Probability of Tornado Occurrence across Canada*

VINCENT Y. S. CHENG,[+] GEORGE B. ARHONDITSIS,[#] DAVID M. L. SILLS,[@] HEATHER AULD,[&] MARK W. SHEPHARD,[&] WILLIAM A. GOUGH,[**] AND JOAN KLAASSEN[&]

[+] *Ecological Modelling Laboratory, Department of Physical and Environmental Sciences, and Climate Laboratory, Department of Physical and Environmental Sciences, University of Toronto, and Adaptation and Impacts Research Section, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, Toronto, Ontario, Canada*

[#] *Ecological Modelling Laboratory, Department of Physical and Environmental Sciences, University of Toronto, Toronto, Ontario, Canada*

[@] *Cloud Physics and Severe Weather Research Section, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, Toronto, Ontario, Canada*

[&] *Adaptation and Impacts Research Section, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, Toronto, Ontario, Canada*

[**] *Climate Laboratory, Department of Physical and Environmental Sciences, University of Toronto, Toronto, Ontario, Canada*

(Manuscript received 11 February 2013, in final form 11 June 2013)

ABSTRACT

The number of tornado observations in Canada is believed to be significantly lower than the actual occurrences. To account for this bias, the authors propose a Bayesian modeling approach founded upon the explicit consideration of the population sampling bias in tornado observations and the predictive relationship between cloud-to-ground (CG) lightning flash climatology and tornado occurrence. The latter variable was used as an indicator for quantifying convective storm activity, which is generally a precursor to tornado occurrence. The CG lightning data were generated from an 11-yr lightning climatology survey (1999–2009) from the Canadian Lightning Detection Network. The results suggest that the predictions of tornado occurrence in populated areas are fairly reliable with no profound underestimation bias. In sparsely populated areas, the analysis shows that the probability of tornado occurrence is significantly higher than what is represented in the 30-yr data record. Areas with low population density but high lightning flash density demonstrate the greatest discrepancy between predicted and observed tornado occurrence. A sensitivity analysis with various grid sizes was also conducted. It was found that the predictive statements supported by the model are fairly robust to the grid configuration, but the population density per grid cell is more representative to the actual population density at smaller resolution and therefore more accurately depicts the probability of tornado occurrence. Finally, a tornado probability map is calculated for Canada based on the frequency of tornado occurrence derived from the model and the estimated damage area of individual tornado events.

1. Introduction

Tornadoes are one of nature's most hazardous phenomena, capable of causing significant property damage and economic disruption as well as human injuries and fatalities. The tornadic events in Barrie 1985 (Etkin et al. 2001), Edmonton 1987 (Charlton et al. 1995), and southern Ontario 2009 (Ashton et al. 2010a,b) are amongst the most significant and costly tornado events in Canadian history. Multidisciplinary forensic analysis of a number of tornado-damaged areas in eastern Canada revealed that buildings in which more than 90% of the occupants were killed or seriously injured did not have anchorage of house floors into the foundation or anchorage of the roof to the walls (Allen 1992, 1986, 1984; Carter et al. 1989). As a result, the 1995 National Building Code of Canada (NBCC 2005) was updated to include provisions that ensure basic structural resilience under low-end tornadic loads

*Supplemental information related to this paper is available at the Journals Online website: http://dx.doi.org/10.1175/JCLI-D-13-00093.s1.

Corresponding author address: Vincent Y. S. Cheng, Ecological Modelling Laboratory, University of Toronto, 1265 Military Trail, Scarborough ON M1C 1A4, Canada.
E-mail: chengv@geog.utoronto.ca

DOI: 10.1175/JCLI-D-13-00093.1

in areas of Canada defined as ''tornado prone.'' However, the definition of tornado-prone regions requires a rigorous assessment of the spatial frequency of tornado occurrence, and to date these areas have not been clearly identified in Canada and thus these life-saving structural measures are not necessarily being implemented in areas at risk. In light of this need, Environment Canada has recently compiled an updated 30-yr national tornado database, founded upon an expert meteorological assessment of all existing tornado records. This database may serve as the basis for the implementation of tornado resiliency measures in the National Building Code of Canada (Sills et al. 2012).

Despite the notable efforts to consolidate a national tornado database, it is impossible to document all tornado occurrences in every single area, as tornadic events could easily be missed due to their relatively small spatial extent and time duration, as well as due to the absence of observers, structures, or daylight. In fact, tornado observations in any tornado dataset are often biased and underreported (Doswell and Burgess 1988), and the number of observations depends not only on the meteorological factors related to tornado occurrence (e.g., King et al. 2003) but also on nonmeteorological factors such as the monitoring network, proximity to populated areas, and/or radar locations, landscapes, and topography (Schaefer and Galway 1982; Grazulis and Abbey 1983; Ray et al. 2003; King 1997). Thus, the development of tornado climatology and the delineation of tornado-prone areas that could conceivably dictate security standards and infrastructural investments are very challenging. Of all the nonmeteorological factors investigated (e.g., obscured vision due to density of trees and hills, absence of roads, and buildings), population density is documented as the key nonmeteorological factor in determining the observation bias of tornadoes and nontornadic severe thunderstorms in many North American studies (e.g., Anderson et al. 2007; Ray et al. 2003; King 1997; Etkin and Leduc 1994; Paruk and Blackwell 1994; Snider 1977; Tescon et al. 1983; Schaefer and Galway 1982). The population sampling bias is typically accounted for with statistical models that correct tornado observations in areas with high sampling bias using nearby sites with low sampling bias and reliable tornado records. These statistical approaches are generally limited to relatively small spatial domains due to their underlying assumption that the tornado climatology is homogeneous or that there is significant areal coverage of reliable sites to derive credible adjustments (e.g., Anderson et al. 2007; Ray et al. 2003; King 1997). However, this assumption is profoundly violated in countries like Canada, where extensive areas with unreliable data exist, and thus the observational uncertainty can significantly compromise our predictive power.

The latter problem can be overcome when simultaneously considering *meteorological* covariates of the spatial variability of tornado occurrence. Finding potentially meaningful predictive relationships between meteorological factors and tornado occurrence could be a valuable tool for assessing tornado climatology and risk, particularly for large areas with significant observational uncertainty. There are many studies that have investigated possible linkages between meteorological covariates with tornado observations, for example, sounding-derived parameters such as convective available potential energy, vertical wind shear (Brooks et al. 2003b), and lightning flash polarity (Carey et al. 2003), but none of these relationships has been used in conjunction with predictive frameworks that explicitly accommodate the observation error associated with the tornado data. In addition, most studies typically assume that the tornado observations are independent and ignore other possible sources of spatial correlation within the domain modeled (e.g., Anderson et al. 2007; King 1997). Like in any spatially distributed modeling exercise, the explicit consideration of the error covariance in space is essential for drawing correct statistical inference and for identifying factors unaccounted for by the model (Anderson et al. 2007; Wikle and Anderson 2003). In this study, our thesis is that the Bayesian paradigm offers an effective means to accommodate all the aforementioned error sources and thus impartially communicate the total uncertainty associated with the forecasting of tornado occurrence (Arhonditsis et al. 2007, 2008a,b). Further, the use of hierarchical Bayes offers a conceptually plausible way for addressing the complexity pervading natural systems (Clark 2005; Cheng et al. 2010). In particular, the Bayesian hierarchical modeling can be an indispensable methodological framework to disentangle complex environmental patterns, to exploit disparate sources of information, to accommodate tightly intertwined processes operating at different spatiotemporal scales, and to explicitly consider the variability pertaining to latent variables or other inherently unmeasurable quantities (Clark and Gelfand 2006; Clark 2005; Zhang and Arhonditsis 2009).

The objective of this paper is to predict tornado occurrence across Canada using a Bayesian modeling approach and to shed light on the Canadian tornado climatology. The problem of tornado occurrence assessment is dissected into a two-pronged strategy in which we first consider the covariance between lightning flash density climatology and tornado occurrence, and then we postulate that the likelihood to observe a tornado is closely related to the population density. Our

approach also explicitly accommodates the fact that the model residual variability is likely to be characterized by distinct spatial patterns, arising from the model structural uncertainty. This paper is organized as follows: section 2 describes the sources and treatment of the data prior to the modeling analysis. Section 3 describes the rationale and basic features of our Bayesian modeling approach. Section 4 presents the results and discussion, including a sensitivity analysis of the model predictions to the grid resolution. Section 5 presents the conclusions and future perspectives of the present modeling framework.

2. Data sources and processing

a. Tornado data

Our analysis is based on the updated national tornado database for the period 1980–2009 (Sills et al. 2012). This database includes an intensity F-scale rating (Fujita 1981) and a confidence rating of ''confirmed'' (direct evidence of a tornado, e.g., visual evidence), ''probable'' (all available evidences point to the likelihood of a tornado but without direct evidence), and ''possible'' (ambiguous or unreliable tornado evidence). We included only the confirmed and probable events for all F-scale ratings (F0–F5) to ensure that our analysis is not based on nontornadic events that most likely are included in the possible category. The 30-yr study considered herein can also be compared with the last national tornado study based on the previous 30-yr period (1950–79; Newark 1984). To achieve a more accurate representation of the tornado occurrence for the United States–Canada border regions, U.S. tornado data for the same period were also obtained from the U.S. National Weather Service Storm Prediction Center (SPC).

