

A Bayesian Hierarchical Modeling Framework for Correcting Reporting Bias in the U.S. Tornado Database

COREY K. POTVIN

Cooperative Institute for Mesoscale Meteorological Studies, and School of Meteorology, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

CHRIS BROYLES

NOAA/NWS/NCEP/Storm Prediction Center, Norman, Oklahoma

PATRICK S. SKINNER

Cooperative Institute for Mesoscale Meteorological Studies, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

HAROLD E. BROOKS

NOAA/OAR/National Severe Storms Laboratory, and School of Meteorology, University of Oklahoma, Norman, Oklahoma

ERIK RASMUSSEN

Cooperative Institute for Mesoscale Meteorological Studies, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 14 August 2018, in final form 17 October 2018)

ABSTRACT

The Storm Prediction Center (SPC) tornado database, generated from NCEI's *Storm Data* publication, is indispensable for assessing U.S. tornado risk and investigating tornado–climate connections. Maximizing the value of this database, however, requires accounting for systemically lower reported tornado counts in rural areas owing to a lack of observers. This study uses Bayesian hierarchical modeling to estimate tornado reporting rates and expected tornado counts over the central United States during 1975–2016. Our method addresses a serious solution nonuniqueness issue that may have affected previous studies. The adopted model explains 73% (>90%) of the variance in reported counts at scales of 50 km (>100 km). Population density explains more of the variance in reported tornado counts than other examined geographical covariates, including distance from nearest city, terrain ruggedness index, and road density. The model estimates that approximately 45% of tornadoes within the analysis domain were reported. The estimated tornado reporting rate decreases sharply away from population centers; for example, while >90% of tornadoes that occur within 5 km of a city with population > 100 000 are reported, this rate decreases to <70% at distances of 20–25 km. The method is directly extendable to other events subject to underreporting (e.g., severe hail and wind) and could be used to improve climate studies and tornado and other hazard models for forecasters, planners, and insurance/reinsurance companies, as well as for the development and verification of storm-scale prediction systems.

1. Introduction

Tornado climatologies have a wide range of valuable applications, including developing economic loss models and benefit–cost analyses for tornadoes (e.g., [Simmons](#)

[et al. 2015](#); [Grieser and Terenzi 2016](#); [Romanic et al. 2016](#)); evaluating tornado warning performance and how it has changed over time (e.g., [Brooks and Correia 2018](#)); statistically modeling physical influences on tornado frequency (e.g., [Kellner and Niyogi 2014](#); [Elsner et al. 2016a](#)); investigating relationships between tornado length, width, and intensity (e.g., [Brooks 2004](#); [Agee and Childs 2014](#));

Corresponding author: Dr. Corey K. Potvin, corey.potvin@noaa.gov

DOI: 10.1175/WAF-D-18-0137.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

analyzing tornado risk (Tecson et al. 1983; Schaefer et al. 1986; Coleman and Dixon 2014); and identifying climate signals in tornado activity (e.g., Brooks et al. 2014; Allen et al. 2015; Guo et al. 2016; Cook et al. 2017). The success of such efforts, however, can be severely compromised by database errors arising from reporting biases. Of particular interest to the present study is the tendency for tornadoes in rural areas to go unreported due to a lack of observers, which produces artificial spatial and temporal trends in analyses of tornado frequency. Much effort has been devoted to estimating and removing this underreporting bias (e.g., Ray et al. 2003; Feuerstein et al. 2005; Verbout et al. 2006; Anderson et al. 2007; Widen et al. 2013; Agee and Childs 2014; Elsner et al. 2016b).

While the influence of reporting errors on the tornado database is readily recognized, quantifying the resulting biases is complicated by at least three problems arising from general limitations of statistical modeling. First, the period of the database is short enough that sampling error produces locally significant departures of the observed tornado frequency from the underlying (i.e., expected) distributions at smaller spatial scales. This is because tornado reports cluster in space and time owing to 1) the random clustering that occurs even for completely spatially random processes and 2) the additional clustering that results from tornado outbreaks and families (i.e., groups of tornadoes spawned by the same thunderstorm). The resulting sampling errors must be carefully accounted for to avoid data overfitting (i.e., fitting noise rather than the underlying distribution).

Second, the underlying tornado distributions may themselves be substantially nonuniform in space or time at the smaller scales of interest, making it difficult to distinguish real from artificial gradients in tornado frequency. For example, it has been hypothesized that tornado frequency maxima exist that are much smaller scale than the “tornado alleys” defined in the literature due, for example, to local topography (Broyles and Crosbie 2004). It is also possible that tornado distributions have changed during the database period due to anthropogenic climate change or natural climate variability (e.g., Allen et al. 2015; Brooks et al. 2014; Tippett et al. 2015). Departures from spatial or temporal homogeneity therefore cannot be immediately attributed to geographical or temporal trends in reporting bias.

Third, while many variables have already been implicated in tornado reporting bias, including population (e.g., Anderson et al. 2007), distance from nearest city (e.g., Widen et al. 2013), and distance from nearest Weather Surveillance Radar-1988 Doppler (WSR-88D; Ray et al. 2003), there may be other variables that are yet unknown to be important or that are difficult to measure and include in the analysis. For example, it is

plausible that cultural differences between National Weather Service (NWS) Weather Forecast Offices (WFOs) impact the probability of suspected tornadoes in distant rural areas being surveyed (e.g., Doswell 2007), but it is unclear how to quantify this effect. Omitting explanatory variables results in a portion of the variation in the sample population being misattributed to the model covariate(s), producing errors in the diagnosed dependencies of the predicted variable. When this “omitted-variable bias” is suspected to be large, care must be taken in interpreting the results.

With these considerations in mind, we develop a Bayesian hierarchical model to estimate tornado reporting rates (TRRs) and actual expected tornado counts from the SPC tornado database. The method incorporates elements of previously published Bayesian approaches to this problem but adds a novel feature that addresses a serious solution nonuniqueness issue that may have impacted previous studies. The nonuniqueness arises from the fact that similar numbers of tornado reports can occur with high actual tornado counts and low TRRs as with low actual tornado counts and high TRRs. Our analysis domain covers most of the central United States for the period 1975–2016. We motivate and describe our Bayesian model in section 2. In section 3, we describe the development and cross validation of the model. The results of the final model-predicted TRR and expected tornado counts are presented in section 4. Finally, potentially valuable extensions of this work are summarized in section 5.

2. Data and Bayesian model

a. Gridding reported tornado counts

We use the 1975–2016 portion of the SPC tornado database for analysis. Omitting earlier years from the analysis avoids the systematic overrating of tornadoes that occurred prior to the adoption of the Fujita scale in 1975 (Brooks and Craven 2002). Following Elsner et al. (2016b), tornado records having the same starting location, date and time, length, and width as another record in the database are flagged as duplicates and removed. This results in the omission of 329 out of 45 954 reports (~0.7%). The remaining reports are then tallied within 10-km-diameter grid cells on an 1800 km × 1800 km Mercator grid centered at 38.7°N, 92.0°W (Fig. 1a). The northeastern 600 km × 600 km corner of the domain is excluded from the analysis to avoid difficulties with grid cells within the Great Lakes. The analysis domain includes ~62% of all (nonduplicate) 1975–2016 tornado reports in the database. Tornadoes are treated as points located at the midpoint of the line

