Revisiting the Predictability of the Haicheng and Tangshan Earthquakes

Didier Sornette 1,2,3,*, Euan Mearns 1 and Spencer Wheatley 1

1 ETH Zurich, Chair of Entrepreneurial Risks, D-MTEC, Scheuchzerstrasse 7, CH-8092 Zurich, Switzerland; emearns@ethz.ch (E.M.); spencerwheatley@gmail.com (S.W.)
2 Institute of Risk Analysis, Prediction and Management, Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen 518055, China
3 Tokyo Tech World Research Hub Initiative (WRHI), Institute of Innovative Research Tokyo, Institute of Technology, Yokohama 4259, Japan
* Correspondence: dsornette@ethz.ch

Abstract: We analyze a set of precursory data measured before but compiled in retrospect of the $M_{S}7.5$ Haicheng earthquake in February 1975 and the $M_{S}7.6–7.8$ Tangshan earthquake in July 1976. We propose a robust and simple coarse-graining method that aggregates and counts how all the anomalies together (levelling, geomagnetism, soil resistivity, earth currents, gravity, earth stress, well water radon, well water level) develop as a function of time. We demonstrate strong evidence for the existence of an acceleration of the number of anomalies leading up to the major Haicheng and Tangshan earthquakes. In particular for the Tangshan earthquake, the frequency of occurrence of anomalies is found to be well described by the log-periodic power law singularity (LPPLS) model, previously proposed for the prediction of engineering failures and later adapted to the prediction of financial crashes. Using a mock real-time prediction experiment and simulation study, based on this methodology of monitoring accelerated rates of physical anomalies measured at the surface, we show the potential for an early warning system with a lead time of a few days.

Keywords: Tangshan; earthquake; LPPLS; anomalies; acceleration; prediction

Highlights:
- Archived records of physical precursors to the Tangshan earthquake follow a log-periodic power law singularity (LPPLS) model.
- Had it been possible to assemble these abundant records into a central database ahead of the Tangshan earthquake, it may have been able to be predicted.
- Physical and observational surface records trump seismic records in earthquake prediction.

1. Introduction

Earthquakes are vivid expressions of the forces of plate tectonics and have contributed greatly to the geomorphology and organization of rock structures within the Earth’s crust and upper mantle. For physicists, mathematicians, and geologists, in addition to seismologists, earthquakes are also fascinating due to their remarkable power-law statistics:

(i) the (exponential) Gutenberg–Richter (GR) distribution of earthquake magnitudes, which translates into a power law distribution of seismic moments, which is valid over several decades, i.e., over a span of scales larger than most known power laws in physical and social sciences;

(ii) the Omori law (the rate of aftershocks decays as $\sim \frac{1}{(t-t_m)^p}$ after a main shock that occurred at a time $t_m$);

(iii) a spatial Green function quantifying the power-law decay of the influence of the main shock as a function of the distance to its aftershocks;
(iv) a fertility law of the average number of aftershocks triggered as a function of the magnitude of the main shock, which translates into a power law fertility as a function of the seismic moment or energy of the earthquake;
(v) power-law distributions of the lengths of the faults on which earthquakes occur;
(vi) fractal, multifractal, or hierarchical scaling of the set of earthquake epicenters and fault networks, etc.

For physicists and modelers more generally, these laws suggest the existence of underlying physical mechanisms associated with some kind of criticality in the sense of phase transitions and field theory with zero mass. For mathematicians, these laws reflect the symmetry of scale invariance [1]. Scale invariance refers to the invariance of an object (material or mathematical) over changes of scales of observation. A convenient mathematical means to express scale invariance symmetry for a scale-dependent observable \( O_L \) is to require that, when scales are changed by a factor \( \lambda \), the observable is changed by a factor \( \mu(\lambda) \) such that:

\[
O_L = \mu(\lambda) O_{\lambda L}
\]

It is straightforward to verify that the solution of Equation (1) is a power law \( O_L = C L^\alpha \), with \( \alpha = -\frac{\ln \mu}{\ln \lambda} \). Note that, in the log-space (\( \ln L \), \( \ln O_L \)), scale invariance becomes simply a translational invariance. Power laws are the hallmark of scale invariance as the ratio \( O_{\lambda L}/O_L \) does not depend on \( L \), i.e., the relative value of the observable at two different scales only depends on the ratio \( \lambda \) of the two scales.

Remarkably, one can break partially this (continuous) scale invariance into a “discrete” scale invariance, in which Expression (1) only holds for specific values of \( \lambda \) that are integer powers of a preferred scaling ratio (see [2] for an extensive review). The goal of the present paper is to revisit previous suggestions that such discrete scale invariance could be relevant to describe the dynamics of earthquake activity. Specifically, we complement previous investigations by analyzing in detail the dynamics of precursors of an infamously destructive earthquake, the 1976 Tangshan earthquake. This is offered as an illustration of how insights emerging from deep symmetry considerations that seem, a priori, to be far from concrete applications can actually be useful at a practical and operational level.

