The razor in the waterfall: using longitudinal data to sharpen the analysis of cascading disaster risk

T. J. Huggins¹, K. Chen¹, W. Gong¹, and L. Yang¹

1 Southern University of Science and Technology, Shenzhen, Guangdong, China
E-mail: yangll@sustech.edu.cn

Abstract. Cascading disasters progress from a triggering disaster event to a diverse range of consequent disasters. Disasters following the Great East Japan Earthquake of 2011 highlight how these cascades also progress to multiple geographical locations. However, the very low frequency of these events means their analysis has usually excluded base-rate data. This common practice risks overestimating future likelihoods. A simplified approach to base-rates for less catastrophic cascades, following the rule of Occam’s razor, may help develop more accurate predictions of future likelihoods. The current research hypothesized that an intuitively relevant (0.05) probability of cascading flood-related disasters could be derived from a large, generic register of disaster events. A threshold-based analysis of transitions between phases of a hydrologic flood-related cascade was performed using ten years of data from the USA state of Florida. This analysis identified a 0.05 probability of flood-related cascading disasters. The same analytical methods were reliable when applied to subsequent data, from the year 2000. Similar approaches to information extraction and probability analysis can be applied to climatic data collected at more regular intervals. This will improve the usefulness of analytical results, which can then be added to expert analyses of more contemporary events and scenarios.

1. Introduction
Cutter [1] used the term cascading to define combinations of disaster events that progress from initial triggers to subsequent disasters, across a range of locations. For example, the September 11th terrorist attacks in the USA involved a plane hijacking that led to an attack at a high altitude of the World Trade Centre in New York. This attack subsequently caused a complete collapse of the entire building. Triggers for these events, and the potential for these triggers to set off more than one major event, were not detected despite substantial investments in intelligence gathering and analysis [1]. Cascading disasters have remained very difficult to manage, predict, or even conceptualize. They are easily confounded with compound disasters, even though the latter progress through linear chains between disaster events, rather than non-linear interactions between diverse events and event locations [2]. Zuccaro, Gregorio and Leone [3] concluded that particularly data-intensive analysis is required to predict cascading disasters. The data required for this analysis is often unavailable [3]. This is one reason why cascading disasters and their impacts have been uniquely challenging to define and manage; within national and international strategies for disaster risk reduction [4].

It is clear that the difficulties of identifying, predicting and managing cascading disasters do not make them any less important. The potentially catastrophic impacts of these cascades mean that they still form an important part of disaster research agendas. The catastrophic 2011 cascade progressing from the Great East Japan Earthquake and Tsunami to the Fukushima Daiichi nuclear disaster provides another clear example. Similar cascades of disaster events could occur in many developed economies.
throughout the world [5]. This possibility equates to substantial levels of overall risk, because of the severe impacts of comparably cascading disasters [6].

Kumasaki, King, Arai, and Yang [7] took a longer-term perspective on Japanese cascading disasters. It was demonstrated how press coverage of disaster events, among other documents, can be used to analyze the relationship between cascading disaster components. This analysis highlighted opportunities for analyzing cascading disasters that occur much more frequently than the September 11 and Great East Japan cascades. The impacts of more frequent cascading disasters tend to be smaller and less severe. However, with reference to United Nations concept of overall risk [6], it is the increased likelihood of transitions from one disaster component to another, that makes these cascades just as relevant. Frequent transitions can also amount to more robust sets of base rate data, for the purposes of analysis.

Many disaster events result from interactions between multiple natural and human systems. According to Huggins, Hill, Peace and Johnston [8], this means that disaster risk scenarios result from definitively complex systems. Some aspects of more frequent cascading disasters may nonetheless be much simpler than they appear. The principal of Occam’s razor dictates that scientists should begin by looking for simple explanations, rather than starting with more complex, but potentially pointless, approaches. This is particularly true when analyzing disaster events, which have often been recorded in large, pre-existing databases. The current research asked whether the substantial volume of this data can be harnessed to help resolve persistent issues concerning the analysis of relatively low frequency, cascading disasters.

Base-rates have often been excluded when analyzing the likelihood of cascading disasters. For example, in the combination of Cross Impact Analysis (CIA) and Interpretative Structural Modeling (ISM) by Ramirez de la Huerga, Silvera and Turoff [9]. This combination was used to develop subjective expert ratings of correlations between cascading disaster components [9], without being mired in the very infrequent occurrence of relevant cascades.