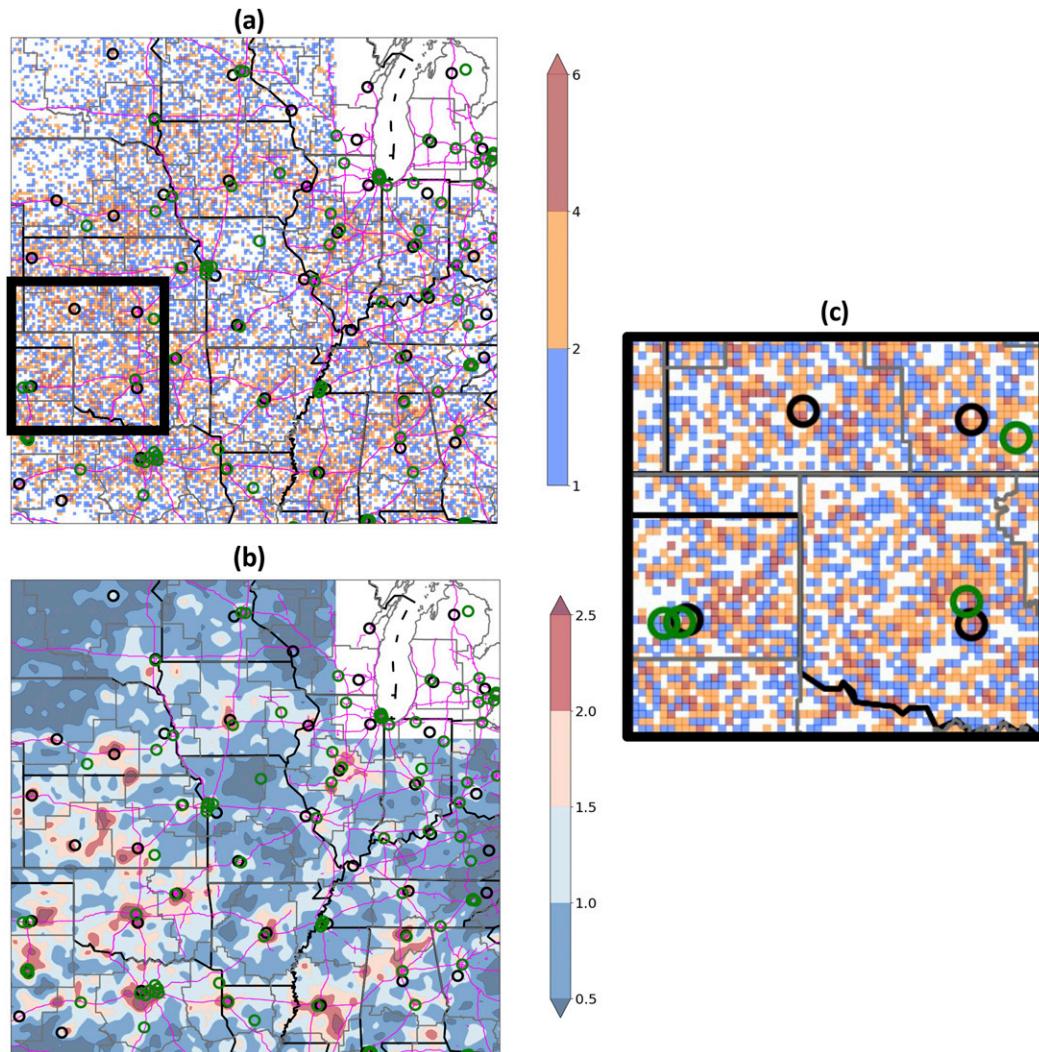


FIG. 1. Reported 1975–2016 tornadoes (a) tallied over 10-km grid cells, then (b) smoothed with a Gaussian kernel of width $2\sigma = 20$ km to aid interpretation. Tornado counts in the northeastern corner of the domain are not shown since they are not used in the analysis. In this and subsequent figures, NWS county warning area (CWA) borders are represented in gray, WFOs by black circles, cities with population $\geq 100\,000$ (“100K cities”) in 1990 by green circles, and interstate highways in magenta. (c) The region enclosed in the black square in (a) is magnified to better illustrate the noisiness of the data.

segment bounded by the recorded start and end coordinates. If no end location is recorded (typically because the tornado track is too short), the tornado location is set to the start coordinate.

Visual inspection of the gridded tornado counts (Fig. 1) reveals two important considerations. First, the noisiness of the counts at smaller scales (Figs. 1a,c) suggests much of the variance therein arises from sampling error, which is expected given the brevity of the record. Second, the strong correlation of local count maxima with cities (Fig. 1b) indicates that underreporting of tornadoes in rural areas introduces severe bias even at larger scales. It is critical to account

for both errors when interpreting and analyzing reported tornado counts.

b. TRR model

Following previous studies, we use a Bayesian hierarchical model (Davidson-Pilon 2016) comprising a TRR model and a spatial process model to calculate posterior distributions of TRRs and expected tornado counts λ conditional on the reported counts N . TRR is modeled as a function of a single geographical covariate (e.g., population density). The spatial process model is necessary to account for the stochastic nature of tornado occurrence. We use the Python pymc3 module (Salvatier et al. 2016)

to obtain Markov chain Monte Carlo (MCMC) samples of the posterior distributions. Parameter trace plots and the Geweke (1992) convergence diagnostic are used to confirm the suitability of the prior distribution assigned to each model parameter and of the prescribed “burn in” period for each Markov chain.¹

We experimented with several geographical covariates for the TRR model, some of which are computed at the center of the grid cell (distance-related parameters) and others of which are averaged over the grid cell (remaining parameters). The tested covariates are population density P ; terrain ruggedness index (TRI; Riley et al. 1999); road density R ; tree canopy percentage T , and distance to nearest 100 000 resident city C , 5000 resident city c , WFO W , interstate I , WFO or 100 000 resident city CW , and interstate or 5000 resident city cI . Several of these covariates are plotted in Fig. 2. City population data, required for computing C , c , CW , and cI , were available for 1990 and 2000. Gridded population data, required for computing P , were available for 1990, 2000, and 2010. Temporal interpolation was not used to estimate population in intervening years; rather, the city and gridded population datasets valid nearest the mean year of the analysis period were used. The gridded population data were averaged from their native 1-km grids onto the 10-km analysis grid. As will be shown in section 3a, P explains more variance in reported counts than any of the other tested covariates.

The relationship of TRR to each geographical covariate can be crudely estimated using the following procedure:

- 1) To account for large-scale gradients in actual tornado frequency, divide the reported tornado count within each grid cell by the mean of the gridded counts within a suitable (herein, a 250-km square) neighborhood.
- 2) Bin each covariate (e.g., $C = 0\text{--}5$ km, $5\text{--}10$ km, and so on) and identify the set of grid cells lying within each bin.
- 3) Average the normalized counts within each bin.
- 4) Consider the ratios of average normalized counts between bins; these provide estimates of the ratios of domain-mean TRR between bins.

¹ The burn-in period is the series of initial Markov chain states that is considered unrepresentative of the distribution being sampled and is, therefore, discarded from the final Markov chain output. This period can be determined subjectively by inspecting trace plots of the Bayesian parameters and estimating how many iterations were required for their means to converge, or objectively by using the Geweke (1992) diagnostic to determine whether a statistically significant shift in each parameter mean has taken place by a prescribed number of iterations.

The analysis reveals that TRR is quite sensitive to each of the geographical covariates (Fig. 3). An important implication of the resulting large TRR gradients is that making the analysis grid spacing too large, and therefore the gridded tornado counts too smooth, artificially decreases (increases) model-predicted TRR in urban (rural) areas. On the other hand, reducing the analysis grid spacing rapidly increases the computational cost of the model. We found that 10-km grid spacing provides an acceptable trade-off between these effects for our dataset. Rerunning the final model (described in section 3) with 5-km grid spacing produces mean posterior TRR and λ results (not shown) very similar to those produced with 10-km grid spacing (section 4b).

We model TRR as either a negative exponential function or a fourth-degree polynomial function² of the geographical covariate x :

$$\text{TRR} = \text{TRR}_{\min} + (1 - \text{TRR}_{\min})\exp(-\beta x), \quad \text{and} \quad (1)$$

$$\text{TRR} = 1.0 + ax + bx^2 + cx^3 + dx^4, \quad (2)$$

where β is assigned a lognormal prior with $\mu = 0.5$ and $\sigma = 2$; a , b , c , and d are assigned normal priors with $\mu = 0$ and $\sigma = 10$; and TRR_{\min} is a constant that prevents the TRR posteriors from falling below a prescribed limit. Both TRR models are designed such that $\text{TRR} \rightarrow 1$ as $x \rightarrow 0$; the reason for this is described in section 2c. Since $x \geq 0$ and $\beta \geq 0$, $\text{TRR} \in [\text{TRR}_{\min}, 1]$ for the negative exponential model. When using the polynomial model, TRR is clipped to $[\text{TRR}_{\min}, 1]$ during the MCMC sampling.

The use of TRR_{\min} in the TRR models was motivated by preliminary experiments with P in which the TRR posterior locally approached zero in the (sparsely populated) western part of the domain when a lower limit was not imposed. While this unrealistic behavior of the TRR models did not degrade the posterior N in those (or other) regions, it did inflate the posterior λ to compensate for the negatively biased TRR. The source and our solution for this parameter aliasing problem are explained in section 2c. Since P explains more variance in reported counts than any of the other tested covariates, we are only interested in TRR estimates from the TRR(P) model, and therefore we set $\text{TRR}_{\min} = 0$ in experiments with other x .

To stabilize the MCMC sampling, especially when the polynomial model is used, we transform each x such that $x_i \in [0, 1]$. For covariates other than P , we obtain x_i by dividing x by a constant. For P , which has a larger

² In preliminary tests, third- and fifth-degree polynomials tended to underfit and overfit the data, respectively.