The tornado data were classified into grids as follows: tornado reports of all F scales (F0–F5) were aggregated for the 30-yr period and were plotted as points (tornado touchdown location) or paths (track of the tornado on the ground), if the path information was available. Using the 1-km polygon vertices previously created by Burrows and Kochtubajda (2010) for their lightning study, three separate grids with cell sizes 25×25, 50×50, and $100 \times 100\,\text{km}^2$ were overlaid onto Canada to cover areas from the southern Canada–United States border to 70°N. The grid cells that were completely over water bodies (e.g., at the center of lakes and oceans) were removed. We tallied the tornado points and paths that occurred within or passed through each cell i, resulting in a tornado count for each of the cells. In the case of tornado paths, a tornado is counted more than once when it crosses more than one cell, but it cannot be counted multiple times within the same cell. The tally was done separately for the three grids examined. Tornado densities were then calculated by dividing the tornado counts per cell by the cell area. Finally, it should be noted that as the gridcell size increased, the cells at the United States–Canada border extended farther south and thus more U.S. area and SPC data were included in the analysis.

b. Lightning data

There is a body of literature relating the polarity of lightning flashes and, in particular, the predominately positive cloud-to-ground (CG) lightning flashes to tornadic activity (e.g., Reap and MacGorman 1989; Branick and Doswell 1992; Knapp 1994; MacGorman and Burgess 1994). However, recent empirical evidence suggests that a large majority of tornadic storms during the warm season (April–September) can also be associated with predominately negative cloud-to-ground lightning flashes in the contiguous United States (Carey et al. 2003). A similar linkage is found between predominately negative cloud-to-ground lightning and violent tornadic (F4 and F5) storms (Perez et al. 1997). The relationship between lightning polarity and severe storms appears to demonstrate significant regional variability (Carey and Rutledge 2003; Carey et al. 2003), and therefore our understanding is still considered to be at a developing stage (Carey et al. 2003). Given this knowledge gap, we did not attempt to use lightning polarity to model tornado occurrences. Rather, we used the mean annual cloud-to-ground flash density, a typical proxy for the frequency of thunderstorm occurrence and duration (Huffines and Orville 1999; Burrows and Kochtubajda 2010). This implies that we may be introducing an additional source of uncertainty in regards to our capacity to discern the actual tornado occurrence *per event* recorded, as our predictor is not directly related to the response variable targeted. Nevertheless, our aim is to use the *climatology* of CG lightning to identify areas in Canada that are prone to thunderstorms, which are typically required for the generation of tornadoes. Mean annual lightning flash density has been shown to be an ideal surrogate variable for thunderstorms (Huffines and Orville 1999), and the lightning network data are spatially richer and more accurate than human-based thunderstorm observations (Huffines and Orville 1999; Changnon 1993; Reap and Orville 1990; Changnon 1988a,b, 1989). We obtained the Canadian Lightning Detection Network (CLDN) CG lightning flash density data for the 1999–2009 period. The 11-yr CG flash density data were originally sorted into $1 \times 1\,\text{km}^2$ cells (Burrows and Kochtubajda 2010; Shephard et al. 2013). Using the gridded data, mean annual CG flash densities were calculated in three separate grids of $25 \times 25\,\text{km}^2$ ($625\,\text{km}^2$), $50 \times 50\,\text{km}^2$ ($2500\,\text{km}^2$), and $100 \times 100\,\text{km}^2$ ($10\,000\,\text{km}^2$) (see also Fig. S1 in the supplementary material).

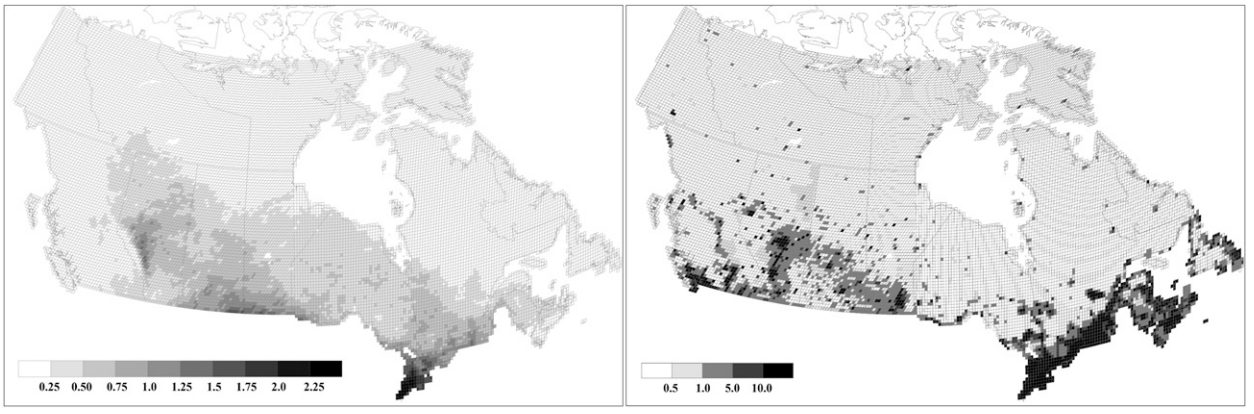

FIG. 1. (left) Mean annual CG lightning flash density (number of flashes $km^{-2}\,yr^{-1}$), based on data from 1999–2009 in a $25 \times 25\,km^2$ grid. (right) Areal interpolated population density (persons km^{-2}) from Canadian census 2001 subdivision boundary in $25 \times 25\,km^2$ grid.

c. *Population density*

Initial reporting of tornado observations is typically an interactive process between the general public and the meteorological service personnel, which underscores the presence of a strong positive relationship between tornado observations and population density. However, there are a few variations of the most sensible population density expression, such as (i) the total county population density, where all of the population within a county is included, and (ii) the rural population density, where the densely populated areas (e.g., cities or towns) within a county are excluded. The rationale for using rural population density is that by excluding cities or towns that are densely populated but small in areal extent, the calculated population density provides a more accurate portrayal of the real areal coverage of a study site (e.g., King 1997; Paruk and Blackwell 1994). On the other hand, the rural population density approach would not work in cases where a substantial areal extent of the county is populated (e.g., large metropolitan areas; Anderson et al. 2007). To address this issue, one can use a dataset that has a higher spatial resolution than the county boundary units. For this reason, we obtained Canada's census subdivision boundary units (Statistics Canada 2001), which have more refined spatial resolution with over 40 political administrative boundary unit types, such as cities, towns, villages, Indian reserves, Indian settlements, and unorganized territories. They are all defined in separate entities, such that most small population centers are separated from larger rural communities. Further, since the model grid cells cover both countries at the United States–Canada border, we also obtained the U.S. 2000 census county population density. Population densities in the original boundary units were recalculated into the model grids of $25 \times 25\,km^2$ (Fig. 1, right), $50 \times 50\,km^2$, and $100 \times 100\,km^2$ (see Fig. S2 in the supplementary material) based on the areas of overlap between census subdivision units and model grid cells, assuming that the population density from the original boundary units of the census subdivision is spatially homogeneous (Goodchild et al. 1993). The spatially detailed population density from the census subdivision would only be retained in the smaller model grids, whereas when the grid size increases (e.g., $100 \times 100\,km^2$ grid), more population clusters would be averaged within one cell, having an effect similar to that of small densely populated areas for the calculation of county population density.

3. Methodology

The tornado cell data are discrete counts that predominantly contain zeros, high observational uncertainty, and intricate spatial covariance driven by nonmeteorological and meteorological factors. These issues have made classical (frequentist) regression methods intractable. In contrast, the use of Bayesian inference techniques offers a flexible means to decompose the problem of tornado occurrence into a series of conditional models coherently linked together via Bayes' rule (Anderson et al. 2007; Wikle and Anderson 2003; Wikle 2003):

$$\underbrace{(\text{explanatory factors}, \text{parameters} \mid \text{data})}_{[1]\text{Posterior model}} \propto \underbrace{(\text{data} \mid \text{explanatory factors}, \text{parameters})}_{[2]\text{Data model}} \times \underbrace{(\text{explanatory factors} \mid \text{parameters})}_{[3]\text{Explanatory model}} \times \underbrace{(\text{parameters})}_{[4]\text{Parameter model}}, \quad (1)$$

where the posterior model [1] reflects our beliefs related to the relative importance of the explanatory (or predictive) factors examined and the underlying parameter values after the consideration of the available data. This in turn can be thought of as the product of the data model [2], specifying the dependence of the observed data on the processes of interest and parameters, with the explanatory model [3] describing the nature of the relationship between response variable and explanatory factors, and the parameter model [4] quantifying the uncertainty in parameter values. This formula will serve as the basis for estimating the probability of tornado occurrence.

a. Data model

Our model postulates that the population density is the primary factor that determines the accuracy of tornado observations in our dataset. In particular, we assume that there is a threshold population density above which all tornadoes are expected to be observed, whereas below this threshold level the probability of tornado detection is proportional to the population density (Newark 1983; King 1997; Anderson et al. 2007). Similar to the Anderson et al. (2007) study, we specify a binomial model in which the observed tornado counts Tobs_i for each grid cell i are conditioned upon the actual (but unobserved) tornado occurrences Tlatent_i and the probability of detection $p_i(\beta)$:

$$\text{Tobs}_i \,|\, \text{Tlatent}_i, \lambda_i, p_i(\beta) \sim \text{Binomial}[\text{Tlatent}_i, p_i(\beta)]. \tag{2}$$

