methods

a. Gridded tornado counts and population densities

We divide our 1800 km × 1800 km central U.S. analysis domain (Fig. 1a) into 10-km grid cells as in P19. This analysis grid is nearly identical to that in P19.[1] As in P19, we exclude the northeastern 600 km × 600 km corner of the domain from the analysis since many of the grid cells therein are located within the Great Lakes. We compute the reported tornado count N within each 10-km grid cell using the recorded start points of the tornado reports from the Storm Prediction Center Severe Weather Database.[2] Prior to doing so, however, we identify and remove duplicate tornado reports by using the method of Elsner et al. (2016). A separate N is computed for each of 10 tornado categories examined in this study (Table 1). The breakpoints between categories were selected to maintain similar domain-total N between corresponding categories of the three examined tornado attributes. For example, the largest category for each attribute—(E)F0 (rating), 0–1 mi (length; 1 mi ≈ 1.6 km), and 0–50 yd (width; 1 yd ≈ 0.9 m)—contains 15 356, 16 016, and 17 072 tornadoes, respectively. Herein, (E)F indicates a rating using either the enhanced or original Fujita scale.

We compute the population density (PD) within each 10-km grid cell in our analysis domain by averaging over the PD of the encompassed 1-km grid cells. The 1-km PD are obtained from U.S. census data. The midpoint year of the analysis period determines which of the three available U.S. censuses (1990, 2000, or 2010) is used to compute PD. For the full-period analyses, valid 1 January 1975–31 December 2018, the midpoint year is considered to be 1996 (the precise midpoint of 1975–2018 is 1996.5). For 5-yr analyses valid within the full analysis period, the midpoint year ranges from 1977 to 2016. For midpoint years prior to 1990 or following 2010, we compute PD from the 1990 or 2010 census, respectively. To compute PD for midpoint years 1991–2009, we interpolate between the 1990 and 2000 censuses or between the 2000 and 2010 censuses.

b. Bayesian hierarchical model

Following previous studies (e.g., Anderson et al. 2007; Elsner et al. 2016), we use a Bayesian hierarchical model to simultaneously estimate TRR and the expected actual tornado counts λ. The primary novelty of the Bayesian hierarchical models used in this study and in P19 is the mitigation of parameter confounding (solution nonuniqueness). The present model improves upon the P19 model, principally by incorporating a sophisticated conditional autoregressive (CAR) model. We implement the Bayesian model using the Python pymc3 probabilistic programming framework (Salvatier et al. 2016).

To accommodate the known overdispersion of tornado occurrence data (e.g., Elsner and Widen 2014), reported tornado counts N_i within each 10-km grid cell i are assumed to

---

[1] The analysis grid origin is shifted 1.15° westward of that in P19 to include more of the Great Plains.
[2] In principle, these data are identical to the National Centers for Environmental Information Storm Data, especially for events that occurred more than 1–2 years ago (P. Marsh 2020, personal communication).
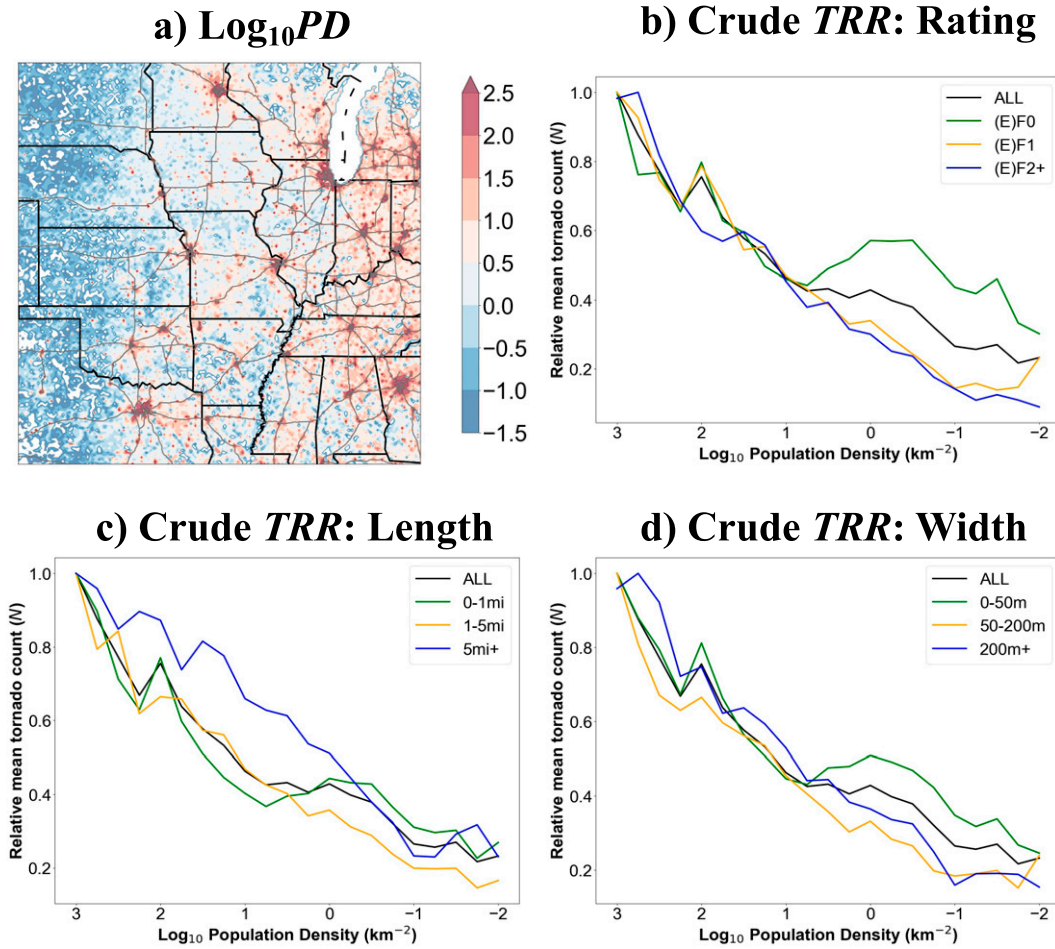
a) Log₁₀PD

b) Crude *TRR*: Rating

c) Crude *TRR*: Length

d) Crude *TRR*: Width

FIG. 1. (a) $Log_{10}PD$ over the analysis domain, with interstate highways in gray. Also shown are simple analyses of normalized mean N vs $log_{10}PD$, stratified by (b) damage rating, (c) damage pathlength, and (d) damage path width.

follow a negative binomial distribution NegBin. The pymc3 module implements NegBin as a Poisson distribution whose rate parameter, rather than being a constant, follows a gamma distribution with prescribed mean μ and shape parameter $\alpha = \alpha_{\text{disp}}$. Thus, we model the reported tornado counts as

$$N_i \sim \text{NegBin}(\mu, \alpha = \alpha_{\text{disp}}).$$

The expected reported tornado counts $\mu_i \equiv \langle N \rangle_i$ are the product of the (unknown) λ_i and TRR_i:

$$\mu_i = \lambda_i \text{TRR}_i.$$

Therefore, we can express our reported tornado counts model as

TABLE 1. Domainwide mean N per year (averaged over the 1975–2018 analysis period), and mean and credible interval of the domainwide posterior TRR and λ yr^{-1}.

Tornado category	Description	N yr^{-1}	Mean posterior TRR (90% CI)	Mean posterior λ yr^{-1} (90% CI)
All	All tornadoes	680	0.45 (0.42–0.48)	1544 (1440–1648)
(E)F0	Rated (E)F0	349	0.53 (0.47–0.57)	691 (626–756)
(E)F1	Rated (E)F1	220	0.39 (0.35–0.42)	562 (510–619)
(E)F2+	Rated ≥ (E)F2	111	0.34 (0.25–0.42)	329 (248–428)
0–1 mi	Damage pathlength ≤ 1 mi	364	0.42 (0.38–0.46)	897 (809–996)
1–5 mi	Damage pathlength > 1 mi and ≤ 5 mi	178	0.45 (0.39–0.49)	403 (360–451)
5 mi+	Damage pathlength > 5 mi	137	0.56 (0.48–0.62)	245 (215–282)
0–50 yd	Damage path width ≤ 50 yd	388	0.47 (0.42–0.50)	859 (791–943)
50–200 yd	Damage path width > 50 yd and ≤ 200 yd	201	0.42 (0.37–0.46)	481 (424–532)
200 yd+	Damage path width > 200 yd	92	0.35 (0.25–0.45)	268 (194–356)

$$N_i \sim \text{NegBin}(\mu_i = \lambda_i \text{TRR}_i, \alpha = \alpha_{\text{disp}}). \quad (1)$$

We model TRR as an exponential function of a transformed population density covariate P (e.g., Anderson et al. 2007; Elsner et al. 2016). As in P19, we mitigate confounding between λ and TRR by transforming PD such that TRR = 1 for P exceeding a threshold P_{max}:

$$\text{TRR}_i = \text{TRR}_{\text{min}} + (1 - \text{TRR}_{\text{min}}) \exp(-\beta_1 P_i) \quad \text{and} \quad (2)$$

$$P_i = \begin{cases} 0 & \text{if } \log_{10}(\text{PD}_i) \geq P_{\text{max}} \\ P_{\text{max}} - \log_{10}(\text{PD}_i) & \text{if } \log_{10}(\text{PD}_i) < P_{\text{max}} \end{cases}. \quad (3)$$

Rather than prescribing P_{max} as in P19, we include it as a parameter in the TRR model. This modification removes the need to empirically tune P_{max}, which we will show varies considerably with tornado type (e.g., short vs long path; section 3b) and allows the uncertainty in this parameter to be represented in the posterior distributions of TRR and λ. To prevent TRR from approaching zero in very rural areas (which we judge to be unrealistic), we also introduce the model parameter TRR_{min}, which is the lower bound of TRR. This is another improvement over the P19 model, which prescribed TRR_{min} on the basis of trial and error; as with P_{max}, TRR_{min} varies considerably with tornado attributes (section 3b). The β_1 parameter in Eq. (2) controls the rate at which TRR decreases with P.

The expected actual tornado counts λ_i are modeled as

$$\lambda_i = \exp(\beta_0 + \omega_i), \quad (4)$$

where β_0 is the domain-mean λ and ω_i captures spatially correlated random residuals, which is necessary to account for the spatial heterogeneity in λ. In maps of λ shown later, a Gaussian smoother has been applied to mitigate the influence of tornado overdispersion (appendix A) and thereby better reveal the underlying tornado climatology. No smoothing is applied to nonspatial representations of λ.

CAR models are often used to account for residual spatial autocorrelations in natural data, which can otherwise lead to biased and overconfident parameter posteriors. CAR models, however, can induce confounding between predictor variables and the spatial process (e.g., Reich et al. 2006), potentially leading to biased, underconfident posteriors for fixed-effect parameters. This is a substantial danger in the present application, where we seek to minimize the aliasing between TRR and λ, since spatial correlations in N can arise from spatial autocorrelations in both actual tornado counts and population density. To model ω_i, therefore, we adopt the restricted spatial regression (RSR; Hughes and Haran 2013) model, which mitigates this parameter aliasing by enforcing orthogonality between the spatial process and potentially confounding covariates. The RSR is a restricted version of the intrinsic CAR (ICAR; Rue and Held 2005) model in that it attempts to estimate only spatial structure that is not present in predictive covariates, and on sufficiently large scales to be meaningful, as judged by the analyst. This dimensional reduction further motivates the RSR approach by making it much more

computationally efficient than traditional CAR models, which can be prohibitively expensive on large analysis grids like the one used herein (32 400 grid cells). Our RSR model implementation is described in appendix B. Our use of an RSR model is a statistically rigorous alternative to the P19 strategy of dividing the analysis domain into subregions within which λ is held uniform.

The priors for each model parameter are (see appendix B for RSR model priors)

$$\alpha_{\text{disp}} \sim \text{HalfNormal}(10\,000); \quad \beta_0 \sim \text{Normal}(0, 10\,000);$$
$$\beta_1 \sim \text{LogNormal}(0.5, 4); \quad P_{\text{max}} \sim \text{Normal}(6.5, 0.25); \text{ and}$$
$$\text{TRR}_{\text{min}} \sim \text{Uniform}(0, 1)$$

where Normal denotes the normal distribution with specified μ and variance σ^2, HalfNormal denotes the half-normal distribution with specified σ^2, LogNormal denotes the lognormal distribution with specified μ and σ^2, and Uniform denotes the continuous uniform distribution with specified lower and upper limits. While most of the priors are noninformative, we achieved more reliable Markov chain Monte Carlo (MCMC) sampling convergence using the narrow P_{max} prior given above. In all of our experiments, we generate four chains of 1000 MCMC samples using the no-U-turn sampler (Hoffman and Gelman 2014) and discard the first 500 samples of each chain. Parameter convergence is confirmed through direct inspection of the MCMC traces and posterior distributions. Parameter uncertainty estimates shown in the paper are the 90% Bayesian credible intervals (CI) computed from the MCMC traces.

c. Estimating temporal variability

Following Elsner et al. (2016), we tested several expansions of our Bayesian model to account for temporal variability in TRR and λ over the full analysis period. To capture temporal trends in TRR and their interactions with population density, we expanded the TRR model as follows:

$$\text{TRR}_{i,j} = \{\text{TRR}_{\text{min}} + (1 - \text{TRR}_{\text{min}}) \\ \times [-\beta_1 P_{i,j} + \beta_3 P_{i,j}(y - y_0)] \exp[\beta_2(y - y_0)]\}, \quad (5)$$

where j is the year index, $\beta_2 \sim \text{Normal}(0, 10\,000)$; $\beta_3 \sim \text{Normal}(0, 10\,000)$; and $y_0 = 1996$. To account for interannual variability in domainwide tornado frequency, we expanded the λ model as follows:

$$\lambda_{i,j} = \exp(\beta_0 + \omega_i + \upsilon_j), \quad (6)$$

where $\upsilon_j \sim \text{Normal}(0, 1/\tau_j)$, and $\tau_j \sim \text{LogNormal}(0, 4)$.