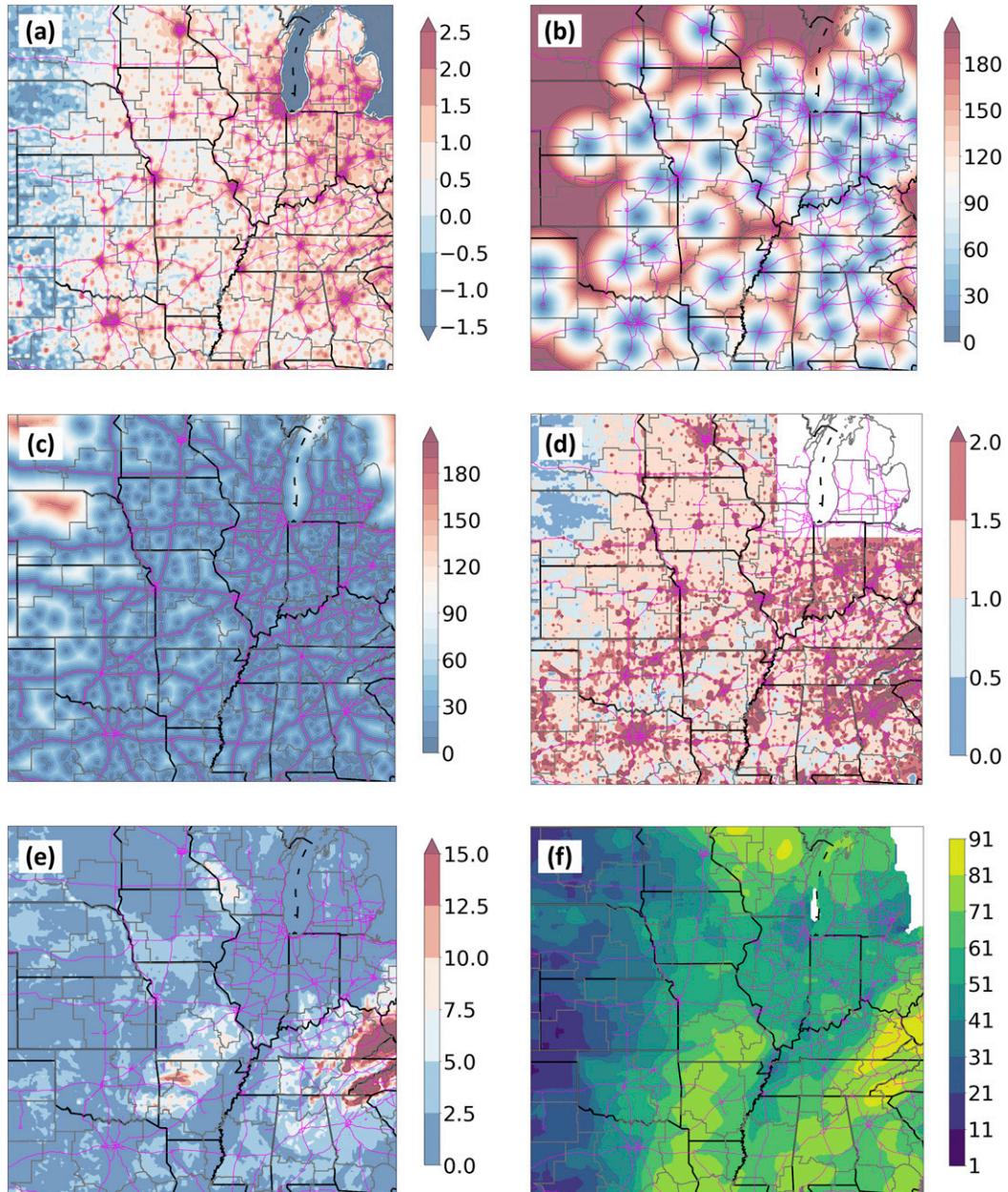


FIG. 2. Maps of several of the geographical variables tested as covariates in the Bayesian model: (a) base-10 logarithm of population density (km^{-2}), (b) distance from nearest 100K city (km), (c) distance from nearest 5K city or interstate (km), (d) road density [$^{\circ} (0.1^{\circ})^{-2}$], (e) terrain ruggedness index [$\text{m} (0.1^{\circ})^{-2}$], and (f) percent tree canopy. The 1990 population data were used for (a)–(c). The top-right corner in (d) is masked since the road density dataset excludes part of this region.

“dynamic range” than the other covariates, we first impose a lower bound of 0.001, then transform the result as follows:

$$P_t = [\log_{10}(P) + 4]/10. \quad (3)$$

Preliminary experiments with multivariable models (both negative exponential and polynomial) for TRR

failed to produce nontrivial improvements over the simple P model. It seems unlikely, however, that TRR is completely determined by P (or any single variable). This suggests that the noise in the reported tornado counts inhibits the use of more complex models.

Some previous studies have modeled reported tornado counts with a spatial Poisson distribution (e.g.,

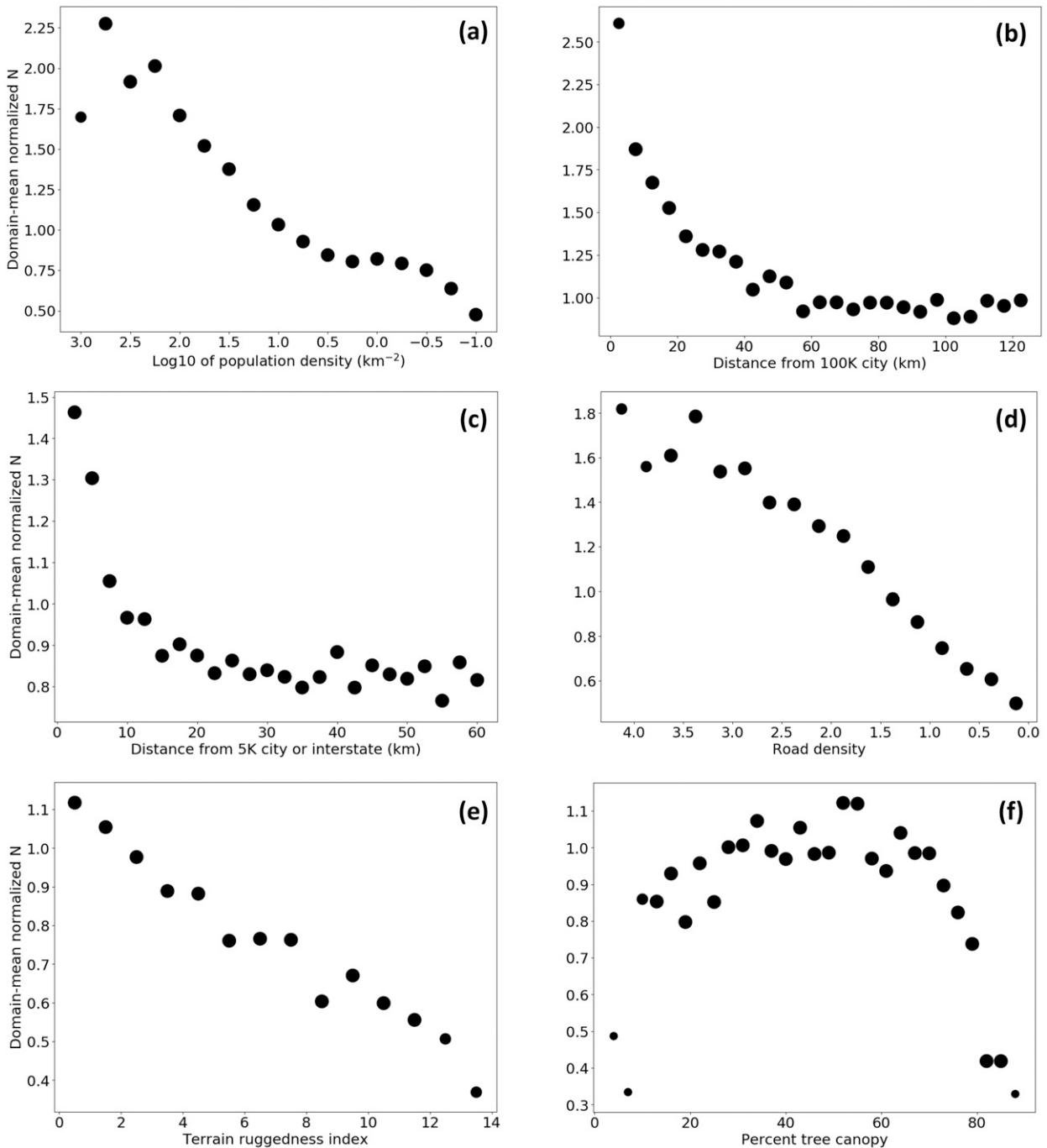


FIG. 3. Domain-mean normalized N vs (a) base-10 logarithm of population density (km^{-2}), (b) distance from nearest 100K city (km), (c) distance from nearest 5K city or interstate (km), (d) road density [$^{\circ} (0.1^{\circ})^{-2}$], (e) terrain ruggedness index [$\text{m} (0.1^{\circ})^{-2}$], and (f) percent tree canopy. Large, medium, and small dots represent gridpoint counts > 100 , 30–100, and < 30 , respectively.

Anderson et al. 2007). This distribution describes a spatial point pattern generated by a completely spatially random process with prescribed frequency λ , which in our application is the expected actual tornado count per grid cell. Tornadogenesis locations are not completely

spatially random, however; tornadic thunderstorm outbreaks and tornado families lead to considerable overdispersion, as evidenced by variance-to-mean ratios well above unity in both reported (Elsner and Widen 2014) and model-predicted actual tornado counts (not shown).

