Predicting the Severity of Tornado Events by Learning a Statistical Manifold for Tornado Property Losses

Preprint · February 2023

DOI: 10.13140/RG.2.2.34754.96963

3 authors, including:

Thilini V. Mahanama
University of Kelaniya
20 PUBLICATIONS **9** CITATIONS

SEE PROFILE

Dimitri Volchenkov
Texas Tech University
162 PUBLICATIONS **1,009** CITATIONS

SEE PROFILE

Predicting the Severity of Tornado Events by Learning a Statistical Manifold for Tornado Property Losses

Thilini Mahanama* Pushpi Paranamana† Dimitri Volchenkov‡

Abstract

We examine the relationship between property losses caused by tornadoes and their physical parameters, namely the tornado path length and width, using data reported by the National Oceanic and Atmospheric Administration in the United States. We observe that the statistics of property losses cannot be described by a single distribution but rather by a two-dimensional statistical manifold of distributions that may reflect two different mechanisms of property loss compensations. Assessing the difference between distributions of losses caused by tornadoes using Kolmogorov-Smirnov's distance, we construct the 2-D manifold using the method of multi-dimensional scaling. Then we define a curvature coefficient that characterizes the contraction and expansion of the derived manifold to explain the complex dynamics of the probability distributions of losses. The regions with expansions identify the ranges of physical parameters for which the extreme tornado events may occur, which helps in assessing compensation strategies.

Keywords: Risk assessment, tornado property losses, statistical manifold learning

1 Introduction

Actuarial science uses statistics to define ambiguity risks, analyze likelihood of uncertain future events, and to determine funds needed to pay connected claims. Statistical methods have a long-standing focus on inference, which is achieved through the creation and fitting of project-specific probability models (distributions) []. It is noteworthy that statistical methods are essentially top-down approaches, as it is assumed that we know the distribution from which the empirical data have been generated, and then the unknown parameters of the model should be estimated from the available data [1]. Although the use of statistical methods in actuarial science is consecrated by tradition, the obvious pitfall of such an approach is that the link between input and output variables is user chosen that may result in a less accurate prediction model especially if the relationships between variables are highly nonlinear and not easy to understand.

As the increase in data complexity makes classical statistical inference less tractable, the machine learning approach may be used to recognize patterns and create data clusters which share common characteristics that may influence the outcome. Machine learning methods constitute a bottom-up approach, as no particular model is assumed, but one begins with the data and then

*Department of Industrial Managment, Faculty of Science, University of Kelaniya, Kelaniya, 11600, Sri Lanka, thilinim@kln.ac.lk (Corresponding Author)

†Department of Mathematics & Computer Science, Saint Mary's College, IN 46556, USA

‡Department of Mathematics & Statistics, Texas Tech University, Lubbock TX 79409-1042, USA

an algorithm develops a model with prediction as the main goal [1]. The key concept of machine learning methods is a manifold hypothesis, suggesting that natural data lie along low-dimensional manifolds in high-dimensional data embedding space [2, 3, 4]. Algorithms of machine learning (based on some manifold related metric) are designed in a way to separate tangled data manifolds representing meaningful clusters in the data [5]. However, despite convincing prediction results, the lack of an explicit model (that would allow us to derive a quantitative measure of confidence) often makes machine learning solutions difficult to directly relate to already existing knowledge []. Many of machine learning applications result in inaccurate or irrelevant research results, as proper research protocols are not fully followed [6].

With the rapid increase in rich and unwieldy data arriving from the many industries, it is important to complement machine learning by statistical modeling to obtain a desired convergence that would take advantage of the best of both approaches for advancing actuarial science [7]. In our work, we intend to make a prominent step towards such a convergence of the statistical and machine learning approaches by defining a hybrid learning algorithm resulting in a two-dimensional manifold of different statistical models pertained to the physically disparate scales of tornado events. The proposed novel synthetic technique bearing the certain features of both machine learning and traditional statistical methods may help to understand and describe rare events and, especially, so called black swan events that come as a surprise from the point of view of a standard statistical approach [8].

The risk assessment and compensation strategies for the damage caused by a tornado depend on its severity [9]. For typical tornadoes, the private insurance companies compensate the property losses claimed by their clients [10]. However, they fail to cover the property losses attributed to severe (extreme) tornado events, as the scale of losses is rather high [11]. For such cases, the United States Department of Homeland Security introduced Federal Emergency Management Agency (FEMA) as a special coverage program [12, 13]. When a catastrophic tornado occurs in a state, the state governor proclaims a state of emergency [14], and, upon presidential approval, FEMA then distributes funds to the state and local governments [15]. Henceforth, the local governments redistribute money to counties and municipalities but not directly to individual tornado victims.

Enhanced Fujita Scale [16, 17] is commonplace to measure the intensity (degree of damage) of a tornado. This is limited to a six-degree damage (EF0, EF1,..., EF5) scale based on wind speed estimates [18, 19]. In [20, 21, 22], the relationship of tornado intensities to their sizes is modeled to predict tornado intensity when the path length and width are given. Section 1 depicts the physical parameters (path length and width) of a tornado rated as EF1. [20] used Weibull distributions for estimating tornado path lengths, widths, and intensities, but these distributional assumptions fail for high-end tornadoes. We address these limitations by proposing a non-parametric tornado severity scale on a continuum of its physical parameters.