However, these kinds of subjective estimates are likely to neglect, rather than systematically consider, longer-term trends. This is due to cognitive biases that lead humans to neglect base rate frequencies, in favor of more recent information [10, 11, 12]. This means that individuals predicting a cascade running for example, from gale strength wind, to a rain-storm, and then to flooding, are more likely to focus on recent events, rather than considering longer term antecedents. According to Hertwig, Barron, Weber and Edo [13], the common focus on recent events leads to overestimating the likelihood of rare events.

There are ways to mitigate the neglect of prior frequencies, towards improving the accuracy of likelihood estimates [14]. These remedies include selecting experts who have been personally exposed to base-rate information [15, 16, 17]. This exposure to base-rates can be improved by providing information in a format that highlights the relevance of longer-term base rates [18] - exemplifying how information formats can promote naturalistic cognitive adaptations to complex disaster-related scenarios, as outlined by Huggins et al. [8].

Base-rates are required in any case. Rather than continuing to assume that they are too difficult to calculate, the current research took an Occam’s razor to large registers of diverse disaster events and locations. As outlined below, a relatively simple approach appeared to suit the analysis of cascading flood-related disasters in particular.

2. Cascading Flood-Related Disasters

The generic hydrologic cycle makes it particularly simple to build a cascading model of flood-related disasters. Van Brahna [19] (p. 728) defined the following “dominant processes” of the generic hydrologic cycle, in terms of km$^2$ transported per year:

1. Evaporation from oceanic bodies of water, at 425,000 km$^2$;
2. Rainfall onto oceans, at 383,000 km$^2$;
3. Net transport to land through wind and pressure changes, at 42,000 km$^2$;
4. Land-based plants’ conversion of water into water vapor, at 71,000 km$^2$;
5. Rainfall onto land, at 113,000 km2;
6. River flow to oceans, at 42,000 km². Metrics outlined by Van Brahana do not account for dominant processes concerning ground-water flow and infiltration. However, it is clear that water evaporated from the oceans makes a substantial and generic contribution to rainfall onto land. This occurs once a large part of that water volume (42,000 km²) has been transported to land via air currents, otherwise referred to as wind. In other words, there is a clear relationship between coastal winds and rainfall onto land. This relationship is an integral part of the basic physics of energy transfer between atmosphere, oceans and land masses, also outlined by Van Brahana [19] and generally referred to as climate. Flooding occurs when the eventual rate of rainfall exceeds the interception, evaporation, infiltration, and storage capacities of land-based geography [20]. It is most commonly associated with flows that exceed the bank height of river channels [20]. It follows that severe flooding is likely to result from intense rainfall (dominant process 5), exceeding the capacities outlined. The importance of this dominant process is highlighted in black, in Figure 1.

As also highlighted in Figure 1, the intensity of dominant process number 5 results from the transport of atmospheric humidity through dominant process number 3 (net transport to land through wind and pressure changes). A gross hydrologic contribution of 42,000 km² from the later process generally adds over 50% to the 71,000 km² atmospheric water volume produced via land-based plants. This means that water volume transported inland from oceanic areas increases rainfall to intensities and durations exceeding the flood-reducing capacities of land-based geography.

In sum, strong wind can be a trigger for many heavy rain and flooding scenarios. In terms of the standard hydrologic cycle, the intensity of dominant process 3 (net wind-related transport) increases the intensity of dominant process 5 (rainfall onto land). The intensity of the latter dominant process can then increase the likelihood of intense flooding, beyond capacity thresholds which include dominant process 6 (river flow to oceans). Figure 1 shows how these interactions can drive cascading disasters characterized by the intensity of dominant processes 3 and 5 in particular.

Elements of the hydrologic cycle summarized in Figure 1 illustrate how flooding is generally exacerbated by heavy rainfall. However, heavy rainfall is not necessary and sufficient to constitute a disaster. This means that rainfall is generally not commonly registered as a disaster type, in data sets such as the United Nations Office for Disaster Risk Reduction (UNISDR) Desinventar databases [21]. Severe storms typically include heavy rainfall and are registered as disasters. Severe gales may also be associated with heavy rainfall. It is therefore important to associate both types of disasters with heavy rainfall when using generic disaster data from Desinventar and similar databases.

The current, initial research into cascading disasters aimed to use the same kind of generic disaster event data to calculate a simple probability of flood-related cascading disasters. This type of cascading disaster progresses indeterminately, from severe gales to flooding events at diverse locations across an extended geographic area. It was therefore hypothesized that there would be an intuitively relevant 0.05 probability of at least one cascading flood-related pathway, triggered by a widespread gale disaster.
3. Methods

The Desinventar [21] database for the USA State of Florida was selected because it covered a particularly long period, from 1972 to 2001. It was also selected because event duration had either been registered by repeated entries (e.g. flooding registered in the same county on subsequent days) or by a duration value for each disaster occurrence. Table 1 shows the result of a basic query for gale, storm and flooding disasters from this database. In this example, a duration of 0 indicates that a disaster occurred within the day in question.