The probability of detection $p_i(\beta)$ represents the likelihood to observe a tornado and is associated with the population density y_i by the following exponential expression:

$$p_i(\beta) = \exp[-\beta/\exp(y_i)], \tag{3}$$

where β is the population effect parameter and $\exp(y_i)$ is the exponential transformation of the original population density data. The actual occurrence of tornadoes Tlatent_i in the model domain is specified as a Poisson process, conditional on the average or expected tornado occurrence rate per grid cell λ_i, provided by the predictive model

$$\text{Tlatent}_i \,|\, \lambda_i \sim \text{Poisson}(\lambda_i). \tag{4}$$

b. Explanatory model

The explanatory (or predictive) model expresses the logarithm of the expected tornado rate λ_i in given grid cell i as a linear function of the corresponding mean annual CG flash density x_i:

$$\log(\lambda_i) = \alpha_0 + \alpha_1 x_i + a_i, \tag{5}$$

where a_i is a gridcell-specific random effect, capturing the residual variability of the tornado frequency in grid cell i, stemming from other explanatory factors/processes unaccounted for by the model. For example, Brooks et al. (2003b) found that certain thresholds of convective available potential energy and vertical wind shear at the 0–6-km layer show a strong correlation with the severity of thunderstorms, and this pattern cannot be explained by the mean annual lightning flash density alone. Moreover, the inclusion of a_i aims to address the possibility that the signature of the lightning flash density may not be consistently evident in the tornado frequency records throughout the model domain. That is, the globally common parameterization of our predictive model may be violated in smaller geographic areas due to, for example, orographic-influenced thunderstorms (Taylor et al. 2011) or lake-breeze convergence-influenced convective processes (King et al. 2003). For the same reason, it is reasonable to assume that the random effects of the unaccounted for factors have a regionalized/localized character and thus are spatially correlated.

The characterization of the spatially correlated random terms a_i was based on the Bayesian conditional autoregressive (CAR) model (Besag et al. 1991). The random error terms are jointly distributed as a multivariate normal distribution with mean 0 and an unknown covariance matrix (Besag and Kooperberg 1995). In particular, the model postulates that the spatial random effect in cell i depends only on the neighboring cells of i (N_i) and that all of the neighbors have equal influence (weight of 1) on i. The term a_i is defined by the conditional normal distribution $a_i \sim N(\mu_i, \sigma^2/n_i)$, where

$$\mu_i = \frac{1}{n_i} \sum_{j \in N_i} a_i \tag{6}$$

and n_i is the number of adjacent grid cells. Because we used a first-order neighborhood approach and squared cells, N_i represents the eight immediately adjacent cells.

c. Parameter model

In Bayesian inference, model parameters are treated as random variables rather than fixed quantities. As such, prior distributions can be formulated to depict our knowledge on the relative plausibility of their values before the consideration of the observed data (Gelman et al. 2004). Here, we opted for ''noninformative'' or

TABLE 1. Posterior means and standard deviations (SD) of the parameters and model predictions.

Model resolution		β	α_0	α_1	σ	$\sum$Tlatent$_i$ Total (30 yr)	$\sum$Tlatent$_i$ Mean annual (yr^{-1})	$\sum$Tobs$_i$ Total (30 yr)	$\sum$Tobs$_i$ Mean annual (yr^{-1})
$25 \times 25\,km^2$	Mean	3.129	−3.536	1.543	1.574	4773	159.1	2069	69.0
($625\,km^2$)	SD	0.143	0.2	0.195	0.074				
$50 \times 50\,km^2$	Mean	3.102	−2.298	2.078	1.59	4395	146.5	2088	69.6
($2500\,km^2$)	SD	0.183	0.228	0.256	0.09				
$100 \times 100\,km^2$	Mean	3.334	−1.07	2.6	1.778	4287	142.9	2140	71.3
($10\,000\,km^2$)	SD	0.26	0.263	0.353	0.131				

"flat" priors reflecting no prior knowledge of the model parameters. In particular, the prior distributions of β, α_0, α_1, and σ were specified as follows:

$$\beta \sim N(0, 1000), \tag{7}$$

$$\alpha_0 \sim N(0, 1000), \tag{8}$$

$$\alpha_1 \sim N(0, 1000), \quad \text{and} \tag{9}$$

$$\sigma^2 \sim \text{IG}(0.001, 0.001), \tag{10}$$

where N and IG denote the normal and inverse gamma distributions for the regression coefficients and conditional variance among the spatial random error terms σ^2, respectively.

A sequence of realizations from the model posterior were obtained using Markov chain Monte Carlo (MCMC) simulations. Specifically, we used the general normal proposal Metropolis algorithm as implemented in the WinBUGS software (Lunn et al. 2000); this algorithm is based on a symmetric normal proposal distribution, whose standard deviation is adjusted over the first 4000 iterations such that the acceptance rate ranges between 20% and 40%. We collected 20 000 samples from two independent chain runs for each model configuration. The posterior statistics were calculated using a thin of 10, which yielded a sample size of 4000 for all the model configurations considered. The entire modeling process is undertaken in three grid sizes (25-, 50-, and 100-km grid squares) separately in order to assess model sensitivity to the selection of the grid size.

4. Results and discussion

Posterior means and standard deviations of the model parameters for all three model grids are shown in Table 1. The discussion of our results refers to the 25-km model, while the comparison of the inference drawn among the different grid configurations is presented in a following section.

a. Comparison between predicted tornado occurrence and tornado observations

The total predicted tornado occurrence (calculated as the total sum of the posterior means of Tlatent$_i$) for the entire 30-yr period of our study for all three model grids are at least twice the number of recorded tornadoes, based on the counts of tornado points and paths by grid cells (Table 1; Fig. 2). Tornado observations were on average close to $70\,yr^{-1}$, whereas the model-predicted tornado occurrence was close to $150\,yr^{-1}$ (Table 1). Some areas with no tornado observations in the 30-yr study period, for example, northern parts of Alberta and Manitoba and northwestern Ontario, are predicted to have a low (~0.10 tornadoes per $10\,000\,km^2$) probability of annual tornado occurrence with the 25-km grid model. Similarly, the southern border of Alberta and Saskatchewan are predicted to experience even higher probabilities of tornado occurrence [~(0.5–1.5) tornadoes per $10\,000\,km^2$]. Considering the sparse population density and the lack of radar coverage in some of those areas, the discrepancy between observed and predicted tornado occurrence is not surprising. Generally, we note that extensive areas across the prairies are corrected by our model, whereas the more heavily populated areas in the Edmonton–Calgary corridor, Regina, Saskatoon, and Winnipeg are subjected to a lower observation error (Fig. 2). Adjustments also continue eastward to northwestern Ontario and to a lesser extent in the area between northeastern Ontario and Quebec (Fig. 2). There are no major adjustments in southern Ontario or southern Quebec because of the more reliable tornado detection assumed by the much denser population.

In a similar manner, the prairies were characterized by the highest standard deviation values of the true tornado occurrences Tlatent$_i$ (Fig. 2, right panels). The values decrease proportionally as we move northward to areas with lower predicted tornado occurrences. Further, our analysis suggests that the standard deviations slightly increase as the grid size decreases, which is an expected result as the posterior means also increased slightly with

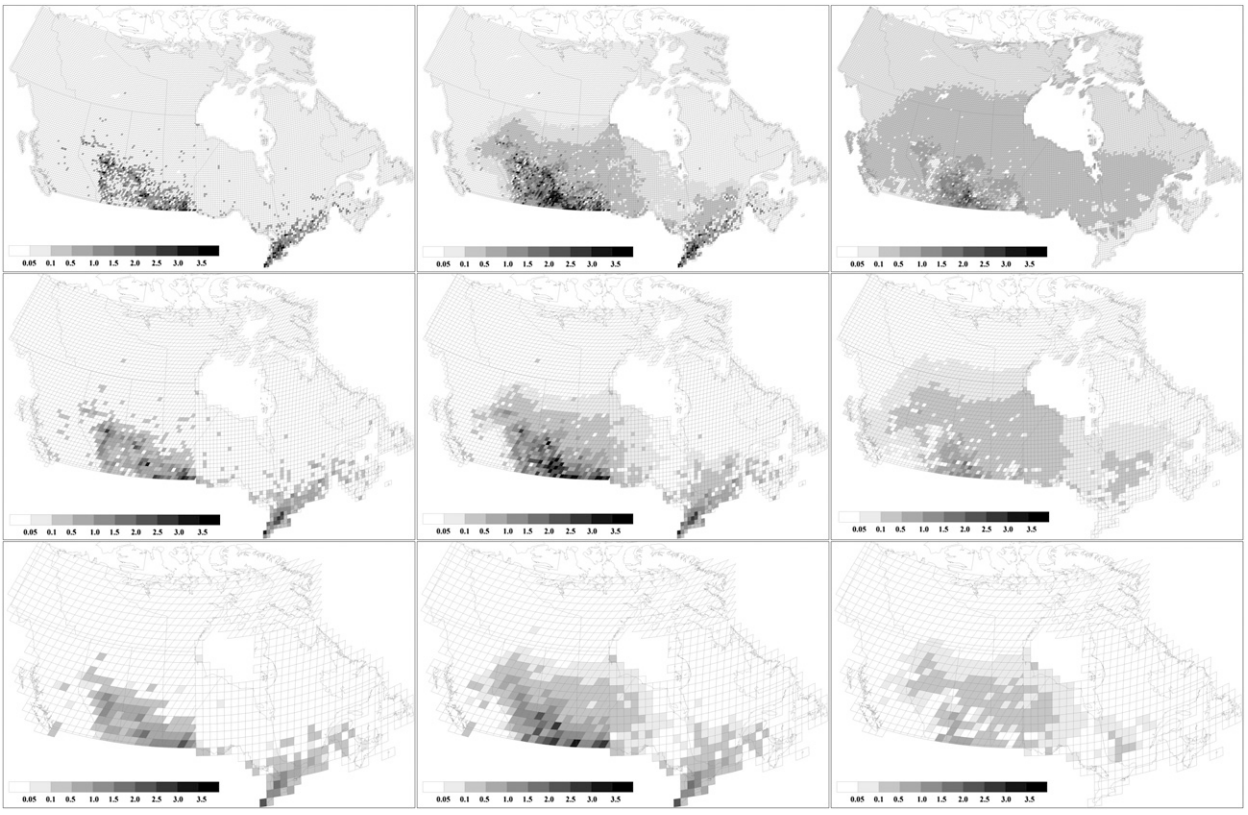

FIG. 2. (left) Gridded observed tornado counts standardized to $10\,000\,\mathrm{km}^{-2}\,\mathrm{yr}^{-1}$. (center) Posterior mean of tornado occurrence standardized to $10\,000\,\mathrm{km}^{-2}\,\mathrm{yr}^{-1}$. (right) Posterior standard deviation of tornado occurrence standardized to $10\,000\,\mathrm{km}^{-2}\,\mathrm{yr}^{-1}$. (top)–(bottom) The 25×25, 50×50, and $100 \times 100\,\mathrm{km}^2$ models, respectively.