As discussed in section 4b, we ultimately found this time-dependent model to be inadequate for estimating the evolution of TRR and λ. Given the limitations of the time-dependent model, we adopted an alternative approach to estimating temporal variability in TRR and λ: running the time-independent model (with one modification, described in section 4a) for each 5-yr subinterval of the full (1975–2018) analysis period. We use the Hamed and Rao (1998) modification of the nonparametric Mann–Kendall test (provided by the pyMannKendall package;

Hussain and Mamhud 2019) to assess the statistical significance of the linear trend in the point estimates of each parameter. This modified test accounts for the serial autocorrelation in the point estimates arising from their overlapping time intervals. We use the Theil–Sen method to produce point estimates of the slopes of the linear trends. The Theil–Sen method is nonparametric, it is insensitive to influential outliers, and its point estimates are insensitive to autocorrelation. As will be shown in section 4, statistically significant (defined herein as $p < 0.05$) trends exist in λ for many of the examined tornado categories. Attributing these trends, however, is complicated by changes in damage assessment practices during the analysis period (e.g., Agee and Childs 2014; Edwards et al. 2021).

d. Interpreting TRR estimates

When modeling occurrences of all tornadoes, TRR is simply the fraction of tornadoes that are reported (as in P19). This follows from our definition of $\langle N \rangle_{i,j}$ in Eq. (1) and our assumption that TRR = 1 for sufficiently large P [Eqs. (2) and (3)]. When modeling *subsets* of the tornado dataset based on tornado attributes that themselves are subject to bias, the meaning of TRR changes. To see why this is, consider that even if all tornadoes capable of producing (E)F2+ damage were reported, we would still expect fewer (E)F2+ tornado reports in rural areas because of the sparsity of damage indicators.

Thus, when we filter the tornado dataset on attributes like damage rating, pathlength, and path width, the relationship between TRR and P is now governed by both the detection rate of tornadoes actually satisfying the prescribed attribute condition *and* the sensitivity of reports of those attributes to P. If we adopt an interpretation of TRR that is strictly analogous to that for the full tornado dataset, then the TRR for tornadoes that actually satisfy the prescribed attribute condition is the probability of such tornadoes being both reported *and* correctly rated as meeting the prescribed criteria.

Even this improved interpretation of TRR, however, does not account for two potentially substantial effects. The first effect is the inflation of N and therefore of TRR caused by tornado attributes being systematically underestimated. For example, we know that (E)F0 tornado counts are inflated by systematic underrating of tornadoes with (E)F1+ winds. The second effect is the potential inflation of TRR, and corresponding deflation of λ, for certain tornado categories arising from our assumption that TRR = 1 in sufficiently populous areas (i.e., where $P \geq P_{max}$). This assumption does not account for the fact that damage indicators, particularly the types required to justify significant and violent damage ratings, are not uniformly densely distributed even in the most urban areas. It is plausible that substantial attribute reporting bias (especially for damage rating) exists even in the most populated areas of the analysis domain. Given these two effects, the model-predicted domainwide TRR for a given tornado category estimates the ratio of the number of reports in that category (i.e., N) to the number of tornadoes that would have been assigned to that category were $P \geq P_{max}$ throughout the analysis domain.

This clarified definition of TRR has two important implications for how to interpret the Bayesian model estimates. First, for tornado categories whose N may be underestimated even where $P \geq P_{max}$ [e.g., (E)F2+ tornadoes], the model-predicted λ should be viewed as a lower-limit estimate of the true λ. Second, the full-period TRR estimates for tornadoes satisfying prescribed attribute criteria (e.g., width exceeding a threshold) should not be sensitive to changes in damage assessment practices (e.g., the switch from reporting mean path width to maximum path width) unless those changes had different effects on reported tornado attributes in areas with $P \geq P_{max}$ versus $P < P_{max}$.

3. Full-period results

a. Simple analysis

Before presenting our Bayesian model results, we will demonstrate how much insight into tornado reporting bias can be gained from simple analysis of the U.S. tornado database. We first bin the N for each tornado category (section 2a) by $\log_{10}PD$, where the range and spacing of the $\log_{10}PD$ bins is chosen such that each bin contains > 50 grid cells (to mitigate sampling error). For each tornado category, we average the N within each $\log_{10}PD$ bin, then divide the domain-mean N for each bin by the maximum domain-mean N (which for most tornado categories corresponds to the highest bin). The resulting normalized N (Figs. 1b–d) provide crude estimates of TRR, under the assumption that TRR = 1 for $\log_{10}PD$ exceeding the lower bound of the $\log_{10}PD$ bin containing the maximum domain-mean N.

This simple analysis reveals that (E)F1 and (E)F2+ reports are strongly concentrated in urban areas, whereas (E)F0 reports are somewhat more evenly distributed between rural and urban areas (Fig. 1b). This result is consistent with the expectation that while (E)F1+ tornadoes are presumably more likely to be reported than (E)F0 tornadoes, they are less represented in the database due to the severity of the underrating bias in rural areas, which decreases the number of (E)F1+ reports and *increases* the number of (E)F0 reports. Similarly, 0–50-yd reports are more concentrated in rural areas (Fig. 1d), which suggests tornado width is also substantially underestimated. Conversely, 5-mi+ reports decrease much less sharply than 0–1-mi tornado reports with PD at high PD (Fig. 1c), which suggests that tornado pathlength is less severely underestimated than intensity and width.

Of course, this rudimentary analysis must not be interpreted too strictly. The simple procedure, for example, makes no provision for the large spatial gradients in the true tornado frequency across the central United States nor for the noise in the tornado record owing to its brevity. Neither does this procedure provide uncertainty estimates. Hence, rigorous statistical methods are required to produce accurate point and uncertainty estimates of TRR and λ. The simple analysis does, however, provide a plausibility check for the Bayesian model results presented in section 3b.

Another simple and valuable way to analyze the tornado report database is to compute mean tornado attributes as a
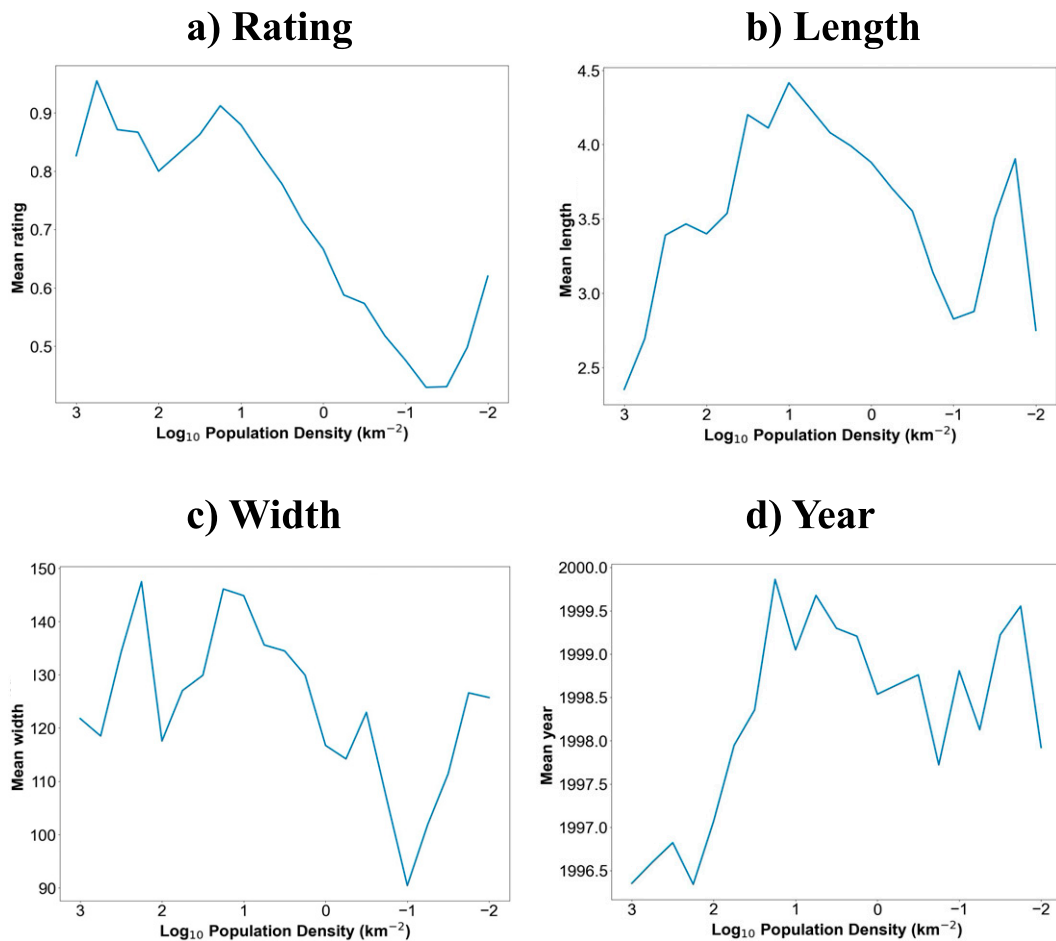
a) Rating

b) Length

c) Width

d) Year

FIG. 2. Simple analysis of mean reported tornado attributes vs $\log_{10}$PD: (a) damage rating, (b) damage pathlength, (c) damage path width, and (d) year.

function of PD (Fig. 2). In the absence of biases in reported tornado attributes, and assuming that stronger, longer-track, or wider tornadoes are progressively more likely to be observed than weaker, shorter-track, or narrower tornadoes as PD decreases (a detection bias), the mean reported attributes would increase as PD decreases. The mean reported values of all three attributes, however, sharply *decrease* as PD decreases from ~10 to ~0.1 km^{-2} (Fig. 2), suggesting that the effect of the progressively (as PD decreases) greater underestimation of tornado attributes exceeds the effect of the progressively larger probability of detecting higher-end tornadoes. This result together with the larger sensitivity of tornado frequency to PD for higher reported damage ratings and path widths (Figs. 1b,d) indicate serious underestimation of tornado attributes, especially intensity and width. Given the severity of these attribute biases, careful interpretation of model-predicted TRR is critical when analyzing cross-sections of the tornado database (section 2d). At high PD, the attribute and detection biases appear to approximately balance each other for tornado damage rating (Fig. 2a) and, to a lesser degree, path width (Fig. 2c). For pathlength, however, the detection bias seems to dominate the attribute bias

at high P (Fig. 2b), suggesting that tornado pathlength is better estimated than tornado intensity and path width in populous areas.

In addition to damage rating, pathlength, and path width, we also examine how mean tornado report *year* varies with population density (Fig. 2d). The mean tornado report year is very close to the midpoint of the analysis period, 1996.5, within the highest three PD bins, then sharply increases to ~1998–2000 for the remaining PD bins. This result suggests that TRR increased most substantially outside of high-population areas during the analysis period, which is consistent with our expectation that the full-period TRR is relatively close to unity in densely populated regions of the domain.

b. Bayesian analysis

We now analyze the Bayesian model posteriors obtained by training the model (section 2b) on the N for each tornado category. Domain-mean posterior distributions of TRR for each PD bin (Fig. 3) are consistent with the preliminary conclusions from our simple analysis of the tornado report database (Fig. 1). The TRR for (E)F0 tornadoes is less
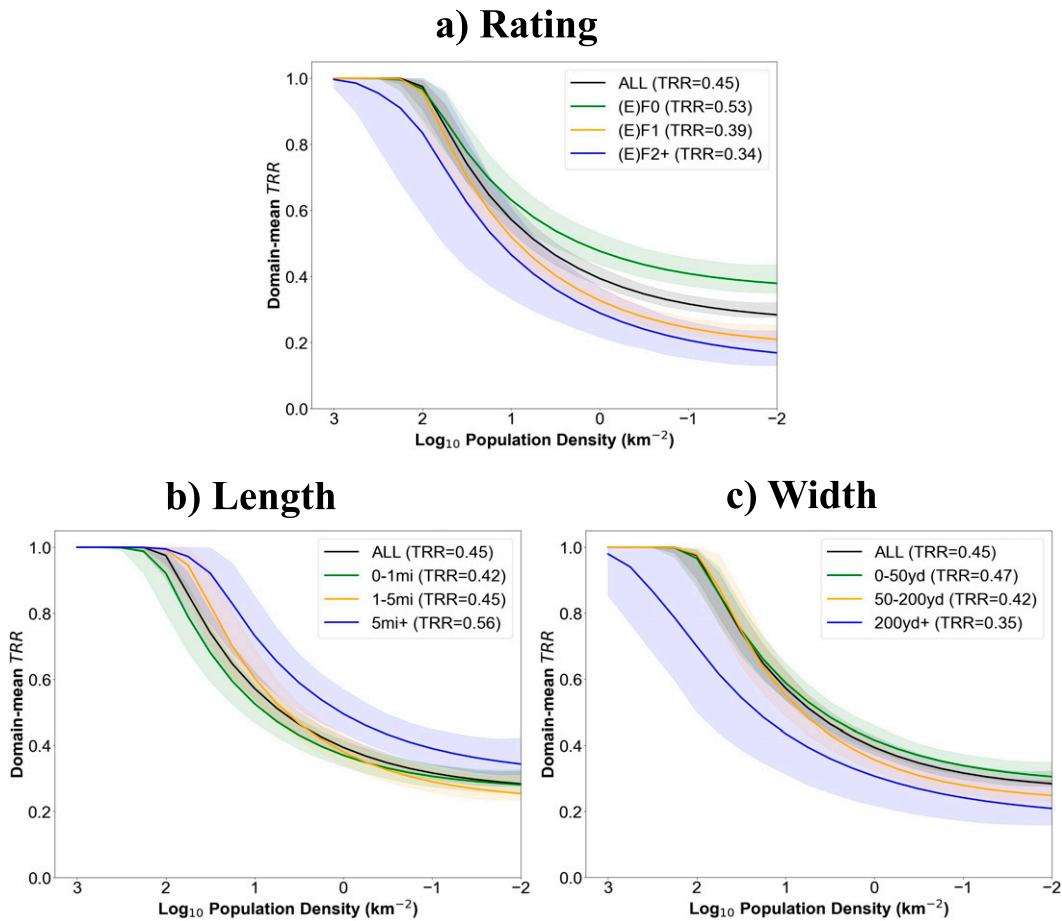
a) Rating

b) Length   c) Width

FIG. 3. Mean posterior TRR vs $\log_{10}$PD for each tornado category: (a) ALL (black), (E)F0 (green), (E)F1 (orange), and (E)F2+ (blue); (b) ALL (black), 0–1 mi (green), 1–5 mi (orange), and 5 mi+ (blue); and (c) ALL (black), 0–50 yd (green), 50–200 yd (orange), and 200 yd+ (blue). The 90% CIs are shaded. The domainwide TRR is listed next to each tornado category in the legend.