**Table 1. Database query result**

| Serial | Start date | Type of event | Geography Name | Duration |
|--------|------------|---------------|----------------|----------|
| 4442   | 3/10/1992  | Inundation    | St. Johns      | 2        |
| 4670   | 28/05/1993 | Gale          | Lee            | 0        |
| 4672   | 28/05/1993 | Gale          | Pinellas       | 0        |
| 4671   | 28/05/1993 | Gale          | Sarasota       | 0        |
| 4673   | 29/05/1993 | Gale          | Collier        | 0        |

The state of Florida is anecdotaly affected by a distinct rainy season, from the 15th of May to the 15th of October every year [22]. This required an examination of how weather-related disaster events were temporarily distributed over the course of each year. The temporal distribution was summarized using an extended version of Figure 2, where periods affected by rain-related disasters have been marked with vertical black bars. Relevant disaster events appear to have been concentrated between May and October of each year. This confirmed the need to focus on rainy season data, rather than data from entire calendar years. The seasonal concentration was particularly marked from the year 1990, perhaps resulting from the way disasters were being registered at that time. This observation led to a specific, decade-long sample of data from 1990 to 1999.

![Figure 2](image.png)

**Figure 2.** Seasonal concentration of rain-related disasters in Florida, USA.

To ensure that disasters could realistically cascade from one county location to another, a single Water District was selected instead of working with data for the entire State of Florida. The Southwest (SW) Florida Water Management District, shown in Figure 3, was selected because it experienced the highest number of seasonal flooding disasters \( (n = 104) \) between 1990 to 1999. The other four water management districts respectively experienced 29, 17, 60, and 44 floods over the same periods.

Data from SW Florida data was then converted to daily sequence of prevalent disaster types. Electric storms accounted for nearly all storm events and were used as a proxy for all storm disasters. Prevalent disaster types were selected by the number of counties affected, and then by time of onset. For example, where one day marked two counties affected by gales and only one by flooding, gale became the prevalent event. Where the same day was marked one county affected by a gale and then another affected by flooding, gale still became the prevalent event, because the gale event started first.

To avoid systematically underestimating future likelihoods, the current analysis of base rate information did not focus on the frequency of complete cascades. This would have amounted to a probability of close to nil, even for a theoretically common cascade between gale, storm and flooding disasters. Instead, the current analysis focused on the frequency of transitions between one component disaster state and another. The frequencies of these transitions were then used to calculate the basic probability of the same transitions combining into a more complex, flooding-related cascade.

Data was then re-analyzed to incorporate a trigger threshold of gale disaster affecting at least three counties. The incorporation of this threshold was roughly based on the model of disaster spreading in
networks, by Buzna, Peters and Helbing [23]. In climatic terms, it accounted for how extended cold fronts move larger quantities of atmospheric water from oceanic areas to land. The generic hydrologic model [19] used for the current research and summarized in Figure 1 accounts for these extended fronts, and the way they transfer atmospheric water, in terms of pressure changes.

![Figure 3. SW Florida water management district as at 1990](image)

This threshold for gale events affecting at least 3 SW Water Management District counties was applied by excluding all gale events affecting < 3 counties. The prevalence of sub-threshold gales was replaced by either storms or floods occurring on the same days or, in the absence of other relevant disasters, by a null state.

This approach to extracting and structuring the Florida data resulted in a continuous sequence of prevalent gale(o), storm, flooding or null disaster events (e.g. gale (o), gale (o), flood, flood) for each calendar day of the 1990-1999 rainy seasons. The analysis of this sequence was based on procedures outlined by E and Yang [24]. This involved using the below formula in Matlab software (version R2014a), to produce an initial transition matrix, where \( i \) is the prevalent disaster state for a given day and \( A_{ij} \) is the frequency of transition to the subsequent daily state, \( j \). Constitutive fractions were then adjusted, to exclude: numerator values for transitions leading from the end of each rainy season to the start of the following one and; denominator values for the last day of each year in the period 1990-1998.

\[
\hat{a}_{ij} = \frac{A_{ij}}{\sum_{j=1}^{N} A_{ij}}, \quad i = 1,2,\ldots,N, \quad j = 1,2,\ldots,N
\]  

(1)