the decrease of the gridcell size. Conversely, in the most populated areas, most notably from southern Ontario to southern Quebec, the standard deviations are zero. That is, the predicted tornado occurrences are simply equal to Tobs_i, given that our model postulated high tornado detection efficiency in densely populated areas.

b. *Importance of the lightning covariate, population density, and spatial random effects*

The posterior estimate for the parameter α_1 is 1.53 ± 0.19, indicative of a distinct relationship between the CG flash density and the underlying expected tornado rate (Table 1). The random effects term a_i, representing explanatory factors–processes unaccounted for by the model, demonstrates strong spatial covariance with the $\mathrm{Tlatent}_i$ posterior estimates, in that positive a_i values typically coincide with higher $\mathrm{Tlatent}_i$ and vice versa (Fig. 3). The main exception to this spatial pattern was in southwestern Ontario, where despite the higher tornado occurrence rates relative to the central prairies, the relative a_i values are distinctly lower, especially when a larger gridcell size is used (see Figs. S3,S4 in the supplementary material). It is also worth noting that southwestern Ontario experiences the most intense CG lightning flash densities in Canada (Fig. 1, right). We believe that the lower a_i values for southwestern Ontario could be a reflection of the nature of the two structural components of our model. First, the lower a_i values could stem from the underlying assumption that the lightning–tornado relationship is identical throughout the study domain. Namely, the assumption of a globally common α_1 value may oversimplify the regional variability in the nature of the explanatory linkage between CG lightning densities and tornado frequency, which in turn it is depicted by the systematic trends of the a_i values in the central prairies and southern Ontario. The problems arising from a geographically constant lightning–tornado relationship can be exemplified by one facet of the spatial variability of the lightning densities: namely, the length of lightning season. Different regions may experience similar mean annual lightning densities, but the length of the lightning season can be significantly different, reflecting the potential differences in the underlying thermodynamic processes (Burrows and Kochtubajda 2010). For example, the Pacific coastal region and southern Nova Scotia appear to have higher

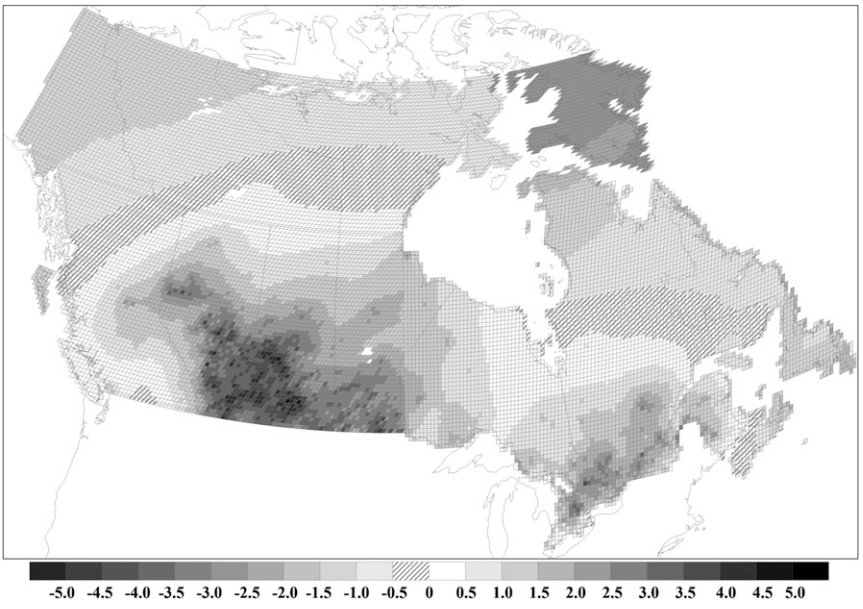

FIG. 3. Posterior mean of the conditional autocorrelation coefficient a_i on a $25 \times 25\,\text{km}^2$ model grid.

lightning flash densities due to a virtual year-round lightning season, as winter lightning commonly occurs when Arctic air masses pass over much warmer water (Burrows and Kochtubajda 2010). It is significantly different than in the Edmonton area, where over 95% of the annual lightning occurred in May–August (Burrows and Kochtubajda 2010). Hence, a different lightning–tornado relationship may be expected between these regions.

Second, the spatial patterns of the a_i posteriors could similarly be a reflection of the assumption of globally common population effects on the likelihood to observe a tornado $p_i(\beta)$. Possible regional population bias may exist because of differences in the quality of the monitoring networks, number of trained weather spotters, training of meteorological service staff, and general public awareness to tornadoes, but these factors are not explicitly considered in the model (Anderson et al. 2007). The omission of this potential bias may influence the probability of tornado detection and subsequently the uncertainty estimates of Tlatent_i. In such a case, it is conceivable that the estimation of the random terms a_i may offset this systematic error, regardless of the CG lightning flash densities.

An appealing approach to investigate the effects of potential region-specific relationships is a Bayesian hierarchical configuration of the present model. Under this framework, the assumption of globally common relationships is relaxed, and the model is dissected into levels (hierarchies) that explicitly account for the role of significant sources of spatial variability (e.g., geographical locations, climatic regimes, particular features of regional monitoring networks, road density, or landscape), thereby allowing for site-specific parameter estimates (Gelman and Hill 2007; Cheng et al. 2010). Finally, the posterior estimate of the model intercept α_0 is -3.536 ± 0.201 (or 0.030 ± 0.006 in the original scale), suggesting that the baseline expected tornado rate across the study area, when we account for the flash density and the spatial random effects, is very low (Table 1).

The posterior estimate of the population effect parameter β is 3.129 ± 0.143 (Table 1). Substituting the mean value back to Eq. (2), we infer that the probability of tornado detection $p_i(\beta)$ is approximately equal to 1 when the population density reaches the level of 6 individuals km^{-2} (Fig. 4, left). In effect, the model allows identifying a population threshold when all tornadoes can be observed, and thus the variables Tlatent_i and Tobs_i are identical. Interestingly, our analysis suggests a population threshold similar to the value reported by King (1997) using data from southern Ontario. When the population density is below 6 individuals km^{-2}, the tornado observational uncertainty is partly reflected in the uncertainty estimates of Tlatent_i. As the Tlatent_i posteriors are also dependent on the meteorological submodel, it is worth noting that the standard deviations of Tlatent_i stem from the relationship between lightning and tornado frequency as well as the population density effects on the probability of tornado detection (Fig. 2).

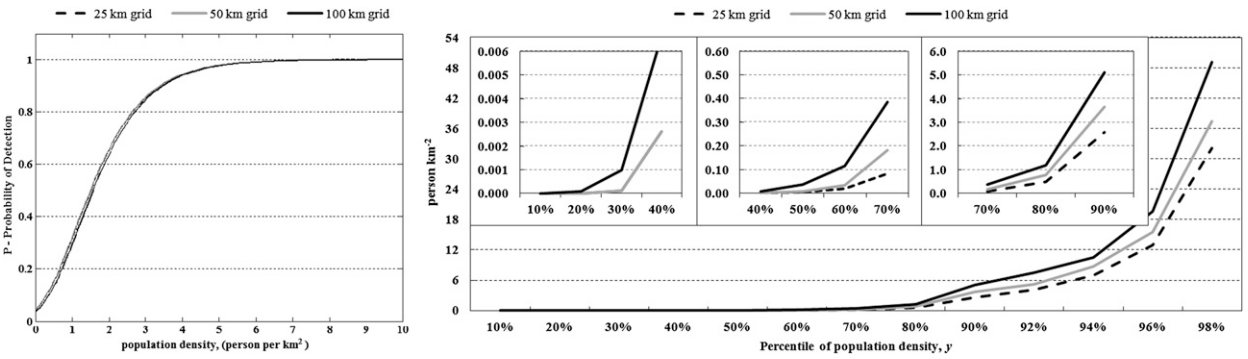

FIG. 4. (left) Probability of tornado detection estimated from the model as a function of population density. (right) Percentile of the gridded population density. Models on a 25 × 25, 50 × 50, and 100 × 100 km^2 grid are shown in dashed, gray, and black solid line, respectively. The three inset panels show an expanded y axis for percentiles from 0% to 90%.