In [18], we proposed a novel tornado event classification (Tornado Property Loss Scale, TPL-Scale) accounting for the property losses reported in the storm database published by the National Oceanic and Atmospheric Administration [24]. We also modeled the non-linear dependence between property losses and the area affected by tornadoes using a Gaussian copula approach [25]. The resulting correlation coefficients were not monotonic over time and location, and this parametric model failed to capture the extreme tornado events.

In this study, we intend to predict extreme tornado events to restructure the two-pronged compensation strategy for tornado-induced damages by learning a statistical manifold for the reported

Figure 1: The physical parameters (path length and width) of a EF1 tornado occurred in Raleigh, NC on April 19, 2019, according to National Weather Service, NOAA [23].

tornado events in the NOAA storm database [24]. We categorize tornado data with respect to the physical parameters (tornado path lengths and tornado path widths) and discern the kernel densities [26] of their property losses. By analyzing them, we observe that a single statistic can not be found to describe all possible losses and that the dynamics of statistics are difficult to interpret (Sec. 2.2).

The objective of this research is to determine potential tornado scales that can cause extreme tornado events to restructure the tornado compensation strategies based on physical parameters. We propose a tornado statistical manifold which will help public and private stakeholders to claim compensations for property losses caused by tornadoes in the United States, see Fig. 2.

We describe a tornado statistical manifold algorithm in detail and utilize it for predicting the compensation strategies for tornado property losses in the following sections: In section 2, we study the reported tornado events and define tornado scales based on the physical parameters. We describe our algorithm for learning a statistical manifold in section 3. Based on the manifold, we identify the physical parameters for potential extreme tornado events in section 4. Finally, we make concluding remarks in section 5 and extend our proposed two-dimensional manifold learning to multidimensional manifold learning by generalizing the algorithm for high-dimensional data.

2 Exploratory Analysis

We analyze the tornado events reported in the U.S. national tornado database [24] and distinguish them based on their physical parameters. As a preliminary study, we discern the probability distributions of property losses to describe different compensation strategies for tornado damages used in the United States.

