

ARTICLE

Received 3 Jul 2014 | Accepted 11 Feb 2015 | Published 25 Mar 2015

DOI: 10.1038/ncomms7599

A Bayesian modelling framework for tornado occurrences in North America

Vincent Y.S. Cheng^{1,2}, George B. Arhonditsis¹, David M.L. Sills³, William A. Gough² & Heather Auld⁴

Tornadoes represent one of nature's most hazardous phenomena that have been responsible for significant destruction and devastating fatalities. Here we present a Bayesian modelling approach for elucidating the spatiotemporal patterns of tornado activity in North America. Our analysis shows a significant increase in the Canadian Prairies and the Northern Great Plains during the summer, indicating a clear transition of tornado activity from the United States to Canada. The linkage between monthly-averaged atmospheric variables and likelihood of tornado events is characterized by distinct seasonality; the convective available potential energy is the predominant factor in the summer; vertical wind shear appears to have a strong signature primarily in the winter and secondarily in the summer; and storm relative environmental helicity is most influential in the spring. The present probabilistic mapping can be used to draw inference on the likelihood of tornado occurrence in any location in North America within a selected time period of the year.

¹Ecological Modeling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, Canada M1C 1A4. ²Climate Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, Canada M1C 1A4. ³Cloud Physics and Severe Weather Research Section, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada, Toronto, Ontario, Canada M3H 5T4. ⁴Risk Sciences International, Ottawa, Ontario, Canada K1N 7G2. Correspondence and requests for materials should be addressed to G.B.A. (email: georgea@utsc.utoronto.ca).

Tornadoes represent one of nature's most hazardous phenomena, capable of causing significant destructions and devastating injuries and fatalities. In May of 2011, the single deadliest tornado event in history occurred with 158 deaths in Joplin, Missouri¹. Over the same year, tornadoes in the United States caused a total of over \$28 billion in property damages¹. Canada has also experienced significant losses caused by tornado activity, with significant tornado events occurring in recent years (for example, 14 tornadoes on 2 August 2006; 18 tornadoes on 20 August 2009)². These events have brought the study of tornado activity to the forefront of climatological research, raising questions regarding the relative influence of large-scale climatic signals and the importance of delineating the role of the underlying atmospheric processes at the appropriate spatiotemporal scale. Tornado climatology represents a significant source of information used by a wide variety of groups, such as weather forecasters and researchers, emergency managers, civil engineers and insurance companies alike. Generally, the probability of tornado occurrence is derived from gridded values of tornado observations coupled with smoothing methods (for example, kernel densities) that are typically applied to delineate large-scale spatial patterns and to subsequently improve estimates for any single location^{3,4}. However, these approaches rarely associate large-scale tornado variability with atmospheric variables that vary on relevant climatic timescales^{5,6}. Predicting tornado activity from surrogate variables of the associated atmospheric processes can overcome issues related to data quality, particularly in areas where tornado observations are unreliable to form the basis of robust inference, due to either population density biases and/or the rarity of occurrence^{2,7}. More importantly, the explicit consideration of large-scale climatic processes can shed light on the principal mechanisms of tornadogenesis and assess tornado frequency under different climate change scenarios or climate analogue studies⁶.

Characterization of the interplay between large-scale atmospheric processes and local conditions leading to tornado formation has been examined in the severe storm research community^{7–9}. Atmospheric variables such as convective available potential energy and vertical wind shear have been shown to be most relevant to mechanisms related to severe thunderstorm and tornado formation⁷. Conditions conducive to more intense tornadoes (F2–F5) appear to be more distinguishable than lower-end tornadoes (F0–F1), particularly by vertical shear parameters^{9–11}. It has also been recognized that favourable conditions for tornado formation tend to vary seasonally. Namely, cool-season tornadoes are predominantly related to high vertical wind shear/low buoyancy regimes,

particularly for more intense (F2–F5) tornadoes, relative to the considerably greater buoyancy/high shear conditions generating tornadoes in the warm season^{12–17}. Recent research has also shown that the convective storm morphology and near-storm environment associated with tornadoes have clear seasonal and geographic patterns^{11,18,19}; notably, spring tornadoes in the Great Plains are produced by discrete, supercellular storms as compared with the predominantly quasi-linear convective systems in the Ohio Valley and the southeast United States^{11,18,20}. Evidence that the prevailing conditions conducive to tornado occurrence exhibit clear seasonal and geographical variability leads us to hypothesize that the detailed characterization of their spatiotemporal patterns may address the current predictive deficiencies and consequently advance our capacity to effectively develop prognostic tools for tornado occurrence.

In this study, the main objective is to delineate the seasonal and geographic variability of the relationship between large-scale climatological/atmospheric variables and both F0–F5 and F2–F5 tornado occurrence. Our analysis identifies the optimal combination of atmospheric variables and explicitly considers the sampling biases in tornado observations to collectively predict the monthly and seasonal tornado activity in North America. Specifically, we use Bayesian inference techniques to dissect the problem of tornado occurrence into a two step-process, in which we first consider the causal linkages between atmospheric variables and tornado activity in space and time, and we then postulate that the likelihood to observe a tornado is closely related to the population density. A conditional autoregressive term is also introduced in the statistical formulation to explicitly accommodate the serial correlation patterns of model residuals. Our ultimate goal is to develop an empirical modelling tool that can be used for drawing inference on the frequency of tornado occurrence or exceedance of a threshold number of tornadoes in any location in North America within a selected time period of the year.

Results

Model performance and posterior parameter patterns. We investigated seven combinations of predictor variables to predict the tornado occurrence (F0–F5 and F2–F5) for each calendar month and season independently (Supplementary Table 1). All of the predictor variables were standardized to examine their relative importance and possibly control any multicollinearity issues. We used two criteria to identify the optimal model configuration: (i) model performance as depicted by the deviance values or $-2 \cdot \log[\text{model likelihood}]$ (Supplementary Tables 2 and 3) and

Table 1 | Posterior parameter mean and s.d. values of the highest-performing model that considers the variables CAPE, HLCY and VWSH to predict F0–F5 tornadoes.

F0F5 Model	α_0		α_1 —CAPE		α_2 —HLCY		α_3 —VWSH		β		σ	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
January	−8.50	1.04	0.23	0.10	0.02	0.23	4.46	1.04	5.57	1.46	2.66	0.16
February	−7.93	0.90	0.43	0.09	0.25	0.20	3.98	0.90	5.63	1.21	2.39	0.15
March	−4.62	0.40	0.14	0.10	0.51	0.12	1.41	0.44	3.06	0.46	1.95	0.09
April	−2.94	0.25	0.19	0.11	0.33	0.09	0.69	0.37	3.25	0.33	1.67	0.06
May	−1.81	0.14	0.25	0.11	0.26	0.07	0.10	0.21	2.82	0.19	1.62	0.05
June	−1.92	0.14	0.49	0.10	0.25	0.07	0.71	0.31	2.32	0.16	1.72	0.05
July	−2.12	0.13	0.72	0.11	0.13	0.07	0.98	0.16	2.27	0.17	1.82	0.06
August	−2.46	0.15	0.53	0.12	0.14	0.10	0.67	0.20	2.22	0.20	2.15	0.08
September	−3.18	0.21	0.44	0.11	0.30	0.12	0.47	0.29	2.81	0.31	2.10	0.09
October	−3.99	0.35	0.08	0.10	0.16	0.14	0.33	0.42	3.84	0.51	2.43	0.11
November	−4.25	0.50	0.12	0.08	0.44	0.17	0.09	0.64	6.22	1.12	2.19	0.11
December	−6.72	0.89	0.06	0.10	0.01	0.22	2.11	1.01	6.14	1.59	2.44	0.16

(ii) the identifiability patterns of the parameter posteriors. When combined with CAPE, models with SHEAR and to a lesser extent with VWSH tended to perform better than the ones with HLCY in February to May. HLCY and SHEAR improved performance in June and July, respectively. VWSH tended to improve model performance in late summer-early autumn, while SHEAR outperformed the rest of the potential predictors throughout autumn until December. Interestingly, when reproducing F2–F5 tornadic activity, model performance with HLCY became better in January and VWSH improved model fit in June. On the basis of the two criteria of model performance and parameter identification, we inferred that the optimal model considers the variables CAPE, HLCY, VWSH to predict monthly values of F0–F5 and F2–F5 tornado counts. (Notably, the *post hoc* implementation of the deviance information criterion values suggests that the same model is not only the highest performing, but also the most parsimonious one in the majority of the cases examined.) The corresponding parameter posterior means and s.d. are shown in Tables 1 and 2, respectively. Results for the rest of the models examined are shown in Supplementary Tables 4–15.