The organization of this manuscript is as follows. In the next section, we provide some general background of the Tangshan earthquake, in particular explaining why it plays a special role in the history of seismicity, which is also deeply intertwined with societal and political developments, and in the science of earthquake forecasting and predictions. Section 3 then presents relevant information on the different anomaly types and the spatial-temporal organization of the hypothesized precursors to the 1975 Haicheng and 1976 Tangshan earthquakes. Section 4 presents our analysis of the accelerated aggregate rate of these precursors before the Haicheng and Tangshan earthquakes, including an ex-post forecasting exercise and statistical tests of the predictability of the Tangshan earthquake. Section 5 concludes by stressing the role of quantitative data aggregation and skilled display of technical information for diagnostic impending ruptures.

### 2. Background on the Tangshan Earthquake

In the period 1966 to 1976, the Chinese engaged in the biggest earthquake monitoring program ever conceived. It began with the Xingtai earthquake sequence, of which the two most powerful events occurred on (UTC) 7th March (M\(_W\) = 6.5) and 22 March 1966 (M\(_W\) = 6.8). The program more or less ended with the sequence of Tangshan earthquakes that began on 28 July 1976 and ended on 27 November 1977 after 30 M > 5 shocks had occurred (Tables 2 and 6 in [3]). Following Xingtai, Chinese premier Enlai Zhou was concerned by the scale of casualties and instructed Siguang Li, China’s most senior geologist, to establish a monitoring program with the aim of predicting earthquakes and offering some advance warning to the people.

Below, we present a novel quantitative analysis of precursory data of two earthquakes, the M\(_S\)7.5 Haicheng earthquake in February 1975 and the M\(_S\)7.6–7.8 Tangshan earthquake in July 1976 in China. One of our motivations is to provide novel insights to the question of
whether there was a degree of predictability for these two events, had the precursory data measured beforehand been compiled into a central, regional database. The technology for such data aggregation did not exist in China in the mid-1970s. Had such data gathering technology existed, could the Tangshan Earthquake have been predicted and hundreds of thousands of lives spared?

The style of the investigation is in the form of two case studies, which is complementary to the presently more prevalent approaches using a sufficient amount of high quality and reliable data enabling statistical testing. In the latter style of scientific works, one requires the existence of similar data at the time of the earthquakes from other regions, and similar data from the epicentral regions at times when there were no earthquakes, to demonstrate that the alleged precursors only occur before large earthquakes and not at other places and/or times. This allows one to avoid artifacts and selection bias, and to quantify the rate of false positives and false negatives. However, there are cases, such as those studied here, for which easily accessible data from a control region does not exist and a large-scale statistical analysis of precursors over a large set of earthquakes is simply impossible. Indeed, the Chinese were familiar with the rigors of the scientific method and in the early 1970s established two test areas at Xinjiang and Shanxi where measurements of a range of physical phenomena were made. Unfortunately, nothing happened at these test sites and after a year or so they were abandoned.

Nevertheless, in such cases as those we report here, data exist and we take the attitude that it would be akin to “throwing the baby out with the bath water” to avoid attempting to extract insights from it. Moreover, the progress of science is also made by accumulating case studies, until the weight of evidence supports or falsifies hypotheses. Ultimately, the sequence of case studies should be complemented by systematic statistical tests using well-designed hypotheses forged on quantitative models, as in [4,5]. We are also aware that the present study may be viewed with skepticism by some scientists who seek to exclude non-seismic observational evidence from the earthquake forecasting debate and then go on to claim that reliable and accurate prediction of earthquakes is not possible at present (e.g., [6–8]). Our goal is more modest in documenting, in a novel manner, what we consider to be two interesting cases that can be useful in the future when combined with other cases and much large data sets.

Stories of strange phenomena preceding earthquakes were engrained in Chinese folklore, such as strange behavior of animals, turbid water in wells, and strange fogs and clouds. The idea was to combine a range of observations and simple measurements made by amateur groups and to combine these with more sophisticated measurements made by professional groups. In this way, China could capitalize on one of its main resources, i.e., people, but this would also engage the population in the prediction program, thus giving the people heightened awareness of earthquake risks.

At its peak, the program engaged 35,000 amateur groups located in schools, factories, and public buildings, and consequently a large and unique set of physical measurements of anomalies and human observations were acquired over the decade. Some of this data is published in graph form in post-mortem reports published after the Haicheng, Tangshan, and Songpan earthquakes [3,9,10], but accessing the raw data that lies under the charts is not straightforward. Amongst other things, the Chinese language presents a barrier to non-Chinese.

In 1981, Jian-zhong Zheng, a Chinese scientist previously affiliated with the Tokyo Institute of Technology (employed at the time of publication at the Institute of Geophysics, Chinese Academy of Sciences, Beijing), published a paper that described some of the anomalies recorded prior to the Haicheng (February 1975) and Tangshan (July 1976) earthquakes [11]. Published in Japanese, the paper has an English abstract and two English language appendices documenting anomalies from the Haicheng and Tangshan earthquakes respectively. One weakness stems from the fact that Zheng [11] does not provide a description of how anomalies are defined. We therefore attempt to verify Zheng’s tables where possible via comparison with other publicized anomaly data that is in the form...
of charts and maps. Nevertheless, we saw here an experimental opportunity to perform a mathematical analysis of the precursor data using the Finite Time Singularity model initially proposed for the prediction of engineering failures [12–17]. Our aim is to determine if these earthquakes could have been predicted based on acceleration in anomalies leading up to the climax.