sensitive to population density than the TRR of (E)F1 and (E)F2+ tornadoes, which are chronically underrated, especially in rural areas (Fig. 3a). As a result, while (E)F0 tornadoes have been undercounted by a factor of 1.9 (90% CI:1.8–2.1) despite count inflation by underrating of stronger tornadoes, (E)F2+ tornadoes have been undercounted by a factor of 2.9 (90% CI: 2.4–4.0), with 6000–14 000 such tornadoes having been rated (E)F0–1 or not reported over the 1975–2018 analysis period (Table 1). Similarly, wider tornadoes are substantially more undercounted than narrower tornadoes (Fig. 3c). According to the model, 200-yd+ tornadoes are 2.9 (90% CI: 2.2–4.0) times as common as the tornado database suggests. The relationship between TRR and damage pathlength is opposite the relationship between TRR and damage rating or path width: TRR *increases* as damage pathlength increases (Fig. 3b), confirming that pathlength is historically better estimated than intensity and width.

The domain-mean λ predicted by the Bayesian model have a flatter intensity distribution than the reported tornado climatology (Fig. 4a). This is a direct consequence of the underrating bias and correspondingly more pronounced undercounting of stronger tornadoes than of weaker tornadoes. Stratifying by damage path width likewise reveals a flatter λ distribution, though the difference from the reported distribution is smaller given the weaker dependence of TRR on width than on intensity (Fig. 4c). Unlike for the other two tornado attributes, the model predicts a more skewed (not flatter) track length distribution than the official climatology (Fig. 4b), since longer-track tornadoes are less undercounted than shorter-track tornadoes.

Since TRR varies substantially with PD, and the large-scale PD varies substantially throughout the central U.S. analysis domain, so also does the large-scale TRR (Fig. 5). The geographic distribution of TRR obtained with our improved model is very similar to that obtained with the more ad hoc model of P19 (cf. our Fig. 5a with their Fig. 10). The large differences in TRR-PD sensitivity between weaker and stronger tornadoes and between shorter-track and longer-track tornadoes is strikingly represented in these TRR maps (cf. Figs. 5b,c and Figs. 5d,e, respectively). The expected actual tornado counts exhibit similar large-scale spatial patterns as the reported tornado counts (Fig. 6), indicating that
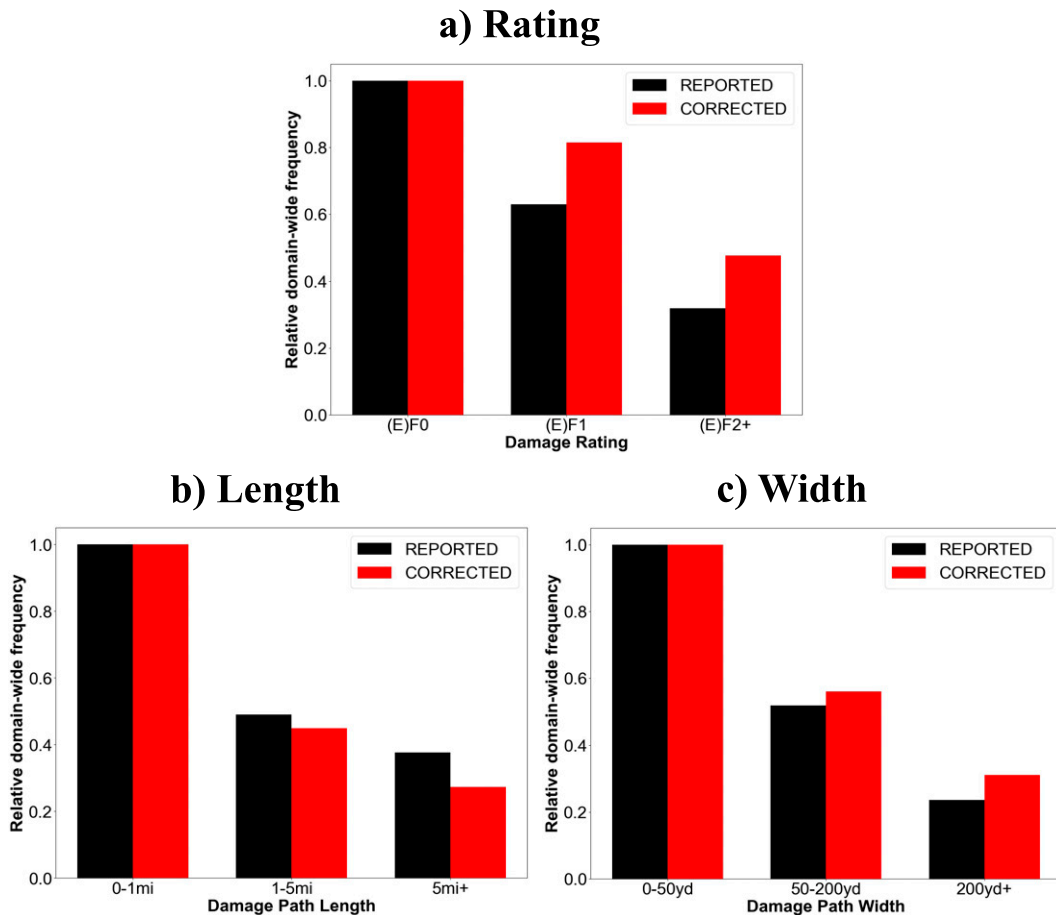
a) Rating

b) Length   c) Width

FIG. 4. Normalized domainwide N (black) and mean posterior λ (red) stratified by (a) damage rating, (b) damage pathlength, and (c) damage path width.

gradients in the latter are dominated by the underlying tornado climatology rather than gradients in TRR. On smaller scales, the model's correction for the reporting biases is evident in the mitigation of local maxima over population centers for ALL tornadoes (cf. Figs. 6a,b); the same is true for (E)F2+ and 5-mi+ tornadoes, but the effect cannot be seen in these heavily smoothed plots (Figs. 6c–f). The differences in magnitude between the reported and expected actual tornado counts are much larger for (E)F2+ tornadoes (Figs. 6c,d) than for 5-mi+ tornadoes (Figs. 6e,f), consistent with the much higher TRR of the latter.

The differences in mean posterior P_{max} and TRR_{min} between tornado categories appear plausible (Table 2), providing additional evidence that the model provides reasonably accurate estimates of these parameters. For example, P_{max} increases with damage rating and path width, consistent with the expectation that more urban areas (with denser damage indicators) are required on average to minimize the probability of underrating stronger or wider tornadoes. Similarly, TRR_{min} decreases with damage rating and path width, which presumably arises from a severe underrating bias in very sparsely populated areas. Conversely, P_{max} *decreases* as damage pathlength increases,

consistent with our hypothesis that the underreporting bias for shorter-track tornadoes exceeds the underrating bias for longer-track tornadoes at high PD. This explanation is supported by our other analyses (Figs. 1c, 2b, 3b) that indicate reported damage pathlength is less sensitive than the other two damage attributes to population density. The mean posterior P_{max} for ALL, 2.00, is consistent with our simple analysis of mean tornado report year versus PD (Fig. 2d) but may be too low given the sharp decreases in tornado counts as PD decreases from ~3 to ~2 (Figs. 1b–d).

For selected full-period analyses, the Bayesian model is evaluated using 10-fold cross validation, where each fold is a randomly selected set of 300-km subregions of the analysis domain (as in P19). For each fold, the nine other folds are used to train the model, then posterior predictive samples are generated for the validation fold. The similarity of the actual N and mean out-of-sample point estimates of N binned by population density (Fig. 7) suggests the model is sufficiently flexible to capture most of the dependence of TRR on PD, and is not unduly overfitting the noisy N. Substantial biases exist, however, in the predicted N at the lowest and (especially) the highest PD. The latter result is further evidence that the P_{max} predicted by the model are too low. A more complex TRR model might
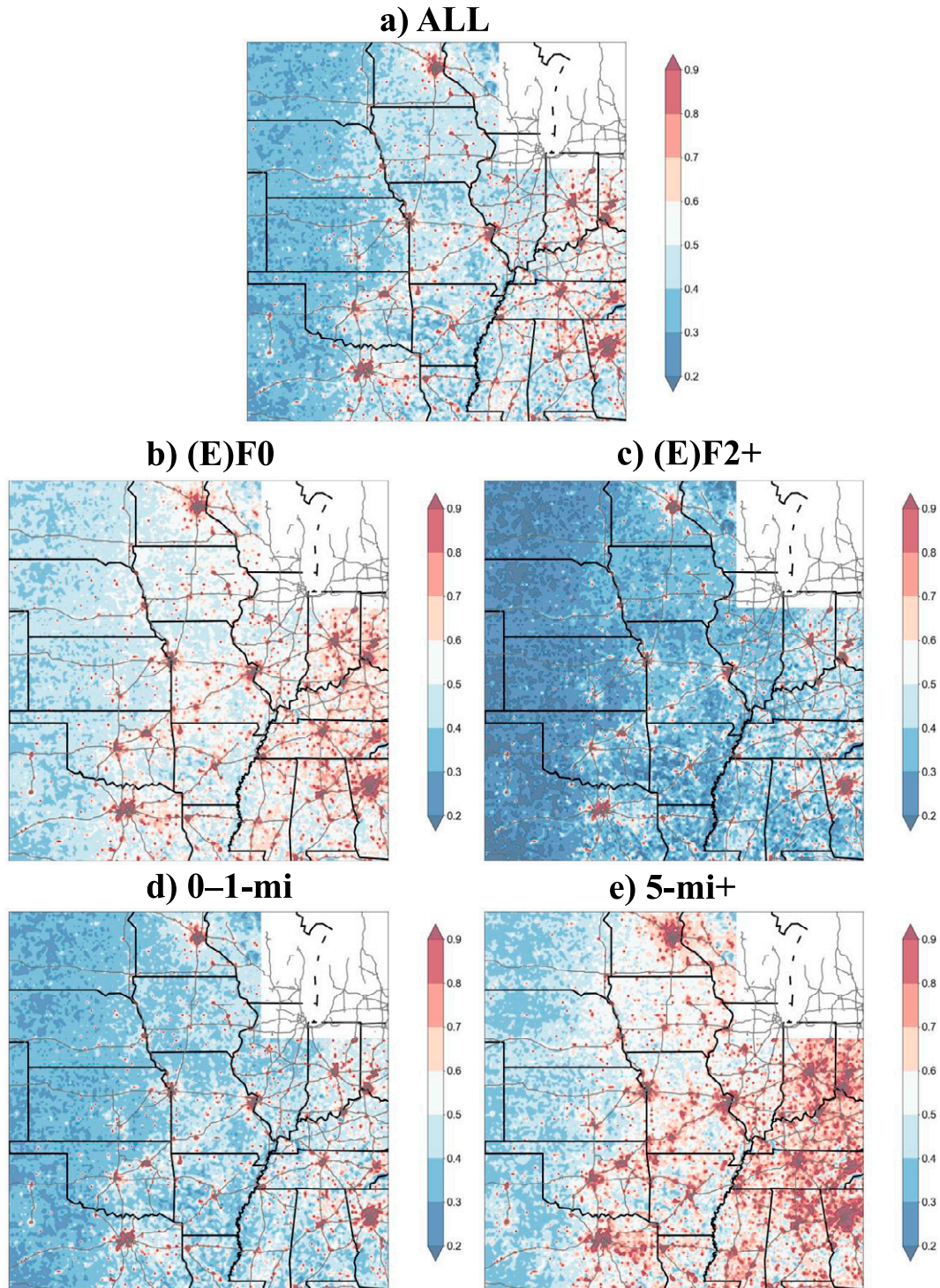
a) ALL

b) (E)F0   c) (E)F2+

d) 0–1-mi   e) 5-mi+

FIG. 5. Mean posterior TRR for (a) ALL, (b) (E)F0, (c) (E)F2+, (d) 0–1 mi, and (e) 5 mi+.