c. *Model sensitivity to gridcell size*

Posterior means of the population effect parameter β (Table 1) are quite consistent across the three model gridcell sizes used, although a slight increase for the posterior standard deviation is observed with the coarser spatial resolution. This is not surprising as the population densities calculated for larger grid cells are less accurate compared to the smaller grid cells. The posterior means of the parameters α_0 and α_1 are larger as the gridcell size increases (Table 1), which simply stems from the scale-dependent units of Tobs_i (counts of tornadoes per cell) relative to the normalized units of the mean annual CG lightning flash density (flash km^{-2}). The conditional variance σ of the random terms α_i is similar for the 25 × 25 and 50 × 50 km^2 grids but increases slightly with the 100 × 100 km^2 grid. Some notable differences among the three grid configurations are as follows: (i) as the grid size increases, the proportion of cells with zero tornado observations is lower (i.e., 93%, 87%, and 80% of cells for the 25 × 25, 50 × 50, and 100 × 100 km^2 grids, respectively); (ii) the population density is somewhat higher with the coarser resolution (e.g., the proportion of cells with population density greater than 6 individuals km^{-2} is 6%, 7%, and 9% for the 25 × 25, 50 × 50, and 100 × 100 km^2 grid cells, respectively; Fig. 4); and (iii) each population center is associated with Tlatent_i projections that are spread over a larger areal extent with the coarser resolution relative to the finer grid, and thus the predicted tornado occurrence rates are higher with the smaller grid size (Fig. 5). Overall, our analysis shows that the predictive statements supported by the model are robust and do not depend strongly on the grid sizes, although the smaller gridcell resolution more accurately portrays the actual population densities and therefore better represent the probability of tornado occurrence.

d. *Extent of tornado occurrence probabilities*

Based on the Tlatent_i predictions with the 25 × 25 km^2 grid configuration (Fig. 2), we used the ordinary kriging method (see, e.g., Ray et al. 2003) to create a map that depicts a smoothed pattern of the regions with different tornado occurrence rate (Fig. 6). Based on the full range of tornado occurrence rate and with reference to other Canadian studies (Newark 1984; Hage 2003), we identified four levels [≥0.05, ≥0.10, ≥1.0, and ≥2.0 (per 10 000 km^2 yr)] to classify regions of very low, low, medium, and high probability of tornado occurrence, respectively. The very low probability areas begin from central British Columbia in the west and move northward to northern Alberta, before dipping southward in Saskatchewan and extend northward again to the northeastern tip of Manitoba. The same very low probability area extends to northern Ontario, southern Quebec, eastern New Brunswick, and Prince Edward Island. The low probability areas follow a smaller in extent but similar pattern, except for a discontinuity north of Lake Superior. The medium probability areas have a distinctly different pattern. The area starts near the foothills of the Alberta Rocky Mountains and extends eastward before dipping south to central Saskatchewan and ending near southern Manitoba. For the eastern part of Canada, the area of medium probability of tornado occurrence begins between Georgian Bay and Lake Ontario in southern Ontario and extends across southwestern Ontario, where there are also some regions with higher probabilities. Some of the significant tornado events that have occurred in the medium probability areas are listed in Table 2. One of the high probability areas of tornado occurrence resides in the central prairies, with a broad clockwise rotated ''J-shaped'' region over south-central Saskatchewan and southern Manitoba. The highest tornado occurrence

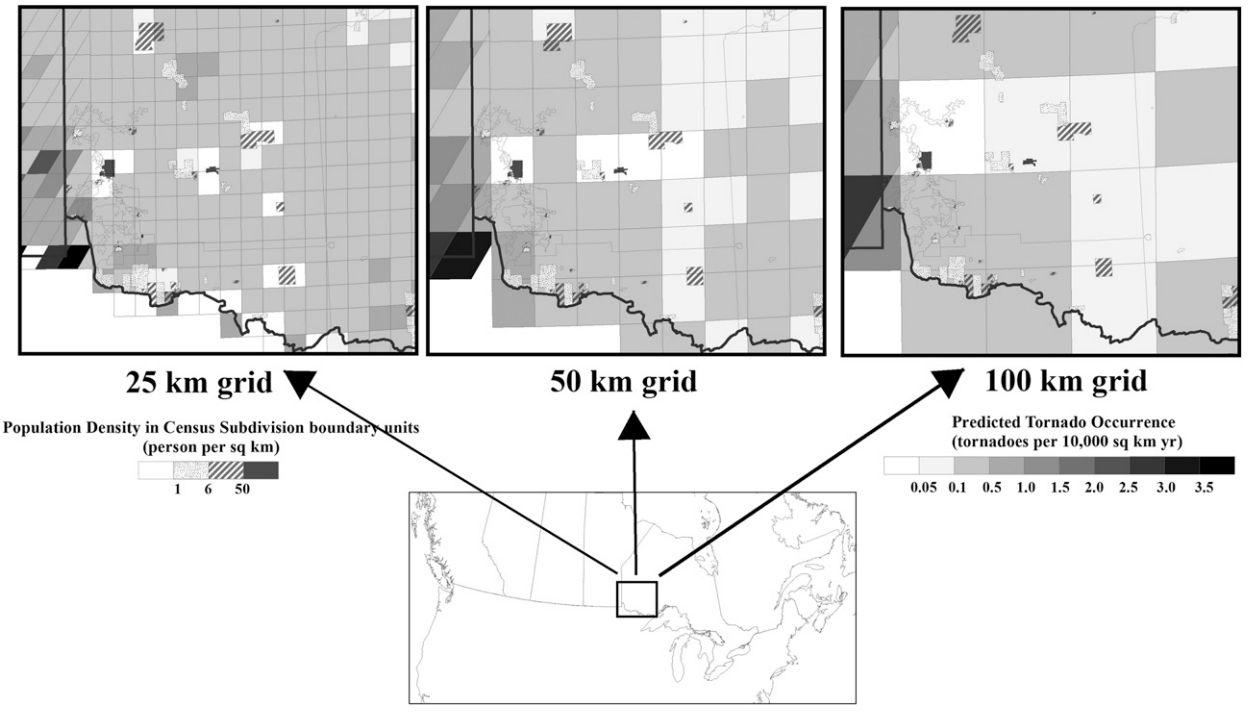

FIG. 5. Posterior mean of Tlatent_i on 25×25, 50×50, and $100 \times 100\,\text{km}^2$ model grids overlaid on the original census subdivision boundary layer with the population density (person km^{-2}) of each census subdivision shown with gray scales. The gray scale for Tlatent_i is the same as those in Fig. 2.

predictions within this region are found in the center of southern Saskatchewan, southeastern Saskatchewan, and southwestern and southeastern Manitoba, with the latter three lying just north of the broad ''C-shaped'' tornado alley in the central plains of the United States (Brooks et al. 2003a). Because of the absence of Canadian data along with the smoothing effect of the Brooks et al. (2003a)'s study, the C-shaped tornado alley did not extend to the northern plains/Canada–United States border (e.g., Dakotas and Montana). Our analysis suggests that the broad C-shaped tornado alley likely extends north into Manitoba and Saskatchewan. Other high probability areas are located within southern Ontario. One area can be found in the southernmost part of Canada near Windsor, Ontario, and appears to be a northeastern extension of the aforementioned tornado alley identified by Brooks et al. (2003a). The other area is located in the inland area of southern Ontario between Lakes Erie, Huron, and Ontario. This region is largely influenced by the lake-breeze circulations generated by these lakes; the tornado activity tends to be suppressed in regions near the lakes regularly visited by lake-modified air, while it is enhanced at inland locations along lake-breeze fronts as well as at sites where lake-breeze fronts interact (King et al. 2003). All the high probability areas have experienced deadly tornadoes in the past (Table 2).

When tornado-affected areas are known, tornado occurrence in area and time ($10\,000\,\text{km}^{-2}\,\text{yr}^{-1}$) can be expressed as tornado probabilities P_T in unit per time (yr^{-1}). Tornado probabilities are typically calculated as follows:

$$P_T = \left(\sum_{t=1}^{T} l_t w_t \right) \Big/ AY, \tag{11}$$

where l is the length of the tornado t, w is the width of the tornado t, A is the area of the grid, Y is the number of years, and T is the number of tornadoes in area A. However, for most of the observations in this dataset, the tornado-affected areas are unknown, and thus it would be impossible to use Eq. (11). In an attempt to predict tornado probabilities, we assume our sampled tornado-affected areas are similar to the large dataset of Schaefer et al. (1986). We applied their sampled median affected areas of all tornadoes (F0–F5) of $0.10\,\text{km}^2$ along with our Tlatent_i to predict tornado probabilities, which are calculated as

$$P_T(\text{yr}^{-1}) = \text{Tlatent}_i(\text{occurrence} \times 10,000\,\text{km}^{-2}\,\text{yr}^{-1}) \times 0.10\,\text{km}^2. \tag{12}$$

