Characterizing Disaster Resistance and Recovery using Outlier Detection

Christopher W. Zobel
Virginia Tech
czobel@vt.edu

Steven A. Melnyk
Michigan State University
melnyk@msu.edu

Stanley E. Griffis
Michigan State University
griffis1@msu.edu

John R. Macdonald
Michigan State University
johnmac@msu.edu

ABSTRACT

Most definitions of disaster resilience incorporate both the capacity to resist the initial impact of a disaster and the ability to recover after it occurs. Being able to characterize and analyze resilient behavior can lead to improved understanding not only of the capabilities of a given system, but also of the effectiveness of different strategies for improving its resiliency. This paper presents an approach for quantifying the transient behavior resulting from a disaster event in a way that allows researchers to not only describe the transient response but also assess the impact of various factors (both main and interaction effects) on this response. This new approach combines simulation modeling, time series analysis, and statistical outlier detection to differentiate between disaster resistance and disaster recovery. Following the introduction of the approach, the paper provides a preliminary look at its relationship to the existing concept of predicted disaster resilience.

Keywords

Simulation, Time Series Analysis, Outlier Detection, Predicted Resilience.

INTRODUCTION

Disaster events can impact many different types of socio-technical systems. From businesses (Melnyk and Piper, 1981) and their extended supply chains (Cavinato, 2004), to civil infrastructure systems (Zobel and Khansa, In Press-a) and networks of information systems (Zobel and Khansa, In Press-b), many aspects of our modern society are subject to the effects of both natural and man-made disasters. The notion of *disaster resilience* (Rose et al., 2007; Shinozuka et al., 2004; Tierney and Bruneau, 2007) is often used to describe the capacity of such systems both to withstand the initial impact of disaster events and to recover from them in a timely manner. Measuring such resilience can thus provide an indication not only of the inherent ability of a system to resist the effects of a disaster, but also of the role that mitigation or response policies can play in strengthening it under different circumstances.

This paper discusses a new time-series based approach to measuring and analyzing system resilience. The approach combines simulation modeling, time series analysis, and statistical outlier detection to statistically capture the transient behavior reflected in the response curve of a system impacted by a disaster event. It enables researchers to identify, quantitatively and with statistical confidence, critical characteristics such as the time at which the disruption is first felt, the time at which it reaches its largest effect, and the magnitude of the effect on the time series response at each time interval. These parameter values can be used not only to characterize the profile of the time series response curve, but also to support a variety of statistical analyses on the transient response, including ANOVA studies and regression analysis.

We begin our discussion with an overview of the approach, and then discuss its use in calculating a quantitative measure of resilience. This is followed by an examination of its relationship with the existing notion of predicted resilience (Zobel, 2010), and a brief look at some of the time series characteristics that illustrate the difference between a disaster and a simple disruption. Finally, we complete the discussion by considering appropriate next steps towards further developing and applying the new approach in this context.

TIME SERIES OUTLIER DETECTION APPROACH

We start this section by noting that many aspects of the new time-series based approach for analyzing resilience procedure are also described in detail in (Melnyk et al., Under Review), although with a focus on supply chain disruptions. The following discussion extends this previous work by positioning it in the context of disasters and disaster operations management. It also explores a number of characteristics of the approach in greater detail.

The foundation of the new approach is the statistical technique of outlier detection (OD). Outlier detection involves using a general set of time series analysis procedures to detect, assess, and display the statistically significant impact of transients (disruptions) in time series data (Liu, 2005; Tsay, 1988). Pankratz (1991) identified several reasons for using outlier detection and adjustment in time series analysis, including a better overall understanding of the time series under study, given its ability to detect their impacts and durations. Because disruptive events can alter the structure of the statistics used for model estimation, uncovering outliers should help pinpoint the effects of this alteration, thus simplifying the structure of the model used. Accordingly, these benefits can lead to both improved intervention analysis and better forecasting performance.

OD helps to accomplish these objectives by first assessing the time series prior to the disaster. This baseline time series is then compared to the post-disruption time series, allowing identification of the presence of statistical outliers and of the type of transient behavior that each one indicates. In particular, OD recognizes four types of outliers, namely: Additive Outliers (AO), Innovational Outliers (IO), Level Shifts (LS), and Temporary Shifts (TS) (Liu, 2009:Chapter 7). Additive Outliers (AO) affect a series for a single time period and are sometimes called pulses. Innovational Outliers (IO) represent events whose effect is propagated across to the time series model and that affect all values observed after their occurrence, typically due to an external cause. Level Shift (LS) events cause the mean level of a series to shift (permanently) to a new value. Finally, Temporary Shift (TS) events, or transient changes, are events that, after an initial impact, exhibit decaying effects according to some dampening factor, δ (Tsay, 1988). Of these four types of outliers, the most straightforward to evaluate is that of a level shift (LS); it is also the type that we expect to deal with when facing a disaster. Consequently, we focus on this type of outlier.