produce a better fit to N at very low and very high PD, but at the risk of overfitting the data and at the computational expense of estimating a larger set of model parameters. It is also not clear whether a more complex TRR model could be constructed that places an upper bound of unity on TRR (which is our mechanism for mitigating TRR–λ confounding). Concerns about the N predictions at extreme PD notwithstanding, the success of the cross validation increases our
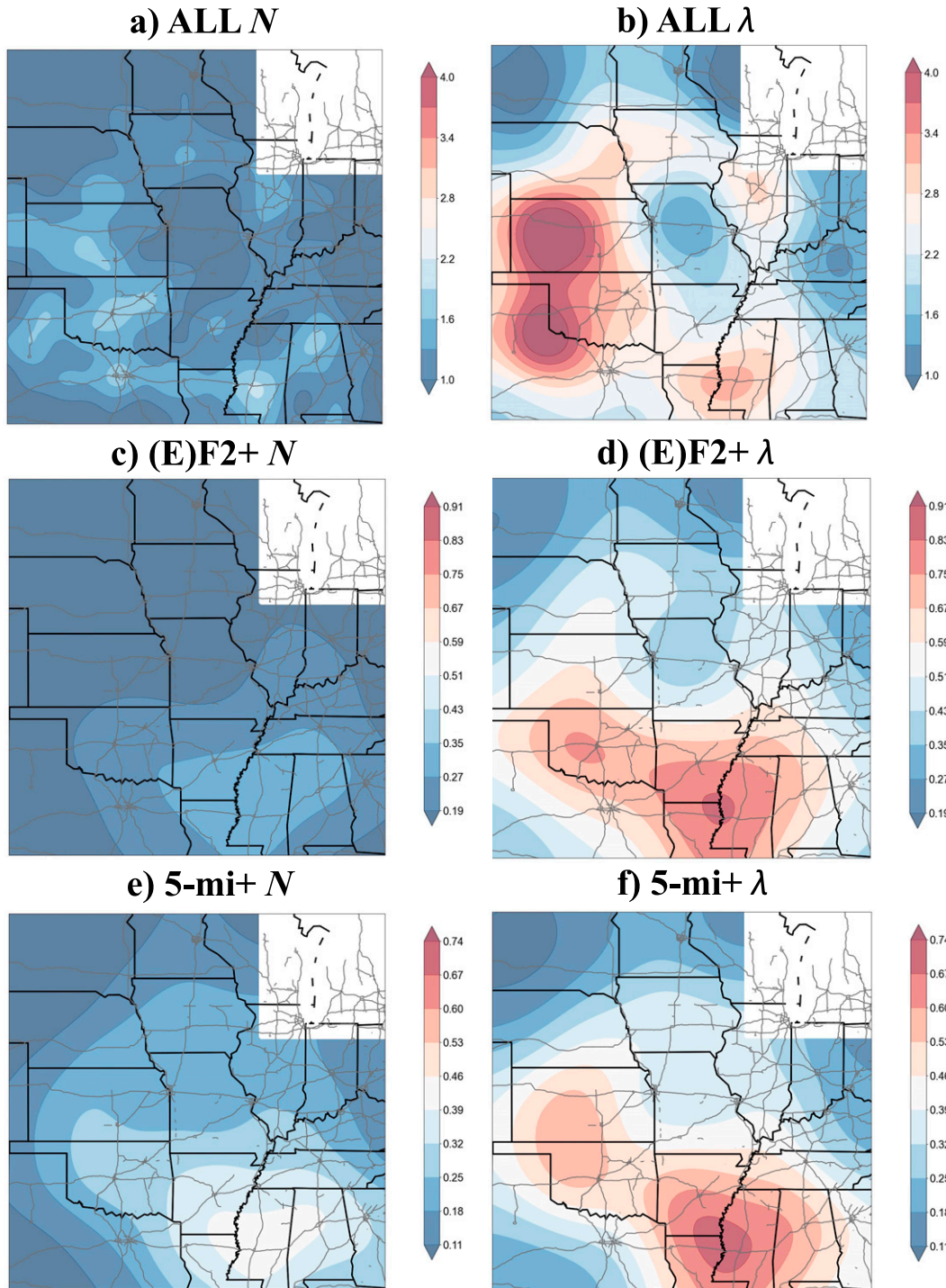
FIG. 6. (left) N and (right) mean posterior λ for (a),(b) ALL; (c),(d) (E)F2+; and (e),(f) 5 mi+. All fields have been smoothed using a Gaussian kernel to mitigate effects of tornado overdispersion (appendix A).

confidence in the TRR and λ posteriors. Failing to account for this TRR dependence and instead naïvely setting N at each 10-km grid cell to the average N within the corresponding 300-km subregion (i.e., assuming TRR = 1 everywhere) greatly increases the mismatch between the actual and estimated P-binned N (Fig. 7). The substantial inferiority of these naïve estimates to the Bayesian model predictions underscores the importance of accounting for reporting bias to accurately reproduce the tornado record.

TABLE 2. Mean posterior P_{max} and TRR_{min} for each tornado category.

Tornado category	P_{max}	TRR_{min}
All	2.00	0.26
(E)F0	2.01	0.36
(E)F1	2.03	0.18
(E)F2+	2.29	0.13
0–1 mi	2.12	0.26
1–5 mi	1.79	0.23
5 mi+	1.62	0.30
0–50 yd	2.02	0.28
50–200 yd	1.93	0.23
200 yd+	2.70	0.17

4. Time series results

a. 5-yr analyses

In our initial 5-yr analyses, unrealistic temporal fluctuations occurred in P_{max} that were negatively correlated with fluctuations in TRR_{min} (Fig. 8a), positively correlated with fluctuations in TRR, and negatively correlated with fluctuations in λ (Fig. 8b). These patterns appear to be yet another instance of parameter confounding. Noting that the P_{max} in all of the initial 5-yr analyses (i.e., for every tornado category) were approximately trendless (e.g., Fig. 8a), we repeated each 5-yr analysis with its P_{max} fixed to the corresponding mean posterior P_{max} from the full-period analysis (section 3b; Table 2). Adding this model constraint mitigated the spurious parameter fluctuations (Figs. 8c,d) while preserving the qualitative trends in TRR, λ, and TRR_{min} obtained with the original model (cf. Figs. 8c,d; Figs. 8a,b).

Results of the 5-yr retrievals with the updated model for various tornado categories are shown in Fig. 9. In all categories but for 200-yd+ (Fig. 9g), the mean posterior TRR exhibits a statistically significant upward trend over the 1975–2018 analysis period. For most categories, however, the TRR does not substantially increase over approximately the last decade of the analysis. For example, while the TRR for ALL approximately doubled over the period, it did not increase over the final 10–15 years, during which nearly one-half of all tornadoes were unreported (Fig. 9a).

No statistically significant trend is found in the 5-yr λ for ALL (Fig. 9a), suggesting that most or all of the upward trend in N arose from the concurrent increase in TRR. The same is true for (E)F0 tornadoes (Fig. 9b), which make up nearly one-half of all tornadoes (Table 1), and for (E)F1 tornadoes (not shown). For (E)F2+ tornadoes, on the other hand, a statistically significant decreasing trend occurs in λ (Fig. 9c). The sharp decrease of (E)F2+ λ during the 1980s coincides with decreasing (E)F0 and (E)F1 λ, suggesting that the decreasing (E)F2+ λ are not due primarily to increases in tornado intensity underrating, or to *decreases* in tornado intensity *overrating*, which was common prior to our analysis period (e.g., Schaefer and Edwards 1999). The abrupt decline in (E)F2+ λ during the 2010s coincides with a decrease in (E)F0 λ but relatively stationary (E)F1 λ. The 0–1-mi λ (Fig. 9d) exhibits a statistically significant downward trend, whereas the 5-mi+ λ

(Fig. 9e) and 1–5-mi λ (not shown) exhibit statistically significant *upward* trends. A similar pattern occurs for the tornado width categories: whereas 0–50-yd λ significantly decreases over the analysis period (Fig. 9f), the 200-yd+ (Fig. 9g) and 50–200-yd λ (not shown) significantly increase. The pronounced decreases in 0–1-mi λ and 0–50-yd λ contrast with the *increases* (albeit statistically insignificant) in the corresponding N.

To assess how well the 5-yr retrievals collectively predict the full-period N, we sum the observed N and means of the posterior predictive distributions of N for the (1979–83, 1984–88, …, 2014–18) analyses and then bin them by PD (Fig. 10). The observed and predicted N match reasonably well, with the largest discrepancies again occurring at extreme PD. The success of the 5-yr predictions in reproducing the full-period observed N increases our confidence in the time-dependent analyses presented in Fig. 9. Additional analysis suggests that the use of 1990 population data in the early period analyses (which produces temporal errors in PD of up to 13 years) does not unduly degrade the TRR and λ estimates (appendix C).

It has recently been shown that while trends in large-scale U.S. tornado frequency are likely small over the last several decades, regional trends are more substantial and correspond in part to an eastward shift in tornado frequency (e.g., Gensini and Brooks 2018). Given the large increases in TRR over the analysis period, and the expected spatial differences in the magnitudes of these changes due to spatial differences in population density, we compared the trends in the 5-yr N and 5-yr λ for ALL over our 1975–2018 analysis period to determine whether the spatial shift in reported tornado frequency is substantially affected by reporting bias. To reduce noise in the analyzed trends, and to be more consistent with Gensini and Brooks (2018), we upscaled the 5-yr N and 5-yr λ to a 100-km grid collocated with our original (10 km) analysis grid. As in our previous time series analyses, we computed the Theil–Sen slope and used the Hamed and Rao (1998) modification of the Mann–Kendall test to assess statistical significance.

The linear trends in 5-yr N and λ are qualitatively similar to each other throughout most of the analysis domain (Fig. 11), adding credence to the previously identified spatial shift in tornado frequency. However, while the trends in N are positive through most of the domain, negative and positive trends in λ approximately balance each other. These results are consistent with the rapid increase of domainwide N and the stationarity of domainwide λ over the analysis period, respectively (Fig. 9a). The linear trends in λ and N suggest that tornado frequency has decreased faster over certain western portions of our domain than the tornado record indicates and that the general eastward shift of tornado frequency is slightly more pronounced than currently recognized.

b. Limitations of the time-dependent model

Using the ALL tornadoes dataset, we applied different versions of the time-dependent Bayesian model described in section 2c to the full analysis period to see if we could improve upon our full-period or 5-yr analyses with the original (time independent) model. Including the interannual variability
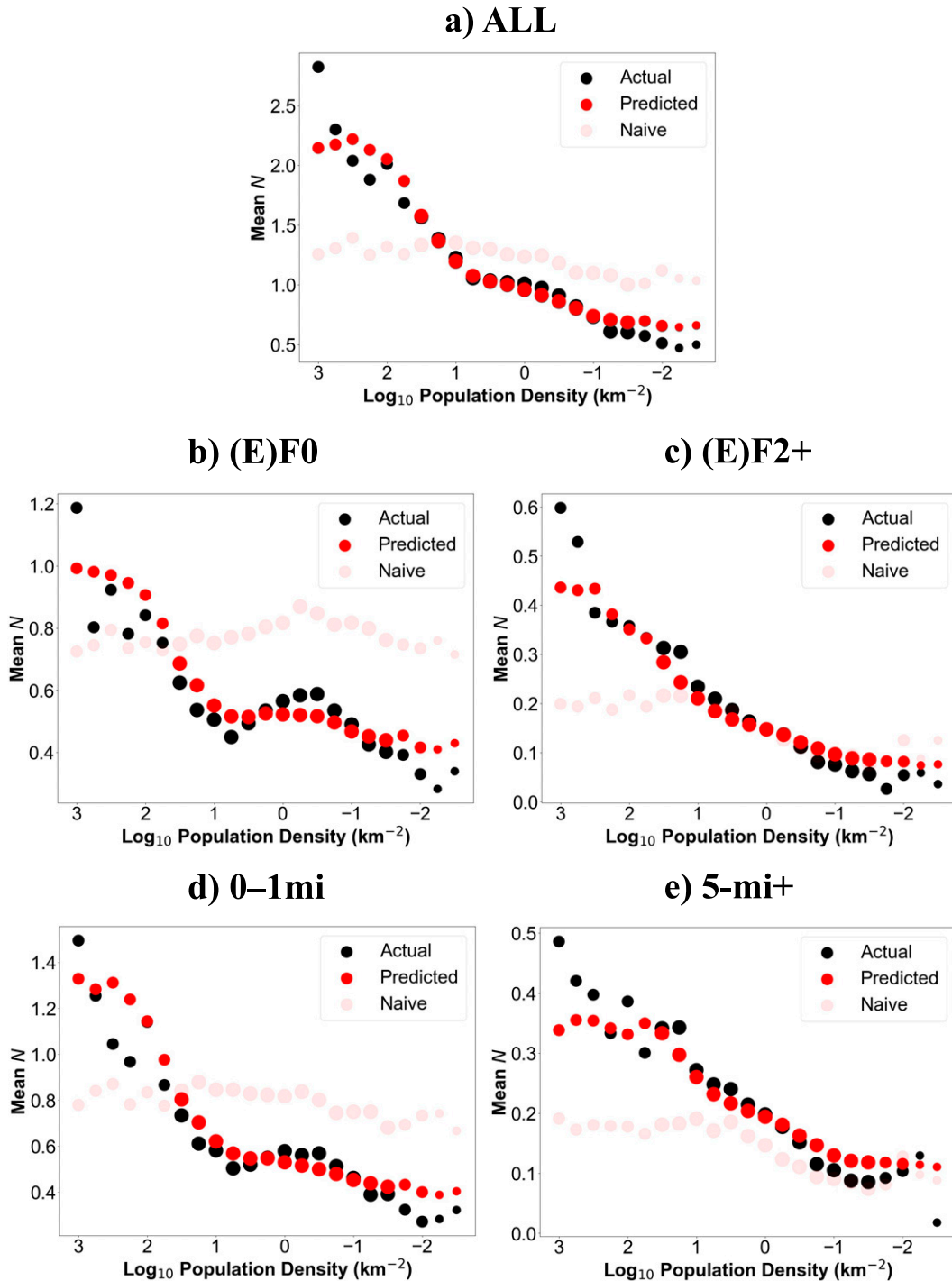
a) ALL

b) (E)F0   c) (E)F2+

d) 0–1mi   e) 5-mi+

FIG. 7. Mean out-of-sample N (black), mean posterior N (solid red), and naively estimated N (transparent red) binned by $\log_{10}$PD for (a) ALL, (b) (E)F0, (c) (E)F2+, (d) 0–1 mi, and (e) 5 mi+. Large, medium, and small dots represent gridpoint counts >500, 100–500, and <100, respectively.