Thus, following [Elsner and Widen \(2014\)](#) and [Elsner et al. \(2016a\)](#), we adopt a negative binomial (NegBin) distribution model for reported tornado counts, which allows the variance of the counts to vary independently of the mean. The pymc3 formulation of the negative binomial distribution follows a gamma distribution with $\mu = \lambda$ and shape parameter α . We assign α a half-normal prior with $\sigma = 100$ and set the expected reported tornado count to $\text{TRR} \times \lambda$, as in [Elsner et al. \(2016a\)](#). Thus, the distribution of reported tornado counts N within each grid cell over the analysis period is modeled as

$$N = \text{NegBin}(\mu = \text{TRR} \times \lambda, \alpha = \alpha). \quad (4)$$

The λ are assigned uniform priors over $[0, 10]$.

c. Addressing solution nonuniqueness

In preliminary experiments, the Bayesian model was run independently within individual subdomains similar to [Anderson et al. \(2007\)](#), except that the subdomains were square with diameter D set to 100 or 300 km. The resulting TRR and λ posteriors contained unrealistic discontinuities between adjacent subdomains (not shown). This problem arises primarily from a solution nonuniqueness problem that, to the authors' knowledge, has not been addressed in the literature. Specifically, the product $\text{TRR} \times \lambda$ in the Bayesian model [Eqs. (4) and (5)] means similar N values are predicted for low TRR and high λ as for high TRR and low λ (i.e., aliasing occurs between TRR and λ ; the end of [section 2b](#) alluded to this problem). Thus, the TRR and λ posteriors will be unrealistically sensitive to the vagaries of the data and model parameters and are, therefore, meaningless. Obtaining high-confidence estimates of TRR and λ therefore requires constraining one or the other in some reasonable way.³ We have already explained ([section 2b](#)) that we impose a lower limit on TRR (TRR_{\min}) to address this solution nonuniqueness. Fully resolving the problem, however, additionally required imposing an upper limit on TRR. We did this by adding a fixed parameter P_{\max} to the population density covariate formula [Eq. (4)] such that $P_t \geq P_{\max} = 0$ and, thus, $\text{TRR}(P_t \geq P_{\max}) = 1$. That is, we assume all tornadoes are reported within grid cells where the population density exceeds a prescribed threshold. The final P input to the model is therefore

$$P_{\text{final}} = \begin{cases} 0, & P_t \geq P_{\max} \\ P_{\max} - P_t, & P_t < P_{\max} \end{cases}. \quad (5)$$

³ We experimented with including a conditional autoregressive model for λ and training the model over the entire analysis domain simultaneously, as in [Elsner et al. \(2016b\)](#). The resulting TRR posteriors, however, were often obviously biased, indicating the additional constraint did not mitigate the solution nonuniqueness.

The procedure for selecting P_{\max} is described in [section 3b](#).

Constraining TRR in this way greatly reduced, but did not totally eliminate, the spurious intersubdomain discontinuities in model-predicted TRR (not shown). Attributing the remaining discontinuities primarily to insufficient sample sizes, we then reformulated the hierarchical model such that TRR is trained over the entire analysis domain simultaneously but λ is still trained independently within (i.e., allowed to vary between) square analysis subdomains. The combination of all three approaches mostly eliminated the remaining intersubdomain discontinuities. With these important modifications, our final Bayesian model produces posteriors of domain-constant β or (a, b, c, d) , depending on the TRR model used; a corresponding TRR posterior valid on the analysis grid; and a λ posterior valid on a coarser grid with spacing D .

3. Model development

a. Prediction metrics

Given how noisy the reported tornado counts are ([Fig. 1a](#)), there is serious potential for fitting meaningless patterns in the data, especially with our polynomial TRR model. We therefore use k -fold cross validation ([Efron and Tibshirani 1993](#)) to assess and compare the predictive capabilities of different models. The method proceeds as follows. First, the original sample (of analysis subdomains, in our case) is divided into equally sized (to the extent possible) subsamples. The model is trained on $k - 1$ of the subsamples and then validated on the remaining subsample. This training–testing process is repeated until each of the k subsamples has served as the testing set. We chose $k = 10$ for our tests; this is a common choice of k since it allows most of the sample to be used in each training fold and provides reasonably low-variance error estimates without being unduly computationally expensive. We select as validation metrics the coefficient of determination R^2 , sample Pearson correlation coefficient r , and relative root-mean-square error (RRMSE). We also compare domain-mean reported and out-of-sample predicted counts as functions of each geographical variable to further assess how well the model captures the geographical dependencies of N .

b. Sensitivity to geographical covariate

Cross validation was performed for each TRR model covariate using both the polynomial and negative exponential models. While $D = 100$ km was adopted for the final model ([section 3c](#)), these tests used $D = 300$ km for computational expediency and because the relative performance of the TRR models with different covariates

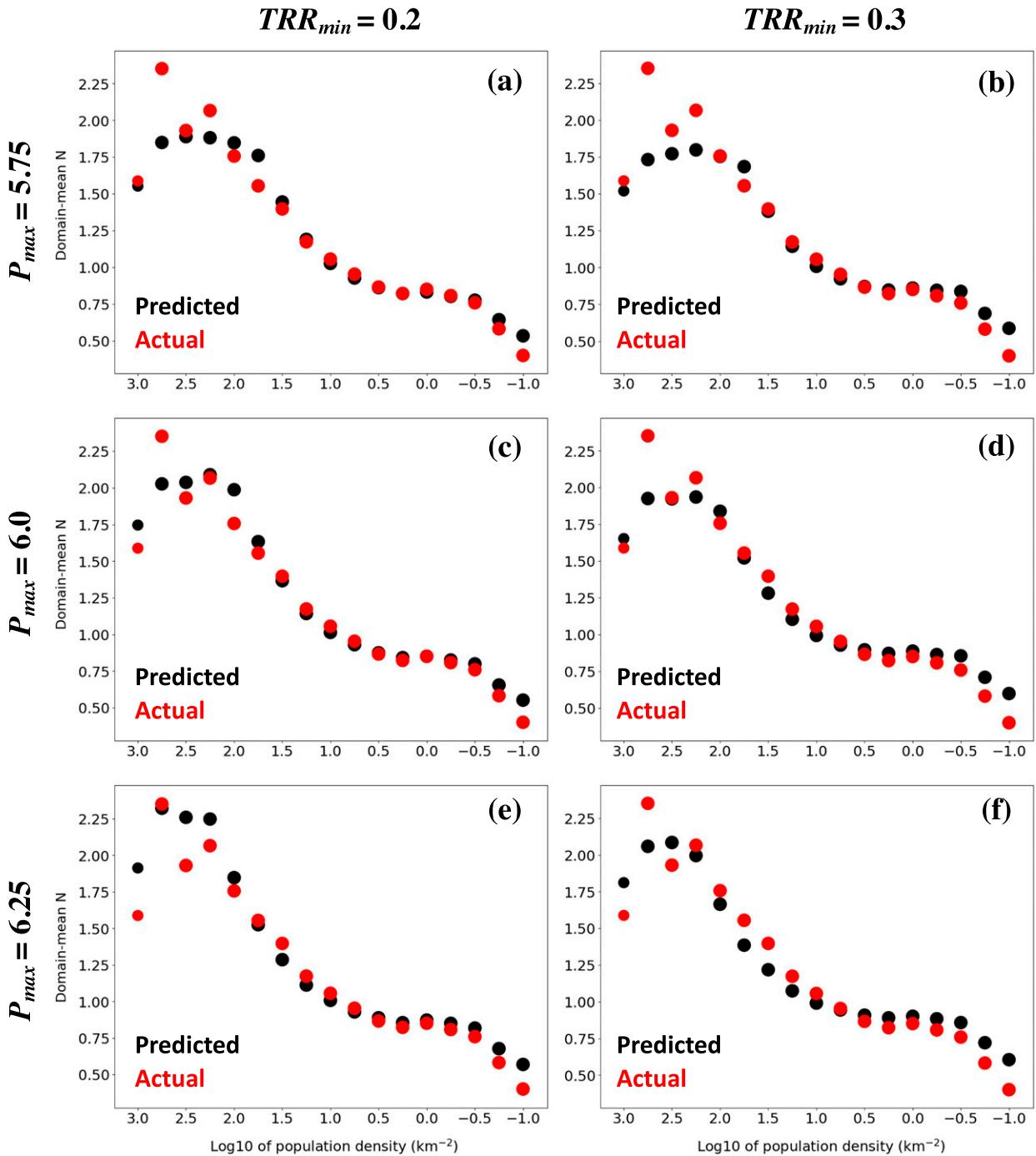


FIG. 4. Domain-mean, out-of-sample predicted N (black) and actual N (red) vs base-10 logarithm of population density (km^{-2}) for (a) $P_{\max} = 5.75$, $\text{TRR}_{\min} = 0.2$; (b) $P_{\max} = 5.75$, $\text{TRR}_{\min} = 0.3$; (c) $P_{\max} = 6.0$, $\text{TRR}_{\min} = 0.2$; (d) $P_{\max} = 6.0$, $\text{TRR}_{\min} = 0.3$; (e) $P_{\max} = 6.25$, $\text{TRR}_{\min} = 0.2$; and (f) $P_{\max} = 6.25$, $\text{TRR}_{\min} = 0.3$.

decrease just outside of population centers and become quite low in very rural areas (Figs. 9 and 10), which is expected given the crude analysis presented earlier (Fig. 3). This has the desired effect of reducing the artificial maxima over cities in the mean posterior λ (cf.