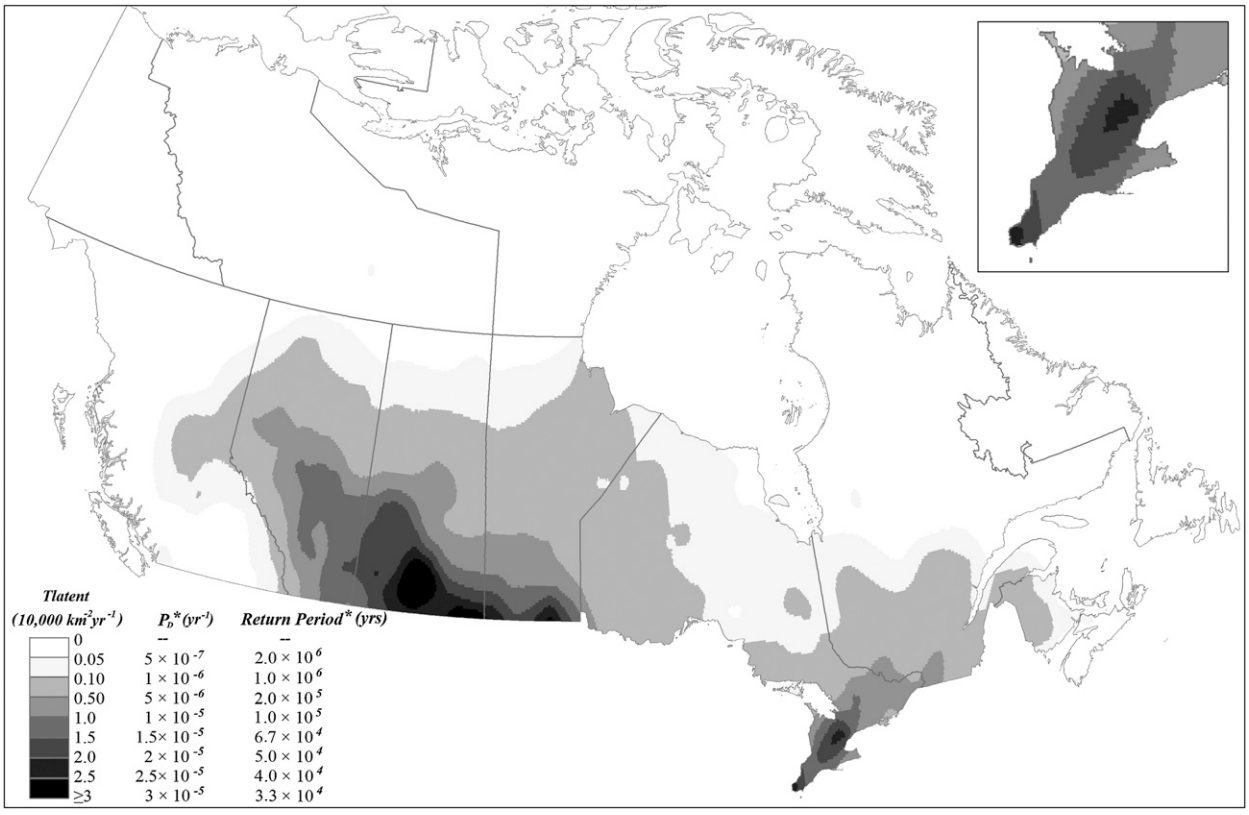

FIG. 6. Predicted tornado occurrence ($10\,000\,km^{-2}\,yr^{-1}$), tornado probability ($yr^{-1}$), and the return period (yr), based on the posterior mean values of $Tlatent_i$ of the $25 \times 25\,km^2$ model. Asterisk implies that the median tornado-affected area (F0–F5) of $0.10\,km^2$ from the Schaefer et al. (1986) records is used. An inset map of southwestern and south-central Ontario is shown in the upper right corner.

Based on this assumption, the tornado probabilities and the return periods are directly proportional to the predicted tornado occurrences (Fig. 6): for example, areas where the tornado occurrences are predicted to be 1 tornado $10\,000\,km^{-2}\,yr^{-1}$ have a tornado probability of $10^{-5}\,yr^{-1}$, which is equivalent to a return period of 10^5 yr. Our results in tornado probabilities can be compared with the F0–F5 projections reported by Schaefer et al. (1986). In Schaefer et al. (1986), the areas with tornado probabilities greater than $10^{-6}\,yr^{-1}$ covered most of the United States east of the Rocky Mountains. Based on our results, the same frequency areas should extend well

TABLE 2. Regions of ≥1 and ≥2 tornado occurrences per $10\,000\,km^2 \times yr^{-1}$ and historically significant tornado events occurred within each region.

Tornadoes $10\,000\,km^{-2}\,yr^{-1}$	Regions	Significant tornado events (fatalities or F-scale)
≥1	Central Alberta	Edmonton 1987 (27), Pine Lake 2000 (11)
	Western-Central Saskatchewan	Lloydminster 1983 and 2000 (F3s)
	Southern Manitoba	Elie 2007 (F5)
	Southern Ontario	Waterloo–Wellington 1967 (F3), Gray–Dufferin Counties 1996 (2 F3s), southern Ontario outbreak 2005 (2 F2s)
≥2	Southern Saskatchewan	Regina 1912 (28)
	Southern Manitoba–United States border	Portage La Prairie 1922 (5)
	Tip of southwestern Ontario–United States border	Sarnia 1953 (7), Windsor/Tecumseh 1946 (17), Windsor 1974 (9)
	Southwestern Ontario lake convergence zone	Woodstock 1979 (2), Hopeville—Barrie 1985 (12)

northward to the Canadian plains, east of the Alberta Rocky Mountains, and also to northwestern Ontario. The same frequency also characterizes areas that extend from southwestern Ontario northward to the east of Lake Superior and eastward to southwestern Quebec and New Brunswick. Regarding the tornado probabilities of greater than $10^{-5}\,\mathrm{yr}^{-1}$, Schaefer et al. (1986)'s projections lie closer to southwestern Ontario than to the Canadian Prairies. However, our result suggests that a significant portion of the Canadian Prairies close to the U.S. northern prairies regions should also be included, as well as a much larger portion of southwestern Ontario (i.e., from Windsor, Ontario, extending to the east of Georgian Bay).

5. Conclusions and future perspectives

Newark (1984) provided a first national view of the tornado climatology of Canada. In that study, the tornado observations were assumed to be reliable in areas with population density ≥ 1 person km^{-2}, while areas with lower population density were extrapolated based on subjective meteorological knowledge. In this study, we developed a Bayesian modeling approach founded upon the explicit consideration of the population sampling bias in tornado observations and the predictive relationship between cloud-to-ground lightning flash climatology and tornado occurrence, in order to predict the probability of tornado occurrence in Canada. The key findings of our analysis are as follows:

- There is a population density threshold of 6 individuals km^{-2}, below which there is uncertainty in the detection of tornadoes. Nonetheless, regional variability may be associated with differences in the quality of the monitoring networks, number of trained weather spotters, training of meteorological service staff, and general public awareness.
- Mean annual CG lightning density is an important meteorological covariate for partially explaining tornado occurrence, although substantial spatial variability exists in regards to the strength of this predictive relationship.
- In sparsely populated areas, our analysis shows that the probability of tornado occurrence is significantly higher than what is represented in the 30-yr data record. Areas with low population density but high lightning flash density demonstrate the greatest discrepancy between predicted and observed tornado occurrence.
- The total predicted tornado occurrences for the entire 30-yr period of our study were at least twice the number of recorded tornadoes, based on the counts of tornado points and paths by grid cells. Tornado observations were on average close to $70\,\mathrm{yr}^{-1}$, whereas the model predicted tornado occurrence close to $150\,\mathrm{yr}^{-1}$.

The sensitivity analysis shows that the predictive statements supported by the model are fairly robust to the grid configuration; hence, the modeling framework should be a sound foundation for objectively defining tornado-prone areas for the National Building Code of Canada. The ability of the model to delineate the importance of meteorological factors and the observational error associated with the population density lays the groundwork for a more detailed investigation of these relationships. Future research goals would be to examine additional lightning flash properties (e.g., lightning polarity, total lightning density, multiplicity, and first stroke peak current) and other atmospheric meteorological parameters (e.g., convective potential available energy and deep-level wind shear) that could potentially be relevant to tornado occurrence and their intensity (Brooks et al. 2003b). Moreover, future investigation of the lightning–tornado relationship via a Bayesian hierarchical framework will be valuable in elucidating the underlying drivers of spatial variability. Similarly, this modeling framework would be helpful in identifying potential regional population bias associated with the likelihood of tornado detection. Such a hierarchical configuration can conceivably augment the capacity of tornado occurrence models to effectively support hazard assessment (Banik et al. 2007), as the realistic representation of spatial heterogeneity is critical in developing tornado hazard maps (Tan and Hong 2010). Further, manufactured structures, such as mobile homes and school portables are vulnerable even to weaker (F0 and F1) tornadoes (Ashley 2007; Sutter and Simmons 2010), whereas critical and hazardous facilities, such as hospitals and nuclear plants, may be subjected to a greater risk with more intense (F2 and higher) tornadoes. It is thus critical to develop models for various tornado intensities, so that the associated risks with different types of infrastructure can be evaluated. Such a modeling exercise will be presented in a future study.

From a socioeconomic perspective, one of most direct impacts of tornadoes is their financial cost on individuals and infrastructure. The large financial losses suffered from recent tornadic events, such as the Goderich 2011 (CAD 75 million) and the Leamington 2010 (CAD 120 million) Ontario tornadoes (Insurance Bureau of Canada 2012), have shown that the destruction of infrastructure and losses to individual owners and insurers can be overwhelming. These events have highlighted the importance of developing tornado forecasting tools and

risk information systems for engineering and financial loss models that will improve our capacity to cope with damage control and minimize the societal risk of tornadoes. Canada is undergoing significant population growth and expansion, centering mainly in Alberta and southern Ontario, and thus it is expected that societal impacts of tornadoes will increase. It is critical that the tornado risk to Canada's society is properly assessed, for example, through tornado casualty models, and that adequate tornado vigilance from the Meteorological Service of Canada, emergency groups, and by the general public are in place. Our predictions of tornado occurrence should be essential in integrating tornado risk with future socioeconomic studies.