The posteriors of the standardized regression coefficients were characterized by clear seasonal patterns, which were consistent across all models tested. According to our modelling analysis, CAPE has a predominantly positive influence on tornado activity during the summer months (June to August) and a secondary (but distinct) one in February. VWSH demonstrated its strongest positive linkage with the tornado occurrence during winter months (December to March) and to a lesser extent in summer. The signature of HLCY was more distinct during the spring (March–April) and somewhat weaker in late fall and winter for the F0–F5 and F2–F5 models, respectively. Interestingly, although it was not included in the highest-performing model among the combinations examined, SHEAR was characterized by fairly similar seasonal patterns with VWSH with respect to its influence on model predictions (see Supplementary Tables 4–15). Regarding the relative importance of the different predictors, the standardized regression coefficients suggest that VWSH is most closely related to the variability of tornado occurrence (F0–F5 and F2–F5) in the winter, and together with CAPE appear to shape the tornadic activity in the summer. HLCY is more influential in early spring, although its posterior standardized mean values are slightly lower than those for VWSH (and SHEAR).

Our analysis revealed several identifiability issues with interesting implications about the relative significance of the different predictor variables, when considered collectively as predictors in modelling frameworks of tornadic activity. For example, the signature of VWSH was significantly compromised

after the addition of SHEAR with F2–F5 models in spring and fall months (see posteriors in Supplementary Tables 9 and 13). On the other hand, SHEAR is well identified with a distinct signature across all the F2–F5 models, with September being the only exception. The HLCY posteriors remain fairly similar with respect to their absolute magnitude and identifiability by the addition of VWSH (Tables 1 and 2 versus Supplementary Tables 4 and 5), whereas the addition of SHEAR influences HLCY more significantly during the fall and winter (Supplementary Tables 4 and 5 versus Supplementary Tables 10 and 11). The posteriors of CAPE are well identified for all F0–F5 models as well as for F2–F5 models during the summer maximal activity. The signature of CAPE remains fairly well delineated and practically unaltered by the different permutations of the other (wind shear-related) predictor variables.

The differences of the random effect terms, ϕ , between the two-predictor CAPE–HLCY and the three-predictor CAPE–HLCY–VWSH F0–F5 (F2–F5) models for each month are shown in Fig. 1 (Supplementary Fig. 2). Substantial differences are found in January and February (southeast United States) and from June to August (Eastern Atlantic Canada/United States to the Great Lakes), where the magnitude of the autoregressive term ϕ from the CAPE–HLCY–VWSH model is smaller. Interestingly, these are also the months when the impact of VWSH is more distinct, and thus its inclusion into the model allows us to effectively capture the variability of tornado occurrence in the winter and summer. On the other hand, the spatiotemporal patterns of the ϕ term with the two models are practically identical during the spring and autumn months.

Seasonal variability characterized the posterior values of the population effect parameter, β . The highest values were derived for the winter months and the lowest ones during the summer, that is, a higher population density is needed to observe tornadoes during the former rather than the latter months (Tables 1 and 2). Substituting β back into equation (2) to identify the population thresholds where nearly all tornadoes can be observed ($p_i(\beta) \geq 0.995$), the threshold levels ranged from 6.2 to 7.2 persons km⁻² for F0–F5 tornadoes, which is consistent with the findings of other studies^{2,21}, and 6.4 to 14.0 persons km⁻² for F2–F5 tornadoes (Supplementary Figs 3 and 4). These predicted patterns may be related to the seasonal variability in the day length, which in turn affects the ability to observe tornadoes²². Moreover, because of the higher proportion of nocturnal tornadoes occurring in the winter months (10% in July to 40% in February for F2–F5 tornadoes), the daylight effect on the ability of observing tornadoes may be compounded²². Regarding the higher population threshold range derived for the F2–F5

Table 2 | Posterior parameter mean and s.d. values of the highest-performing model that considers the variables CAPE, HLCY and VWSH to predict F2–F5 tornadoes.

F2F5 Model	$\alpha 0$		$\alpha 1$ —CAPE		$\alpha 2$ —HLCY		$\alpha 3$ —VWSH		β		σ	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
January	−9.84	1.38	0.17	0.22	0.19	0.39	4.59	1.44	5,980	2,882	2.92	0.38
February	−9.85	1.17	0.24	0.12	0.49	0.29	4.39	1.03	163.0	128.8	2.04	0.32
March	−6.12	0.62	−0.13	0.16	0.62	0.21	1.26	0.77	4.04	1.15	2.41	0.17
April	−4.60	0.44	−0.02	0.17	0.78	0.16	0.06	0.62	4.57	1.00	2.12	0.15
May	−3.41	0.28	0.19	0.17	0.49	0.11	−0.40	0.32	4.65	0.61	1.93	0.11
June	−4.07	0.33	0.56	0.20	0.51	0.13	0.28	0.46	3.49	0.54	2.32	0.14
July	−4.43	0.33	0.52	0.23	0.24	0.15	1.04	0.31	2.72	0.50	2.21	0.22
August	−4.96	0.45	0.36	0.23	0.48	0.19	0.61	0.32	3.53	0.72	2.77	0.29
September	−5.61	0.50	0.11	0.23	0.43	0.24	−0.25	0.60	5.12	1.44	2.97	0.27
October	−6.11	0.62	0.04	0.18	0.25	0.22	0.69	0.67	4.30	1.32	2.93	0.28
November	−5.75	0.87	0.00	0.16	0.44	0.29	−0.45	1.13	11.07	4.72	2.69	0.21
December	−9.30	1.66	−0.16	0.19	0.56	0.42	2.29	1.65	79.87	72.50	2.91	0.33

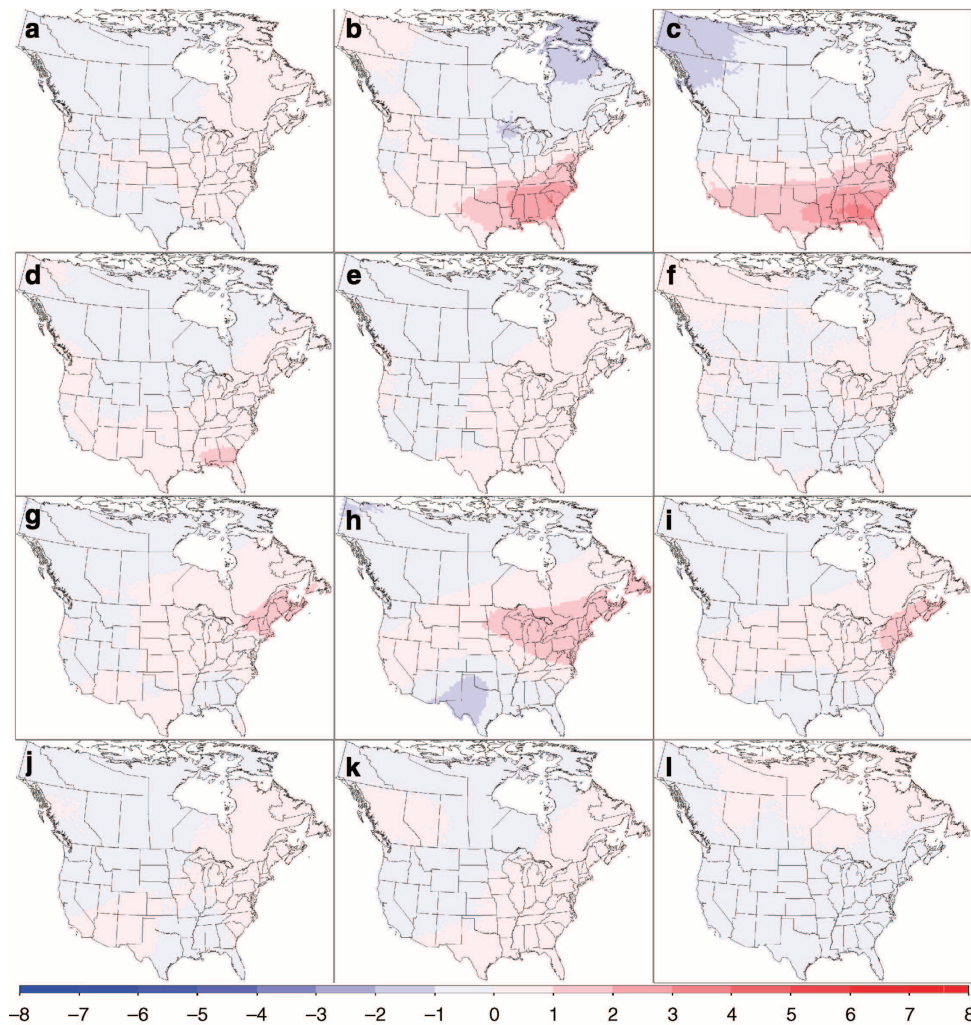


Figure 1 | Differences in the conditional autoregressive term between the CAPE-HLCY F0-F5 and CAPE-HLCY-VWSH F0-F5 monthly models. (a) December; (b) January; (c) February; (d) March; (e) April; (f) May; (g) June; (h) July; (i) August; (j) September; (k) October; (l) November.