As the next section illustrates, the precursors that were monitored in real time are very diverse, and exhibit strong heterogeneity in space and time. The large variability of their spatial distributions, times of occurrence, and amplitudes confused the Chinese seismologists in charge of the earthquake monitoring program. The guiding idea of the present study is to see if coarse-graining this wealth of noisy information might have brought some meaning to their development suggesting that, together, they might have been sufficient to provide useful predictions for decision makers to act upon. In other words, the key idea is to avoid being swamped by the details and the corresponding overwhelming variability and try to extract a robust signal. This strategy has been successful for the prediction of failures of engineering structures, which are also characterized by enormous variability of the recorded signals [12,16]. The coarse-graining strategy is inspired by the Renormalisation Group theory of critical phenomena [18,19].

Our approach is reminiscent of the accelerated moment release method [20–22], based on the hypothesized increase in seismicity prior to a major earthquake. We stress, however, the fundamental difference in the fact that the accelerated moment release method only uses seismicity, whereas our study is built on eight non-seismic variables (levelling, geomagnetism, soil resistivity, Earth currents, gravity, Earth stress, well water radon, well water level). The present work is thus more in the spirit of Johansen et al. [23,24] of aggregating non-seismic data for testing the existence of earthquake precursors, but with a much larger set of physical variables to aggregate upon.

3. Description of Precursors for the Haicheng and Tangshan Earthquakes

For the Haicheng earthquake, Zheng [11] reports six anomaly classes (see below) where the numbers in brackets give the number of reported anomalies per class:

- Geomagnetism (2)
- Resistivity (2)
- Earth currents (13)
- Gravity (2)
- Earth stress (8)
- Radon (16)

Of these, only three have a reasonable quantity of data, namely earth currents, earth stress and radon. There is not really sufficient information from Haicheng to conduct meaningful analysis and we shall therefore focus on Tangshan where more data are available:

1. Geodimeter (2)
2. Levelling (11)
3. Geomagnetism (11)
4. Resistivity (20)
5. Earth currents (15)
6. Gravity (14)
7. Earth stress (31)
8. Radon (29)
9. Well water level (21)

The two geodimeter readings fall outside of the time range of our analysis (1800 days) and are therefore excluded.

Zheng [11] reports precursor time in days, epicentral distance in kilometers, and the name of the observation station. This has enabled us to compile maps of anomaly distributions that we discuss shortly. First, we clarify what some of these anomalies mean.
Resistivity is measured between four steel electrodes hammered into the ground. An AC current is applied and the soil resistivity is computed.

Earth currents, also known as telluric currents, record spontaneous electric currents in the ground. Two steel electrodes are hammered into the ground, joined by a wire, and an ammeter records the current, if any.

Earth stress is measured by a special transducer planted in the soil and records changes in the stress condition of the soil.

Radon in China was measured in well water and not air. $^{222}$Rn concentrations are measured based on ionizing radioactivity using a scintillation counter. This is a relatively simple hand-held device enabling measurements to be made in amateur monitoring posts. The simplicity is derived from the fact that Rn is a gas that is separated from the water prior to analysis.

Most of the anomaly classes listed above may be viewed as trending anomalies, i.e., the value measured may vary with time. Ideally, these should be set against baseline values for which a baseline is established prior to earthquake conditions and then something happens to cause a change in value from that baseline. Once established, then observations may be made about the change in trend, e.g., jumps, spikes, and turning points. Zheng [11] sometimes reports more than one anomaly per station, for example Changli has four resistivity anomalies. Zheng did not publish graphs that would allow assessment of the judgements applied.

Qian et al. [25] published resistivity logs from the Tangshan area and this offers the opportunity to scrutinize Zheng’s recordings against reality (Figure 1). We find that each locality has a relatively flat baseline three years prior to the event, followed by a baseline drift towards lower values thereafter. In each case, Zheng’s recordings mark the beginning of baseline drift. Subsequent recordings sometimes mark spikes or turning points but it is not always obvious what observations are being recorded.

Figure 1. Resistivity logs from six stations in the Tangshan area as reported in Figure 3 of Qian et al. [25]. Each station has N-S and E-W orientated devices. Zheng [11] reports data for five of these stations. He does not report data for Mafang because there is no anomaly at that station. The blue arrows mark the anomaly times reported by Zheng ([11] in each case, Zheng’s data marks the first appearance of a trend on the logs.
One important observation to make from Figure 1 is clear seasonal variation in some of the localities, for example Changli and Baodi. This is perhaps not surprising since resistivity may respond to rainfall (soil moisture), temperature, and frozen ground. This caused us to temporarily doubt that resistivity, well water level, and radon concentrations were responding to earthquake conditions at all. All may have been responding to climatic conditions, and an uncommonly dry period did precede the Tangshan earthquake that would have caused well water to fall and radon concentrations to rise. However, when we plot the data on a map (see Figure 5 below and [26]), we find that these anomalies also align with structural elements, and so conclude that earthquake signals may be superimposed upon other effects such as seasonal weather variations.

One further question is to what extent the data of Zheng [11] is complete. This was discussed in his paper where he thought that it should be relatively so. We can confirm this is not the case. Zheng reports Earth current data from 13 measuring stations around Tangshan and presumably only from stations that registered an anomaly. In the State Seismological Bureau (SSB) post-mortem report, Jiang and Chen [3] publish a map showing Earth current anomaly stations and distinguish between those with an anomaly and those without. This map shows 41 measuring sites, 26 with anomalies and 15 without. This is roughly twice as many sites with anomalies as reported by Zheng [11]. However, the distributions of the anomalies from the two sources are very similar. Zheng’s localities appear on the SSB map.