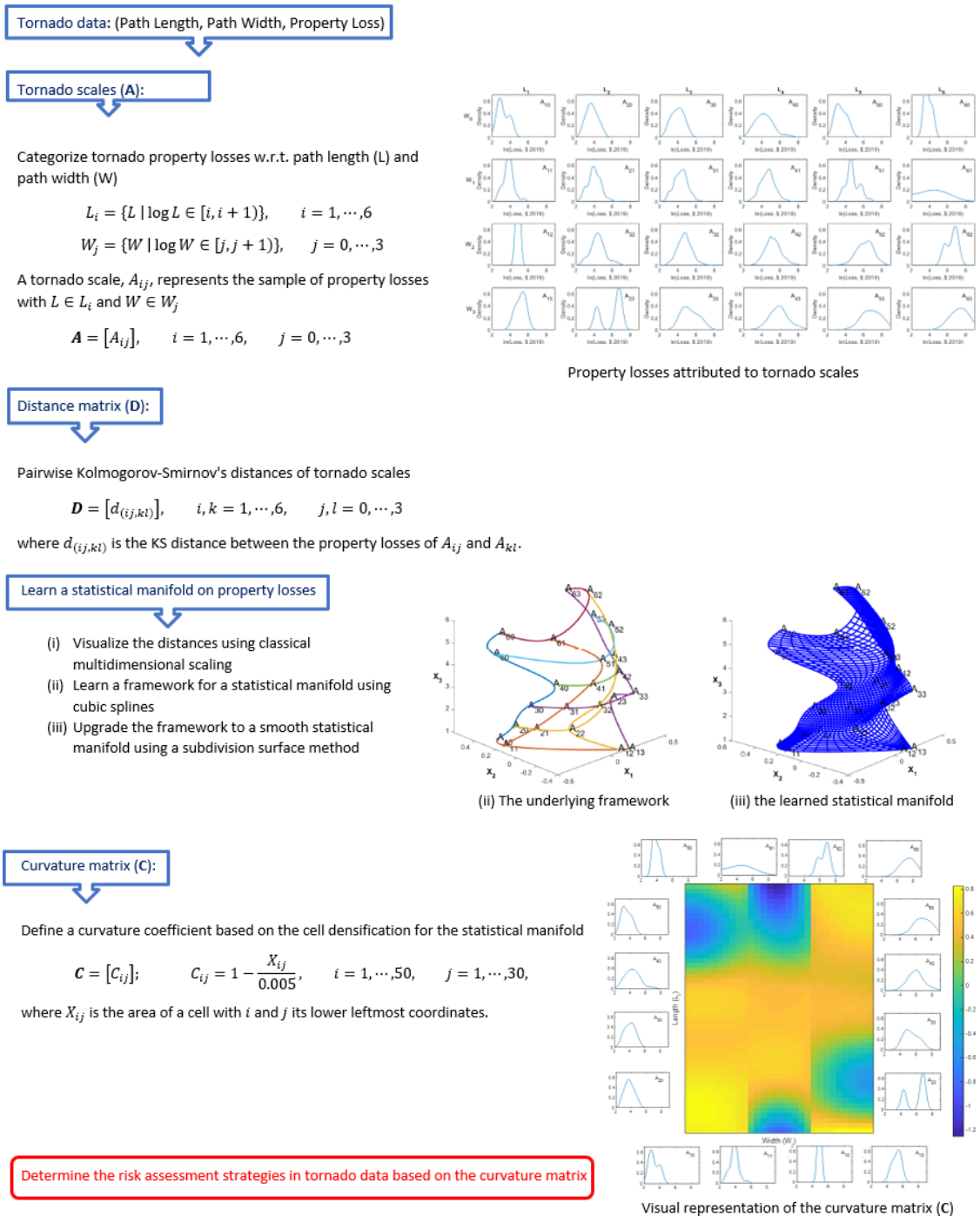


Figure 2: Statistical Manifold Learning Algorithm for Tornado Risk Assessment

2.1 Categorizing Tornado Data with respect to Physical Parameters

In this section, we define tornado scales by taking different physical parameters (tornado path length and path width) into account. We use the tornado property losses (measured in USD 2019) and

their physical parameters (measured in ft) for each tornado reported between 1993 and 2018 in the storm database published by the National Oceanic and Atmospheric Administration [24]. Then, we define the following scales of physical parameters using logarithms of path lengths ($L_i,\ i = 1, \cdots, 6$) and path widths ($W_j,\ j = 0, \cdots, 3$):

$$\begin{aligned} L_i &= \{L \mid \log\ L \in [\ i\ ,\ i+1)\}, \qquad i = 1, \cdots, 6 \\ W_j &= \{W \mid \log W \in [\ j\ ,\ j+1)\}, \qquad j = 0, \cdots, 3. \end{aligned} \tag{1}$$

We further categorize the tornado data using the combinations of length and width scales. Then, we examine the tornado-induced property losses for each combined scale given below:

$$A = [A_{ij}], \quad i = 1, \cdots, 6, \ \ j = 0, \cdots, 3, \tag{2}$$

where A_{ij} represents the sample of property losses with path lengths in i^{th} length interval, L_i, and path widths in j^{th} width interval, W_j.

2.2 Exploratory Analysis on Determining Bilateral Coverages for Tornado Damages

This section qualitatively analyzes the distributions of tornado-induced property losses of tornado scales introduced in section 2.1. We provide the kernel density plots of 24 tornado scales in Fig. 3 and discern the inconsistent distribution shapes. Every tornado scale is characterized by a unique distribution.

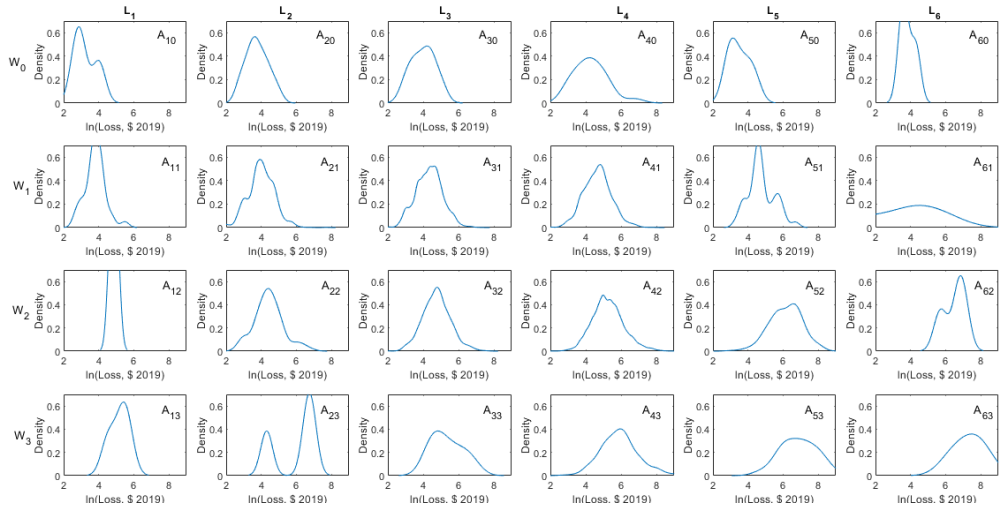


Figure 3: Property losses attributed to tornado scales in A. A_{ij} represents the property losses of the tornadoes with path lengths in $L_i,\ i = 1, \cdots, 6$ and path widths in $W_j,\ j = 0, \cdots, 3$.

Since the distribution densities (statistics) in A_{31}, A_{41}, A_{32}, A_{42} cells are similar, any of the statistics can be used to predict the losses in neighboring tornado scales. Due to this relatively high predictability, the insurance companies are mainly prepared to compensate the claims for such tornado property losses.

The two tornado scales A_{60} and A_{61} have the same length scale but different width scales. Even though the difference between their path widths is only 90 ft, the statistics seem to be

significantly different. For instance, A_{60} shows one peak whereas A_{61} shows several peaks in the density curves. The tornado scale A_{61} seems to illustrate an extremity and insurance companies fail to compensate such losses. For such cases, the government introduced coverage programs to fund states of emergency. This funding is distributed by the Federal Emergency Management Agency, which is a part of Homeland Security [27]. They fund the states and local governments, but they do not compensate the losses of individuals directly.