tornadoes, we believe that it likely reflects the fact that significant tornadoes tend to be reported in more populated areas comparing with F0–F5 events. The latter assertion reflects the ‘damage-related’ nature of the tornado intensity estimates in the F-scale rating. Tornadoes are more likely to have a high F-scale rating in populated areas, where there is more infrastructural capital in their path²³.

Assessment of tornado occurrence in North America. The spatial distribution of the 30-year observations, T_{obs} , along with the corresponding predictions, T_{latent} , for F0–F5 (F2–F5) tornadoes are shown in Fig. 2a–d and Fig. 2e–h (Fig. 3a–d and Fig. 3e–h), respectively. Additional information related to the results of the CAPE-HLCY F0-F5 and F2-F5 models is provided in Supplementary Figs 5 and 6. The spatial patterns for T_{obs} and T_{latent} are qualitative similar, indicative of the satisfactory model performance. In quantitative terms though, there are areas where the predicted tornado occurrences were considerably larger than the observed ones, primarily in the central to northern Plains and the Canadian prairies. The discrepancies were more pronounced during the spring and summer months (Table 3 and Fig. 2e–h and Fig. 3e–h; see also Supplementary Figs 8 and 9 for differences related to the CAPE-HLCY-VWSH model). Notably, the predominance of T_{latent} , relative to T_{obs} , is more evident in Canada than in the United States, suggesting a potentially significant

sampling bias of tornado observations in the Canadian Prairies² (Table 3).

In springtime, the global maximum for F0–F5 and F2–F5 tornado occurrences is located in the southern to central Plains, and extends eastward from Illinois to western Tennessee/Kentucky. F0–F5 T_{latent} predictions extend to southern Ontario and portions of Prairies in Canada but less so for F2–F5 tornadoes. Tornado occurrences in the summer shift northward to the central Plains and to the Prairies, and extends eastward to the Midwest, southern Ontario and the northeastern seaboard states. Note that there are peaks in eastern Colorado and Florida peninsula for F0–F5 tornadoes in the summer that are absent for F2–F5 tornadoes. These are predominately non-supercell tornadoes that are mostly of the lower F-scales¹⁸. It has been hypothesized from the tornado data that a sizable fraction of tornadoes across Canada are of the non-supercell type²⁴. In the fall, areas of high tornadic activity shift south of the Prairies to the United States, east of the Rocky Mountains, characterized by a clear local minimum in the Appalachian Mountains. In the winter, the F0–F5 tornado maximum activity is located in the Gulf coast states, northward to Arkansas/Kentucky and in northern and central California coast.

The random effects terms in the seasonal models demonstrate strong spatial covariance with the T_{latent} posterior estimates in that the positive ϕ values typically coincide with higher T_{latent}

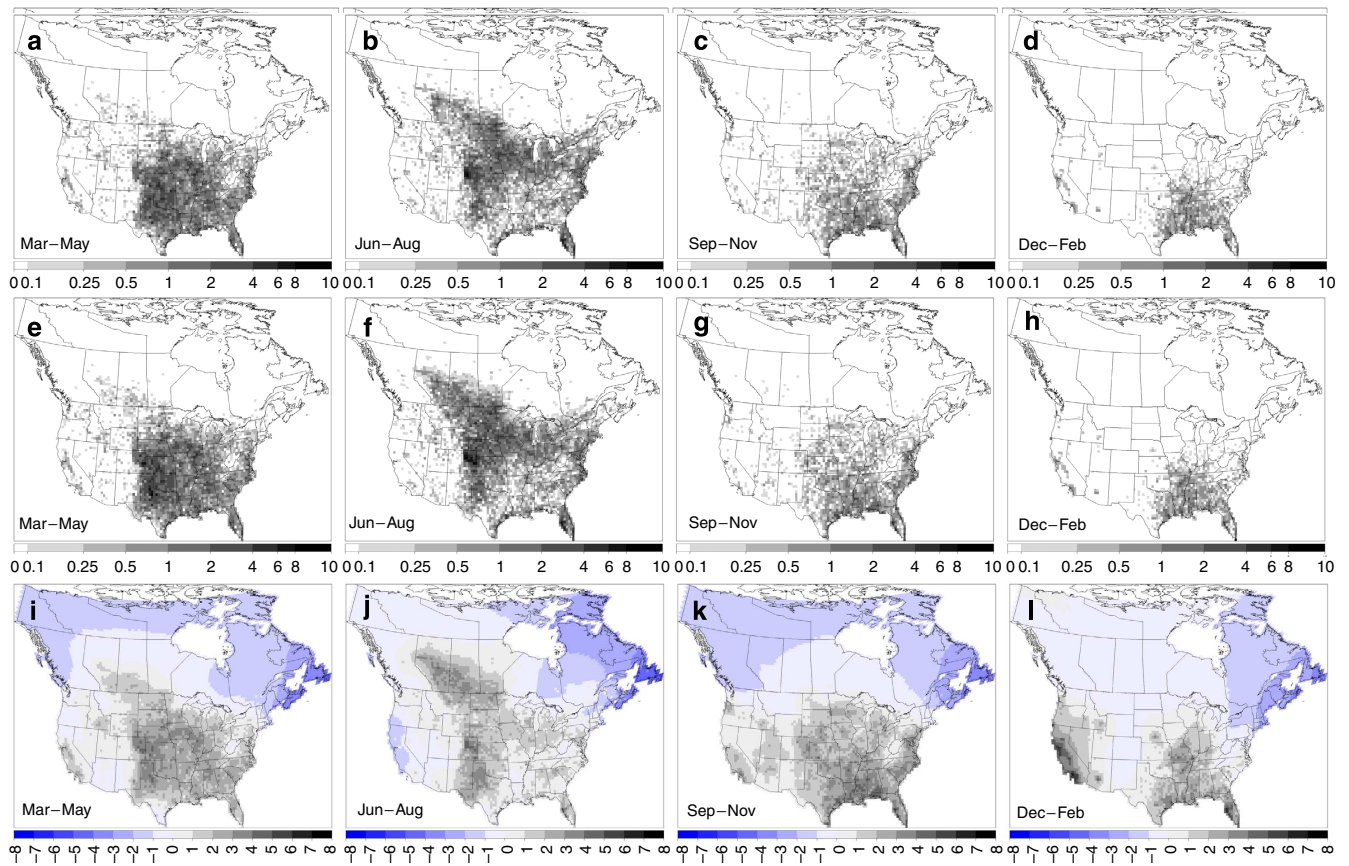


Figure 2 | Results of the CAPE-HLCY-VWSH F0-F5 model. The observed tornado counts per 100 km² · yr⁻¹, *T*_{obs}: (a) Spring; (b) Summer; (c) Autumn; (d) Winter; the predictions of the total number of tornado occurrence per 100 km² · yr⁻¹, *T*_{latent}: (e) Spring; (f) Summer; (g) Autumn; (h) Winter; values for the conditional autoregressive term, φ : (i) Spring; (j) Summer; (k) Autumn; (l) Winter.

and vice versa (Fig. 2i–l and Fig. 3i–l). In other words, φ tends to compensate for the residual variability unaccounted for by the predictors included in the current model, although there were deviations from that pattern. For example, in the eastern United States and southern Ontario, where the tornado activity peaks up in the summer, φ 's are considerably lower than the Central United States (Colorado maximum) and Prairies. In the winter F0–F5 models, φ 's in the southwest United States have distinctly higher values than the southeastern region, despite the lower number of tornadoes in the west coast. While the likelihood of addressing the structural inadequacies of our model with the inclusion of additional predictors cannot be ruled out, we believe that an equally effective strategy may be to relax the assumption of common parameter values over the entire geographic domain. For modelling purposes, the delineation of regions that demonstrate greater homogeneity in their tornado activity and the subsequent configuration of a hierarchical structure, under which site-specific parameters will be used to describe the causal relationship between tornadic activity and atmospheric/climatological variables, may increase our capacity to accommodate the observed variability in space and time².