We now proceed to examine the temporal and spatial distribution of the anomalies. Figure 2 shows the time of the first appearance of the anomalies on the y-axis and the time distribution of the anomalies as listed by Zheng [11] along the x-axis. Panel (a) plots all the data for Tangshan (Zheng [11] Appendix 2), and panel (b) plots a subset (see below). The order in the legend records the sequential order of first appearance, with ground water anomalies appearing first and Earth currents appearing last.

One notable feature from Figure 2a is the bi-modal distribution of water, radon, resistivity, gravity, and levelling anomalies. There is an earlier group that disappears at ~900 days to re-appear again at ~539 days, coincident with the time of the Haicheng earthquake. Chinese seismologists have told us that, following the Bohai earthquake of 1969, stress indicators appeared in the vicinity of both Yingkou (close to Haicheng) and Tangshan. This was called the “aftereffect anomaly field” (Chengmin Wang, personal communication), a term used to represent the appearance of stress in one area as a result of a previous earthquake. What the anomaly distribution may show is the reduction in this “aftereffect anomaly field” in Tangshan as it became focused on Haicheng. Then, after the Haicheng earthquake, the aftereffect anomaly field re-appeared in Tangshan.

For our analysis and on the maps below, we have set a filter at 540 days so that we only use anomaly data that is directly linked to the gestation of the Tangshan earthquake. Setting the filter at 540 days, we re-ordered the data in Figure 2b, thus providing a new order for the first appearance of anomalies.

Figure 3 shows the distribution of gravity and geomagnetic anomalies. There is a slight concentration of anomalies around Tangshan that may follow the fault lines as we reconstruct them [27]. Equally, these anomalies are spread all over the map. There is a tendency for these anomalies to follow the basin margin and for the geomagnetic anomalies to be distributed towards the east. The inset graph shows both of these anomaly classes have a weak inverse correlation between time and distance. In other words, as the earthquake approached, the anomalies became more distant. This observation confused the Chinese seismologists at the time and contributed to the failure to predict the Tangshan earthquake [28].
As seen in Figure 4, although there is a cluster of Earth stress anomalies on Tangshan, there is no clear focus on the city. If anything, there is a greater focus on Beijing. Once again, there is a tendency for these anomalies to focus on the basin margin and this may show that there was a change in stress between the basin and the basin margin. The inset graph of Figure 4 shows that Earth stress anomalies drew closer to the epicenter as the time of the earthquake approached.
Figure 3. Map showing the spatial distribution of gravity and geomagnetic anomalies. The yellow numbers give the anomaly time in days before the main shock of the Tangshan earthquake (28 July 1976). The stars show the locations of the three main shocks, Tangshan (28 July 1976) in the middle, Luanxian to the NW (also on 28 July 1976), and Ninghe to the SE (15 November 1976). The dashed lines show our interpretation of the Tangshan transfer fault system, where pink denotes a dextral strike-slip leg and dark blue denotes a normal fault leg [27]. Base image from Google Earth. The dappled beige color reflects the flat surface of the densely populated Bohai Bay rift basin. The dark green reflects the mountainous, forested, and dissected Yanshan Range. The boundary between the two marks the boundary of the Bohai Bay basin. The inset graph plots the distance as a function of time of the anomalies. The straight lines suggest a weak inverse correlation between time and distance.

Figure 4. Map showing the distribution of Earth stress and levelling anomalies. See caption to Figure 3 for explanation of map annotations.
Well water level, well water radon, and resistivity anomalies show a clear focus on the Beijing–Tianjin–Tangshan area but no clear focus on Tangshan itself (Figure 5). It is possible that weak lineaments in the distribution could be controlled by deep structure. The inset graph shows that near term radon anomalies appeared quite distant to the epicenter, >400 kms away. This contributed to confusion within the SSB.

Figure 5. Map showing the distribution of well water level, well water radon, and resistivity anomalies. See caption to Figure 3 for explanation of map annotations.

Earth currents are the only anomaly class to be closely associated with the Tangshan earthquake in time and space (Figure 6). There are three spatial outliers at Xianghe, Qinhuangdao, and Penzhuang. Earth current anomalies first appear 100 days before the earthquake, but at quite distant locations—Qinhuangdao, Baidaihe, and Penzhuang—but then move towards the epicenter.

Earth current anomalies also show good spatial and temporal relationships with the Haicheng and Songpan earthquakes and could have been used to predict the Tangshan earthquake. Unfortunately, at the time, the Chinese seismologists were unaware of this relationship.

Although the Earth currents provide good spatial resolution, questions remain about timing. The objective of this paper is to determine if a finite time singularity model based on all the data may be used to improve temporal resolution.
4. Accelerated Rates of Precursors before the Haicheng and Tangshan Earthquakes

4.1. Qualitative Analysis

The Haicheng (1975/2/4) and Tangshan (1976/7/28) earthquakes were separated in time and space by 540 days and 450 km. For each event, about 5 years of observation of anomalies are available leading up to the earthquake. For convenience, Table 1 summarizes the list and number of anomalies by type in the 5 year period prior to the Tangshan earthquake, which are already discussed at the beginning of Section 3. Such information is presently not available for Haicheng.