Thus, we cannot use a single probability distribution (a unique statistic) to examine all the tornado-induced property losses in the database as their distributions are distinct variants with reference to the physical parameters. For example, the bimodal density curves in A_{23} , A_{62} and A_{63} might display the two-fold compensation strategies used in the United States: 1. private insurance 2. government coverages. We substantiate this claim using the characteristics of the forthcoming statistical manifold on tornado property losses. As the initial step, we quantitatively measure the differences between distributions of tornado scales in section 3.1.

3 Learning a Statistical Manifold

In this section, we learn a statistical manifold for describing the difference between distributions of tornado-induced property losses with respect to different physical parameters. We determine the pairwise differences between the statistics of tornado scales using Kolmogorov-Smirnov's distance and visualize them using classical multidimensional scaling [28, 29, 30]. Based on the resultant configuration, we learn a statistical manifold by outlining the inherent patterns of physical parameters. Then, we use a subdivision surface method to smoothen the proposed manifold. We define a curvature coefficient based on the expansion and contraction of the manifold to identify the physical parameters for potential extreme tornado events.

3.1 Constructing a Distance Matrix based on the Differences between Statistics in Tornado Scales

This section quantitatively compares the distributions of property losses for the tornado scales introduced in section 2.1. We quantify the pairwise differences between the 24 distributions of property losses attributed to A using Kolmogorov-Smirnov's distance [28, 31]. to obtain symmetric non-parametric distances between two tornado scales. Kolmogorov-Smirnov's distance between a pair of tornado scales provides the maximum distance between the empirical distribution functions of their property losses. We define the empirical distribution function of the property losses of A_{ij} at a specified loss (z), $\widehat{F}_{ij}(z)$, as the proportion of property losses in A_{ij} less than or equal to z:

$$\widehat{F}_{ij}(z) = \frac{1}{n}\sum_{a=1}^{n} I[Z_a \leq z], \tag{3}$$

where n is the number of tornadoes recorded in A_{ij} and $I(.)$ is the identity function. Then, the Kolmogorov-Smirnov's distance between the property losses of category A_{ij} and A_{kl} is given by

$$d_{(ij,kl)} = \max_{z}\left(\left|\widehat{F}_{ij}(z) - \widehat{F}_{kl}(z)\right|\right) \quad i,k = 1,\cdots,6, \;\; j,l = 0,\cdots,3, \tag{4}$$

where $\widehat{F}_{kl}(z)$ is the proportion of property losses in A_{kl} less than or equal to z. Consequently, we define a distance matrix (D) which consists of these pairwise distances.

$$\begin{aligned} D &= [d_{(ij,kl)}], \quad i,k = 1,\cdots,6, \quad j,l = 0,\cdots,3 \\ &= [d_{pq}], \quad p,q = 1,\cdots,24. \end{aligned} \tag{5}$$

where p and q denote the position of the distance (d) in the matrix based on its row (p) and column (q). We utilize this distance matrix to investigate a geometrical representation of tornado scales.

3.2 Visualizing Tornado Scales using Classical Multidimensional Scaling for the Distance Matrix

We apply classical multidimensional scaling (CMDS), also known as principal coordinate analysis [29, 30, 32] to obtain a visual representation for the distance matrix (D). In particular, we choose two-dimensional CMDS as higher-dimensional CMDS provide chaotic configurations. In CMDS, a two-dimensional coordinate matrix for the tornado scales is obtained by minimising the following loss function (residual sum of square) [33]:

$$\mathrm{S}_D(X) = \sqrt{\sum_{p\neq q=1,\cdots,24} (d_{pq} - ||x_p - x_q||)^2}, \quad d_{pq} \in D. \tag{6}$$

By implementing CMDS, we find a new configuration (X) for tornado scales such that their distances, d_{pq}, are well-approximated by the distances between the corresponding elements in this new configuration $||x_p - x_q||$. That is, this quadratic optimization problem finds the best possible coordinates based on the distances in D. The coordinate matrix (X) is constructed using the following steps [34, 35]:

1. Use the distance matrix (D) to calculate the inner product matrix (B),

$$B = -\frac{1}{2} JDJ,$$

   where $J = I - \frac{1}{n}11^T$, I is the identity matrix, 1 is a vector of all ones and n is the number of tornado scales (n=24).

2. Decompose B using

$$B = V\Lambda V^T,$$

   where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ such that $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$, is the diagonal matrix of eigenvalues of B, and $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, is the matrix of corresponding unit eigenvectors.

3. Extract the first and second eigenvalues $\Lambda_2 = \mathrm{diag}(\lambda_1, \lambda_2)$ and corresponding eigenvectors $V_2 = [\mathbf{v}_1, \mathbf{v}_2]$.