Transition values from the resulting matrix were transposed onto the network model of potential pathways shown in Figure 4. The cumulative probability of all pathways was then calculated using the following formula, where \( P_{gf} \) is the probability of a transition from gale (o) to storm, \( P_{gg} \) is from gale to gale (o), \( P_{gf} \) is from gale to flood, \( P_{gs} \) is from gale (o) to gale (o), \( P_{g} \) is from storm to flood, and \( P_{ff} \) is from flood to flood.

\[
P_{gf} = \sum \left( \frac{P_{gs} \times P_{sf}}{P_{gg} \times P_{gf}} \right)
\]  

(2)
4. Results

Table 2 shows the overall transition matrix for SW Florida gale (ø), storm and flood events during the rainy seasons of 1990-1999. Figure 5 shows how values from this matrix were transposed onto the network model of potential pathways introduced in Figure 4.

![Network model with transition values.](image)

**Figure 5.** Network model with transition values.

**Table 2.** Overall transition matrix.

| From   | Gale (ø) | Storm | Flood | Null |
|--------|----------|-------|-------|------|
| Gale (ø) | 0.1111   | 0.0667 | 0.0889 | 0.7333 |
| Storm   | 0.0833   | 0.1917 | 0.0250 | 0.7000 |
| Flood   | 0.0000   | 0.1136 | 0.3864 | 0.5000 |
| Null    | 0.0227   | 0.0674 | 0.0151 | 0.8948 |
The cumulative probability calculation, outlined at the end of Section 3, resulted in a 0.05 probability of cascading flood-related disasters following each threshold gale, when rounded to a standard two decimal places. The corresponding Desinventar database only included two years of Florida data following the original 1990-1999 timeframe. This did not allow for validating the calculated probability against actual occurrences over the longer term. A rudimentary test for reliability was performed, by applying the current methods to rainy season data from the year 2000. This result also rounded up to a probability of 0.05, and was only 0.0003 less than the original 1990-1999 probability.

5. Conclusion
Cascading disasters present complex challenges for affected populations because of their rapidly compounding impacts. Cascading combinations of disaster events also pose a persistent challenge for disaster risk analysis. The very low frequency of particularly notorious cascades, such as the cascade triggered by the Great East Japan Earthquake, makes it very difficult to incorporate base-rate data for calculating future probabilities [6]. This is problematic because future likelihoods based on recent events alone are likely to be over-estimated [13, 15, 16, 17, 18].

The current research therefore aimed to calculate a simple base-rate probability for a relatively common type of cascading disaster, using a large set of disaster frequency data. It was hypothesized that the calculated probability of cascading, flood-related disasters would reach an intuitively relevant level of 0.05 per trigger event (widespread gales). This hypothesis was supported by a resulting simple probability of approximately 0.05. The methods used to generate this probability were found to be reliable when applied to data from the year 2000. The current results show that it is possible to arrive at an intuitively relevant, base-rate probability for developing cascading disaster models and scenarios. This was achieved by using generic, open access, data which was analyzed using a relatively simple approach to complex environmental dynamics and trigger thresholds. As outlined, the current data was limited to a broad register of which counties were affected by which disasters, on which day. This, and other limitations of the current research, can be overcome through further research into cascading disaster probabilities.

5.1 Implications for further research
As outlined, the Desinventar data set for Florida State only included start dates, rather than start times. This restricted the current research to a discrete analysis, of transitions occurring from one day to the next. Many climatic transitions nonetheless occur within the same day or even the same hour. As exemplified by flash flooding, nature’s hydrologic cycle often progresses much more rapidly than the arbitrary distinctions of generic disaster registers, i.e. between one day and another. This means that many natural transitions are effectively excluded by the discrete structure of common data sources.

Further efforts to build probabilistic models of cascading disasters will benefit from continuous approaches to Markovian transitions and chain modelling, using data collected at more regular intervals. This more inclusive approach to disaster-related data is likely to identify base rate frequencies in excess of 0.05 cascades per trigger. Where relevant data concerns climatic components such as wind-speed, rain-fall, and river flow, localized thresholds will need to be set, to determine what constitutes a disastrous state for each variable. This is in addition to thresholds that need to be set for the initial, trigger state.

As at the time of finalizing the current paper, critical infrastructure failure had become a more definitive aspect of cascading disasters [25]. However, the current paper was limited to a series of simple, climatic transitions that appeared to flow from one type of natural hazard event to another. Future research can also define transitions in terms of “striking”, “undermining”, “compounding”, and “blocking”, which led to infrastructure impacts outlined by Kumasaki et al. [7] (p. 1438). These and other transition types can help account for geographic shifts, from events progressing from one location to another. Accounting for shifts in space and scale will help provide more valuable long and short-term predictions to the agencies tasked with managing cascading disaster risk in terms of particular geographic zones.