Table 1. For Tangshan earthquake, the count of anomaly events by type, for those types with more than two anomaly events. Types with a single anomaly event include: “energy of minor earthquake”, “hypocentral migration mechanism of minor earthquake”, “microseismicity”, and “oil flow”. We have removed the Geodimeter (2 in total, one 9.8 years before and another 5.6 years before) and tiltmeter anomalies (1 in total, 2.6 years out). Keeping or removing these anomalies does not affect the statistical analysis, which uses data from more recent years.

| Earth Stress | Radon | Underground Water | Resistivity | Earth Currents | Gravity | Leveling | Geomagnetism |
|--------------|-------|-------------------|------------|----------------|---------|----------|-------------|
| 31           | 29    | 21                | 20         | 15             | 14      | 11       | 11          |

As visible from Figures 7 and 8 and Table 2, the frequency of anomalies clearly increases towards the earthquakes. More precisely, the null hypothesis that the 38 anomalies in the year prior to the Haicheng earthquake come from the same underlying frequency parameter as the 10 anomalies in the year before is rejected in favor of the alternative hypothesis that the frequency parameter underlying the 38 events is greater, with \( p = 3 \times 10^{-5} \).
Figure 7. Normalized cumulative sum of anomalies in Tangshan (black) and Haicheng (blue dashed). Time goes from left to right. 0 corresponds to the occurrence time of the Tangshan (respectively Haicheng) earthquake. Inset: Six month time window version of these curves, which shows how Haicheng has a less clean LPPLS trajectory compared to Tangshan. See text for implications.

Figure 8. Replot of Figure 2 of the time of anomalies leading up to the Tangshan earthquake, by type (excluding types with only a single anomaly event). Time goes from left to right as in Figure 7, in units of years. The data points are slightly offset vertically and scattered horizontally to improve visibility where data points are superimposed. The time of the Haicheng event (540 days) is given by the vertical dashed line.

Table 2. Number of anomalies in the ten 365 day periods up to the earthquake, for the prior 10 years, where the earthquakes are defined to occur at the end of year 0. The total number of anomalies is also given. Years 7–9 have zero anomalies and year 6 may be not representative because levelling and sea level are aggregated into these years.

| Years Prior | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | Total |
|-------------|---|---|---|---|---|---|---|---|---|---|-------|
| Tangshan    | 0 | 0 | 0 | 3 | 1 | 5 | 8 | 15| 25| 102| 159   |
| Haicheng    | 0 | 0 | 0 | 5 | 0 | 0 | 2 | 3 | 10| 38 | 58    |

The Tangshan and Haicheng anomalies exhibit a similar acceleration, quantified by the share of their total anomalies in the final years: 16% and 17%, respectively, for one year before, and 63% and 65% for the year of the earthquake. Whether the total number of
anomalies being quite different between the two cases is due to varying density of sensors or a lower actual rate of anomalies is unknown. Note that there is no clear indication of a spike in anomalies in the Tangshan series due to the Haicheng event, 540 days prior. However, e.g., resistivity and levelling anomalies do occur around that time.

Some of the anomalies appear to occur in clusters (Figure 8), which could reflect a seasonal overprint in, for example, underground water and resistivity. Earth current anomalies, however, only appear 100 days prior to the Tangshan event. For all of the anomalies plotted in Figure 8, with the exception of underground water, there is a significantly higher frequency of events in the year prior to the earthquake relative to the four years prior to that ($p < 0.05$, using the R:poisson.test function). This sample is inadequate to properly characterize baseline behavior, and data on additional earthquakes would therefore be of great value.

4.2. Quantitative Analysis

To enable further statistical analysis of the anomaly data, we combine all of the anomalies for each earthquake and simply treat them as a single class. As a flexible model of the acceleration of the rate of anomaly appearance, we employ the log-periodic power law singularity (LPPLS) model [12–17,29] adapted to a formulation in terms of point processes. We thus write the event/point hazard rate,

$$h(t) \propto (t_c - t)^{-\alpha}(1 + \beta \cos(\omega \ln(t_c - t) + \phi)), \quad 0 < \alpha < 1,$$

with a finite time singularity at $t_c$. This function, $h(t)$, is the intensity of a Poisson process, having counting process $N(t)$, being the number of points up to and including time $t$. Parameter $\beta$ quantifies the amplitude of the log-periodic oscillations decorating the overall power law increase in the hazard rate and is a parameter to be calibrated to the data. The expected value of $h(t)$ is given by integrating the intensity,

$$E[N(t)] = \int_{-\infty}^{t} h(s)ds = \gamma + \delta \cdot (t_c - t)^{1-\alpha}(1 + \zeta \cos(\omega \ln(t_c - t) + \phi')), \quad t \leq t_c.$$

Note that the condition $\alpha < 1$ prevents the finite time singularity in (3) from diverging. In other words, the rate diverges but the total number converges to a finite number at the singular time $t_c$. Parameter $\gamma$ is thus the total expected number of events up to the singular time $t_c$. Parameter $\delta$ quantifies the amplitude of the increase in the expected number of events up to $t_c$. Parameter $\zeta$ is equal to $\zeta = \frac{\beta \omega}{\omega^2 + (\alpha - 1)^2}$, from the integration of Expression (2).