The expected tornado frequency, λ_i , for the months when the highest F0–F5 and F2–F5 tornado occurrence is experienced in the United States (May) and Canada (July) are presented in Fig. 4 and Supplementary Fig. 7 for the CAPE-HLCY model. A wide area extending from Central United States to Illinois and onto western Tennessee in May is evident with high likelihood of F0–F5 tornadoes, resembling a clockwise rotated 'C' shape. For F2–F5 tornadoes, high likelihood rates are predicted in an extensive area around Oklahoma and Kansas. Relatively high F2–F5 tornado

frequency rates are also predicted for Kentucky and (to a lesser extent) the Southern Great Lakes region. This pattern of tornadic activity shifts northward and extends east–west in July. Note that tornadoes predicted over the Great Lakes (that is, over water) are based on mechanisms related to tornado environments over land. They are not based on actual waterspout observations, but were included to give an indication of tornado occurrences and transition across the United States–Canada.

Discussion

North America experiences high tornado activity in the southern Great Plains and more broadly in the central United States during the springtime. According to our modelling analysis, CAPE, HLCY and VWSH (or SHEAR depending on the model examined) exert significant control and can offer a fairly accurate depiction of the spatial patterns. Tornado activity in the summer shifts north and Canada experiences its highest tornado occurrences in July. CAPE and VWSH are both important predictors in the summer. This finding reflects the transport of moisture from the Gulf of Mexico, occurrence of the elevated mixed layer²⁵, and the poleward shift of jet stream, which in turn provide sufficient CAPE, large-scale upward motion and vertical wind shear to enhance majority of tornado activity. Non-supercell tornadoes are also common in the summer in Florida, eastern Colorado, northern Plains¹⁸ and may represent a substantial fraction of the total number of occurrences in Canada²⁶. Reflective of the latter assertion is the importance of CAPE (and weak HLCY and SHEAR signatures) in these regions, as updrafts (strong vertical shear) collocated along pre-existing vertical vortices enhance (prevent) non-supercell tornadoes²⁷.

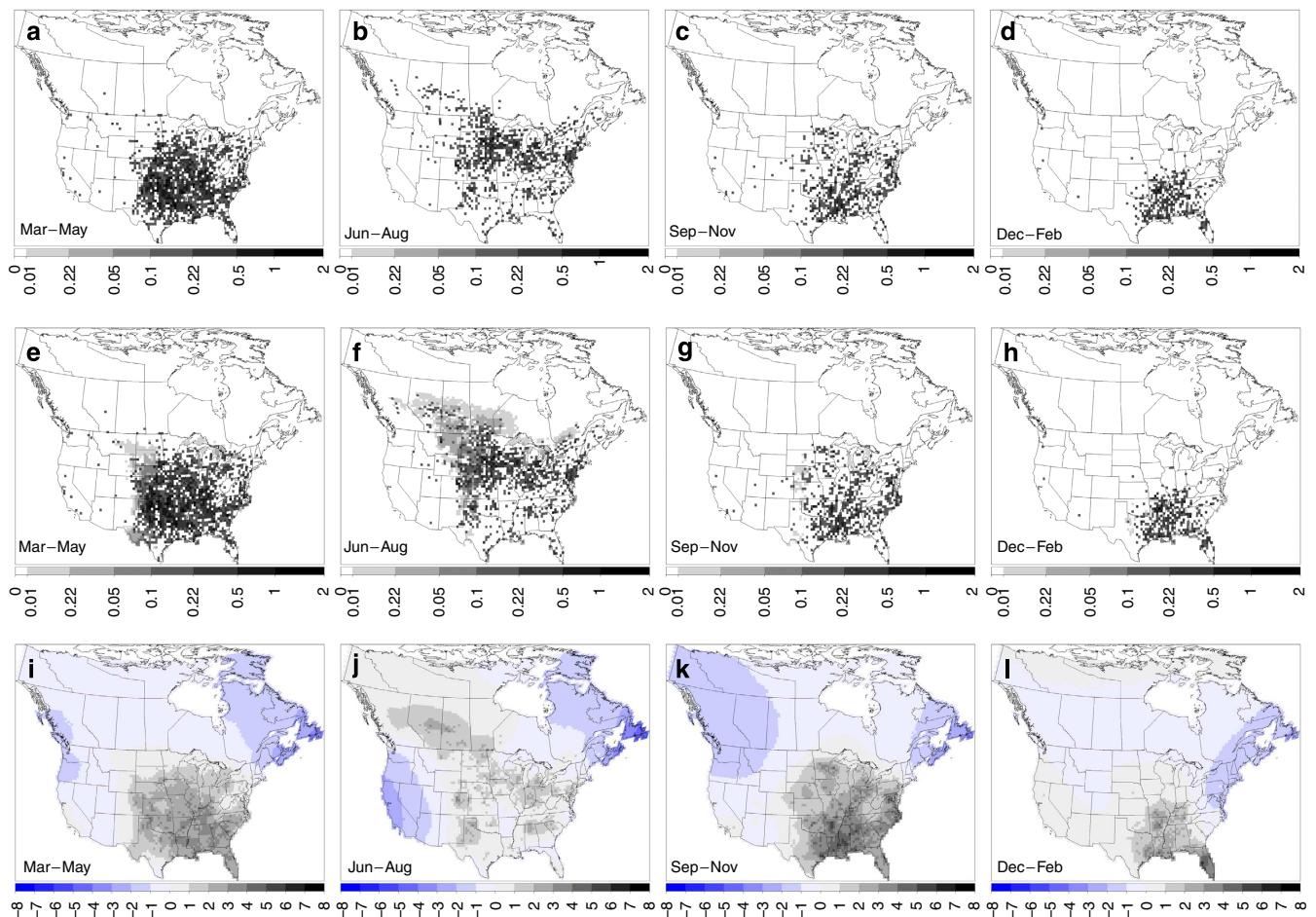


Figure 3 | Results of the CAPE-HLCY-VWSH F2-F5 model. The observed tornado counts per $100 \text{ km}^2 \cdot \text{yr}^{-1}$, *Tobs*: (a) Spring; (b) Summer; (c) Autumn; (d) Winter; the predictions of the total number of tornado occurrence per $100 \text{ km}^2 \cdot \text{yr}^{-1}$, *Tlatent*: (e) Spring; (f) Summer; (g) Autumn; (h) Winter; values for the conditional autoregressive term, ϕ : (i) Spring; (j) Summer; (k) Autumn; (l) Winter.

During the fall season, the role of CAPE is minimized, while the influence of VWSH (and SHEAR) increases and the HLCY signature remains relatively constant. The mechanisms that generate tornadoes in the fall are more complicated and can be linked with different weather systems. In particular, tropical cyclones (TC) account for over 20% of all tornadoes in September (This number is based on the frequency of Gulf-landfalling TC-tornadoes from 1980 to 2008 (ref. 28) and the total number of tornado observations in this study). There are significant differences in the relative magnitudes of moisture, instability, shear and lift in conditions preceding TC tornadoes²⁹. The smaller-scale boundaries associated with TC tornadoes are distinctly different from those of large-scale extratropical cyclones in spring and late autumn³⁰. TC tornadoes are also associated with stronger lower tropospheric vertical shear compared with those on the Great Plains²⁹, which explains the stronger (weaker) and more (less) identifiable influence of SHEAR (VWSH), particularly for F2-F5 tornadoes. By contrast, almost all tornadoes in November stem from strong cold frontal systems, resembling to those in the spring season. Overall, this bimodal pattern may explain the change in the magnitude and/or sign of the standardized regression coefficients between September/October and November, reflecting a shift in the relative role of the different atmospheric variables. In winter, VWSH (and to a lesser extent SHEAR) is characterized by the highest-standardized slopes. This result is on parity with existing evidence from the literature that the influence of the subtropical

jet to tornado activity is strongest in the winter, and can be a major factor of tornado formation and tornado outbreaks under extratropical storms^{12,31}. Moreover, strong vertical wind shear (i) can play a role in organizing quasi-linear convective systems that can include individual or multiple supercells³² and (ii) can produce a large vertical pressure gradient when horizontal vorticity is tilted and thus facilitates nocturnal convection to overcome large negative buoyancy, thereby producing tornadoes³³. Thus, our finding that significant tornado occurrences in winter are closely associated with VWSH and SHEAR is a plausible result.