Figs. 11a and 11b). The mean posterior λ are generally at least twice as large as the smoothed actual N throughout the analysis domain, indicating that literal interpretation of the U.S. tornado database leads to severe underestimation of tornado frequency.

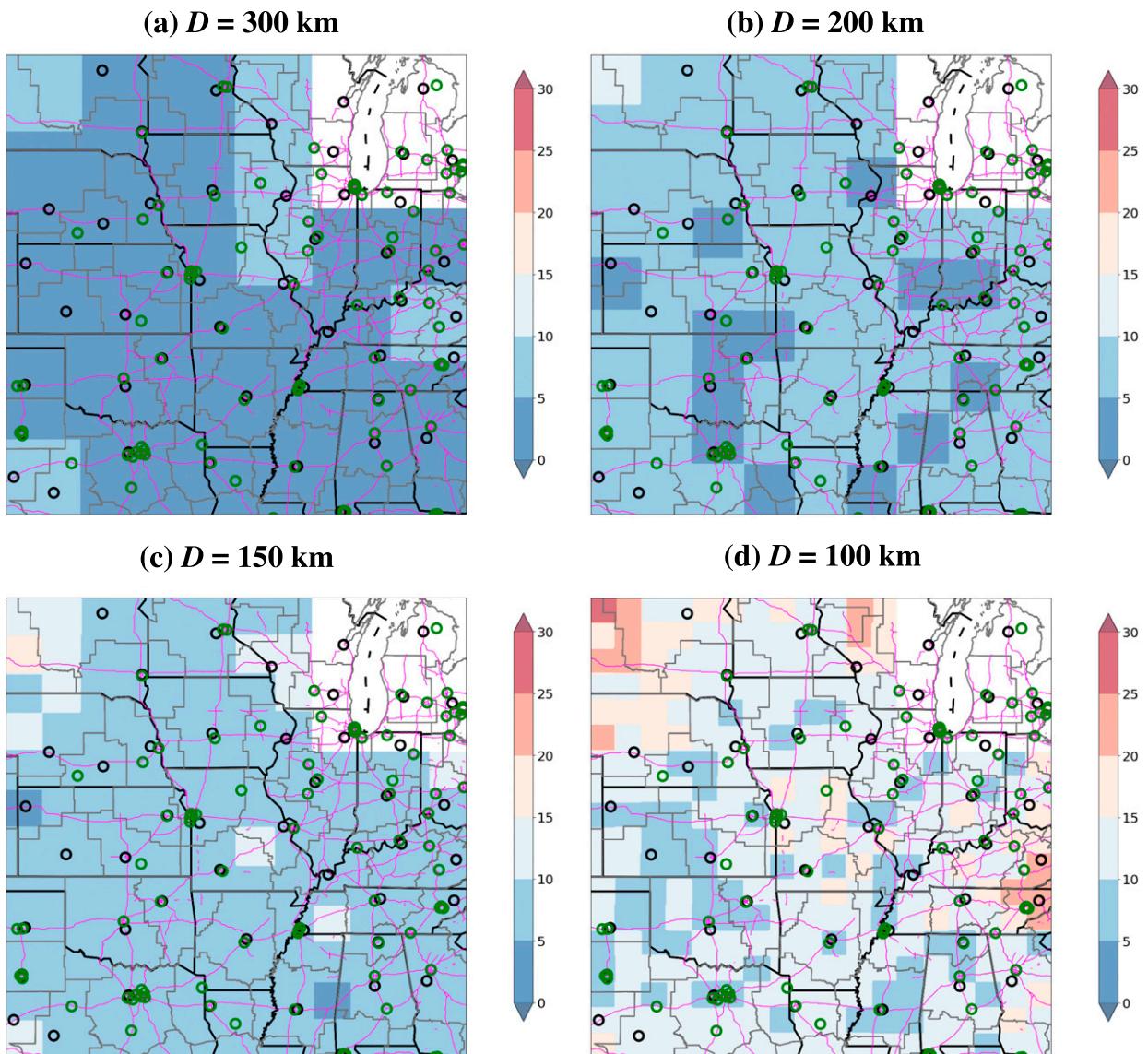


FIG. 5. Relative standard deviation (%) of out-of-sample λ posterior from final TRR model with $D =$ (a) 300, (b) 200, (c) 150, and (d) 100 km.

More specifically, the model predicts that only $\sim 45\%$ of tornadoes that occurred within the analysis domain were reported; that is, the actual tornado rate is ~ 2.2 times the reported rate. One major implication of this result is that the tornado production rates among supercells and/or quasi-linear convective systems have been substantially underestimated (e.g., Trapp et al. 2005). Another implication of this substantial underreporting is that U.S. tornado warning probability of detection (POD) and false alarm rate (FAR) have both been overestimated. Moreover, the presumed increase in TRR with time (in tandem with population density) has introduced a potentially substantial downward trend in POD and FAR.

Using $\text{TRR}_{\min} = 0.2$ (Figs. 12a,b) produces similar results as $\text{TRR}_{\min} = 0.3$ (Figs. 10 and 11a) over most of the domain, but over the western regions, substantially lower TRR and therefore higher λ occur. Again, we judge the TRR inferred by this model to be too low in the western part of the domain. Not imposing a lower limit at all (i.e., $\text{TRR}_{\min} = 0$) allows TRR to become even lower (Fig. 12c) and λ even higher (Fig. 12d) over the western areas. Both $\text{TRR}_{\min} = 0.2$ and $\text{TRR}_{\min} = 0$ cause a marked westward shift of the analyzed tornado alley (cf. Figs. 11a with Figs. 12c,d). This misleading result illustrates the importance of accounting for the TRR- λ aliasing problem by intelligently constraining TRR in low- (and high-) population areas. In an