The current paper outlines the viability and relevance of base-rate calculations for some cascading disasters. However, this does not dispense with the value of expert ratings. As demonstrated in prior
research [15, 16, 17, 18], expert combinations of information concerning both the long and short-term will lead to even more balanced, accurate and useful, predictions. The current authors remain committed to this dually integrated approach, towards better modelling cascading disaster risk.

6. References
[1] Cutter S L 2003 The science of vulnerability and the vulnerability of science Ann. Assoc. Am. Geogr. 93
[2] Cutter S L 2018 Compound, cascading, or complex disasters: What's in a name Environ.: Sci. Policy Sus. Dev. 60 16
[3] Zucccaro G, De Gregorio D and Leone M F 2018 Theoretical model for cascading effects analysis Int. J. Disaster Risk Red. 30 199
[4] McGee S, Frittman J, Ahn S J and Murray S 2016 Implications of cascading effects for the Hyogo framework Int. J. Disaster Resilience Built Env. 7 144
[5] Krausmann E and Cruz A M 2013 Impact of the 11 March 2011, Great East Japan Earthquake and Tsunami on the chemical industry Nat. Hazards 67 811
[6] The United Nations 2016 Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction (Geneva: The United Nations)
[7] Kumasaki M, King M, Arai M and Yang L 2016 Anatomy of cascading natural disasters in Japan: main modes and linkages Nat. Hazards 80 1425
[8] Huggins T, Hill S, Peace R and Johnston D 2018 Extending ecological rationality: Catching the high balls of disaster management Proc. ISCRAM Asia Pacific ed K Stock (Palmerston North, New Zealand: Massey University) 2018 295–309
[9] Ramirez de la Huerga M, Bañuls Silveira V A and Turoff M 2015 A CIA-ISM scenario for analyzing complex cascading effects in operational risk management Eng. Appl. Artif. Intell. 46 289
[10] Kahnemann D and Tversky A 1972 Subjective probability: A judgement of representativeness Cogn. Psychol. 3 430
[11] Tversky A and Kahnemann D 1973 Availability: A heuristic for judging frequency and probability Cogn. Psychol. 5 207
[12] Bar-Hillel M 1984 Representativeness and fallacies of probability judgment Acta Psychol. 55 91
[13] Hertwig R, Barron G, Weber W U and Erev I 2004 Decisions From experience and the effect of rare events in risky choice Psychological Science 15 534
[14] Koehler J J 2010 The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges Behav. and Brain Sci. 19
[15] Christensen- Szalanksi J J J and Beach L R 1995 Experience and the base-rate fallacy Org. Behav. and Hum. Perform. 29 270
[16] Balla J, Iansek R and Elstein A 1985 Bayesian diagnosis in presence of existing disease The Lancet 325 326
[17] Duthie B and Vincent K R 1986 Diagnostic hit rates of high point codes for the diagnostic inventory of personality and symptoms using random assignment, base rates, and probability scales J. Clin. Psychol. 42 612
[18] Gigerenzer G and Hoffrage U 1995 How to improve Bayesian reasoning without instruction: Frequency formats Psychol. Rev. 102 684
[19] Van Brahama J 2014 Hydrologic cycle Encyclopedia of Natural Resources (vol 2) ed Y Wang (London: CRC Press) pp 728–730
[20] Abdullah J, Kim J and Julien P Y 2014 Hydrologic modeling of extreme events Encyclopedia of Natural Resources (vol 2) ed Y Wang (London: CRC Press) pp 731-740
[21] UNISDR 2019 Desinventar UNISDR www.desinventar.net
[22] Miller K 2018 Rainy season start, end dates now fixed Palm Beach Post http://weatherplus.blog.palmbeachpost.com/2018/05/03/breaking-rainy-season-start-end-dates-now-fixed/
[23] Buzna L, Peters K and Helbing D 2006 Modelling the dynamics of disaster spreading in networks Physica A 363 132
[24] E F and Yang L 2018 Cascading natural disasters and prediction (poster) 2018 ISCRAM Asia Pacific Conference

[25] Alexander D and Pescaroli G 2019 What are cascading disasters? (London: UCL Press) p 7

Acknowledgments
The authors gratefully acknowledge that this research was funded by the National Natural Science Foundation of China, Project No. 71771113, and by the National Key Research and Development Program of China, Project No. 2018YFC0807000.