As a first diagnostic of LPPLS behavior, by specifying $N(t_c)$ and $t_c$, one can plot

$$N(t_c) - N(t) = c(t_c - t)^{1-\alpha} \cdot P(t),$$

where $P(t) := 1 + \zeta \cos(\omega \ln(t_c - t) + \phi')$ and look for a linear relationship in log-log scale of $N(t_c) - N(t)$ as a function of $t_c - t$, whose slope is equal to $1 - \alpha$, up to log-periodic oscillations around the straight line. This is done in Figure 9, where for Tangshan a value of $\alpha$ of about 0.75 is suggested, and for Haicheng about 0.9 (but with less data and a visually worse fit). The data and plot for Haicheng is still clean due to the long gap in points around 100 days out, and then the accumulation of points around 0–1 days out from $t_c$. To make it slightly cleaner, and also readable, we have randomized the date of the anomalies uniformly in ($-1, +1$) days.
with the assumption of identically independently distributed (i.i.d.) residuals clearly violated. The anomaly times were randomized with uniform noise on \([\log(\cdot), \log(\cdot)]\) and \((\log(\cdot), \log(\cdot))\), respectively. In both cases, theoretical 95% prediction intervals are plotted with dashed lines, however with the assumption of identically independently distributed (i.i.d.) residuals clearly violated. The anomaly times were randomized with uniform noise on \((-1, 1)\) due to plausible data imprecision.

In Tangshan, having more data points makes further statistical analysis feasible. We approach statistical estimation of the LPPLS (1 model, for sample \(t_i \leq \ldots \leq t_n\) on window \([l, r]\), as a probability density, with likelihood, \(L(\theta) = \prod_{t_i \leq t_n} h(t_i) \exp(- \int_{t_i}^{t_n} h(s)ds)\). This approach avoids: (i) imposing a time scale of aggregation necessary for estimation via GLM regression; and (ii) regression of Function (3) to the observed counting process (cumulative number of anomalies), whose residual errors will be dependent, in violation of the regression assumptions. This method of calibrating the data taken as a point process thus improves on existing methods previously used to qualify the presence of LPPLS and accelerated precursors \([12, 20, 21, 24, 30]\).

The results of the MLE estimation of Equation (2) are summarized in Table 3 where time is measured in days. The results are relatively stable across the windows fitted—the first three fits going up to the time of the Tangshan earthquake but varying the length of anomaly history considered, and the last three varying the data to go up to 3, 5, and 10 days before the earthquake. The confidence intervals for the critical time all contain the true time (defined as \(t_c = 0\), and are quite precise, with standard error of the parameter less than 1 day. In all cases, the log-periodicity component is highly significant (e.g., via t-test against \(\beta = 0\)). The residual diagnostics indicate good fit of the trend (time scale transform test of Ogata [31], checking for uniform residuals). A few of the fits are visualized in Figure 10.

Table 3. Summary of fitted LPPLS hazard model (Equation (2)) to Tangshan data, with estimated parameter value and standard error in parenthesis. Fits are given for six periods, including in the sample all anomalies within that given period, with 0 being the time of the earthquake.

| Sample Size | Period  | \(\alpha\)  | \(\beta\)  | \(\omega\)  | \(\Phi\)  | \(t_c\)  |
|------------|---------|-------------|-------------|-------------|-------------|-------------|
| 126        | (-596,0) | 0.70 (0.06) | 0.50 (0.01) | 4.6 (0.23)  | 1.9 (1.2)   | 0.5 (0.5)   |
| 120        | (-447,0) | 0.61 (0.06) | 0.51 (0.11) | 4.9 (0.41)  | 3.9 (0.7)   | 0.01 (0.01) |
| 100        | (-365,0) | 0.75 (0.08) | 0.53 (0.12) | 4.8 (0.30)  | 2.8 (1.4)   | 0.9 (0.7)   |
| 111        | (-596, -3) | 0.61 (0.07) | 0.61 (0.11) | 5.0 (0.19)  | 3.9 (0.96)  | -0.05 (0.63) |
| 107        | (-596, -5) | 0.63 (0.09) | 0.63 (0.11) | 5.0 (0.18)  | 3.8 (0.87)  | -0.05 (0.30) |
| 101        | (-596, -10) | 0.61 (0.10) | 0.61 (0.11) | 5.2 (0.18)  | 4.7 (0.90)  | 0.04 (0.04) |
Let us say a word on the comparison between the LPPLS Models (2) and (3) and a more traditional model, such as a simple power law finite time singularity, corresponding to \( \beta = 0 \) in Expression (2) and \( \zeta = 0 \) in Expression (3). Figure 9 shows a non-parametric rendering of the simpler model. The corresponding residual analysis strongly rejects this simple model with a \( p \)-value much smaller than \( 10^{-3} \). In contrast, as mentioned above, the residual analysis of the fits shown in Figure 10 cannot reject at standard levels of statistical significance the null that the residuals are identically independently distributed, as they should with the time scale transform test of uniform residuals [31]. This comparison confirms the clear superiority and added value of the LPPLS model over the simpler one. The log-periodic component seems thus necessary to account for the intermittency of the temporal unfolding of the anomalies.