The observations bias stemming from the differences in the population density was particularly evident in the Central/Northern Plains and the Prairies. Population thresholds for minimizing the likelihood of bias and observing all tornadoes ranged from 6 to 7.25 persons km^{-2} and 6 to 14 persons km^{-2} for F0-F5 and F2-F5 tornadoes, respectively. The higher values for F2-F5 may simply reflect the damage-based F-scale rating system, which could be biased towards populated area, where service- or infrastructure-related impacts can frequently be experienced²³. The observation bias related to population density also appears to follow a seasonal pattern. Our modelling study suggests that highest population thresholds are needed to observe tornadoes during the winter months rather than the summertime. The different proportions of nocturnal tornadoes among the various periods of the year as well as the daylight variability may shape this pattern and determine the

Table 3 | Comparison between observed monthly numbers of tornadoes in Canada and the United States and those predicted by the CAPE-HLCY-VWSH model.

Month	Σ Tobs			Σ Tlatent			Difference		
	Canada	United States	Total	Canada	United States	Total	Canada	United States	Total
F0-F5									
January	0.0	25.3	25.3	0.1	26.9	27.1	0.1	1.6	1.8
February	0.0	27.5	27.5	0.1	32.2	32.4	0.1	4.7	4.8
March	0.2	71.4	71.6	0.7	77.2	77.9	0.5	5.7	6.2
April	1.0	134.7	135.7	3.5	148.1	151.6	2.5	13.4	15.9
May	7.4	246.4	253.8	17.7	288.3	306.0	10.3	41.9	52.2
June	19.0	217.6	236.6	35.9	266.6	302.5	16.9	49.0	65.9
July	23.8	116.1	139.8	44.0	140.9	184.8	20.2	24.8	45.0
August	13.2	75.4	88.6	25.0	89.7	114.7	11.8	14.2	26.0
September	2.4	60.4	62.9	5.7	67.2	72.9	3.3	6.7	10.0
October	0.3	46.2	46.4	2.2	52.9	55.1	1.9	6.8	8.7
November	0.1	49.6	49.7	2.2	54.6	56.8	2.1	5.0	7.1
December	0.0	23.0	23.0	0.2	24.6	24.8	0.2	1.6	1.8
F2-F5									
January	0.0	4.6	4.6	0.3	5.8	6.0	0.3	1.2	1.4
February	0.0	5.3	5.3	0.1	5.9	6.0	0.1	0.6	0.7
March	0.0	14.8	14.8	0.3	16.0	16.3	0.3	1.1	1.4
April	0.2	22.1	22.2	1.9	23.6	25.5	1.7	1.6	3.3
May	0.9	33.3	34.2	5.7	38.9	44.6	4.8	5.6	10.4
June	1.1	19.7	20.9	4.4	23.8	28.1	3.2	4.0	7.3
July	1.9	9.4	11.3	4.6	10.9	15.5	2.7	1.6	4.2
August	1.3	5.6	7.0	3.9	6.7	10.6	2.5	1.1	3.6
September	0.2	6.0	6.2	2.3	6.8	9.1	2.1	0.7	2.8
October	0.0	6.5	6.6	0.8	7.2	8.0	0.8	0.7	1.4
November	0.0	11.2	11.2	7.3	11.8	19.1	7.3	0.6	7.9
December	0.0	4.3	4.3	0.2	4.6	4.8	0.2	0.3	0.5

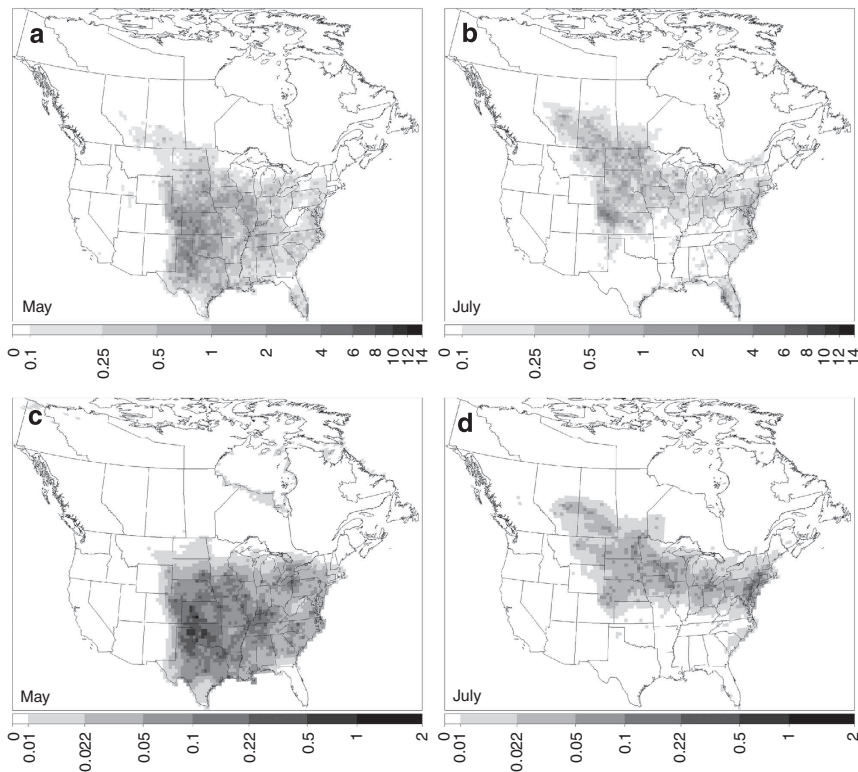


Figure 4 | Predictions of the CAPE-HLCY-VWSH model—Peak months of tornado activity for the United States and Canada. (a) May (F0-F5); (b) July (F0-F5); (c) May (F2-F5); (d) July (F2-F5) tornadoes.