Capturing a sequence of LS outliers in a time series allows one to generate values for parameters such as the *Time of Onset* (TO) of a disruption and the associated *Response* (of the time series) *at Onset* (RO), as well as the *Time of Recovery* (TR) and the *Response at Recovery* (RR) (Melnyk et al., Under Review). Furthermore, it supports capturing values for parameters that reflect the transient behavior of the system, such as the *Time of Climax* (TC) and the *Response at Climax* (RC), along with the *Turning Point* (TP) and *Response at Turning Point* (RP), which indicate the point at which recovery begins (Melnyk et al., Under Review) (see Figure 1). In each case, the time-based measures (TO, TC, TP, TR) are taken *relative* to the time that the disruption took place (TD).

Capturing the parameter values associated with each of a number of different simulated scenarios provides the opportunity for a decision maker to compare the different response curves *analytically*, rather than simply qualitatively. This then supports much more effective decision-making by providing the ability to answer such questions as: How much loss is there? How long did it take to recover? How long did it take until we started to recover? To what level of functionality did we recover?

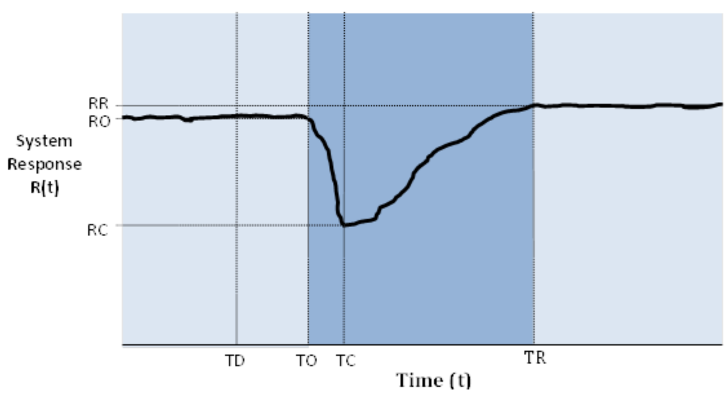


Figure 1. Time Series Signature Critical Points

Analyzing the area above the curve

At the heart of this study is the concept of a disaster, which can be defined as a "serious disruption of the functioning of society, posing a significant, widespread threat to human life, health, property, or the environment, whether caused by accident, nature or human activity, and whether developing suddenly or as a result of complex, long-term processes" (ISDR, 2004:3). As noted by van Wassenhove (2006), disasters can be differentiated by the speed of onset – sudden (earthquake, tornado, tsunami) or slow (famine, drought, poverty). The effects of a disaster on a system can also be characterized by using the area above the response curve to generate a measure of the system's resilience. This idea was originally introduced by Bruneau et al. (2003), and it has been adapted and extended by a number of researchers (Cimellaro et al., 2006; Zobel, 2010; Chang and Shinozuka, 2004; Shinozuka et al., 2004; Miles and Chang, 2011; Zobel, 2011). The measured area above the curve (with respect to the response at the onset of the disruption) represents the total loss experienced by the system, compared to what might be expected in the absence of the disruption.

Zobel (2010) defines a measure of *predicted resilience* that estimates the remaining area *below* the curve, building on Bruneau's (2003) concept of the disaster resilience triangle. The predicted resilience measure allows the visualization of the tradeoffs between the amount of loss and the time to recovery, but it is limited to the case of sudden-onset disasters in which the full impact of the event is felt nearly instantaneously. The assumption of sudden-onset behavior effectively allows the area above the curve to be approximated by a right triangle, and thus simplifies its calculation and representation (Bruneau et al., 2003; Zobel, 2010). In the case of a rapidly progressing disaster such as an earthquake, this is often a reasonable approximation; however, many disasters such as hurricanes are more slow-onset in nature, in that the loss felt by the system is accumulated over some initial period of time.

With the prevalence of such slow-onset behavior in mind, Zobel and Khansa (In Press-b) developed a related approach to approximating the area above a response curve that applies to both sudden-onset and slow-onset events. Defined in the context of a multi-event information system attack, the new approach was also extended to support the use of simulation for characterizing probabilistic behavior. Along with related work on sudden-onset natural disasters that was developed in (Zobel and Khansa, In Press-a), this effort helped to establish the validity of taking a piecewise approach to calculating the area above the curve.