Notably, the data available for the Haicheng earthquake gives much poorer results as shown in Figure 11, which shows that the data does not fit well with the LPPLS model. The fits are unstable due to the high clustering of points combined with large gaps between clusters, and the small sample size. The hazard function would have to be allowed to go negative to allow better fitting. This may come as a surprise, given the fact that the Haicheng earthquake was successfully predicted by the SSB and is the only official prediction that led to a large-scale evacuation in human earthquake history, whereas the Tangshan earthquake was not officially predicted. Post-mortem analyses have shown, however, that the Haicheng forecast was based mainly upon the seismic foreshock sequence [32]. We thus confirm that the sparse precursor anomalies reported by Zheng [11] for the Haicheng earthquake would have been difficult to use for a reliable prediction.

Further extensions of the method of calibration include a Bayesian method using a flat prior distribution. This can naturally give the uncertainty in the estimate of \( t_c \) through its posterior distribution.

4.3. Forecasting (Ex-Post) the Tangshan Earthquake and Testing Its Predictability

To more fully examine the evolution of the predicted critical time when approaching the Tangshan earthquake, we fit the LPPLS hazard function (Equation (2)) on a growing window \( (l = -596, r) \) with \( r = -50, -49, \ldots, 0 \) (unit of days) and again with \( l = -365 \) (unit of days), where the earthquake time is always defined to be at time 0. The 90% profile likelihood confidence interval for the predicted critical time is summarized in Figure 12 for both of these cases, which does a reasonable job of bracketing the true critical time. The prediction is highly uncertain more than 30 days before the event; however, from about 20 days before, the prediction becomes increasingly precise as the earthquake time is approached—with a slight shift towards a too early estimate of the earthquake time (about
5–10 days). With the exception of a few unstable fits, the prediction of an imminent event becomes clear in the week prior to the event.

![Figure 11](image-url)  
**Figure 11.** Haicheng best LPPLS fit, cumulative view: total number of anomalies up to time t shown on the x-axis, where time (in days) flows from left to right with 0 as the time of occurrence of the Haicheng earthquake. The fit is of poor quality. Uniformity of residuals test rejects the LPPLS model with $p = 0.0047$. Estimated parameters are $\alpha = 0.7$, $\beta = 0.2$, $\omega = 5$, $\phi = 3$, $t_c = 0.01$. Standard errors were not numerically possible.

![Figure 12](image-url)  
**Figure 12.** Ninety percent confidence interval of predicted time $t_c$ (earthquake occurrence time) in days vs. days before earthquake. Along the x-axis and y-axis, 0 corresponds to the time of occurrence of the Tangshan earthquake. The x-axis gives the “present” time at which the mock prediction is made. For instance, at 10 days before the earthquake, the predicted time of the earthquake is between $-1$ and $-3$ days, i.e., the method predicts the event between 1 to 3 days before the earthquake actually occurred. The diagonal line gives a lower bound on the estimate, which is where the estimate of the critical time is the end of the fitted data window (i.e., the “present” time). Note that when the confidence interval lines go above 0, the critical time is overestimated and vice versa. The left plot fixes the data starting time at $l = -596$ days prior to the earthquake, whereas the right plot takes it at $l = -365$. Dramatic changes in the confidence interval are possibly due to the periodicity of the function, which can inject bi-modality into the likelihood over the critical time parameter.

We now perform a simulation study to test the feasibility of an early warning system if the anomalies really do follow an LPPLS trajectory. In particular, assuming parameters consistent with the observed data (taking parameters from row 1 of Table 3), we indicate the predictability of the earthquake time with a simple prediction experiment: for a given
realization of anomalies (data), we estimate Expression (2) each day in the 40 days up to \( t_c \), fitting on the growing sample of observed anomalies \( \{t_1, \ldots, t_{N(i)}\} \), for days \( i = 86, \ldots, 126 \). We declare an alarm at the first individual fit in the sequence of 40 where the predicted critical time (estimated parameter \( t_c \)) is sufficiently near and precise. In detail, we declare an alarm if the nearest edge of the 95% confidence interval of \( t_c \) falls within 7 days of the most recent date \( t_i \) within the sample (i.e., “today” in a real-time prediction context), and its width is less than 14 days. We can also count the total number of the 40 fits that satisfy the alarm condition. Simulations indicate that, in ~93% of cases, an alarm will be raised for an earthquake within the 10 days before the actual time. In addition, in ~5% of cases, an alarm will be prematurely raised for earthquakes falling in the interval from 40 to 20 days prior to the actual time. A rate of false alarm is not quantified here, given lack of observation of anomaly data under non-earthquake conditions. However, under a homogeneous Poisson assumption for anomaly times for non-earthquake conditions, the observed super-exponential accumulation of points is exceedingly unlikely: for instance, the Kolmogorov–Smirnov test that the point process of anomalies is uniform rejects this null with \( p < 10^{-16} \).

5. Conclusions

It is not widely known that the Chinese, under the stewardship of Enlai Zhou, in the period 1966 to 1976, developed an earthquake prediction methodology in one of the world’s biggest science and social science projects ever undertaken. The program climaxed with the successful prediction of the M7.6 Haicheng earthquake in February 1975 that saved many thousands of lives. Eighteen months later, the industrial city of Tangshan, 180 kms east of Beijing, was flattened without warning, by an M7.8 earthquake. Somewhere between 240,000 and 650,000 people lost their lives. The jubilation that followed Haicheng success turned to despair. Zedong Mao died in September 1976. China set a new course and the earthquake prediction program was prematurely dismantled and all but forgotten. The recipe for success lay smoldering in hundreds of technical documents and reports.