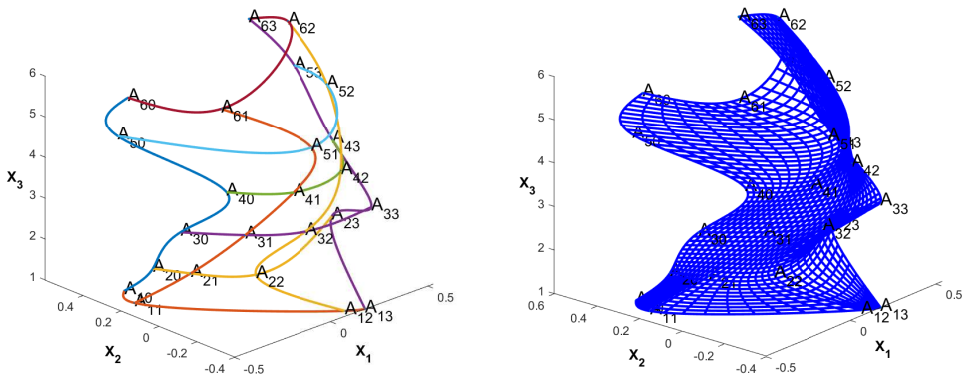
4. The coordinate matrix (X) is given by

$$X = [\mathbf{X}_1, \mathbf{X}_2]^T = V_2 \Lambda_2^{\frac{1}{2}}.$$

3.3 Learning a Statistical Manifold by Defining a Hybrid Algorithm

The two-dimensional configuration obtained using CMDS, (X_1, X_2), provides relative positions for the tornado scales based on their physical parameters and property losses. We comprehend the inherent patterns related to each tornado path length scale $(L_i,\ i = 1, \cdots, 6)$ considering the sequence $A_{i0}, A_{i1}, A_{i2}, A_{i3}$ and for each width scale $(W_j,\ j = 0, \cdots, 3)$ considering $A_{1j}, A_{2j}, A_{3j}, A_{4j}, A_{5j}, A_{6j}$ into account. Using periodic interpolating cubic splines [36], we outline the contours of length scales $(A_{i0}, A_{i1}, A_{i2}, A_{i3})_{i=1,\cdots,6}$ and width scales $(A_{1j}, A_{2j}, A_{3j}, A_{4j}, A_{5j}, A_{6j})_{j=0,\cdots,3}$. To obtain a two-dimensional manifold, we locate the length contours such that consecutive contours preserve unit distance. Thus, we learn an underlying framework for a statistical manifold by defining a hybrid algorithm based on CMDS and ordering with respect for tornado physical parameters, see Fig. 4(a). In this graph, each vertex represents the probability distribution of property losses in the corresponding tornado scale (A_{ij}). This configuration of tornado scales preserves the distances between probability distributions of property losses in A.

The fuzzy structure in Fig. 4(a) can be smoothened by enhancing the degree of edges. We upgrade the underlying framework in Fig. 4(a) to a smooth manifold by subdividing the ambient space of contours. In particular, we increase the cardinality of vertices and subdivide each edge into 10 equally-spaced segments to obtain 10 supplementary vertices. Taking the new vertices into account, we add contours with respect to width $(W_j^*,\ j = 1, ..30)$ and length $(L_i^*,\ i = 1, ..50)$ scales using cubic splines, see Fig. 4(b). As a result, we improve the coarse-grained conformational manifold in Fig. 4(a) to the smooth statistical manifold illustrated in Fig. 4(b). In fact, each point on this manifold potentially represents a distribution of property losses for a given path length and path width. In section 3.4, we delineate the statistics of the manifold by defining a curvature coefficient.



(a) (b)

Figure 4: (a) The underlying framework and (b) the learned statistical manifold based on the tornado-induced property losses constructed by implementing classical dimensional scaling and a method of subdivision surfaces. $A_{ij},\quad i = 1, \cdots, 6,\quad j = 0, \cdots, 3$, represents the tornadoes with lengths in L_i and widths in W_j.

3.4 Defining a Curvature Coefficient Matrix for the Statistical Manifold

Each point on the learned statistical manifold postulates a statistic based on the property losses with respect to the physical parameters. Since this curved manifold seems to be too complex to interpret, we introduce a curvature coefficient to identify the physical parameters for potential extreme tornado events.

In Fig. 4(b), the bottom leftmost (A_{10}-A_{11}-A_{21}-A_{20}) and the top rightmost (A_{52}-A_{53}-A_{63}-A_{62}) zones consist of compressed (dense) cells. The cells in the regions A_{22}-A_{23}-A_{33}-A_{32} and A_{51}-A_{52}-A_{62}-A_{61} illustrate relatively high expansions. Non-trivial curvatures are present in both of these cell types: compressed and expanded cells.

The statistics of property losses in compressed cells hardly change within neighboring cells (see Fig. 3), i.e., that most of the cell statistics have the same form in densification. However, since the statistics tremendously vary from one cell to another in expanded cells, we identify the potential regions for extreme tornado events on the manifold. This variation escalates at the border zones of compressed and expanded regions. Since the statistics in expansion and transition zones change abruptly from the conventional statistics, distinct statistics are essential to determine the significant variations from one cell to another. Thus, the coordinates of expanding cells seem to provide the range of physical parameters of potential extreme tornado events. Correspondingly, the compensation strategies for tornado-induced property losses are related to the cell densification in the statistical manifold, Fig. 4(b).