term ν in the λ model negligibly improved the model fit to the data, with the coefficient of determination R^2 of the mean posterior N increasing by < 1% and the point estimates of TRR and λ changing by < 1%. Including only the population-

density–time interaction term (i.e., the β_3 term) produced a P_{max} point estimate of nearly 4, which is implausibly high (Table 2). Since the posterior P_{max} was too high, the posterior TRR were too low, with a domainwide TRR estimate of 0.21,
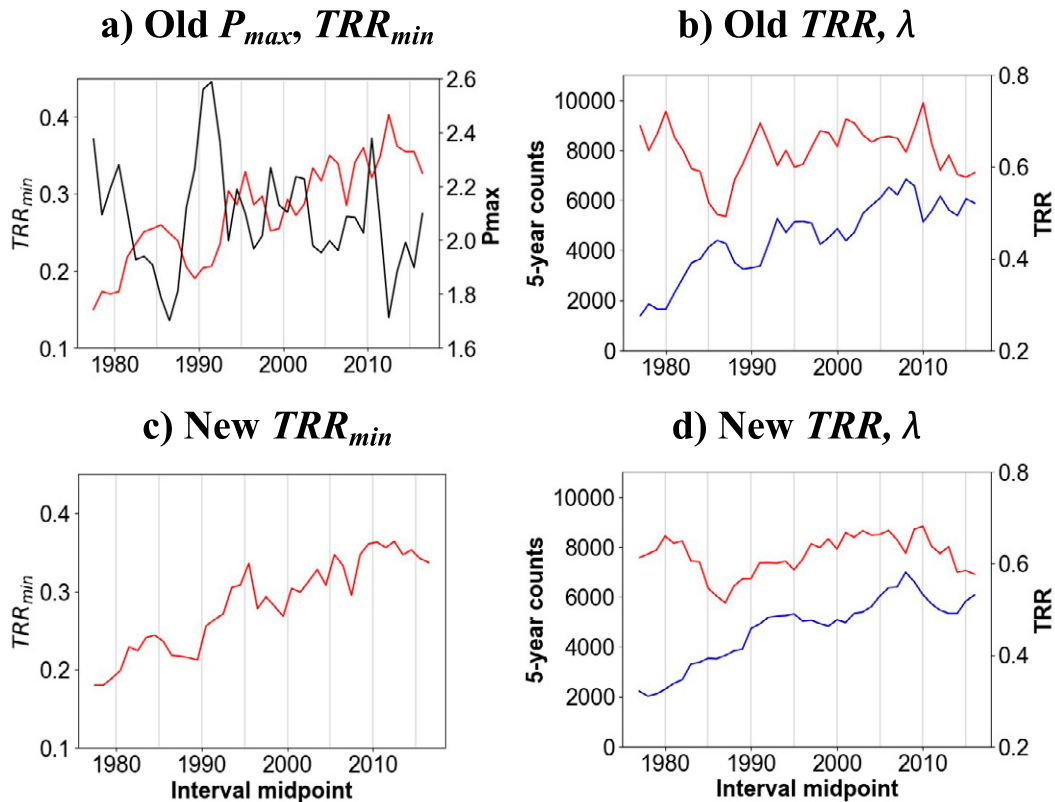
FIG. 8. Five-year mean posterior (a) P_{max} (black) and TRR_{min} (red) and (b) TRR (blue) and λ (red) for ALL using the original TRR model. Also shown are 5-yr mean posterior (c) TRR_{min} and (d) TRR (blue) and λ (red) for ALL using constant P_{max}.

less than one-half of that obtained in our full-period analysis (0.45; Table 1). Fixing P_{max} to 2.00 as in our 5-yr analyses produced a domainwide TRR estimate that was very similar to those from our full-period Bayesian analyses (0.44 vs 0.45). The estimated evolution of domainwide TRR (not shown), however, poorly matched that obtained from our 5-yr analyses, suggesting that the quasi-exponential model for the population-density–time interaction is too simple.

Including the population-density-independent TRR evolution (i.e., β_2) term in addition to the β_3 term produced a domainwide TRR estimate of 0.44 and a P_{max} point estimate of 2.01, both very similar to those obtained in the full-period analysis. However, the maximum TRR was much less than 1 early in the analysis period (unlikely) and much greater than 1 late in the period (impossible). It is unclear how to retain the condition that TRR = 1 for $P > P_{max}$ when the β_2 term is included in the TRR model (whether or not the β_3 term is also included). Moreover, examination of Eqs. (1), (5), and (6) reveals that whether the β_2 term is included in the TRR model or the λ model has no impact on the predicted N. Thus, the Bayesian model cannot distinguish between tendencies in TRR and λ. Indeed, in an experiment with the β_3 term included in the TRR model but the β_2 term moved to the λ model, the TRR tendency was severely underestimated, and the λ was erroneously estimated to have nearly

doubled over the analysis period (whereas a small, statistically insignificant trend was found in our 5-yr analyses; Fig. 9a).

In light of this new manifestation of the TRR–λ confounding problem, and the large errors that could arise from assuming quasi-exponential evolution of domainwide TRR and/or λ, we conclude that applying the original model to series of subintervals of the analysis domain produces more accurate estimates of these parameter than applying any version of the time-dependent model on the full analysis period. The original model is also appropriate for the full-period analyses, since the improvement obtained by including some or all of the temporal terms is trivial (increase in R^2 of the mean posterior $N < 1\%$) yet increases the computational cost by an order of magnitude.

5. Discussion

According to both our simple analysis (section 3a; Fig. 1) and our Bayesian model predictions (section 3b; Fig. 3; Table 1), the population density threshold above which virtually all tornadoes are reported is approximately 100 people per kilometer squared within the central United States. This result sharply contrasts with the conclusion by Cheng et al. (2013, 2015) that this threshold is only about 6 or 7 people per
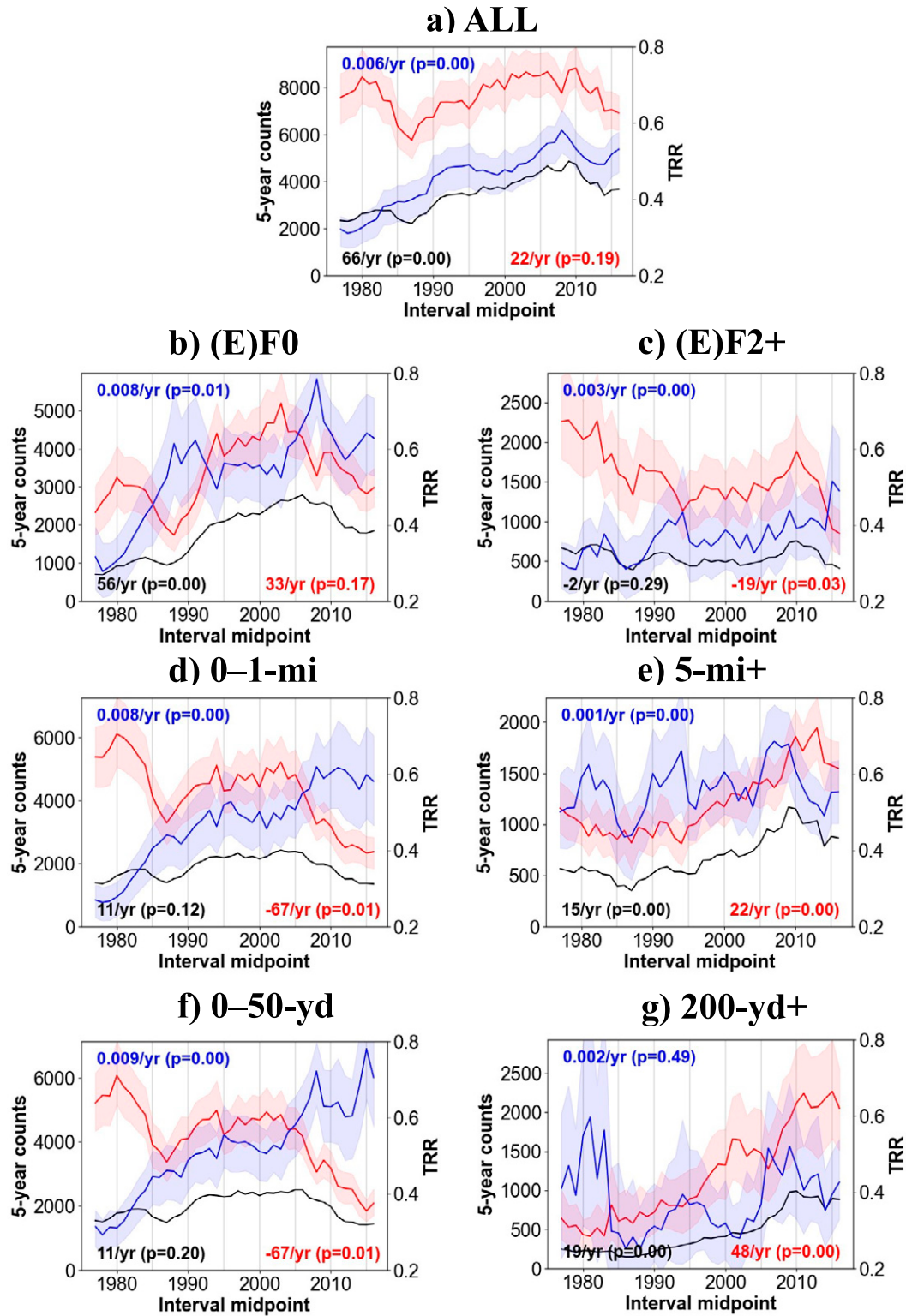
FIG. 9. Five-year N (black curve), mean posterior TRR (blue curve), and mean posterior λ (red curve) for (a) ALL, (b) (E)F0, (c) (E)F2+, (d) 0–1 mi, (e) 5 mi+, (f) 0–50 yd, and (g) 200 yd+. The 90% CIs for TRR and λ are shaded. For each time series, the linear trend and its statistical significance are indicated.
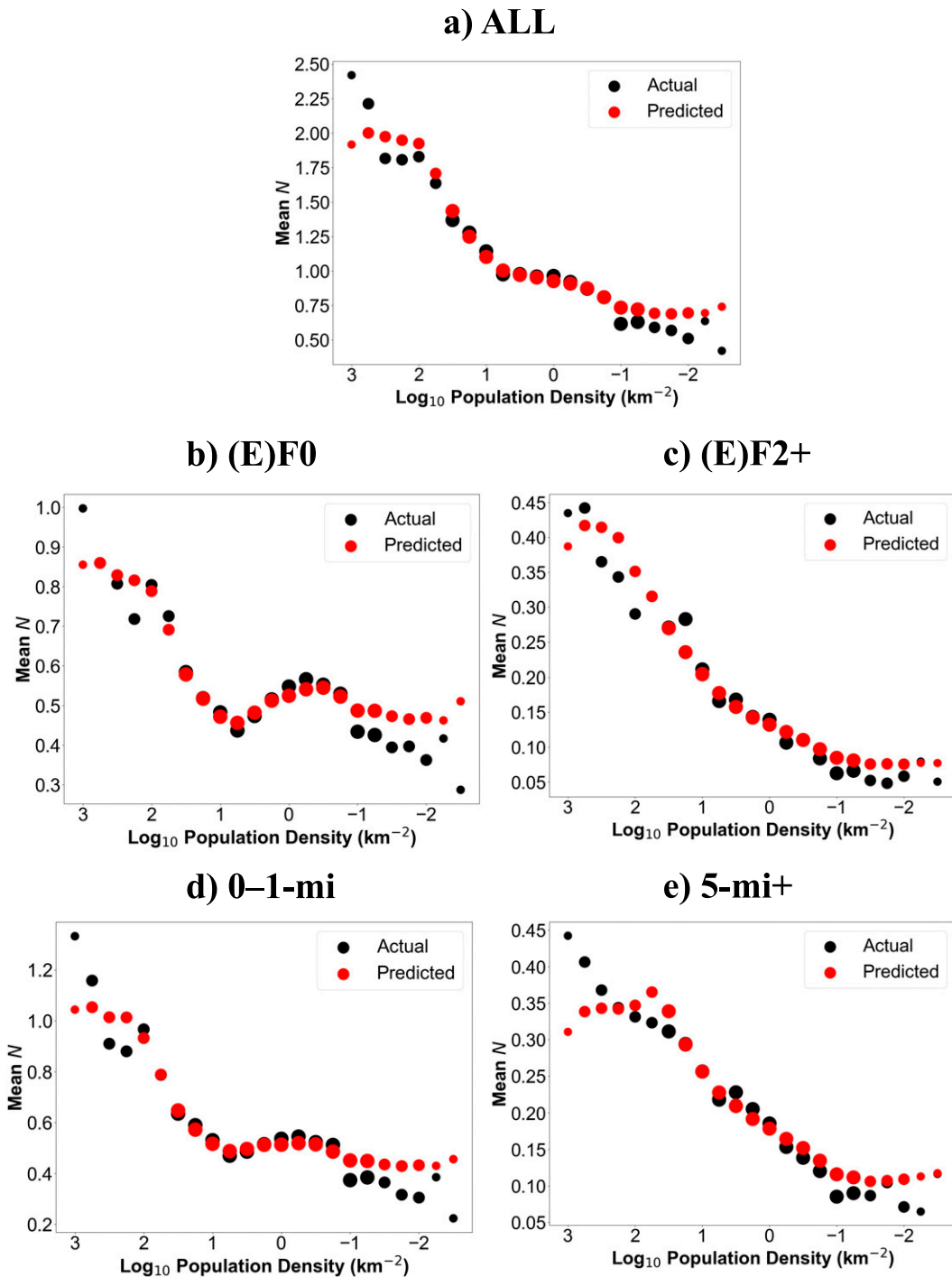
a) ALL

b) (E)F0   c) (E)F2+

d) 0–1-mi   e) 5-mi+

FIG. 10. Mean N (black) and mean posterior N (red) binned by $\log_{10}$PD, accumulated over the 5-yr model predictions, for (a) ALL, (b) (E)F0, (c) (E)F2+, (d) 0–1 mi, and (e) 5 mi+. Large, medium, and small dots represent gridpoint counts >500, 100–500, and <100, respectively.

kilometer squared in Canada. Given that there is no obvious reason why the TRR would be much higher in Canada than in the central United States, it is plausible that the Bayesian models used in those studies suffered the parameter confounding that our model is designed to mitigate.