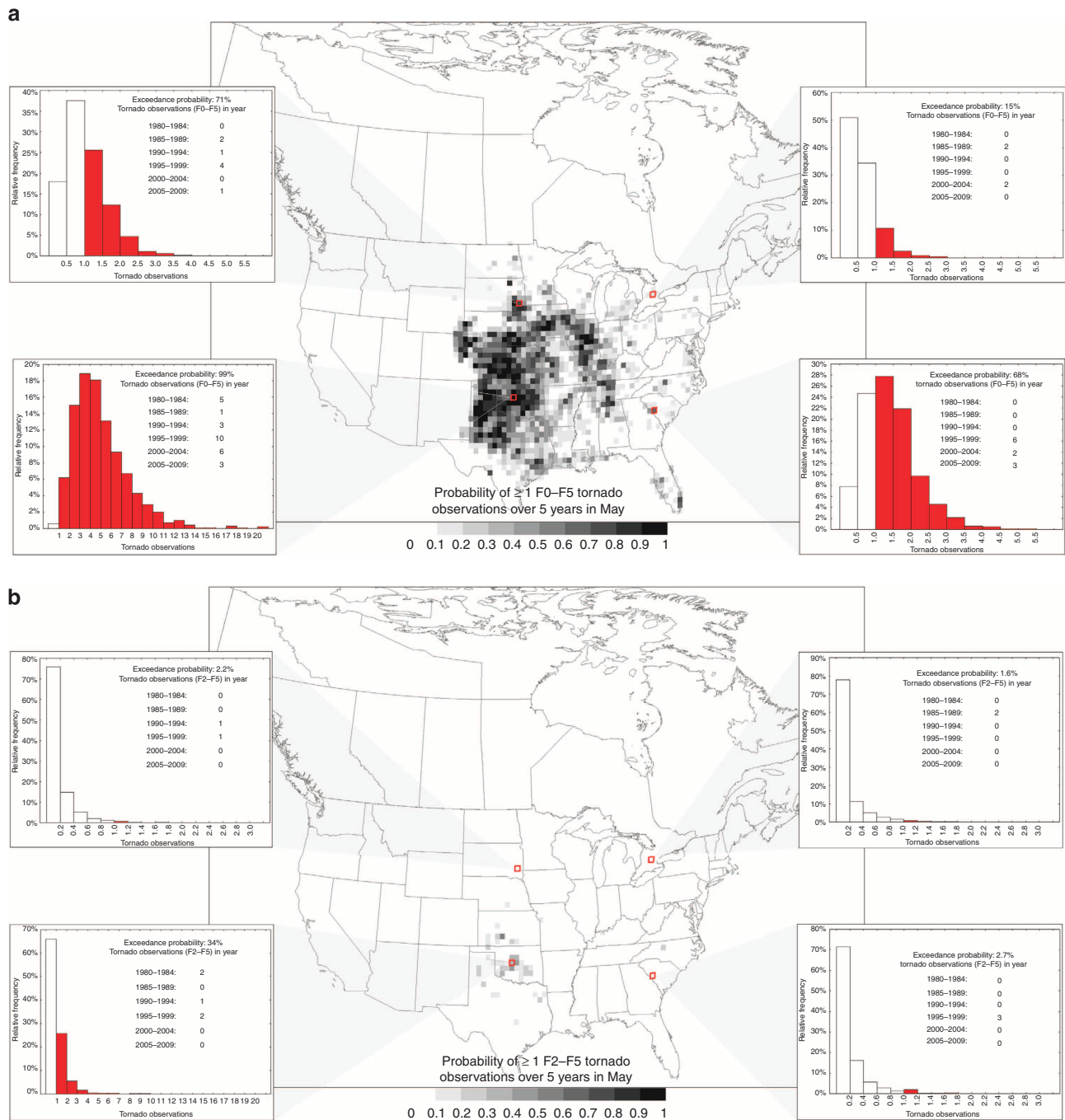


Figure 5 | Exceedance probability of one tornado event in May over the course of 5 years as predicted by the CAPE-HLCY-VWSH model.

The four selected grid cells (indicated by the red boundaries) were Sioux Fall, South Dakota (Top left), Guelph-Waterloo, Ontario (Top right), El Reno, Oklahoma (Bottom left) and Edgefield county, South Carolina (Bottom right). The corresponding tornado events are also presented in 5-year time intervals during the study period 1980–2009. **(a)** F0–F5 tornadoes; **(b)** F2–F5 tornadoes.

efficacy of the current monitoring systems. Overall, we showed that some of the spatial variability found in the actual tornado observations can be accounted for by the population density, and that our Bayesian modelling framework can effectively accommodate this observation bias.

Model performance is better during the spring (winter) when HLCY (VWSH) is included, reflecting the seasonal variability of their causal association with tornado occurrence. Overall, the standardized regression coefficients had higher values with the F0–F5 than F2–F5 models, although substantial regional variability exists under the current model structure with

seasonal or even monthly parameter specifications. The posterior estimates of the conditional autoregressive term delineated regional features that cannot be captured due to the structural inadequacy of a globally common parameterization and/or the absence of regionally relevant predictors/processes during different months of the year. Characteristic examples are the mid-level-specific humidity gradient, representing a measure of dry air intrusions which increases the propensity of tornadoes in TCs, may be important for tornado activity in August and September since a substantial number of tornadoes are related to landfalling TC from the Gulf of Mexico²⁸ or the strength of

the low-level jet in the springtime, influencing the tornado activity in the South to Central Plains³⁴. As recently shown³⁵, there is a clear spatial distribution of the tornado environment that the current globally common parameter specification cannot address. This facet of our modelling framework will be investigated with a hierarchical model configuration in a subsequent study.

In the context of tornado modelling, the adoption of a coarser temporal resolution (seasonal/monthly averages) is a significant distinction from the typically used predictor variables on shorter (subdaily) timescales⁵. This approach effectively postulates that atmospheric quantities varying on climate timescales can be used to draw inference about tornadic events with lifetime of no more than a few hours (and often only a few minutes)^{5,6}. With this strategy, our intent is not to develop an early warning system, but rather to offer a modelling tool that can be used to characterize the frequency of tornado occurrence (or exceedance of a threshold number of tornadoes) in a particular location in North America within a selected period of the year, for example, probability of occurrence of more than one tornado in the selected four locations in May (Fig. 5a,b). In this regard, the present analysis resembles to modelling practices adopted in other disciplines (for example, aquatic ecology), where empirical models are in place to successfully reproduce the average prevailing conditions in a certain period of the year (summer growing season) and subsequently to predict the likelihood of occurrence of episodic events (end-of-summer hypoxia or cyanobacteria outbreaks) in a natural system. Thus, our modelling work is founded on the assumption that the association between average environmental conditions and ecological processes leading to short-term ecosystem shifts bears similarity to the connection between atmospheric quantities varying on climate timescales and tornado activity. The former linkage may be more evident as episodic ecosystem events are typically the escalation of a complex interplay among physical, chemical and biological processes that occur over a longer time period. In the context of tornadogenesis though, the time duration of the latter pattern with the atmospheric processes involved is distinctly shorter, and therefore the credibility of the present probabilistic mapping depends on the capacity of the changes in the moments (central tendency, spread) of the distribution of the environments occurring in the course of a month to faithfully capture the changes in the frequency of extreme subdaily environments (tails of the same distribution) associated with tornado occurrence.

Methods

Tornado counts and atmospheric data. Canada tornado data covering the period 1980–2009 were collected from the updated national tornado database^{2,24}, whereas the United States tornado data for the same period were collected from the Storm Prediction Center³⁶. It is worth noting that the tornado data set used does not include tornadoes over water (that is, waterspouts) unless they made landfall; waterspout is a common occurrence over the Great Lakes in autumn. The annual number of tornadoes in Canada does not demonstrate any systematic long-term patterns, whereas the United States data suggest a slight increase in tornadic activity, stemming from an increase of F0 tornadoes (Supplementary Fig. 1). There are no obvious trends for more intense tornadoes (F2–F5) in both countries. Although the temporal patterns of the total number of tornadoes over the 30-year study period are not substantially affected by non-meteorological factors, tornado reports are still hampered by sampling biases, mainly associated with population density, leading to underestimations in sparsely populated areas^{23,37,38}. We computed a gridded climatology (30-year total) of monthly and seasonal tornado observations by counting the number of reported tornadoes for each calendar month and season via their tornado touchdown points in a geodesic 50 × 50 km grid (Supplementary Data 1). This process was done separately for all tornadoes (F0–F5s) and for intense tornadoes (F2–F5s). The predicted spatial patterns of tornado occurrence, the population effects along with the corresponding population density thresholds are generally robust to the cell size of the grid configuration².

A range of atmospheric variables from 1980 to 2009 were obtained from the North American Regional Reanalysis³⁹ and their 30-year monthly and seasonally

average values were calculated for consideration as potential predictors for our modelling framework (Supplementary Table 22). North American Regional Reanalysis data were provided on a 32-km Lambert conformal grid, which were subsequently interpolated to match our grid. We selected the convective available potential energy (m² s⁻²) or CAPE as a measure of instability through the depth of troposphere that is related to updraft strength in thunderstorms. CAPE is widely used to quantify the subdaily atmospheric instability leading to tornado events⁶. Vertical wind shear is known to be very critical to tornadogenesis and different wind shear variables may be important in different types of tornado development. Thus, we explore the relative importance of three wind shear-related variables: (i) storm relative environmental helicity (m² s⁻²) or HLCY; a measure of the potential for cyclonic updraft rotation in right-moving supercells, it is calculated for the lowest 3 km layers above ground level⁴⁰; (ii) 0 to 6 km wind shear magnitude (m s⁻²) or SHEAR: a measure of wind shear from the surface through the mid-troposphere⁴¹. Strong vertical shear remove precipitation from updrafts, and induce vertical perturbation pressure gradients, so thunderstorms tend to become more organized and persistent; (iii) vertical wind shear from surface to tropopause (s⁻¹) or VWSH: a measure of the upper-level wind speed normalized by the surface wind speed⁴². It is a measure of the upper-level jet strength inducing shear and vertical motion, and is often associated with severe weather. Strong upper-level jet streak tilts the storm as it rises vertically and has similar effects as SHEAR. Tornadoes can occur in a wide range of CAPE and wind shear levels that will be accommodated through systematic stratification by month/season and subsequent examination of the residual geographic variability. The considerable seasonality characterizing the actual values as well as the covariance patterns among CAPE and the three wind shear-related variables renders support to the latter strategy (Supplementary Figs 10–12). Finally, since population-sampling bias creeps into tornado observations and may cause spurious geographical variability^{23,37,38}, we adopted the Cheng *et al.*'s correction method² that uses population density data to quantify the observation error.