Acknowledgments. Vaisala Inc. of Tucson, Arizona, processed the Canadian Lightning Detection Network data and provided telecommunication services to Environment Canada. We thank Bill Burrows (Environment Canada) for processing and providing the $1 \times 1\,\mathrm{km}^2$ gridded lightning data.

REFERENCES

Allen, D. E., 1984: Tornado damage at Blue Sea Lake and Nicabong, Quebec, July 1984. National Research Council of Canada Building Research Note 222, 12 pp.

——, 1986: Tornado damage in the Barrie/Orangeville area, Ontario, May 1985. National Research Council of Canada Building Research Note 240, 23 pp.

——, 1992: A design basis tornado: Discussion of paper by M. J. Newark. *Can. J. Civ. Eng.,* **19,** 361.

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.

Arhonditsis, G. B., S. S. Qian, C. A. Stow, E. C. Lamon, and K. H. Reckhow, 2007: Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecol. Modell.,* **208,** 215–229.

——, D. Papantou, W. Zhang, G. Perhar, E. Massos, and M. Shi, 2008a: Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. *J. Mar. Syst.,* **73,** 8–30.

——, G. Perhar, W. Zhang, E. Massos, M. Shi, and A. Das, 2008b: Addressing equifinality and uncertainty analysis in eutrophication modelling. *Water Resour. Res.,* **44,** W01420, doi:10.1029/2007WR005862.

Ashley, W. S., 2007: Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005. *Wea. Forecasting,* **22,** 1214–1228.

Ashton, A., M. Leduc, and S. Boodoo, cited 2010a: August 20th 2009—The day the skies rained tornadoes. [Available online at http://www.yorku.ca/glomw10/Tues_March_23_AM/Ashton_Leduc_August%2020th_Talk%20One.ppt.]

——, ——, and M. Firmin, cited 2010b: The August 20th 2009 tornado outbreak: Synoptic overview and the challenges of forecasting an extreme event. [Available online at http://www.yorku.ca/glomw10/Tues_March_23_AM/Ashton_LeducAugust%2020th%20Talk%20TWO.ppt.]

Banik, S. S., H. P. Hong, and G. A. Kopp, 2007: Tornado hazard assessment for southern Ontario. *Can. J. Civ. Eng.,* **34,** 830–842.

Besag, J., and C. L. Kooperberg, 1995: On conditional and intrinsic autoregressions. *Biometrika,* **82,** 733–746.

——, J. York, and A. Mollie, 1991: Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Stat. Math.,* **43,** 1–59.

Branick, M. L., and C. A. Doswell III, 1992: An observation of the relationship between supercell structure and lightning ground-strike polarity. *Wea. Forecasting,* **7,** 143–149.

Brooks, H. E., C. A. Doswell III, and M. P. Kay, 2003a: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting,* **18,** 626–640.

——, J. W. Lee, and J. P. Craven, 2003b: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.,* **67–68,** 73–94.

Burrows, W. R., and B. Kochtubajda, 2010: A decade of cloud-to-ground lightning in Canada: 1999–2008. Part 1: Flash density and occurrence. *Atmos.–Ocean,* **48,** 177–194.

Carey, L. D., and S. A. Rutledge, 2003: Characteristics of cloud-to-ground lightning in severe and nonsevere storms over the central United States from 1989–1998. *J. Geophys. Res.,* **108,** 4483, doi:10.1029/2002JD002951.

——, ——, and W. A. Peterson, 2003: The relationship between severe storm reports and cloud-to-ground lightning polarity in the contiguous United States from 1989 to 1998. *Mon. Wea. Rev.,* **131,** 1211–1228.

Carter, A. O., M. E. Millson, and D. E. Allen, 1989: Epidemiologic study of deaths and injuries due to tornadoes. *Amer. J. Epidemiol.,* **130,** 1209–1218.

Changnon, S. A., 1988a: Climatography of thunder events in the conterminous United States. Part I: Temporal aspects. *J. Climate,* **1,** 389–398.

——, 1988b: Climatography of thunder events in the conterminous United States. Part II: Spatial aspects. *J. Climate,* **1,** 399–405.

——, 1989: Relations of thunderstorms and cloud-to-ground lightning frequencies. *J. Climate,* **2,** 897–921.

——, 1993: Relations between thunderstorms and cloud-to-ground lightning in the United States. *J. Appl. Meteor.,* **32,** 88–105.

Charlton, R., B. Kachman, and L. Wojtiw, 1995: Urban hailstorms—A view from Alberta. *Nat. Hazards,* **12,** 29–75.

Cheng, V., G. B. Arhonditsis, and M. T. Brett, 2010: A revaluation of lake-phosphorus loading models using a Bayesian hierarchical framework. *Ecol. Res.,* **25,** 59–76.

Clark, J. S., 2005: Why environmental scientists are becoming Bayesians. *Ecol. Lett.,* **8,** 2–14.

——, and A. E. Gelfand, 2006: A future for models and data in ecology. *Trends Ecol. Evol.,* **21,** 375–380.

Doswell, C. A., III, and D. W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.,* **116,** 495–501.

Etkin, D., and M. Leduc, 1994: A non-tornadic severe storm climatology of southern Ontario adjusted for population bias: Some surprising results. *CMOS Bull.,* **21,** 4–8.

——, S. E. Brun, A. Shabbar, and P. Joe, 2001: Tornado climatology of Canada revisited: Tornado activity during different phases of ENSO. *Int. J. Climatol.,* **21,** 915–938.

Fujita, T. T., 1981: Tornadoes and downbursts in the context of generalized planetary scales. *J. Atmos. Sci.,* **38,** 1511–1534.

Gelman, A., and J. Hill, 2007: *Data Analysis Using Regression and Multilevel/Hierarchical Models.* 2nd ed. Cambridge University Press, 625 pp.

——, J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004: *Bayesian Data Analysis.* 2nd ed. Chapman and Hall/CRC, 696 pp.

Goodchild, M. F., L. Anselin, and U. Deichman, 1993: A framework for the areal interpolation of socio-economic data. *Environ. Plann.,* **25A,** 383–397.

Grazulis, T. P., and R. F. Abbey Jr., 1983: 103 years of violent tornadoes: Patterns of serendipity, population, and mesoscale topography. Preprints, *13th Conf. on Severe Local Storms,* Tulsa, OK, Amer. Meteor. Soc., 124–127.

Hage, K., 2003: On destructive Canadian Prairie windstorms and severe winters. *Nat. Hazards,* **29,** 207–228.

Huffines, G. R., and R. E. Orville, 1999: Lightning ground flash density and thunderstorm duration in the continental United States: 1989–96. *J. Appl. Meteor.,* **38,** 1013–1019.

Insurance Bureau of Canada, cited 2012: Goderich tornado costs $75 million in insured damage. [Available online at http://www.ibc.ca/en/media_centre/news_releases/2011/09-21-2011.asp.]

King, P., 1997: On the absence of population bias in the tornado climatology of southwestern Ontario. *Wea. Forecasting,* **12,** 939–946.

——, M. J. Leduc, D. M. L. Sills, N. R. Donaldson, D. R. Hudak, P. Joe, and B. P. Murphy, 2003: Lake breezes in southern Ontario and their relation to tornado climatology. *Wea. Forecasting,* **18,** 795–807.

Knapp, D. I., 1994: Using cloud-to-ground lightning data to identify tornadic thunderstorm signatures and nowcast severe weather. *Natl. Wea. Dig.,* **19,** 35–42.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter, 2000: WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.,* **10,** 325–337.

MacGorman, D. R., and D. W. Burgess, 1994: Positive cloud-to-ground lightning in tornadic storms and hailstorms. *Mon. Wea. Rev.,* **122,** 1671–1697.

NBCC, 2005: National Building Code of Canada. Institute for Research in Construction, National Research Council of Canada, Ottawa, Ontario.

Newark, M. J., 1983: Tornadoes in Canada for the period 1950 to 1979. Atmospheric Environment Service Rep. CLI 2-83, 88 pp.

——, 1984: Canadian tornadoes, 1950–1979. *Atmos.–Ocean,* **22,** 343–353.

Paruk, B. J., and S. R. Blackwell, 1994: A severe thunderstorm climatology for Alberta. *Natl. Wea. Dig.,* **19,** 27–33.

Perez, A. H., L. J. Wicker, and R. E. Orville, 1997: Characteristics of cloud-to-ground lightning associated with violent tornadoes. *Wea. Forecasting,* **12,** 428–437.

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.

Reap, R. M., and D. R. MacGorman, 1989: Cloud-to-ground lightning: Climatological characteristics and relationships to model fields, radar observations, and severe local storms. *Mon. Wea. Rev.,* **117,** 518–535.

——, and R. E. Orville, 1990: The relationships between network lightning locations and surface hourly observations of thunderstorms. *Mon. Wea. Rev.,* **118,** 94–108.

Schaefer, J. T., and J. G. Galway, 1982: Population biases in tornado climatology. Preprints, *12th Conf. on Severe Local Storms,* San Antonio, TX, Amer. Meteor. Soc., 51–54.

——, D. L. Kelly, and R. F. Abbey, 1986: A minimum assumption tornado-hazard probability model. *J. Climate Appl. Meteor.,* **25,** 1934–1945.

Shephard, M. W., R. Morris, W. R. Burrows, and L. Welsh, 2013: High-resolution Canadian lightning climatology. *Atmos.–Ocean,* **51,** 50–59.