We determine the degree of cell densifications on the manifold, Fig. 4(b), with respect to tornado physical parameters. In particular, we quantify these deformations based on the underlying curvature coefficients of ambient spaces in cells. Then, we define a matrix of curvature coefficients (C) by comparing each cell area with the mean cell area (0.005) on the manifold as follows:

$$C = [C_{ij}]; \qquad C_{ij} = 1 - \frac{X_{ij}}{0.005}, \quad i = 1, \cdots, 50,\ j = 1, \cdots, 30, \tag{7}$$

where X_{ij} is the area of a cell with i and j its lower leftmost coordinates. The curvature coefficients provide positive values for cell contractions and negative values for cell expansions.

In Fig. 5, we provide a visual representation of the curvature coefficient matrix with respect to the positions of the width ($W_j^*,\ j = 1, ..30$) and length ($L_i^*,\ i = 1, ..50$) coordinates. The relative positions of some scales in A are shown in exterior locations. According to its color scale, yellow represents compressed cells (positive curvature coefficients), blue represents expanded cells (negative curvature coefficients) and transitions between them are shown as green phases. In section 4, we describe how the curvature coefficient matrix is utilized for restructuring the compensation strategies with respect to tornado physical parameters.

4 A Prospective Framework for Predicting the Compensation Strategies for Tornado Property Losses

In this section, we identify the physical parameters for predicting potential extreme tornadoes by utilizing the curvature coefficient matrix (7) defined for the learned statistical manifold. Based on the predicted tornado physical parameters, we envisage the appropriate approach for compensating property losses. That is, we analyze how the compensation strategies vary with the physical parameters of a tornado (i.e., from cell to cell on the manifold).

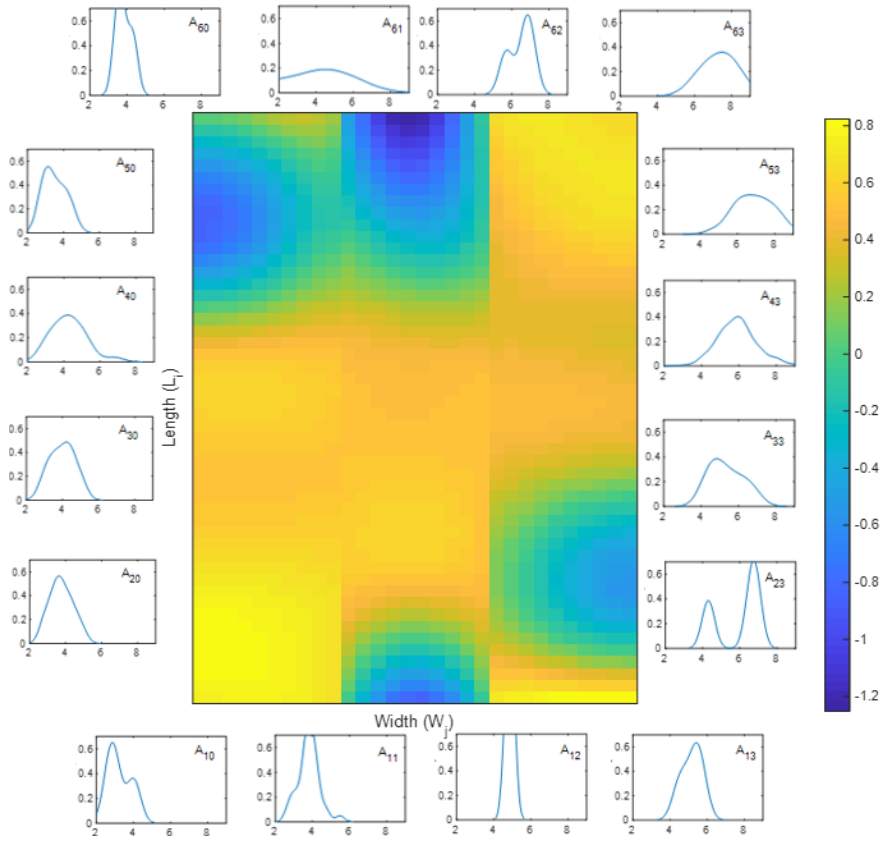


Figure 5: The visual representation of the curvature coefficient matrix with respect to the positions of the path width ($W_j^*,\ j = 1,..30$) and path length ($L_i^*,\ i = 1,..50$) coordinates of a tornado. The relative positions of some scales in A are shown in exterior locations. Yellow and blue regions depict compressed and expanded cells on the manifold, respectively.

We compare the kernel densities of some main tornado scales with the visual representation of the curvature coefficient matrix in Fig. 5. Clearly, the densities corresponding to the monochromic zones seem to have unimodal densities with similar statistics. For example, the neighboring region of A_{11}-A_{10}-A_{20}-A_{30} illustrates a yellow zone in Fig. 5 with moderately similar unimodal densities. However, some regions on the curvature portrait show that the small changes in physical parameters (L_i and W_j) trigger vast changes in the statistics corresponding to their property losses. For example, since the property losses reported for the tornadoes in the zone bounded by A_{40}-A_{50}-A_{60}-A_{61} vary dramatically, we observe relatively rapid changes in color from one cell to another. In some blue zones, we observe the change of modality in densities (unimodal to bimodal). The bimodal density curves might have resulted from the two-fold compensation strategies for tornado damages in the United States. In such scenarios, a single statistic fails to estimate such volatilities in property losses with respect to small changes in tornado path lengths and path widths. Therefore, the blue regions on the manifold seem to demonstrate potential tornado scales for extreme tornado events.