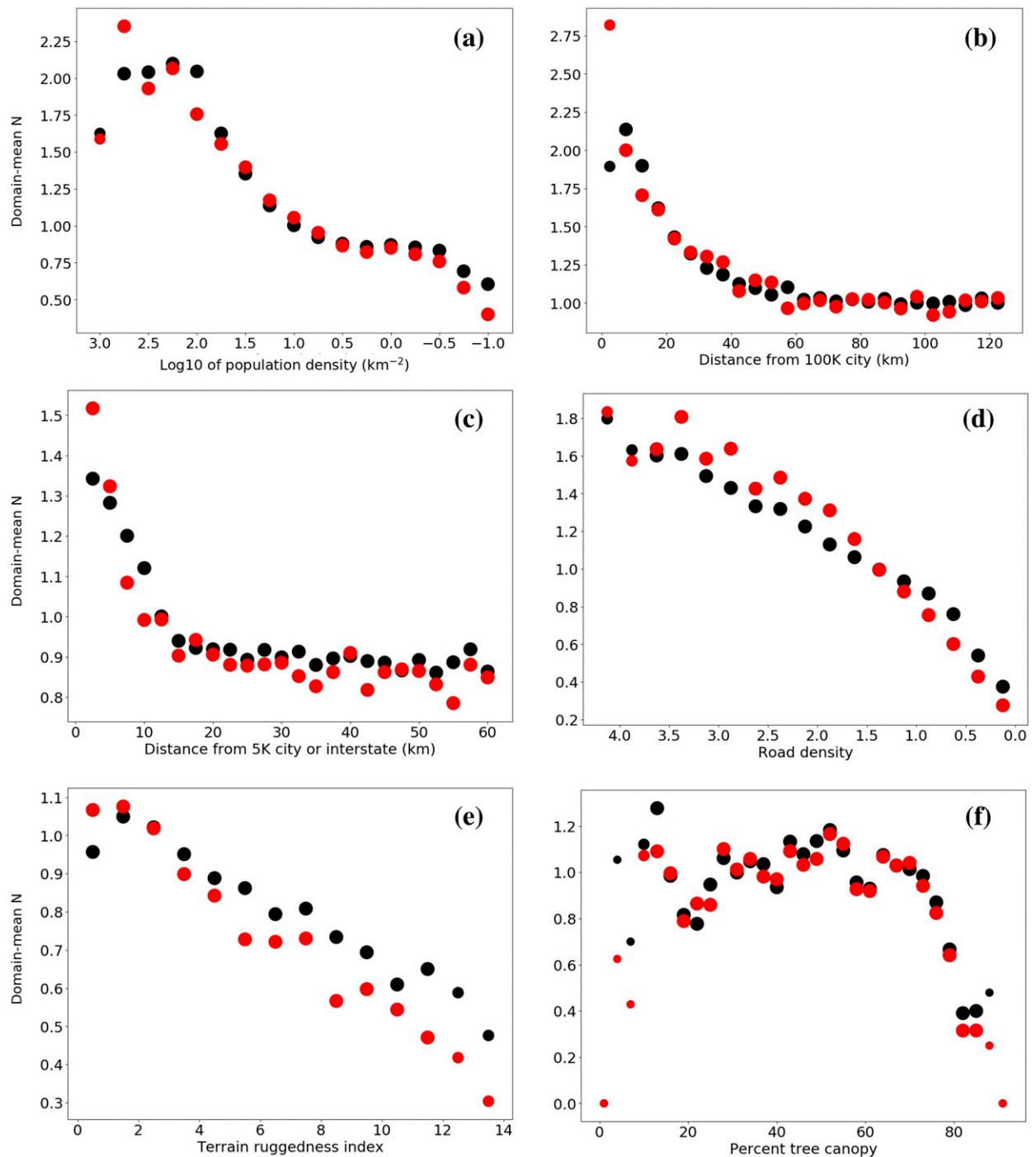


FIG. 6. Domain-mean, out-of-sample predicted N (black) and actual N (red) vs (a) base-10 logarithm of population density (km^{-2}), (b) distance from nearest 100K city (km), (c) distance from nearest 5K city or interstate (km), (d) road density [$^{\circ} (0.1^{\circ})^{-2}$], (e) terrain ruggedness index [$\text{m} (0.1^{\circ})^{-2}$], and (f) percent tree canopy. Large, medium, and small dots represent grid point counts > 100 , 30–100, and < 30 , respectively. Note that the actual N values are not normalized as in Fig. 3.

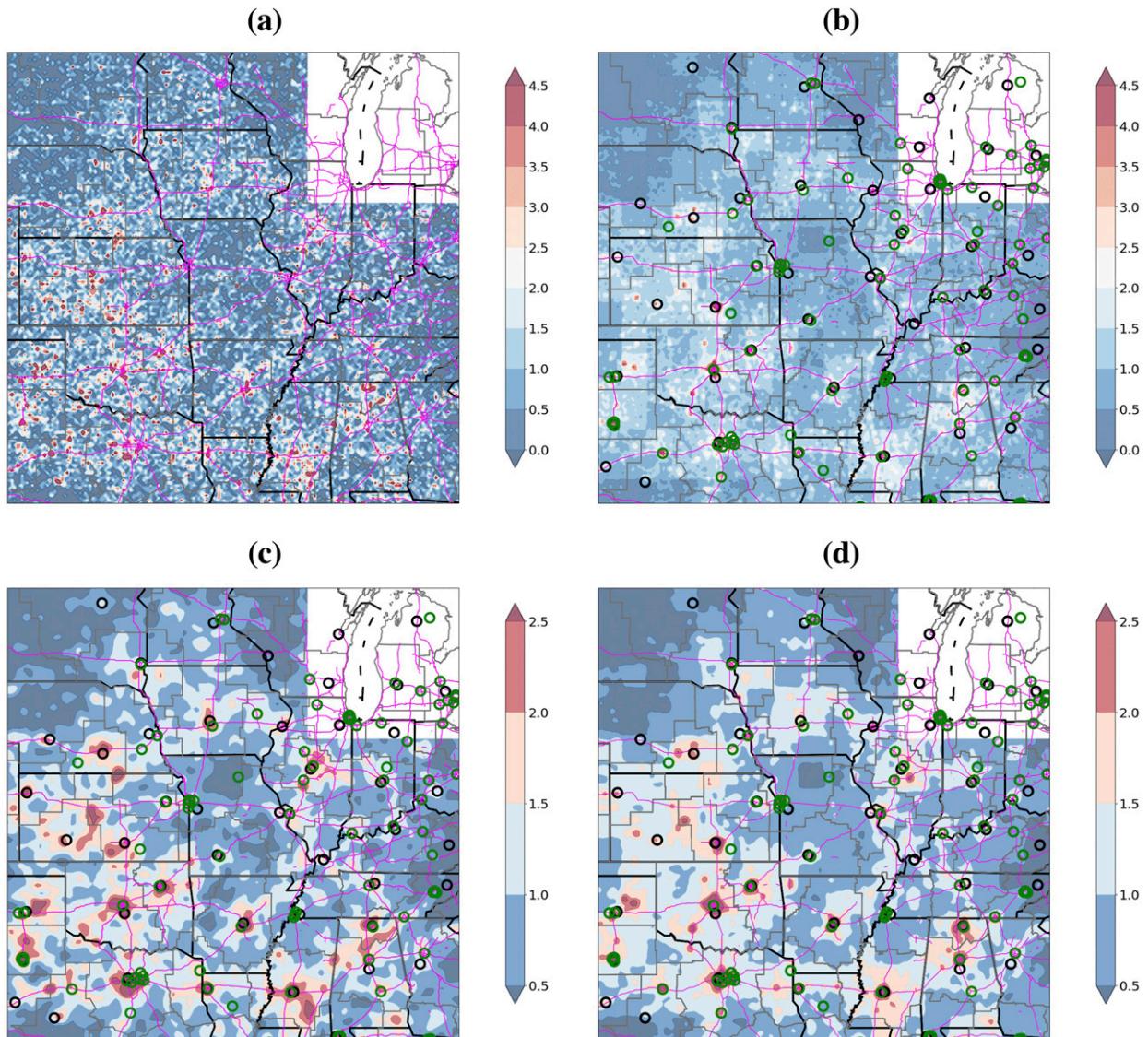


FIG. 7. (a) Actual N and (b) out-of-sample mean predicted N . (c),(d) As in (a) and (b), but smoothed ($2\sigma = 20$ km).

experiment with $\text{TRR}_{\min} = 0.5$ (not shown), spatial trends in the model-predicted λ were quite similar to those obtained using $\text{TRR}_{\min} = 0.3$. Thus, even were our adopted TRR_{\min} too low (which seems unlikely given Fig. 3), this would not degrade qualitative interpretation of the results.

5. Conclusions and future work

Tornadoes are often unreported due to a lack of observers, producing major biases in tornado report databases and climatological analyses derived from them. Hierarchical Bayesian models have been developed to infer tornado reporting rates (TRRs) and expected tornado frequency. Such a model is presented herein

that addresses a serious solution nonuniqueness problem that may have inhibited previous attempts to improve analyses of U.S. tornado climatology. The model is tested with an unprecedentedly large range of geographical covariates, of which population density is found to explain the most variance in reported tornado counts. At scales < 50 km, however, sampling error due to the brevity of the tornado record appears to dominate the underreporting bias. Cross-validation tests confirm that the final model captures real geographical dependencies of TRR without excessively fitting noise in the data. The results suggest only $\sim 45\%$ of tornadoes that occurred within the central U.S. analysis domain during 1975–2016 were reported. This substantial underreporting has major implications for our understanding of the distributions of

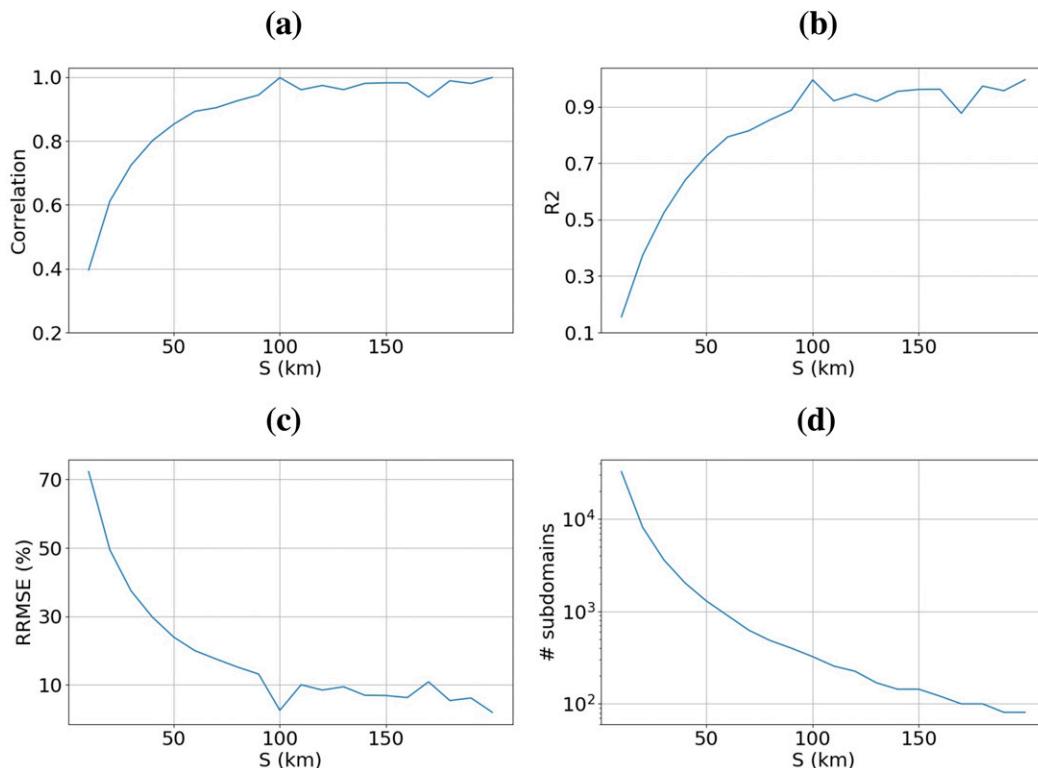


FIG. 8. Cross-validation statistics computed at scales S from 10 to 200 km: (a) r , (b) R^2 , (c) RRMSE (%), and (d) sample sizes for calculations. Statistics at $S > 10$ km were computed by splitting the analysis domain into square subdomains with diameter S , summing the predicted and actual N values over each subdomain, and then comparing the summed counts over the entire domain.