Statistical framework. For each month/season, we used a Bayesian modelling approach to decompose the problem of tornado occurrence into a series of conditional models coherently linked together via Bayes' rule^{2,43}. We specified a binomial model in which the observed (30-year) tornado counts in a particular month/season *t* for each grid cell *i*, *Tobs_{ti}*, are conditioned on the actual (but unobserved) tornado occurrences in the same month, *Tlatent_{ti}*, and the probability of detection, *p_{ti}*:

$$Tobs_{ti} | Tlatent_{ti}, \lambda_{ti}, p_{ti} \sim \text{Binomial}[Tlatent_{ti}, p_{ti}]. \tag{1}$$

The probability of detection *p_{ti}* represents the likelihood to observe a tornado and is associated with the population density *y_i* by the following exponential expression:

$$p_{ti} = \exp[-\beta_t / \exp(y_i)] \tag{2}$$

where *β_t* is the population effect parameter for month/season *t* and $\exp(y_i)$ is the exponential transformation of the original population density data in grid cell *i*. The actual occurrence of tornadoes, *Tlatent_{ti}*, in the model domain is specified as a Poisson process, conditional on the average or expected tornado occurrence rate per grid cell *λ_{ti}*, provided by the predictive model

$$Tlatent_{ti} | \lambda_{ti} \sim \text{Poisson}(\lambda_{ti}). \tag{3}$$

We use a log-linear model for *λ_{ti}* given by the following expression: $\log(\lambda_{ti}) = \alpha_{t0} + \alpha_{t1}x_{1,ti} + \alpha_{t2}x_{2,ti} + \dots + \alpha_{tk}x_{k,ti} + \varphi_{ti}$ where *x_{1,ti}*, . . . , *x_{k,ti}* is a vector of the corresponding standardized monthly/seasonal atmospheric variables, *α_{t0}* is the model intercept, *α_{t1}*, . . . , *α_{tk}* are the regression coefficients and *φ_{ti}* is a cell-specific random effect, capturing the residual variability of the tornado frequency in a particular month/season *t* and grid cell *i*, stemming from other explanatory factors/processes unaccounted for by the model. The inclusion of *φ_{ti}* also aims to address the possibility that the signature of the regression coefficients may not be consistently strong in the tornado frequency records throughout the model domain². It is also reasonable to assume that the random effects of the unaccounted factors have a regionalized/localized character and thus are spatially correlated. The characterization of the spatially correlated random terms *φ_{ti}* was based on the Bayesian conditional autoregressive model⁴⁴. The random error terms are jointly distributed as a multivariate normal distribution with mean 0 and an unknown covariance matrix. In particular, the model postulates that the spatial random effect in cell *i* depends on the neighbouring cells of *i* (*N_i*) and that all of the neighbours have equal influence (weight of 1) on *i*. The term *φ_{ti}* is defined by the conditional normal distribution $\varphi_{ti} \sim N(\mu_{ti}, \sigma^2 n_i)$, where

$$\mu_{ti} = \frac{1}{n_i} \sum_{i \in N_i} \varphi_{ti} \tag{4}$$

and *n_i* is the number of adjacent grid cells. Because we used a first-order neighbourhood approach and squared cells, *N_i* represents the eight immediately adjacent cells. In this study, we opted for non-informative (or 'flat') prior distributions, reflecting no prior knowledge of the model parameters (see model codes in Supplementary Notes 1 and 2). A sequence of realizations from the model posterior were obtained using Markov chain Monte Carlo simulations. Convergence was assessed qualitatively by visually inspecting plots of the posterior Markov chains for mixing and stationarity and by inspecting density plots of the

pooled posterior Markov chains for unimodality. We also assessed convergence quantitatively using the modified Gelman–Rubin convergence statistic⁴⁵. The accuracy of the posterior parameter values was inspected by assuring that the Monte Carlo error for all parameters was less than 5% of the sample s.d. This process is undertaken independently for each month/season to assess the intra-annual variability of the models. We also conducted two additional exercises related to the predictive and structural confirmation of the present modelling framework (Supplementary Figures 13–22). The first skill assessment test was based on splitting the 30-year data set into two subsets; namely, the calibration (1980–1994) and predictive validation (1995–2009) data sets. The former one was used to obtain parameter estimates through Bayesian updating and the derived model predictive posteriors were then tested independently against the latter (Supplementary Table 23). The second skill assessment test aimed to examine the robustness of the inference drawn by the Binomial–Poisson model, given that the analysed tornado data have many zeros. The alternative statistical formulation was the zero-inflated Poisson model, based on a Zero-Inflated probability distribution that allows for frequent zero-valued observations.

Structural confirmation of Bayesian modelling framework. The second skill assessment test aimed to examine the robustness of the inference drawn by the binomial–Poisson model, given that the analysed tornado data have many zeros. The alternative statistical formulation was the zero-inflated Poisson (ZIP) model, based on a Zero-Inflated probability distribution that allows for frequent zero-valued observations⁴⁶. This model is a statistical description of a random event, containing excess zero-count data per unit of time/space or within a fixed interval of a relevant covariate. The model dissects the studied event (tornado occurrence) into two components that correspond to two zero-generating processes. The first process reflects the sampling/observation error and is governed by a binary distribution that generates structural zeros, while the second mechanism represents the tornado occurrence rate and is governed by a Poisson distribution that generates counts, some of which may be zero. In the present model, the probability p of the former process is associated with the population density γ and the mean λ of the latter process depends on the values of the causal factors (atmospheric predictors). The two model components can be described as follows:

$$T_{\text{obs}_{it}} \mid \lambda_{it}(\alpha_{t0}, \alpha_{tj}, x_{tj}, \varphi_{it}); p_{it}(\beta_t, \gamma_i) \sim \begin{cases} \text{Poisson}(\lambda_{it}) & \text{with probability } p_{it} \\ 0 & \text{with probability } 1-p_{it} \end{cases} \quad (5)$$

$$\log(\lambda_{it}) = a_{t0} + \sum_{j=1}^k \alpha_{tj} x_{tj} + \varphi_{it} \quad (6)$$

$$p_{it} = \exp[-\beta_t / \exp(\gamma_i)] \quad (7)$$

where $T_{\text{obs}_{it}}$ denotes the observed tornado counts in grid cell i and month/season t ; p_{it} represents the likelihood to observe a tornado in grid cell i and month/season t ; β_t is the population effect parameter for month/season t ; $\exp(\gamma_i)$ is the exponential transformation of the population density data in grid cell i ; λ_{it} is average or expected tornado occurrence rate per grid cell; x_{tj} represents the standardized value of the j atmospheric variable in grid cell i and month/season t ; α_{t0} and α_{tj} is the intercept and the j regression coefficients of the log-linear model used for λ_{it} , and φ_{it} is a cell-specific random effect, capturing the residual variability of the tornado frequency in a particular month/season t and grid cell i , stemming from other explanatory factors/processes unaccounted for by the model. Similar to the Binomial–Poisson model, we opted for non-informative (or ‘flat’) prior distributions, reflecting no prior knowledge of the model parameters (see Supplementary Notes 1 and 2).