The new time series outlier-based approach introduced above is similar to the work in (Zobel and Khansa, In Press-b), both in its ability to model slow-onset events and in its applicability to simulation-based analysis. In particular, the output parameters generated by the approach can be used directly to create a piecewise approximation of the area under the curve, as in (Zobel and Khansa, In Press-b). Because these parameter values are derived statistically, however, based upon the time series analysis, they can provide a more accurate approximation of the shape of the response curve than what might be available through other estimation techniques.

In order to reflect the importance of being able to model slow-onset events, we choose to define total resilience in the context of the time series outlier approach as being composed of two primary components with a third optional component. The first primary component is the area above the curve associated with the ability of the system to resist the initial impact of the disruptive event (i.e., the system's capacity for *resistance*), and the second is the area associated with the ability of the system to recover from that event (i.e., its capacity for *recovery*) (See Figure 2). The third component, which may not always be present, provides for the possibility that the system may enter a transitory steady-state period at some point during the disruption, during which time it is neither incurring more loss nor commencing recovery from the loss. As illustrated by Figure 2, the size of each specified sub-area is impacted by the maximum amount of loss felt by the system, as well as by the duration of each phase. The area associated with each of the two primary phases also depends on the rate at which the system state is actively changing.

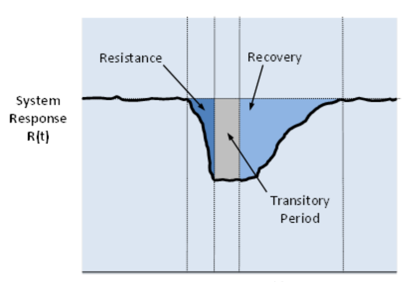
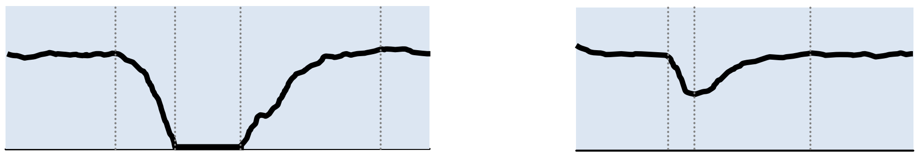


Figure 2. Resilience components



(a) Total loss **(b) Partial loss**

Figure 3. Potential loss scenarios

It is important to note that the general idea of calculating the area above the curve can also be extended to time series that model multi-event disasters, such as one representing an earthquake with aftershocks or one that represents a distributed denial-of-service attack on an information system. In such scenarios, there may be several different periods of resistance or recovery, as well as of transitory steady-state behavior, between the time when the initial disaster event is felt and when full recovery is achieved.

Disaster-specific characteristics

When focusing on a system's response to a disaster, it is natural to consider both total loss scenarios (Figure 3a) and partial loss scenarios (Figure 3b). In the former situation all functionality has been lost for a period of time, whereas in the latter, the system quickly recovers and operations are only partially affected. Because a disaster is a major disruption that causes "widespread human, material, or environmental losses which exceed the ability of [the] affected society to cope using only its own resources" (UNISDR, 2007), one would expect the total loss scenario to be observed more frequently in the case of a disaster than in the case of a more "normal" (and perhaps more expected) disruption.

This does not mean, however, that all processes affected by a given disaster will necessarily follow only the pattern of Figure 3a, since the widespread nature of a disaster implies that it will simultaneously affect a number of different systems in various ways and to varying degrees. It is much more likely that a given disaster will lead to a range of different time series profiles across all affected subsystems, depending on the nature of the disaster event and the processes that are affected by it.

Another characteristic that may be used to differentiate the time series profile of a disaster event from that of a less extreme disruption is the relative size of the RR (response at recovery) value compared to that of the RO (response at onset) value. Due to the significant impacts of a disaster, it simply may not be possible for a system to regain its full functionality again, given the resources available. For example, a major landslide may destroy a rail line that served as a major access route to a warehouse facility, reducing the total throughput available in its supply network even after the facility itself has been repaired. In this case, one might observe a steady-state value for RR that is less than the value of RO.

The opposite may also be true, however, in that the impact of a disaster may actually provide the opportunity to improve the capabilities of a system beyond what was possible before the disaster occurred. This may involve specific actions such as rebuilding a manufacturing facility to make it larger and more modernized so that it can handle more production, or more general policies such as replacing residences in a flooded area with public recreation facilities, in order to improve a flood-prone city's quality of life. In either case, the measured system response after the disaster (RR) would ultimately exceed its value as measured before the disaster occurred (RO).

DISCUSSION AND CONCLUSIONS

The time series-based approach discussed above holds a great deal of promise for helping decision makers to better understand their processes' capacity for resilience. By combining appropriate simulation models with the available empirical data, one can begin to identify different policies for dealing with disasters, and then evaluate the impact of these policies on the system's capabilities for both disaster resistance and recovery. This can be done by quantitatively comparing the values of a number of different calculated parameters as well as by calculating the area above the time series response curve, and it is made possible through the use of statistical outlier detection.