tornadic versus nontornadic storms (e.g., Trapp et al. 2005) and of how tornado warning performance has evolved over time (e.g., Brooks and Correia 2018).

There are at least three important considerations for interpreting our model estimates. First, there are variables that may substantially impact TRR but are not included in the TRR model, which considers only

population density. These include covariates that were individually tested herein, such as road density and distance to nearest city, as well as effects that would be difficult to quantify, such as storm chaser density and differing WFO survey practices. Omitting all of these variables likely introduces locally substantial errors in the predicted TRR, especially at small scales, and may

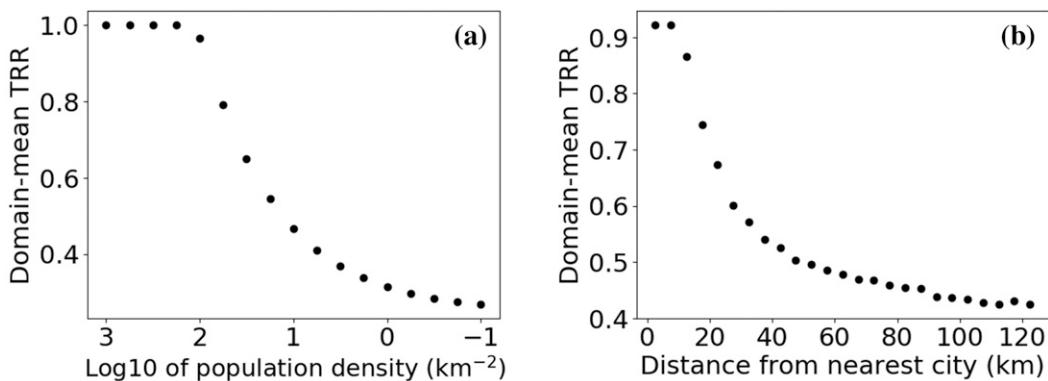


FIG. 9. Domain-mean posterior TRR vs (a) base-10 logarithm of population density (km^{-2}) and (b) distance from nearest 100K city (km).

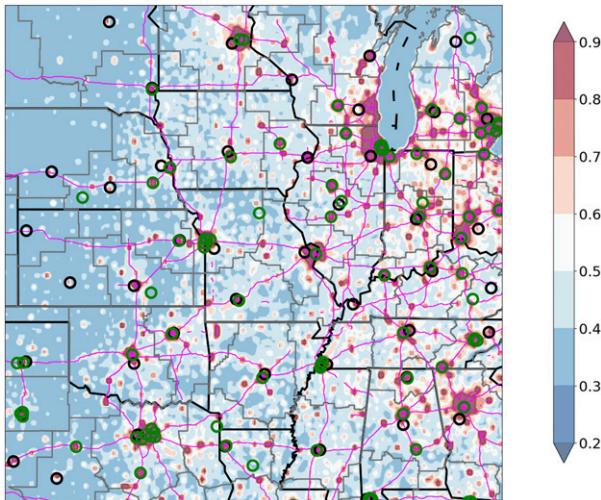


FIG. 10. Mean posterior TRR from the final model. As before, the northeastern corner of the domain is not used in model training (cf. Fig. 1), but model predictions for this region are shown for completeness.

also partly explain why so much of the variance in reported counts remains unexplained even at larger scales (e.g., 50–100 km). Second, locally large errors in expected tornado counts may persist where chance clustering of tornadoes occurred at scales near or exceeding that of the 100-km analysis subdomains (though these errors should be substantially smaller than in the reported tornado counts due to the TRR bias correction). Third, resolving the solution non-uniqueness in TRR and the expected tornado counts requires an ad hoc procedure for constraining one variable or the other. While the procedure we adopted is empirically grounded in the relationship between population density and reported tornado counts and in assessments of model bias, it does not totally eliminate the uncertainty introduced by the initial solution nonuniqueness. We note, however, that adopting other reasonable choices for the tunable parameters in the model (e.g., the population density threshold beyond which TRR is assumed to be unity) produces domain-wide mean TRR predictions within a few percentage points of the prediction by our final model. The sharp decrease in domain-mean reported tornado counts (normalized by large-scale tornado counts) away from urban areas (Figs. 3a,b) adds further plausibility to the model-predicted domain-mean TRR = 0.45.

One of the next steps in this research will be to examine how the model-predicted TRR varies with reported tornado attributes including pathlength, path width, and damage rating. It will be interesting to see, for example, whether the TRR of highly rated

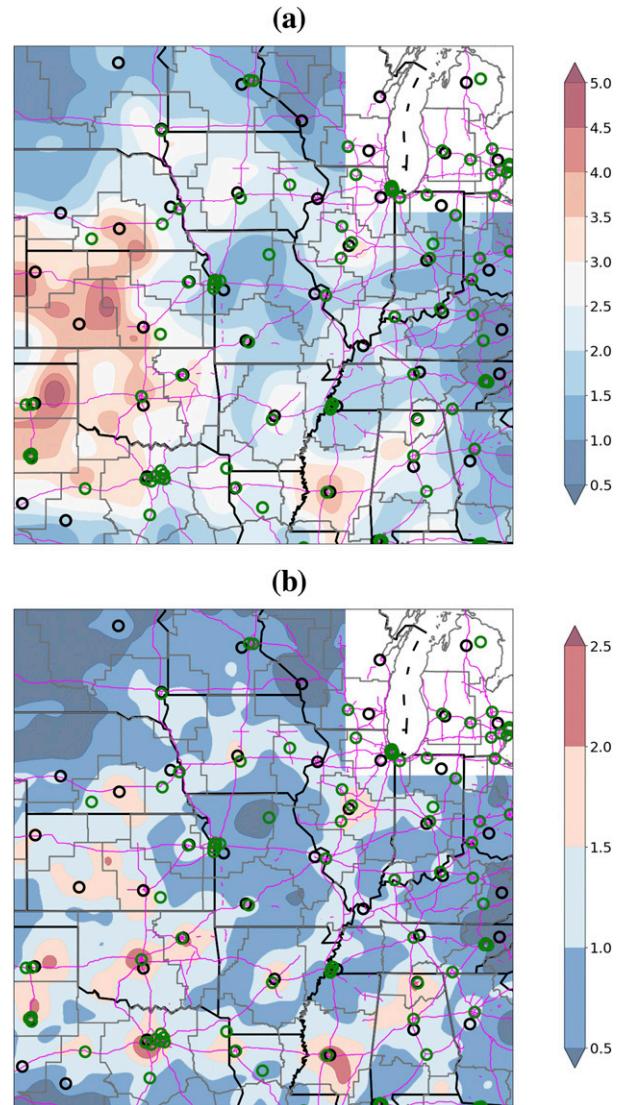


FIG. 11. (a) Mean posterior λ from the final model, smoothed using a Gaussian kernel of width $2\sigma = 65$ km. (b) Reported counts, smoothed using same kernel as in (a). Note the different color bars between (a) and (b).

tornadoes decreases as sharply with population density as that of low-rated tornadoes due to chronic underrating in regions lacking damage indicators. The TRR estimates obtained in the present study could be incorporated into existing Monte Carlo frameworks for examining U.S. tornado climatology (e.g., Strader et al. 2016). We plan to use such a framework to quantify the probability of “mini tornado alleys” (Broyles and Crosbie 2004) in the record arising solely by chance. We also plan to adapt the Bayesian model to inferring biases in reported tornado attributes and in the reported frequency of severe hail. The reporting bias estimates provided by the proposed method could also