References

- Doswell, III C. A., Carbin, G. W. & Brooks, H. E. The tornadoes of spring 2011 in the USA: an historical perspective. *Weather* **67**, 88–94 (2012).
- Cheng, V. Y. S. *et al.* Probability of tornado occurrence across Canada. *J. Clim.* **26**, 9415–9428 (2013).
- Brooks, H. E., Doswell, C. A. & Kay, M. P. Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting* **18**, 626–640 (2003).
- Dixon, P. G., Mercer, A. E., Choi, J. & Allen, J. S. Tornado risk analysis: is dixie alley an extension of tornado alley? *Bull. Am. Meteorol. Soc.* **92**, 433–441 (2011).
- Tippett, M. K., Sobel, A. H. & Camargo, S. J. Association of U.S. tornado occurrence with monthly environmental parameters. *Geophys. Res. Lett.* **39**, L02801 (2012).
- Tippett, M. K., Sobel, A. H., Camargo, S. J. & Allen, J. T. An empirical relation between U.S. tornado activity and monthly environmental parameters. *J. Clim.* **27**, 2983–2999 (2014).
- Brooks, H. E., Lee, J. W. & Craven, J. P. The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.* **67–68**, 73–94 (2003).
- Brooks, H. E., Doswell, C. A. & Cooper, J. On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting* **9**, 606–618 (1994).
- Thompson, R. L., Edwards, R., Hart, J. A., Elmore, K. L. & Markowski, P. Close proximity soundings within supercell environments obtained from the rapid update cycle. *Wea. Forecasting* **18**, 1243–1261 (2003).
- Dupilka, M. L. & Reuter, G. W. Composite soundings associated with severe and tornadic thunderstorms in central Alberta. *Atmos.-Ocean* **49**, 269–278 (2011).
- Grams, J. S. *et al.* A climatology and comparison of parameters for significant tornado events in the United States. *Wea. Forecasting* **27**, 106–123 (2012).
- Hagemeyer, B. C. & Schmocker, G. K. Characteristics of East-Central Florida tornado environments. *Wea. Forecasting* **6**, 499–514 (1991).
- Hanstrum, B. N., Mills, G. A., Watson, A., Monteverdi, J. P. & Doswell, III C. A. The cool-season tornadoes of California and southern Australia. *Wea. Forecasting* **17** (2002).
- Monteverdi, J. P., Doswell, C. A. & Lipari, G. S. Shear parameter thresholds for forecasting tornadic thunderstorms in northern and central California. *Wea. Forecasting* **18** (2003).
- Wasula, A. C., Bosart, L. F., Schneider, R. S., Weiss, S. J. & Johns, R. H. in *20th Conference on Weather Analysis and Forecasting*. Amer. Meteor. Soc. Seattle, WA, USA, Available at <https://ams.confex.com/ams/pdfpapers/71974.pdf> (2004).
- Guyer, J. L., Imy, D. A. & Kis, A. in *23rd Conference on Severe Local Storms*. Amer. Meteor. Soc. St. Louis, MO, USA, Available at <http://www.spc.ncep.noaa.gov/publications/guyer/gulf2f5.pdf> (2006).
- Guyer, J. L. & Dean, A. R. in *25th Conference on Severe Local Storms*. Amer. Meteor. Soc. Denver, CO, USA, Available at <http://www.spc.ncep.noaa.gov/publications/guyer/weakcape.pdf> (2010).
- Smith, B. T., Thompson, R. L., Grams, J. S., Broyles, C. & Brooks, H. E. Convective modes for significant severe thunderstorms in the contiguous United States. Part I: storm classification and climatology. *Wea. Forecasting* **27**, 1114–1135 (2012).
- Thompson, R. L., Smith, B. T., Grams, J. S., Dean, A. R. & Broyles, C. Convective Modes for significant severe thunderstorms in the contiguous United States. Part II: supercell and QLCS tornado environments. *Wea. Forecasting* **27**, 1136–1154 (2012).
- Trapp, R. J., Tessendorf, S. A., Godfrey, E. S. & Brooks, H. E. Tornadoes from squall lines and bow echoes. Part I: Climatological distribution. *Wea. Forecasting* **20**, 23–34 (2005).
- King, P. On the absence of population bias in the tornado climatology of southwestern Ontario. *Wea. Forecasting* **12**, 939–946 (1997).
- Ashley, W. S., Krmenc, A. J. & Schwantes, R. Vulnerability due to nocturnal tornadoes. *Wea. Forecasting* **23**, 795–807 (2008).
- Schaefer, J. T. & Galway, J. G. in *12th Conference on Severe Local Storms*. Amer. Meteor. Soc. (San Antonio, TX, USA, 1982).
- Sills, D. *et al.* in *26th Conference on Severe Local Storms*. Amer. Meteorol. Soc. (Nashville, TN, USA, 2012).
- Banacos, P. C. & Ekster, M. L. The association of the elevated mixed layer with significant severe weather events in the northeastern United States. *Wea. Forecasting* **25**, 1082–1102 (2010).
- King, P. W. *et al.* Lake breezes in southern Ontario and their relation to tornado climatology. *Wea. Forecasting* **18**, 795–807 (2003).
- Wakimoto, R. M. & Wilson, J. W. Non-supercell tornadoes. *Mon. Weather Rev.* **117**, 1113–1140 (1989).
- Belanger, J. I., Curry, J. A. & Hoyos, C. D. Variability in tornado frequency associated with U.S. landfalling tropical cyclones. *Geophys. Res. Lett.* **36**, L17805 (2009).
- McCaul, E. W. Buoyancy and shear characteristics of hurricane-tornado environments. *Mon. Weather Rev.* **119**, 1954–1978 (1991).
- McCaul, E. W., Buechler, D. E., Goodman, S. J. & Cammarata, M. Doppler radar and lightning network observations of a severe outbreak of tropical cyclone tornadoes. *Mon. Weather Rev.* **132**, 1747–1763 (2004).
- Cook, A. R. & Schaefer, J. T. The relation of El Niño–Southern oscillation (ENSO) to winter tornado outbreaks. *Mon. Weather Rev.* **136**, 3121–3137 (2008).
- Weisman, M. L., Klemp, J. B. & Rotunno, R. Structure and evolution of numerically simulated squall lines. *J. Atmos. Sci.* **45**, 1990–2013 (1988).
- Kis, A. K. & Straka, J. M. Nocturnal tornado climatology. *Wea. Forecasting* **25**, 545–561 (2010).
- Weaver, S. J., Baxter, S. & Kumar, A. Climatic role of North American low-level jets on U.S. regional tornado activity. *J. Clim.* **25**, 6666–6683 (2012).
- Thompson, R. L., Smith, B. T., Dean, A. R. & Marsh, P. T. Spatial distributions of tornadic near-storm environments by convective mode. *Electronic J. Severe Storms Meteor.* **8**, 1–22 (2013).
- Schaefer, J. T. & Edwards, R. in *11th Conference on Applied Climatology*. Am. Meteorol. Soc. (Dallas, TX, USA 1999).
- Anderson, C. J., Wikle, C. K., Zhou, Q. & Royle, J. A. Population influences on tornado reports in the United States. *Wea. Forecasting* **22**, 571–579 (2007).

38. Doswell, C. A. & Burgess, D. W. On some issues of United States tornado climatology. *Mon. Weather Rev.* **116**, 495–501 (1988).
39. Mesinger, F. *et al.* North American Regional Reanalysis. *Bull. Am. Meteorol. Soc.* **87**, 343–360 (2006).
40. Davies-Jones, R., Burgess, D. & Foster, M. in *16th Conference on Severe Local Storms. Amer. Meteor. Soc.* 588–592 (Kananaskis Park, Alberta, Canada, 1990).
41. Johns, R. H. & Doswell, C. A. Severe Local Storms Forecasting. *Wea. Forecasting* **7**, 588–612 (1992).
42. Rose, S. F., Hobbs, P. V., Locatelli, J. D. & Stoelinga, M. T. A 10-Yr climatology relating the locations of reported tornadoes to the quadrants of upper-level jet streaks. *Wea. Forecasting* **19**, 301–309 (2004).
43. Wikle, C. K. & Anderson, C. J. Climatological analysis of tornado report counts using a hierarchical Bayesian spatiotemporal model. *J. Geophys. Res.* **108** (2003).
44. Besag, J., York, J. & Mollie, A. Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.* **43**, 1–59 (1991).
45. Brooks, S. P. & Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **7**, 434–455 (1998).
46. Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14 (1992).

Acknowledgements

This project has received funding support from the Natural Sciences and Engineering Research Council of Canada through a Discovery Grant awarded to George Arhonditsis.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interest.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Cheng, V. Y. S. *et al.* A Bayesian modelling framework for tornado occurrences in North America. *Nat. Commun.* 6:6599 doi: 10.1038/ncomms7599 (2015).

Erratum: A Bayesian modelling framework for tornado occurrences in North America

Vincent Y.S. Cheng, George B. Arhonditsis, David M.L. Sills, William A. Gough & Heather Auld

Nature Communications 6:6599 doi: 10.1038/ncomms7599 (2015); Published 25 March 2015; Updated 5 Jun 2015

The captions of Figs 2 and 3 of this Article contain typographical errors that were introduced during the production process. In both Figs 2 and 3, the units for the tornado occurrences should be $10,000 \text{ km}^2 \cdot \text{yr}^{-1}$ rather than $100 \text{ km}^2 \cdot \text{yr}^{-1}$.