Statistical models that fail to resolve the TRR–λ aliasing are likely to produce large errors in both parameters.

One of the more striking predictions of our Bayesian model is that particularly intense or wide tornadoes are severely undercounted (Figs. 3a,c). These results, which are highly
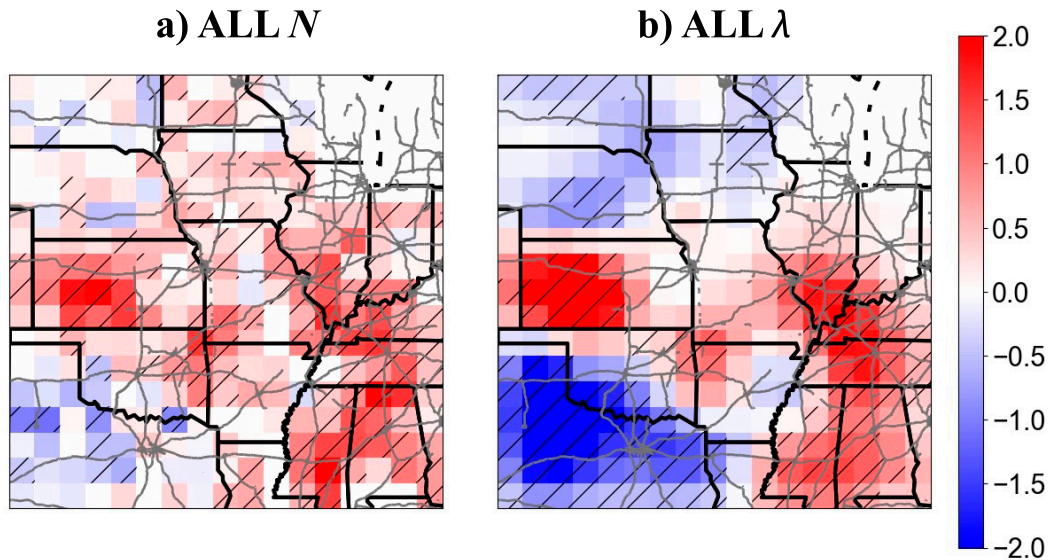
FIG. 11. Linear trend in annual 100-km (a) N and (b) λ for 1977–2016. Hatching indicates statistically significant trends.

plausible in light of our simple analysis of the tornado database (Figs. 1b,d), point to chronic underestimation of tornado intensity and width by damage rating and path width, respectively. The existence of such a large bias in tornado damage ratings is strongly supported by the tornado simulation experiments of Dahl et al. (2017), who found that the near-maximum winds in strong tornadoes are so spatiotemporally localized as to be highly unlikely to be sampled (by anemometers or damage indicators) in rural areas. In fact, the bias-corrected damage rating distribution obtained from the Bayesian model, while shifted toward higher damage ratings than the official tornado record, may still severely underestimate the distribution of tornado-maximum winds, given that the strongest winds in intense tornadoes may frequently fail to intersect sufficiently resilient damage indicators even in highly urbanized areas.

Strong observational evidence that the bias-corrected damage rating distribution remains biased toward lower ratings is provided by recently published wind-based rating distributions computed from Doppler on Wheels (DOW; Wurman et al. 1997) velocity data for 82 supercell tornadoes (Wurman et al. 2021). Although the DOW-based rating distributions are likely positively skewed by selection bias, these ratings are on average 1.5 categories higher than the NWS damage ratings for the same set of tornadoes (Wurman et al. 2021), indicating that the DOW-based rating distributions much better represent true tornado damage potential than does the NWS rating distribution. The fact that our model-predicted rating distribution is much closer to the NWS distribution than to the DOW distributions implies that our bias-correction procedure fails to fully capture the tornado intensity underrating. The corresponding underestimation of the frequency of stronger tornadoes by our model validates our interpretation of the model-predicted TRR (section 2d), which accommodates the possibility that significant tornadoes are systematically

underrated even in the most densely populated regions of the central United States. Wurman et al. (2021) also corroborates our finding of a severe tornado width underestimation bias; their DOW-measured tornado widths were systematically much wider than those assigned by the damage surveys. Tornado damage mitigation strategies and cost–benefit analyses should account for the fact that intense or wide tornadoes are much more common than is implied by the official tornado record.

Our finding that tornado intensity and width are more seriously underestimated than track length (Fig. 3; Table 1) suggests that greater weight should be given to the latter in certain applications. For example, since long-track tornado reports are less biased than high-damage-rating reports, spatial analyses of the former will better represent the true tornado climatology. Given the correlation between tornado intensity and track length (Brooks 2004), it is possible that long-track tornado reports better represent intense tornadoes than do high-damage-rating reports (though it is not clear how to test this hypothesis). The more modest bias in damage path-length also further motivates the use of cumulative damage pathlength in identifying tornado outbreaks or characterizing tornado risk (e.g., Edwards et al. 2004; Broyles and Crosbie 2004; Clark et al. 2012; Fuhrmann et al. 2014; Coleman and Dixon 2014).

It is noteworthy that the model-estimated frequencies of (E)F0 tornadoes are sharply maximized over the Great Plains, whereas (E)F2+ tornadoes are maximized over the Southeast (cf. Figs. 6b,d). One likely meteorological contribution to this pattern is the well-known higher frequency of weak, nonmesocyclonic tornadoes over the Great Plains than over the Southeast (e.g., Lee and Wilhelmson 1997). It is also likely, however, that the longer viewing distances and higher density of storm chasers in the Great Plains causes the reporting biases within low-population areas there to be smaller than in

similarly populated areas of the Southeast. The Bayesian model does not account for these variables and so would overcompensate for reporting bias in the Great Plains if these effects are indeed occurring. Similarly, the sensitivity of reporting bias to population density is likely modulated by Doppler radar coverage, and so the model likely overestimates the frequency of weak tornadoes in areas with dense WSR-88D coverage but sparse population. Such considerations of omitted variable bias motivate the inclusion of additional parameters in the tornado reporting rate model (a nontrivial task; P19) or application of the model to individual geographical regions (although the reduced sample sizes would increase uncertainty in the model parameters).

When interpreting trends in the 5-yr λ (section 4; Fig. 9), it is important to remember that our model makes no accommodation for changes in tornado counts arising from changes in damage assessment practices. The model is designed to correct only for the increases of unreported tornado rates and of the biases in estimated tornado attributes arising from sparse population and damage indicators. There have been many official and unofficial changes in tornado damage assessment that are expected to have introduced temporal variability (likely including both gradual trends and abrupt shifts) into the tornado record. For example, it is safe to assume that the rapid increases in 200-yd+ λ beginning in the mid-1990s and mid-2000s (Fig. 9g) are due at least in part to the switch to reporting maximum instead of average damage path width in 1995 and the implementation of the EF scale in 2007, respectively (Edwards at al. 2021). It is unclear how well we can distinguish such secular variability from true changes (both short and long term) in tornado attributes. Fortunately, the 5-yr λ for ALL tornadoes are not sensitive to changes in damage survey practices, nor to other nonmeteorological effects such as the expansion of Doppler radar coverage, so long as TRR approaches unity in the most populous areas of our central U.S. domain throughout the analysis period.

The linear trend in the 5-yr λ for ALL tornadoes is small (19 yr^{-1}) and statistically insignificant (Fig. 9a), which suggests that long-term climate change has not substantially affected the domainwide tornado frequency during the 1975–2018 analysis period. This result confirms the hypothesis that most or all of the long-term trend in reported U.S. tornado counts is due to nonmeteorological factors. That hypothesis, which is based on the much larger increase in reported (E)F0 than (E)F1+ tornadoes over the tornado record, underlies the use of linear detrending to correct for secular changes in tornado frequency (e.g., Verbout et al. 2006). The Bayesian modeling approach is a far more powerful way to characterize temporal variability in true tornado frequency that does not assume long-term stationarity thereof nor linearity of any secular trend. Our model estimates of TRR for ALL tornadoes suggest the linear secular trend assumption is approximately valid until the last decade of the 1975–2018 analysis period. Much more nonlinear evolution of TRR is estimated for certain tornado categories, however (e.g., Fig. 9g).

The generally nonincreasing TRR near the end of the analysis period for the various tornado categories (Fig. 9) suggests our ability to detect and accurately characterize tornadoes may not improve substantially without major observational innovations (e.g., unpiloted aerial systems; McFarquhar et al. 2020; Wagner et al. 2019, 2021). We therefore expect that statistical techniques to quantify and correct reporting biases will be necessary for the foreseeable future.

Given the large regional trends in tornado frequency identified by previous work (e.g., Gensini and Brooks 2018) and further examined herein (Fig. 11), it is useful to consider the degree to which changes in regional tornado frequency may degrade the Bayesian model estimates. While the model does not explicitly account for regional or domainwide changes in λ over the analysis period, the expected reported tornado counts should be well approximated by the product $\lambda \times$ TRR [i.e., Eq. (1) should be valid] even if a large trend exists in λ. This idea is supported by our finding that explicitly accounting for the large linear trend in domainwide TRR over 1975–2018 (Fig. 9a) via the β_2 term in Eq. (5) does not substantially improve the model (section 4b), which suggests that failure to account for large linear trends in tornado frequency should not substantially degrade the model. The potential effects of highly *nonlinear* changes in regional tornado frequency during the analysis period are less clear, though we suspect the resulting bias in TRR would be small since errors arising in regions with nonlinear λ changes would be substantially offset by errors in regions with approximately opposite λ changes, assuming that changes in tornado frequency can be primarily characterized as spatial shifts. Nonlinear λ changes may be more impactful in our 5-yr analyses since the degree of the aforementioned error offsetting likely decreases for shorter analysis periods. The potential Bayesian model errors arising from nonlinear regional changes in λ could be rigorously investigated using artificially constructed or modified tornado report datasets.

6. Conclusions

Tornado underreporting and underrating have severely contaminated the official U.S. tornado database. Bayesian hierarchical modeling provides a powerful framework for mitigating the impacts of these biases and thereby facilitating studies that depend upon accurate estimates of U.S. tornado frequency. P19 used a Bayesian hierarchical model to estimate the reporting rate and expected frequency of tornadoes over the central United States during 1975–2016. A novelty of the model was that it mitigates a solution nonuniqueness problem that likely degraded estimates of tornado frequency and reporting rate in previous studies. The present study eliminates some ad hoc elements of the P19 Bayesian model. Rather than splitting the analysis domain into subregions within which the true tornado frequency is assumed to be constant, a restricted spatial regression model is used to account for spatial autocorrelations in tornado frequency. In addition, two parameters of the tornado reporting rate model that were prescribed in P19 are now predicted along with the remaining Bayesian model parameters. Using the improved Bayesian model, the present study examines how tornado reporting rate and bias-corrected tornado frequency vary with damage rating, pathlength, path width, and over the 1975–2018 analysis period.

The Bayesian model analysis presented herein indicates that for every (E)F2+ tornado report within the central United States over the 1975–2018 analysis period, 1.9 (90% CI: 1.4–3.0) additional tornadoes with (E)F2+ damage potential occurred but were either unreported or assigned an (E)F0 or (E)F1 rating. In light of recent modeling and observational work (Dahl et al. 2017; Wurman et al. 2021), the tornado intensity underestimation problem is likely even more severe than our analysis suggests. Our model similarly estimates that tornadoes exceeding 200-yd diameter have been undercounted by a factor of 2.9 (90% CI: 2.2–4.0). Reported damage pathlength appears to be less biased than damage rating and damage path width; this advantage should be considered when designing criteria for subsetting the tornado database. The Bayesian model estimates are corroborated by simple analysis of the tornado record; we recommend that future Bayesian model studies of tornado climatology use similar plausibility checks to ensure the veracity of their results.

One particularly valuable application of the Bayesian model is to produce time series of expected actual tornado counts. Our analysis suggests that all-tornado frequency has not changed substantially over the analysis period. For certain tornado categories [e.g., (E)F2+ tornadoes], the time series exhibit large, statistically significant trends over the analysis period. Since the Bayesian model does not account for changes in damage survey practices, however, we cannot isolate meteorological trends in the expected counts from trends arising from changes in reported attribute biases. Given that these residual secular trends over the analysis period are difficult to estimate, detecting true trends in tornado intensity, size, or track length may prove difficult unless they become very large.