As with any analytic approach intended to help with studying disaster resilience and the effectiveness of such policies, the new approach has its constraints and limitations. Among these is the general lack of significant empirical data for describing time series responses in a disaster environment. For some reasonably well-studied processes, such as inventory levels associated with a manufacturing-based supply chain, it is possible to simulate appropriate data and then to study the statistical impacts of different policies on the system's behavior (Melnyk et al., Under Review). For other less concrete processes, however, such as a community's "quality of life," it is often necessary to use indicator variables to represent the process, and it may be possible to measure such variables only infrequently (Birkmann, 2006; Cutter et al., 2003). Consequently, there is still much work to be done not only on continuing to refine the technique in general, but also on adapting the time series outlier approach so that it is also applicable to some of the less technical dimensions of disaster resilience.

REFERENCES

Birkmann, J. (2006) Indicators and criteria for measuring vulnerability: Theoretical bases and requirements. In: *Measuring Vulnerability to Natural Hazards*, pp. 55-77.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A. and von Winterfeldt, D. (2003) A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra,* **19,** 733-752.

Cavinato, J. L. (2004) Supply chain logistics risks: from the back room to the board room. *International Journal of Physical Distribution & Logistics Management*, **34,** 383-387.

Chang, S. E. and Shinozuka, M. (2004) Measuring improvements in the disaster resilience of communities. *Earthquake Spectra,* **20,** 739-755.

Cimellaro, G., Reinhorn, A. and Bruneau, M. (2006) Quantification of seismic resilience. In: *Proceedings of the 8th U.S. National Conference on Earthquake Engineering*, pp. 18-22. San Francisco, CA.

Cutter, S. L., Boruff, B. J. and Shirley, W. L. (2003) Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, **84,** 242-261.

ISDR (2004) *Living with risk: a global review of disaster reduction initiatives,* United Nations Publications, Geneva.

Liu, L.-M. (2009) *Time Series Analysis and Forecasting,* Scientific Computing Associates, River Forest, Illinois.

Liu, L. M. (2005) *Time series analysis and forecasting,* Scientific computing associates.

Melnyk, S. A. and Piper, C. J. (1981) Implementation of Material Requirements Planning: safety lead times. *International Journal of Operations & Production Management,* **2,** 52-61.

Melnyk, S. A., Zobel, C. W., Griffis, S. E. and Macdonald, J. R. (Under Review) An Outlier-based Approach for Assessing Supply Chain Disturbances. *IEEE Transactions on Engineering Management*.

Miles, S. and Chang, S. E. (2011) ResiliUS: A community based disaster resilience model. *Cartography and Geographic Information Science,* **38,** 5-21.

Pankratz, A. (1991) *Forecasting with dynamic regression models,* John Wiley & Sons, New York.

Rose, A., Adosu, G. and Liao, S. Y. (2007) Business interruption impacts on the electric power system resilience to a total blackout of a terrorist attack of Los Angeles: Customer resilience to a total blackout. *Risk Analysis,* **27,** 513-531.

Shinozuka, M., Chang, S. E., Cheng, T.-C., Feng, M., O'Rourke, T. D., Saadeghvaziri, M. A., Dong, X., Jin, X., Wang, Y. and Shi, P. (2004) Resilience of Integrated Power and Water Systems. In: *MCEER Research Progress and Accomplishments: 2003-2004,* (Ed, MCEER), pp. pp. 65-86. Buffalo, NY.

Tierney, K. and Bruneau, M. (2007) Conceptualizing and Measuring Resilience: A Key to Disaster Loss Reduction. In: *TR News*, pp. 14-17.

Tsay, R. S. (1988) Outliers, level shifts, and variance changes in time series. *Journal of Forecasting,* **7,** 1-20.

UNISDR (2007) Terminology (Disaster).

Van Wassenhove, L. (2006) Blackett Memorial Lecture - Humanitarian aid logistics: supply chain management in high gear. *Journal of the Operational Research Society,* **57,** 475-489.

Zobel, C. W. (2010) Comparative visualization of predicted disaster resilience. In: *Proceedings of the 7th International ISCRAM Conference*. Seattle, WA.

Zobel, C. W. (2011) Representing perceived tradeoffs in defining disaster resilience. *Decision Support Systems,* **50,** 394-403.

Zobel, C. W. and Khansa, L. Z. (In Press-a) Characterizing Multi-event Disaster Resilience. *Computers and Operations Research - Special Issue on MCDM in Emergency Management*.

Zobel, C. W. and Khansa, L. Z. (In Press-b) Quantifying Cyberinfrastructure Resilience against Multi-event Attacks. *Decision Sciences.*