Our findings based on the learned manifold and curvature portrait reflect the principle of two-fold compensation mechanisms for tornado property losses in the United States: the private insur-

ance companies compensate property losses claimed by individual clients for small-scale tornadoes, and the government allocate funds for state emergencies due to severe tornadoes. In conclusion, we determine potential extremities in tornado-induced property losses and identify the potential physical parameters for extreme tornadoes to suggest a prospective framework for restructuring compensation strategies via learning a statistical manifold. Our proposed algorithm for tornado risk assessment consists of following process:

1. Partition tornado data (Path Length, Path Width, Property Loss) with respect to log scales of physical parameters, Eq. (1)
2. Combine the partitions of physical parameters to categorize tornado property losses, *tornado scales* , Eq. (2)
3. Calculate pairwise Kolmogorov-Smirnov's distances between tornado scales, Eq. (4), to find the distance matrix, Eq. (5)
4. Implement classical multidimensional scaling to the distance matrix, Eq. (6), to visualize tornado scales in a new configuration
5. Comprehend the inherent patterns of tornado scales and use cubic splines to learn the underlying framework of manifold, Fig. 4(a)
6. Upgrade the underlying framework to a smooth statistical manifold using a subdivision surface method, Fig. 4(b)
7. Define a curvature coefficient for the learned statistical manifold using cell area, Eq. (7)
8. Determine potential extreme tornadoes and tornado loss compensation strategies using the curvature measurements, Fig. 5

In Sec. 5, we extend our proposed two dimensional manifold learning to a multidimensional manifold learning by generalizing the algorithm for high-dimensional data.

5 Conclusion

Since private insurance companies fail to compensate for enormous financial losses due to extreme tornadoes, the U.S. government has introduced coverage programs such as FEMA. In this study, we suggested a framework to predict the strategies for compensating for tornado damages. In order to do that, we learned a statistical manifold by utilizing the reported tornado-induced property losses and their physical parameters in the national tornado database [24].

According to our findings, no single distribution can describe all tornado events in the database. By defining a curvature coefficient matrix based on the manifold, we described how the property losses change with respect to small changes in tornado path lengths and path widths. As a result, we identified the tornado physical parameters for extreme tornado events. The complexity of the learned two-dimensional statistical manifold represents the diversity of compensation strategies in tornado events used for this study.

Our proposed statistical manifold learning algorithm could be generalized for other applications with high-dimensional data. Suppose we want to predict a variable (Y) considering its dependency

on several variables $(X_1, \cdots, X_p)$. First, we create a meaningful partition for each independent variable and combine the partitions of different independent variables to categorize the dependent variable (i.e., scales of Y). Second, we construct the kernel densities of the scales of Y and find the distance matrix based on the pairwise Kolmogorov-Smirnov's distances between scales. Third, we implement classical multidimensional scaling to illustrate a lower dimensional configuration. As described in section 3, we learn a statistical manifold using cubic splines and apply a subdivision surface method to smoothen the manifold. Then, we define a curvature coefficient for the learned statistical manifold based on cell area. Finally, we analyze the dynamics of curvature measurements and utilize them for decision-making.

Although both statistics and machine learning may efficiently endow actuarial science, they have different goals and make different contributions. While machine learning is usually a bottom-up approach that may emphasize prediction, statistics is rather a top-down approach that may focus more on inference. A novel synthetic technique proposed in our work takes advantage of the best of both approaches for improving both inference and prediction of hazardous rare events exposing banks and insurance companies to unpredictable losses. We have demonstrated that the common dialogue between statistics and machine learning can bring improvements in both fields. Regularization and resampling may be required when a single statistic fails to reliably assess risks in the highly curved regions of a statistical manifold. As distinct and even disparate statistics may be applied to assess abruptly changing physical properties of tornado events, both fields may contribute to mutual improvements. https://www.overleaf.com/project/6376c768e858a86888dbce70

References

[] Ij, H. (2018), Statistics versus machine learning. *Nat Methods*, **15**(4), 233.

[1] Ley, C., Martin, R. K., Pareek, A., Groll, A., Seil, R., and Tischer, T. (2022), Machine learning and conventional statistics: making sense of the differences. *Knee Surgery, Sports Traumatology, Arthroscopy*, 1-5.

[2] Cayton, L. (2005), Algorithms for manifold learning. *Univ. of California at San Diego Tech. Rep*, **12**(1-17), 1.

[3] Fefferman, C., Mitter, S., and Narayanan, H. (2016), Testing the manifold hypothesis. *Journal of the American Mathematical Society*, **29**(4), 983-1049.

[4] Sritharan, D., Wang, S., and Hormoz, S. (2021), Computing the Riemannian curvature of image patch and single-cell RNA sequencing data manifolds using extrinsic differential geometry. *Proceedings of the National Academy of Sciences*, **118**(29), e2100473118.

[5] Rathnayake, K., Lebedev, A., and Volchenkov, D. (2022), Deciding on a Continuum of Equivalent Alternatives Engaging Uncertainty through Behavior Patterning. *Foundations*, **2**(4), 1080-1100.