The Bayesian hierarchical modeling framework adopted herein can be applied to a number of other problems. Direct extensions of the present work include examining how tornado reporting bias varies diurnally and seasonally; revisiting previous findings that could have been substantially affected by reporting bias, as we did with the Gensini and Brooks (2018) analysis of changes in U.S. tornado frequency (we found that the eastward shift of tornado frequency is likely more pronounced than was originally reported); and estimating reporting bias in the severe hail and severe wind databases. Ideally, improvements to severe weather climatologies accruing from this and other approaches to correcting reporting bias (e.g., Wurman et al. 2021) will benefit future studies of climate–tornado linkages, cost–benefit analyses of tornado damage mitigation strategies, and evaluations of severe weather prediction tools and operational products. Given the recent flattening of tornado reporting rates, there is likely to be a continuing need for reporting bias correction techniques.

APPENDIX A

Selection of Gaussian Kernel Width for Smoothing λ

A completely spatially random process is modeled by the Poisson distribution:

$$P(k) = E^k \exp(-E)/k!, \tag{A1}$$

where P is the probability of k events occurring within a prescribed region given an expected number of events E. We use Eq. (A1) to calculate the minimum filtering scale required to substantially mitigate sampling errors in spatial plots of the mean λ posteriors from our Bayesian model. First, we determine the smallest E, E_{min}, for which $P(0.8E < k < 1.2E) > 0.95$, that is, for which random sampling errors are < 20% at least 95% of the time. This is $E_{min} = 96$. We round E_{min} to 100, both for simplicity and since, for $E = 100$, k has the additional desirable property of lying within 10% of E more than 70% of the time. We judge these sampling errors to be acceptably small for our application. We then divide E_{min} by the domain-mean reported tornado count $\langle N \rangle$ to obtain the average area A over which $E = E_{min}$ (i.e., over which there are, on average, 100 reported tornadoes). Last, we compute the diameter d of a circle of area A. Thus, if we were to use a circular filtering kernel that evenly weights all grid cells within a radius of $d/2$, we could be confident that sampling errors due to complete spatial randomness would be strongly damped in the analysis.

It is not immediately clear, however, what the corresponding width (2 times the standard deviation) of the *Gaussian* filtering kernel should be. We therefore used the simple circular filter to smooth spatial plots of N and then visually compared the results with Gaussian-filtered N obtained using various kernel widths. We thereby determined that, for each of the tornado categories examined in this study, a Gaussian kernel width $W = 0.6d$ produces analyses that are very similar to those from the simple filter except, of course, less noisy (our primary
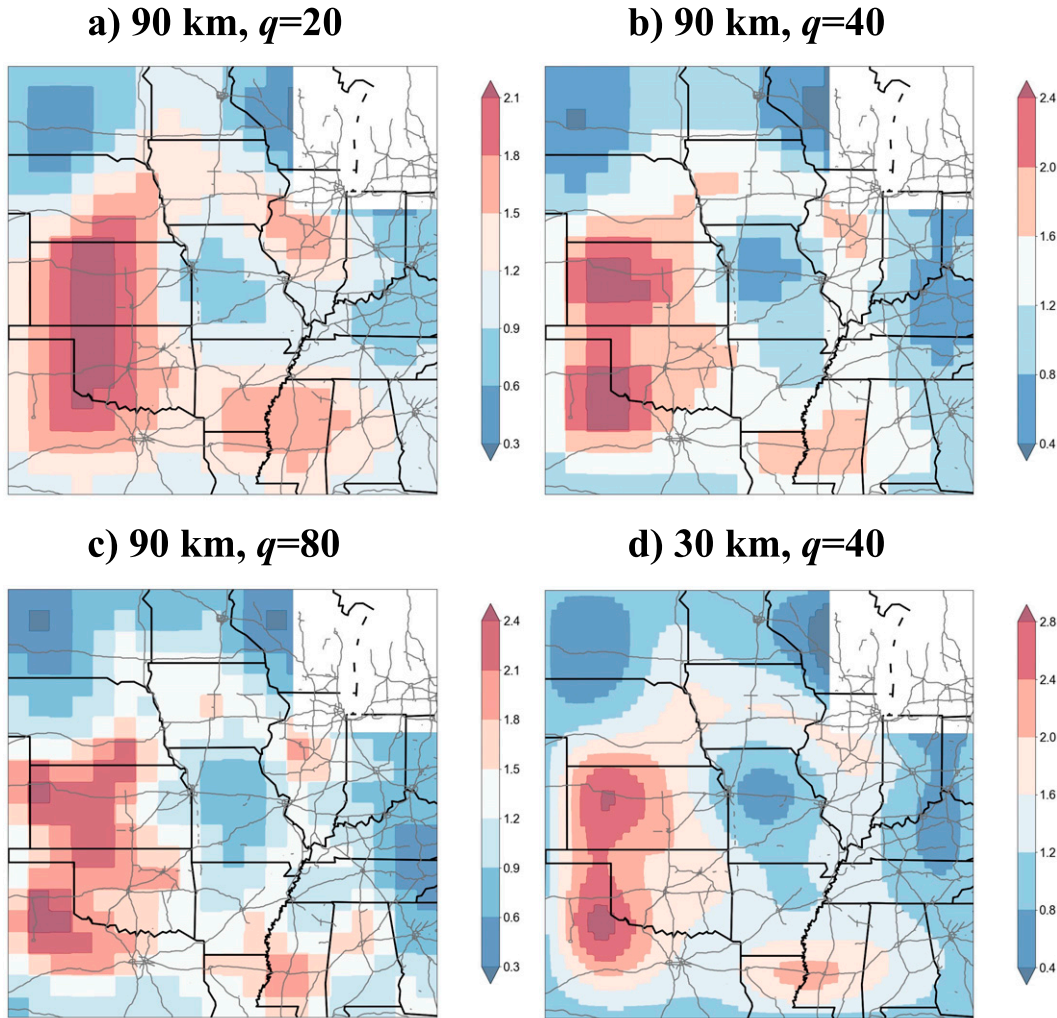
a) 90 km, q=20

b) 90 km, q=40

c) 90 km, q=80

d) 30 km, q=40

FIG. B1. Mean posterior ω obtained using (a) 90-km RSR model and $q = 20$, (b) 90-km RSR model and $q = 40$, (c) 90-km RSR model and $q = 80$, and (d) 30-km RSR model and $q = 40$.

motivation for using a Gaussian filter). Thus, for each tornado category, we calculate W as

$$W = 0.6d = 0.6 \times 2(A/\pi)^{1/2} \approx 0.68(E_{\min}/\langle N \rangle)^{1/2}$$
$$\approx 6.8/\langle N \rangle^{1/2}. \tag{A2}$$

Since tornado occurrences are more clustered than if they were generated by a truly spatially random process (hence our use of the negative binomial distribution in the Bayesian model), W serves as a *lower* estimate of the optimal filter width for reducing sampling errors below a prescribed tolerance. We propose that this approach to selecting an appropriate Gaussian filter kernel width is applicable to any problem involving 2D data that are governed by a spatially quasi-random process.

APPENDIX B

RSR Model Implementation

Implementation of the RSR model required an adjacency matrix for our $n \times n$ correlated random errors grid (which is coarser than our 10-km analysis grid, i.e., $n < 180$; we give more on this later), $\mathbf{A}_{n \times n}$, and a regressor matrix, which for our application is a vector of PD, $\mathbf{X}_{n \times 1}$. We use an eight-connected neighborhood for $\mathbf{A}$. We then compute the projection $\mathbf{P}$ onto the orthogonal component of $\mathbf{X}$:

$$\mathbf{P}_{n \times n} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}},$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. The Moran operator $\mathbf{\Omega}$ for $\mathbf{X}$ with respect to the underlying graph (represented by $\mathbf{A}$) is
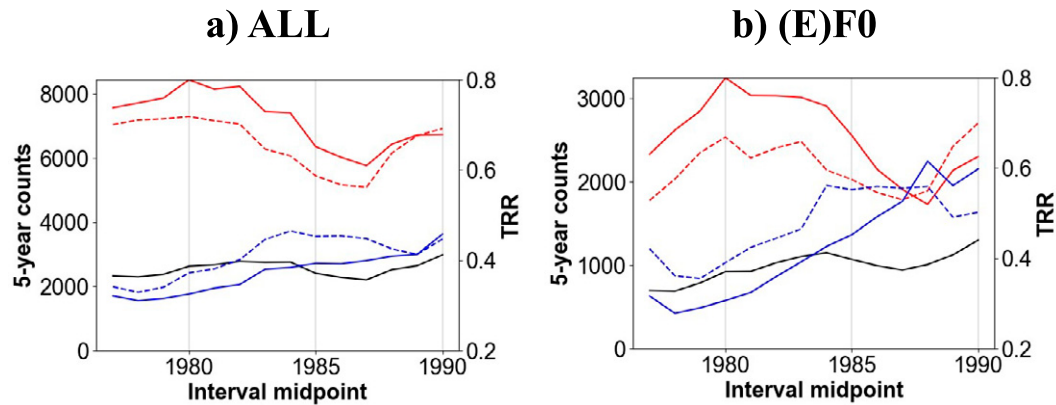
a) ALL

b) (E)F0

FIG. C1. Five-year N (black curve), mean posterior TRR (blue curves), and mean posterior λ (red curves) for (a) ALL and (b) (E)F0. Solid and dashed curves represent the original experiments (i.e., as in Fig. 9) and the new experiments with PD computed from the 2010 census, respectively.

$$\mathbf{\Omega}_{n \times n} = \mathbf{PAP}.$$

We then compute the first q eigenvectors of $\mathbf{\Omega}$ to form $\mathbf{M}_{n \times q}$, which thus comprises mutually distinct clustering patterns in N that are noncollinear with PD. From visual inspection of e^{ω} for different q (Fig. B1), we judged that $q = 40$ captures much of the spatial heterogeneity in λ without excessively fitting finer-scale patterns that are less likely to represent the true long-term climatology.

The correlated random errors variable is given by

$$\omega = \mathbf{M}\alpha_{\mathrm{RSR}},$$

where α_{RSR} controls the variance in ω between neighboring grid cells and is assigned a multivariate normal (MvNormal) prior:

$$\alpha_{\mathrm{RSR}} \sim \mathrm{MvNormal}(\mu = 0, \tau = \tau_s \times \boldsymbol{\alpha}_{\tau}).$$

The smoothing parameter τ_s is assigned a gamma prior:

$$\tau_s \sim \mathrm{gamma}(\alpha = 0.1, \beta = 0.1),$$

and the precision matrix $\boldsymbol{\alpha}_{\tau}$ is given by

$$\boldsymbol{\alpha}_{\tau, q \times q} = \mathbf{M}^{\mathrm{T}} \mathbf{QM},$$

$$\mathbf{Q}_{n \times n} = \mathrm{diag}(\mathbf{A1}) - \mathbf{A},$$

where the "diag" operator creates a diagonal matrix from the input vector and $\mathbf{1}$ is a conformable vector of 1s. Our initial RSR model implementation on the 180×180 analysis grid was prohibitively computationally expensive. We therefore computed the correlated random errors terms on a coarser, 20×20, grid collocated with the analysis grid. Thus, $n = 20$ in the RSR model description above. To obtain ω within each 10-km analysis grid cell, we simply assign it the ω of the corresponding (parent) 90-km RSR model grid cell. Using a 60×60 grid instead (i.e., 30-km RSR model grid cells) did not increase R^2 for model predictions of ALL (Fig. B1) but roughly quadrupled the run time. The

implied negligible impact of smoothing the 10-km ω is not surprising given that we do not attempt to model fine-scale structure in ω (Fig. B1b).

APPENDIX C

Impact on Early Period Analyses of Using 1990 Population Density Data

To qualitatively assess the errors that may arise early in the analysis period due to the unavailability of high-resolution census data prior to 1990, we repeated the (1975–79, …, 1988–92) 5-yr analyses for ALL tornadoes and (E)F0 tornadoes using the PD calculated from the 2010 census (Fig. C1). In the original experiments the PD were valid 13 and 0 years into the future for the 1975–79 and 1988–92 analysis periods, respectively, whereas in the new experiments the PD are valid 35 and 22 years into the future, respectively. Consistent with the general increase of population density during the analysis periods and the general increase of TRR with PD, using the 2010 PD inflates the TRR estimates and correspondingly deflates the λ estimates. The changes in the mean posterior TRR and λ arising from the increased PD errors are substantial for both ALL tornadoes (Fig. C1a) and, especially, (E)F0 tornadoes (Fig. C1b). However, given that the *largest* temporal error in PD in the original experiments is considerably less than the *smallest* temporal error in these new experiments (13 vs 22 yr), the corresponding TRR and λ errors in the original experiments, even at the earliest analysis periods, should be smaller than most of the differences between the original and 2010-PD experiments. Errors arising from the use of the 2010 PD for the (2009–13, …, 2014–18) analyses should be even smaller. We conclude that our use of only 1990, 2000, and 2010 PD data did not substantially degrade our analyses.

REFERENCES

Agee, E., and S. Childs, 2014: Adjustments in tornado counts, F-scale intensity, and path width for assessing significant

tornado destruction. *J. Appl. Meteor. Climatol.*, **53**, 1494–1505, https://doi.org/10.1175/JAMC-D-13-0235.1.

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015: Influence of El Niño–Southern oscillation on tornado and hail frequency in the United States. *Nat. Geosci.*, **8**, 278–283, https://doi.org/10.1038/ngeo2385.

Anderson, C. J., C. K. Wikle, Q. Zhou, and J. A. Royle, 2007: Population influences on tornado reports in the United States. *Wea. Forecasting*, **22**, 571–579, https://doi.org/10.1175/WAF997.1.

Ashley, W. S., and S. M. Strader, 2016: Recipe for disaster: How the dynamic ingredients of risk and exposure are changing the tornado disaster landscape. *Bull. Amer. Meteor. Soc.*, **97**, 767–786, https://doi.org/10.1175/BAMS-D-15-00150.1.