Sills, D. M. L., V. Cheng, P. McCarthy, B. Rousseau, J. Waller, L. Elliot, J. Klaassen, and H. Auld, 2012: Using tornado, lightning, and population data to identify tornado prone areas in Canada. Preprints, *26th Conf. on Severe Local Storms,* Nashville, TN, Amer. Meteor. Soc., P59. [Available online at https://ams.confex.com/ams/26SLS/webprogram/Paper211359.html.]

Snider, C. R., 1977: A look at Michigan tornado statistics. *Mon. Wea. Rev.,* **105,** 1341–1342.

Statistics Canada, cited 2001: Census subdivision reference maps. [Available online at http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=92F0149XIB&lang=eng.]

Sutter, D., and K. M. Simmons, 2010: Tornado fatalities and mobile homes in the United States. *Nat. Hazards,* **53,** 125–137.

Tan, L., and H. P. Hong, 2010: Influence of spatial inhomogeneity of tornado occurrence on estimated tornado hazard. *Can. J. Civ. Eng.,* **37,** 279–289.

Taylor, N. M., and Coauthors, 2011: The Understanding Severe Thunderstorms and Alberta Boundary Layers Experiment (UNSTABLE) 2008. *Bull. Amer. Meteor. Soc.,* **92,** 739–763.

Tescon, 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.

Wikle, C. K., 2003: Hierarchical models in environmental science. *Int. Stat. Rev.,* **71,** 181–199.

——, and C. J. Anderson, 2003: Climatological analysis of tornado report counts using a hierarchical Bayesian spatiotemporal model. *J. Geophys. Res.,* **108,** 9005, doi:10.1029/2002JD002806.

Zhang, W., and G. B. Arhonditsis, 2009: A Bayesian hierarchical framework for calibrating aquatic biogeochemical models. *Ecol. Modell.,* **220,** 2142–2161.

Supplementary Information for Probability of Tornado Occurrence across Canada

VINCENT Y. S. CHENG, GEORGE B. ARHONDITSIS, DAVID M. L. SILLS, HEATHER AULD, MARK W. SHEPHARD, WILLIAM A. GOUGH, AND JOAN KLAASSEN

Figure S1: Mean annual cloud-to-ground light flash density (number of flashes km^{-2} $year^{-1}$), based on data from 1999-2009, in 25 × 25 km^2 grid, 50 × 50 km^2 grid and 100 × 100 km^2 grid, respectively**.**

Figure S2: Population density, y_i, in 25 × 25 km^2 model grid, 50 × 50 km^2 grid and 100 × 100 km^2 grid, respectively.

Figure S3: Posterior mean of the probability of tornado detection, p_i, in 25 × 25 km^2 model grid, 50 × 50 km^2 grid and 100 × 100 km^2 grid, respectively.

Figure S4: Posterior mean of the conditional autoregression coefficient, a_i, in 25 × 25 km^2 model grid, 50 × 50 km^2 grid and 100 × 100 km^2 grid, respectively.

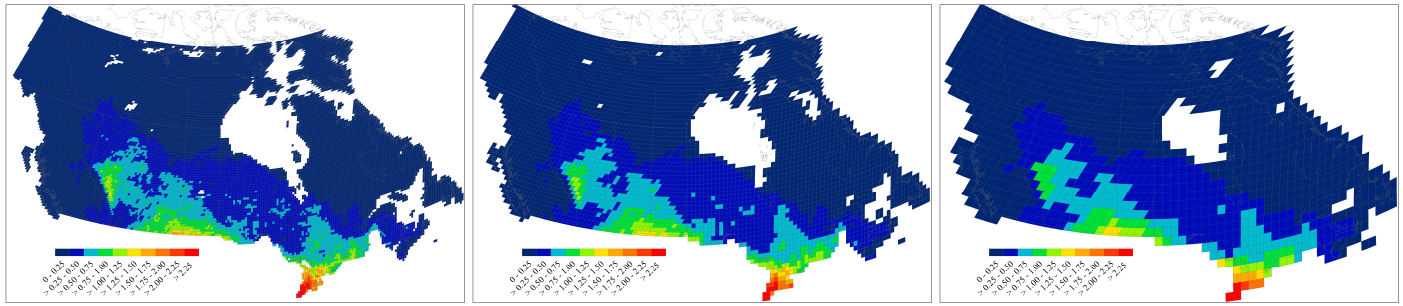

Figure S1. Mean annual cloud-to-ground light flash density (number of flashes km^{-2} $year^{-1}$), based on data from 1999-2009, in 25 × 25 km^2 grid, 50 × 50 km^2 grid, and 100 × 100 km^2 grid, respectively.

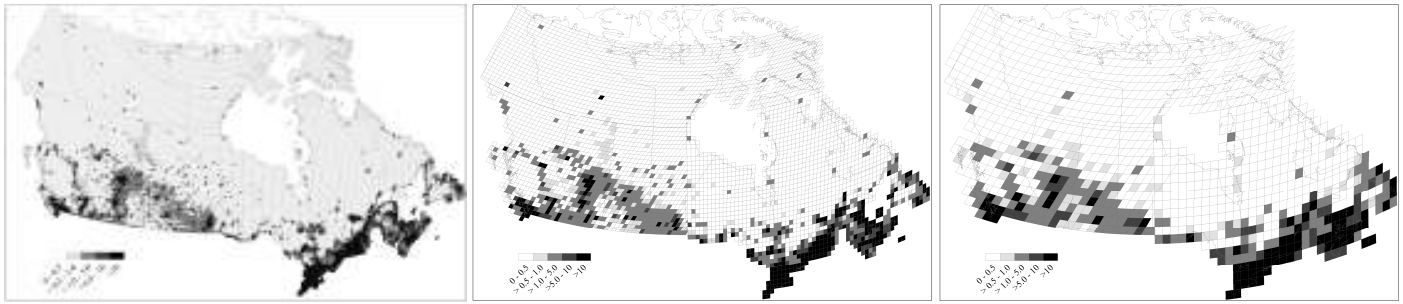

Figure S2. Population density, y_i, in 25 × 25 km^2 model grid, 50 × 50 km^2 grid, and 100 × 100 km^2 grid, respectively.

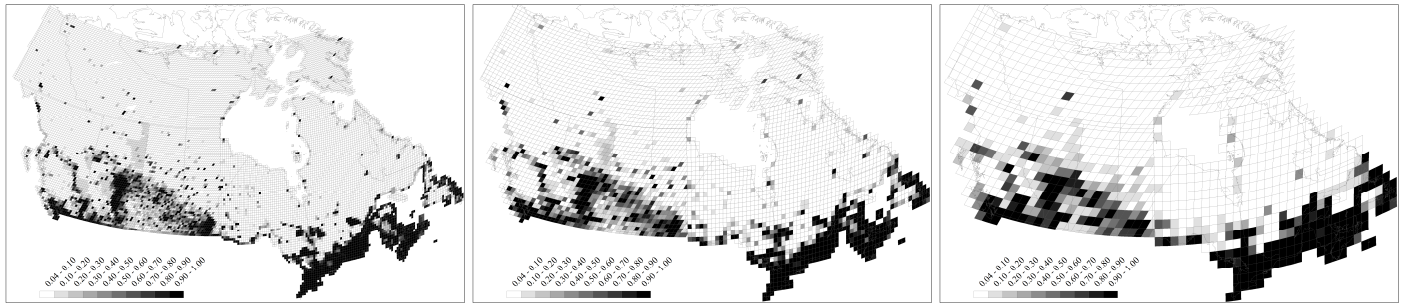

Figure S3. Posterior mean of the probability of tornado detection, p_i, in 25 × 25 km^2 model grid, 50 × 50 km^2 grid, and 100 × 100 km^2 grid, respectively.

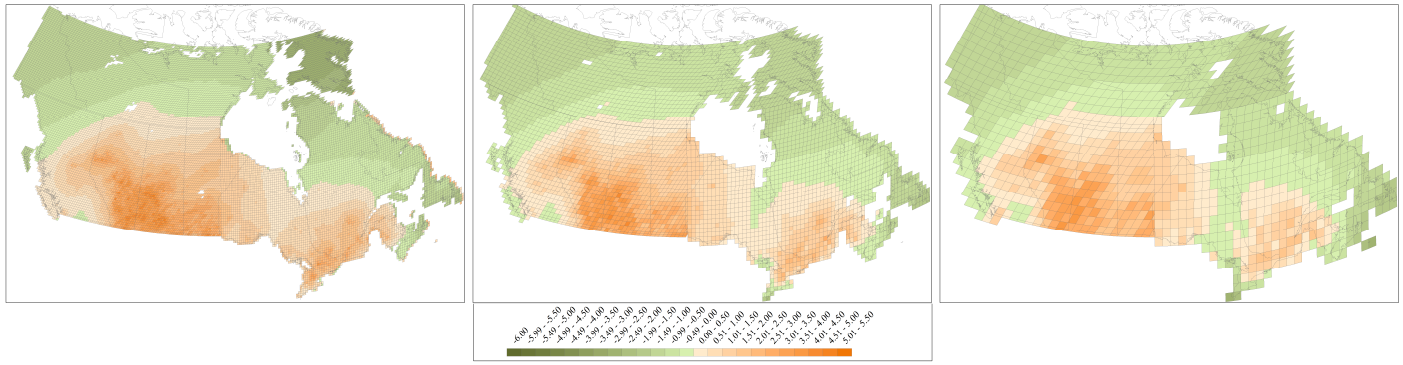

Figure S4. Posterior mean of the conditional autoregression coefficient, a_i, in 25 × 25 km^2 model grid, 50 × 50 km^2 grid, and 100 × 100 km^2 grid, respectively.