[6] Shelmerdine, S. C., Arthurs, O. J., Denniston, A., and Sebire, N. J. (2021), Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. *BMJ Health & Care Informatics*, **28**(1).

[7] Beam, A. L., and Kohane, I. S. (2018), Big data and machine learning in health care. *Jama*, **319**(13), 1317-1318.

[8] Nicholas, N. (2008), The black swan: the impact of the highly improbable. *Journal of the Management Training Institut*, **36**(3), 56.

[9] Daneshvaran, S., and Morden, R. E. (2007), Tornado risk analysis in the United States. *The Journal of Risk Finance*.

[10] Ericson, R., Barry, D., and Doyle, A. (2000), The moral hazards of neo-liberalism: lessons from the private insurance industry. *Economy and society*, **29**(4), 532-558.

[11] Peng, X., Roueche, D. B., Prevatt, D. O., and Gurley, K. R. (2016), An engineering-based approach to predict tornado-induced damage. *In Multi-hazard approaches to civil infrastructure engineering*. Springer, Cham. 311-335

[12] Lindsay, B. R., Kapp, L., Shields, D. A., Stubbs, M., Lister, S. A., McCarty, M., et al, T. (2012), Federal emergency management: A brief introduction. Library of congress Washington DC.

[13] Choi, J., and Wehde, W. (2020), Trust in emergency management authorities and individual emergency preparedness for tornadoes. *Risk, Hazards & Crisis in Public Policy*, **11**(1), 12-34.

[14] Kapucu, N., Augustin, M. E., and Garayev, V. (2009), Interstate partnerships in emergency management: Emergency management assistance compact in response to catastrophic disasters. *Public Administration Review*, **69**(2), 297-313.

[15] Greenough, G., McGeehin, M., Bernard, S. M., Trtanj, J., Riad, J., and Engelberg, D. (2001), The potential impacts of climate variability and change on health impacts of extreme weather events in the United States. *Environmental health perspectives*, **109**(2), 191-198.

[16] NWS. (2019), The Enhanced Fujita Scale (EF Scale). Storm Prediction Center, National Weather Service.

[17] Fujita, T. T. (1971), Proposed characterization of tornadoes and hurricanes by area and intensity (No. NASA-CR-125545).

[18] Mahanama, T., and Volchenkov, D. (2022), Tornado Property Loss Scale: Up to $8 Billion by 2025. Classification, Dependence, and Prediction of Tornado Events in the US. *Journal of Environmental Accounting and Management*, **10**(02), 127-142.

[19] Doswell III, C. A., and Burgess, D. W. (1988), On some issues of United States tornado climatology. *Monthly Weather Review*, **116**(2), 495-501.

[20] Brooks, H. E. (2004), On the relationship of tornado path length and width to intensity. *Weather and forecasting*, **19**(2), 310-319.

[21] Elsner, J. B., Jagger, T. H., and Elsner, I. J. (2014), Tornado intensity estimated from damage path dimensions. *PloS one*, **9**(9), e107571.

[22] Strader, S. M., Ashley, W., Irizarry, A., and Hall, S. (2015), A climatology of tornado intensity assessments. *Meteorological Applications*, **22**(3), 513-524.

[23] NWS. (2019), National Weather Service Reports.

[24] NCEI. (2018), Storm Events Database. National Centers for Environmental Information (NCEI), National Oceanic and Atmospheric Administration.

[25] Nelsen, R. B. (1991), Copulas and association. *In Advances in probability distributions with given marginals*. Springer, Dordrecht. 51-74.

[26] Coleman, T. A., and Dixon, P. G. (2014), An objective analysis of tornado risk in the United States. *Weather and Forecasting*, **29**(2), 366-376.

[27] Schneider, S. K. (1998), Reinventing public administration: A case study of the Federal Emergency Management Agency. *Public Administration Quarterly*, 35-57.

[28] Massey Jr, F. J. (1951), The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, **46**(253), 68-78.

[29] Borg, I., and Groenen, P. J. (2005), Modern multidimensional scaling: Theory and applications. Springer Science & Business Media.

[30] Serva, M., Vergni, D., Volchenkov, D., and Vulpiani, A. (2017), Recovering geography from a matrix of genetic distances. *Europhysics Letters*, **118**(4), 48003.

[31] Heuser, A., Rioul, O., and Guilley, S. (2014), A theoretical study of Kolmogorov-Smirnov distinguishers. In *International Workshop on Constructive Side-Channel Analysis and Secure Design*. Springer, Cham. 9-28.

[32] Chen, C. P., and Zhang, C. Y. (2014), Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, **275**, 314-347.

[33] Mead, A. (1992), Review of the development of multidimensional scaling methods. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **41**(1), 27-39.

[34] Birchfield, S. T., and Subramanya, A. (2005), Microphone array position calibration by basis-point classical multidimensional scaling. *IEEE transactions on Speech and Audio Processing*, **13**(5), 1025-1034.

[35] Wickelmaier, F. (2003), An introduction to MDS. *Sound Quality Research Unit, Aalborg University, Denmark*, **46**(5), 1-26.

[36] Lee, E. T. (1989), Choosing nodes in parametric curve interpolation. *Computer-Aided Design*, **21**(6), 363-370.