——, S. Strader, T. Rosencrants, and A. J. Krmenec, 2014: Spatiotemporal changes in tornado hazard exposure: The case of the expanding bull's-eye effect in Chicago, Illinois. *Wea. Climate Soc.*, **6**, 175–193, https://doi.org/10.1175/WCAS-D-13-00047.1.

Barrett, B. S., and V. A. Gensini, 2013: Variability of central United States April–May tornado day likelihood by phase of the Madden–Julian oscillation. *Geophys. Res. Lett.*, **40**, 2790–2795, https://doi.org/10.1002/grl.50522.

Brooks, H. E., 2004: On the relationship of tornado path length and width to intensity. *Wea. Forecasting*, **19**, 310–319, https://doi.org/10.1175/1520-0434(2004)019<0310:OTROTP>v2.0.CO;2.

——, C. A. Doswell, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640, https://doi.org/10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2.

——, G. W. Carbin, and P. T. Marsh, 2014: Increased variability of tornado occurrence in the United States. *Science*, **346**, 349–352, https://doi.org/10.1126/science.1257460.

Broyles, J. C., and K. C. Crosbie, 2004: Evidence of smaller tornado alleys across the United States based on a long track F3–F5 tornado climatology study from 1880–2003. *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., P5.6, https://ams.confex.com/ams/pdfpapers/81872.pdf.

Bubnicki, J. W., M. Churski, K. Schmidt, T. A. Diserens, and D. P. J. Kuijper, 2019: Linking spatial patterns of terrestrial herbivore community structure to trophic interactions. *eLife*, **8**, e44937, https://doi.org/10.7554/eLife.44937.001.

Cheng, V. Y. S., G. B. Arhonditsis, D. M. L. Sills, H. Auld, M. W. Shephard, W. A. Gough, and J. Klaassen, 2013: Probability of tornado occurrence across Canada. *J. Climate*, **26**, 9415–9428, https://doi.org/10.1175/JCLI-D-13-00093.1.

——, ——, ——, W. A. Gough, and H. Auld, 2015: A Bayesian modelling framework for tornado occurrences in North America. *Nat. Commun.*, **6**, 6599, https://doi.org/10.1038/ncomms7599.

——, ——, ——, ——, and ——, 2016: Predicting the climatology of tornado occurrences in North America with a Bayesian hierarchical modeling framework. *J. Climate*, **29**, 1899–1917, https://doi.org/10.1175/JCLI-D-15-0404.1.

Childs, S. J., R. S. Schumacher, and J. L. Demuth, 2020: Agricultural perspectives on hailstorm severity, vulnerability, and risk messaging in eastern Colorado. *Wea. Climate Soc.*, **12**, 897–911, https://doi.org/10.1175/WCAS-D-20-0015.1.

Clark, A. J., J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012: Forecasting tornado pathlengths using a three-dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113, https://doi.org/10.1175/WAF-D-11-00147.1.

Coleman, T. A., and P. G. Dixon, 2014: An objective analysis of tornado risk in the United States. *Wea. Forecasting*, **29**, 366–376, https://doi.org/10.1175/WAF-D-13-00057.1.

Cook, A. R., L. M. Leslie, D. B. Parsons, and J. T. Schaefer, 2017: The impact of El Niño–Southern Oscillation (ENSO) on winter and early spring U.S. tornado outbreaks. *J. Appl. Meteor. Climatol.*, **56**, 2455–2478, https://doi.org/10.1175/JAMC-D-16-0249.1.

Dahl, N. A., D. S. Nolan, G. H. Bryan, and R. Rotunno, 2017: Using high-resolution simulations to quantify underestimates of tornado intensity from in situ observations. *Mon. Wea. Rev.*, **145**, 1963–1982, https://doi.org/10.1175/MWR-D-16-0346.1.

Doswell, C. A., III, and D. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.*, **116**, 495–501, https://doi.org/10.1175/1520-0493(1988)116<0495:OSIOUS>2.0.CO;2.

——, H. E. Brooks, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563, https://doi.org/10.1016/j.atmosres.2008.11.003.

Edwards, R., R. L. Thompson, K. C. Brosbie, J. A. Hart, and C. A. Doswell III, 2004: Proposals for modernizing the definitions of tornado and severe thunderstorm outbreaks. *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 7.B.2, http://ams.confex.com/ams/pdfpapers/81342.pdf.

——, H. E. Brooks, and H. Cohn, 2021: Changes in tornado climatology accompanying the enhanced Fujita scale. *J. Appl. Meteor. Climatol.*, **60**, 1465–1482, https://doi.org/10.1175/JAMC-D-21-0058.1.

Elsner, J. B., and H. M. Widen, 2014: Predicting spring tornado activity in the central Great Plains by 1 March. *Mon. Wea. Rev.*, **142**, 259–267, https://doi.org/10.1175/MWR-D-13-00014.1.

——, L. E. Michaels, K. N. Scheitlin, and I. J. Elsner, 2013: The decreasing population bias in tornado reports across the central plains. *Wea. Climate Soc.*, **5**, 221–232, https://doi.org/10.1175/WCAS-D-12-00040.1.

——, T. H. Jagger, and T. Fricker, 2016: Statistical models for tornado climatology: Long and short-term views. *PLOS ONE*, **11**, e0166895, https://doi.org/10.1371/journal.pone.0166895.

Fuhrmann, C. M., C. E. Konrad, M. M. Kovach, J. T. McLeod, W. G. Schmitz, and P. G. Dixon, 2014: Ranking of tornado outbreaks across the United States and their climatological characteristics. *Wea. Forecasting*, **2**, 684–701, https://doi.org/10.1175/WAF-D-13-00128.1.

Gensini, V. A., and H. E. Brooks, 2018: Spatial trends in United States tornado frequency. *npj Climate Atmos. Sci.*, **1**, 38, https://doi.org/10.1038/s41612-018-0048-2.

Grieser, J., and F. Terenzi, 2016: Modeling financial losses resulting from tornadoes in European countries. *Wea. Climate Soc.*, **8**, 313–326, https://doi.org/10.1175/WCAS-D-15-0036.1.

Guo, L., K. Wang, and H. B. Bluestein, 2016: Variability of tornado occurrence over the continental United States since 1950. *J. Geophys. Res. Atmos.*, **121**, 6943–6953, https://doi.org/10.1002/2015JD024465.

Hall, S. G., and W. S. Ashley, 2008: Effects of urban sprawl on the vulnerability to a significant tornado impact in northeastern Illinois. *Nat. Hazards Rev.*, **9**, 209–219, https://doi.org/10.1061/(ASCE)1527-6988(2008)9:4(209).

Hamed, K. H., and A. R. Rao, 1998: A modified Mann–Kendall trend test for autocorrelated data. *J. Hydrol.*, **204**, 182–196, https://doi.org/10.1016/S0022-1694(97)00125-X.

Hoffman, M. D., and A. Gelman, 2014: The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, **15**, 1593–1623.

Hughes, J., and M. Haran, 2013: Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. Roy. Stat. Soc.*, **75A**, 139–159, https://doi.org/10.1111/j.1467-9868.2012.01041.x.

Hussain, M., and I. Mamhud, 2019: pyMannKendall: A Python package for non parametric Mann Kendall family of trend tests. *J. Open Source Software*, **4**, 1556, https://doi.org/10.21105/joss.01556.

Lee, B. D., and R. B. Wilhelmson, 1997: The numerical simulation of non-supercell tornadogenesis. Part I: Initiation and evolution of pretornadic misocyclone circulations along a dry outflow boundary. *J. Atmos. Sci.*, **54**, 32–60, https://doi.org/10.1175/1520-0469(1997)054<0032:TNSONS>2.0.CO;2.

Lee, C. C., 2012: Utilizing synoptic climatological methods to assess the impacts of climate change on future tornado-favorable environments. *Nat. Hazards*, **62**, 325–343, https://doi.org/10.1007/s11069-011-9998-y.

McCarthy, D. W., and J. T. Schaefer, 2004: Tornado trends over the past thirty years. *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., 3.4, https://ams.confex.com/ams/pdfpapers/72089.pdf.

McFarquhar, G. M., and Coauthors, 2020: Current and future uses of UAS for improved forecasts/warnings and scientific studies. *Bull. Amer. Meteor. Soc.*, **101**, E1322–E1328, https://doi.org/10.1175/BAMS-D-20-0015.1.

Nouri, N., N. Devineni, V. Were, and R. Khanbilvardi, 2021: Explaining the trends and variability in the United States tornado records using climate teleconnections and shifts in observational practices. *Sci. Rep.*, **11**, 1741, https://doi.org/10.1038/s41598-021-81143-5.

Potvin, C. K., C. Broyles, P. S. Skinner, H. E. Brooks, and E. Rasmussen, 2019: A Bayesian hierarchical modeling framework for correcting reporting bias in the U.S. tornado database. *Wea. Forecasting*, **34**, 15–30, https://doi.org/10.1175/WAF-D-18-0137.1.

Reich, B. J., J. S. Hodges, and V. Zadnik, 2006: Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, **62**, 1197–1206, https://doi.org/10.1111/j.1541-0420.2006.00617.x.

Romanic, D., M. Refan, C.-H. Wu, and G. Michel, 2016: Oklahoma tornado risk and variability: A statistical model. *Int. J. Disaster Risk Reduct.*, **16**, 19–32, https://doi.org/10.1016/j.ijdrr.2016.01.011.

Rue, H., and L. Held, 2005: *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC Press, 280 pp.

Salvatier, J., T. V. Wiecki, and C. Fonnesbeck, 2016: Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.*, **2**, e55, https://doi.org/10.7717/peerj-cs.55.

Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. Preprints, *11th Conf. on Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 603–606, https://ams.confex.com/ams/99annual/abstracts/1360.htm.

——, D. L. Kelly, and R. F. Abbey, 1986: A minimum assumption tornado-hazard probability model. *J. Climate Appl. Meteor.*, **25**, 1934–1945, https://doi.org/10.1175/1520-0450(1986)025<1934:AMATHP>2.0.CO;2.

Simmons, K. M., P. Kovacs, and G. A. Kopp, 2015: Tornado damage mitigation: Benefit–cost analysis of enhanced building codes in Oklahoma. *Wea. Climate Soc.*, **7**, 169–178, https://doi.org/10.1175/WCAS-D-14-00032.1.

Strader, S. M., and W. S. Ashley, 2015: The expanding bull's-eye effect. *Weatherwise*, **68**, 23–29, https://doi.org/10.1080/00431672.2015.1067108.

——, T. J. Pingel, and W. S. Ashley, 2016: A Monte Carlo model for estimating tornado impacts. *Meteor. Appl.*, **23**, 269–281, https://doi.org/10.1002/met.1552.

——, W. S. Ashley, T. J. Pingel, and A. J. Krmenec, 2017: Observed and projected changes in United States tornado exposure. *Wea. Climate Soc.*, **9**, 109–123, https://doi.org/10.1175/WCAS-D-16-0041.1.

Tippett, M. K., J. T. Allen, V. A. Gensini, and H. E. Brooks, 2015: Climate and hazardous convective weather. *Curr. Climate Change Rep.*, **1**, 60–73, https://doi.org/10.1007/s40641-015-0006-6.

Trapp, R. J., and K. A. Hoogewind, 2018: Exploring a possible connection between U.S. tornado activity and Arctic sea ice. *npj Climate Atmos. Sci.*, **1**, 14, https://doi.org/10.1038/s41612-018-0025-9.

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954–2003. *Wea. Forecasting*, **21**, 86–93, https://doi.org/10.1175/WAF910.1.

Wagner, M., R. K. Doe, A. Johnson, Z. Chen, J. Das, and R. S. Cerveny, 2019: Unpiloted aerial systems (UASs) application for tornado damage surveys: Benefits and procedures. *Bull. Amer. Meteor. Soc.*, **100**, 2405–2409, https://doi.org/10.1175/BAMS-D-19-0124.1.

——, ——, C. Wang, E. Rasmussen, M. C. Coniglio, K. L. Elmore, R. C. Balling, and R. S. Cerveny, 2021: High-resolution observations of microscale influences on a tornado track using unpiloted aerial systems (UAS). *Mon. Wea. Rev.*, **149**, 2819–2834, https://doi.org/10.1175/MWR-D-20-0213.1.

Widen, H. M., and Coauthors, 2013: Adjusted tornado probabilities. *Electron. J. Severe Storms Meteor.*, **8** (7), https://ejssm.org/archives/wp-content/uploads/2021/10/vol8-7.pdf.

Wikle, C. K., and C. J. Anderson, 2003: Climatological analysis of tornado report counts using a hierarchical Bayesian spatio-temporal model. *J. Geophys. Res.*, **108**, 9005, https://doi.org/10.1029/2002JD002806.

Wind Science and Engineering Center, 2006: A recommendation for an enhanced Fujita scale (EF-scale), revision 2. Texas Tech University Doc., 95 pp., http://www.depts.ttu.edu/nwi/pubs/fscale/efscale.pdf.

Wurman, J., J. Straka, E. Rasmussen, M. Randall, and A. Zahrai, 1997: Design and deployment of a portable, pencil-beam, pulsed, 3-cm Doppler radar. *J. Atmos. Oceanic Technol.*, **14**, 1502–1512, https://doi.org/10.1175/1520-0426(1997)014<1502:DADOAP>2.0.CO;2.

——, K. Kosiba, T. White, and P. Robinson, 2021: Supercell tornadoes are much stronger and wider than damage-based ratings indicate. *Proc. Natl. Acad. Sci.*, **118**, e2021535118, https://doi.org/10.1073/pnas.2021535118.