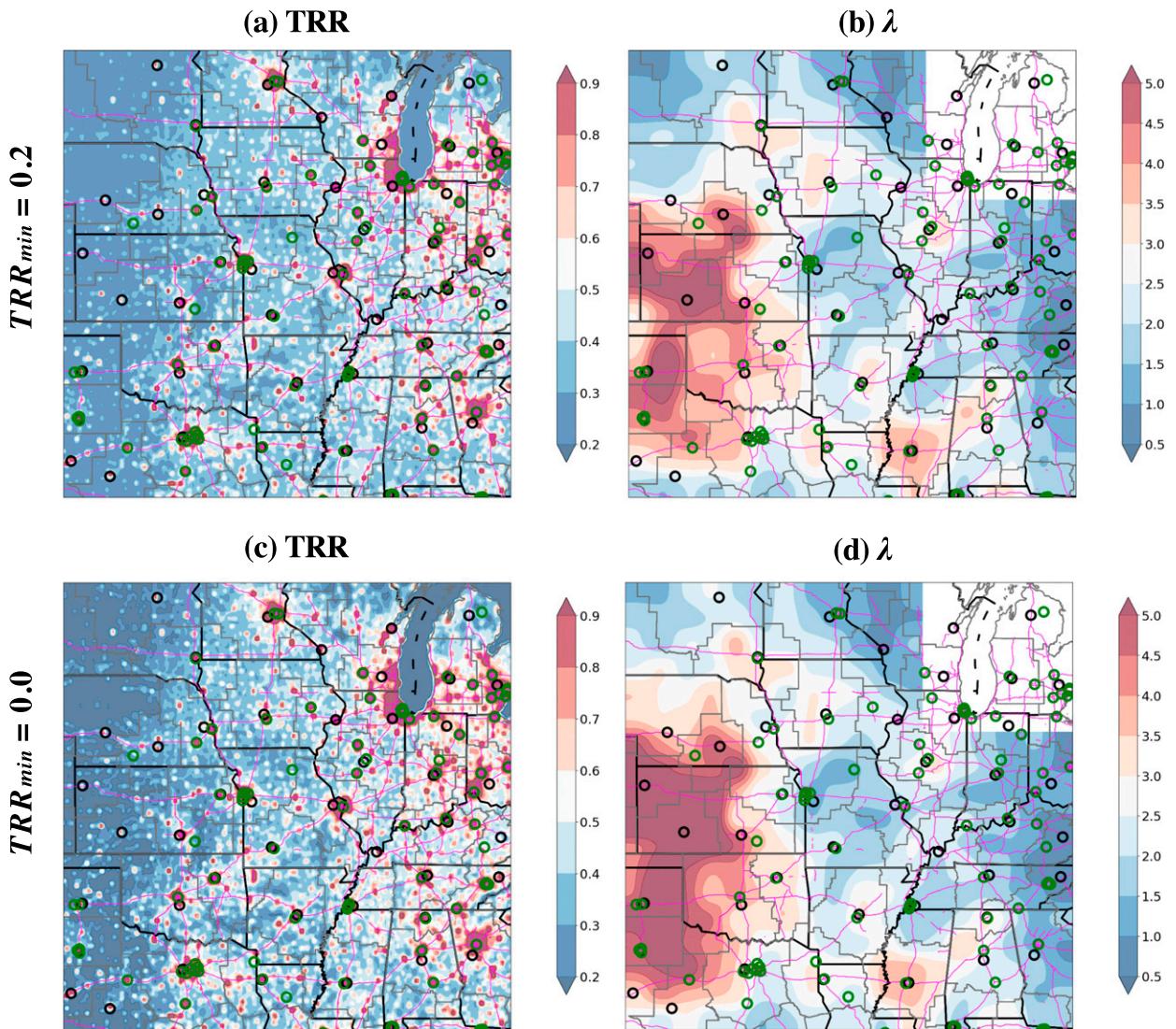


FIG. 12. Mean posterior (a),(c) TRR and (b),(d) λ obtained using (top) $TRR_{min} = 0.2$ and (bottom) $TRR_{min} = 0$.

aid in the development of detection algorithms, machine learning techniques, and convective-scale prediction systems (e.g., Warn-on-Forecast; Stensrud et al. 2009, 2013) that are trained and verified on severe weather reports.

Acknowledgments. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA16OAR4320115, U.S. Department of Commerce. We thank John Allen and Nusrat Yussouf for providing valuable feedback on earlier versions of the manuscript and the anonymous reviewers for their very helpful comments. We thank Andrew Dean and Patrick Marsh for providing the 2010 gridded population

dataset; the 1990 and 2010 datasets were obtained from the Socioeconomic Data and Applications Center (<http://sedac.ciesin.columbia.edu/-data/collection/usgrid/sets/browse>). City, interstate, and CWA boundary shapefiles were obtained from the NWS (<http://www.nws.noaa.gov/geodata/>). The road densities were calculated using data from the OpenStreetMap website (<http://www.openstreetmap.org>). The terrain dataset used to calculate TRI was obtained from the Consultative Group for International Agricultural Research Consortium for Spatial Information (<http://srtm.csi.cgiar.org>) and is described in Jarvis et al. (2008). Both the road densities and TRI were calculated using QGIS (<http://opengeo.qgis.org>). Finally, we thank Doug Speheger for helpful discussion.

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>.
- 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>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0310:OTROTP>2.0.CO;2).
- , and J. P. Craven, 2002: A database of proximity soundings for significant severe thunderstorms, 1957–1993. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 16.2, https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_46680.htm.
- , and J. Correia Jr., 2018: Long-term performance metrics for National Weather Service tornado warnings. *Wea. Forecasting*, **33**, 1501–1511, <https://doi.org/10.1175/WAF-D-18-0120.1>.
- , 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, http://ams.confex.com/ams/11aram22sls/techprogram/paper_81872.htm.
- 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>.
- Davidson-Pilon, C., 2016: *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference*. Addison-Wesley, 252 pp.
- Doswell, C. A., III, 2007: Small sample size and data quality issues illustrated using tornado occurrence data. *Electron. J. Severe Storms Meteor.*, **2** (5), <http://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/26/27>.
- Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.
- 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>.
- , and Coauthors, 2016a: The relationship between elevation roughness and tornado activity: A spatial statistical model fit to data from the central Great Plains. *J. Appl. Meteor. Climatol.*, **55**, 849–859, <https://doi.org/10.1175/JAMC-D-15-0225.1>.
- , T. H. Jagger, and T. Fricker, 2016b: Statistical models for tornado climatology: Long and short-term views. *PLOS One*, **11**, e0166895, <https://doi.org/10.1371/journal.pone.0166895>.
- Feuerstein, B., N. Dotzek, and J. Grieser, 2005: Assessing a tornado climatology from global tornado intensity distributions. *J. Climate*, **18**, 585–596, <https://doi.org/10.1175/JCLI-3285.1>.
- Geweke, J., 1992: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*, J. M. Bernardo et al., Eds., University Press, 169–193.
- 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>.
- Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara, 2008: Hole-filled SRTM for the globe, version 4. CGIAR Consortium for Spatial Information, <http://srtm.csi.cgiar.org>.
- Kellner, O., and D. Niyogi, 2014: Land surface heterogeneity signature in tornado climatology? An illustrative analysis over Indiana, 1950–2012. *Earth Interact.*, **18**, <https://doi.org/10.1175/2013E1000548.1>.
- Ray, P. S., P. Bieringer, X. Niu, and B. Whissel, 2003: An improved estimate of tornado occurrence in the central plains of the United States. *Mon. Wea. Rev.*, **131**, 1026–1031, [https://doi.org/10.1175/1520-0493\(2003\)131<1026:AIEOTO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1026:AIEOTO>2.0.CO;2).
- Riley, S. J., S. D. DeGloria, and R. Elliot, 1999: A terrain ruggedness index that quantifies topographic heterogeneity. *Interm. J. Sci.*, **5**, 23–27.
- 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>.
- 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., 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](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>.
- Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, <https://doi.org/10.1175/2009BAMS2795.1>.
- , and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, <https://doi.org/10.1016/j.atmosres.2012.04.004>.
- Strader, S., T. Pingel, and W. Ashley, 2016: A Monte Carlo model for estimating tornado impacts. *Meteor. Appl.*, **23**, 269–281, <https://doi.org/10.1002/met.1552>.
- Tecson, J. J., T. T. Fujita, and R. F. Abbey Jr., 1983: Statistical analyses of U.S. tornadoes based on the geographic distribution of population, community, and other parameters. Preprints, *13th Conf. on Severe Local Storms*, Tulsa, OK, Amer. Meteor. Soc., 120–123.
- 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., G. J. Stumpf, and K. L. Manross, 2005: A reassessment of the percentage of tornadic mesocyclones. *Wea. Forecasting*, **20**, 680–687, <https://doi.org/10.1175/WAF864.1>.
- 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>.
- Widen, H. M., and Coauthors, 2013: Adjusted tornado probabilities. *Electron. J. Severe Storms Meteor.*, **8** (7), <http://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/131>.