Here, we have revisited the compiled set of precursory data that were recorded before the Haicheng and Tangshan earthquakes, but never gathered into a centralized data base. Rather than sequentially analyzing each anomaly (levelling, geomagnetism, soil resistivity, earth currents, gravity, earth stress, radon, well water level), we proposed a coarse-graining approach of counting the total number of reported anomalies as a function of time. We demonstrated strong evidence for the existence of an acceleration of anomalies leading up to the major Haicheng and Tangshan earthquakes. In particular for the Tangshan earthquake, the frequency of occurrence of anomalies was found to be well described by the log-periodic power law singularity (LPPLS) model but not for Haicheng due to a paucity of data.

Based on a mock real-time prediction experiment and simulation study, we provided an indication of potential for an early warning system based on this methodology of monitoring accelerated rates of anomalies.

Further data from these and additional earthquakes would be of great value to better characterize both baseline and earthquake anomaly behavior. Indeed, the set of relevant anomalies, and the statistics of anomalies near an earthquake, may vary considerably by region and case, notably due to differing geology and tectonic settings.

It is likely that the acceleration that we report here was somehow understood implicitly by the Chinese experts gathering and attempting to understand the accumulating data sets. While such “gut feelings” may explain the human drama of how scientists and decision makers were discussing and quarrelling over the growing feeling of incipient deadly risks, the experts apparently never made the step to quantify this impression of an imminent danger in the simple and transparent way that we have proposed here. In fact, the data from hundreds of monitoring stations was never collated into a central database and the Chinese at the time depended upon amateur staff manning individual stations to draw and report conclusions based upon that single station’s results. In the spirit of
optimally displaying technical information [33], we propose that such a simple robust metric as proposed in the present paper may often provide substantial evidence relating to a developing trend and the impending dangers. This may have convinced the decision makers to act. We see here a parallel to the story of the 1986 Challenger disaster, for which the engineers of the company Thiokol were unable to convey with sufficient clarity to their managers and those of NASA why they were opposed to the take-off on that fateful day. Tufte [33] has shown that an efficient visual display (Figure 13) of the trend of the damage of the O-ring of the boosters on the side of the main rocket could have changed the decisions, which were at that time based on arcane reports and tables full of what appeared as mostly incomprehensible numbers for the managers and decision makers to understand the latent danger of launching at sub-zero temperatures.

Finally, we hope to have convinced the reader of the surprising relevance of deep symmetry considerations to help organize what seems, a priori, to be disorganized heterogenous data collected for a very difficult problem, that of earthquake predictability. This illustrates yet again a fundamental unity of science, from very fundamental concepts to operational applications, which in the present case have the potential to save many lives.

**Author Contributions:** Conceptualization, E.M. and D.S.; methodology, E.M., S.W. and D.S.; software, S.W.; validation, E.M., S.W. and D.S.; formal analysis, E.M., S.W. and D.S.; investigation, E.M., S.W. and D.S.; resources, D.S.; data curation, E.M. and S.W.; writing—original draft preparation, E.M., S.W. and D.S.; writing—review and editing, E.M., S.W. and D.S.; visualization, E.M. and S.W.; supervision, D.S.; project administration, D.S. funding acquisition, D.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors thank the National Natural Science Foundation of China for financial support under grant No. U2039202.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

---

**Figure 13.** Measure of damage of the O-ring rubber on the booster rockets on the side of the main rocket carrying the space shuttle as a function of temperature (in Fahrenheit) at take-off of previous flights. Each black dot is a measure taken for a different previous flight. The red arrow indicates the range of temperature at take-off on 28 January 1986, when the Space Shuttle Challenger exploded 73 s into its flight, killing all seven crew members aboard. Note that this was below 32°F, the temperature at which water freezes. The red ellipsis emphasizes the importance of reporting data even when no damage is observed (“the dog that did not bark”), which occurred for flights taking off at relatively warm temperatures on the ground. The upward red arc outlines the trend of the growing number of previous flights that showed significant damage of the critical O-ring structure as the temperature at take-off was lower. Notice the large gap between the temperature at take-off of all previous flights and that of 28 January 1986. Detailed post-mortem analyses confirmed the causal role of the failure of O-rings in the Challenger disaster (adapted from Tufte [33]).
Conflicts of Interest: The authors declare no conflict of interest.

References
1. Dubrulle, B.; Graner, F.; Sornette, D. (Eds.) Scale Invariance and Beyond; EDP Sciences and Springer: Berlin, Germany, 1997; p. 286.
2. Sornette, D. Discrete scale invariance and complex dimensions. Phys. Rep. 1998, 297, 239–270. [CrossRef]
3. Jiang, H.; Chen, F. (Eds.) The 1976 Tangshan Earthquake, 2nd ed.; Seismological Press Beijing: Beijing, China, 1982. (In Chinese)
4. Nandan, S.; Ouillon, G.; Wiemer, S.; Sornette, D. Objective Estimation of Spatially Variable Parameters of Epidemic Type Aftershock Sequence Model: Application to California. J. Geophys. Res. Solid Earth 2017, 122, 5118–5143. [CrossRef]
5. Nandan, S.; Ouillon, G.; Sornette, D.; Wiemer, S. Forecasting the full distribution of earthquake numbers is fair, robust and better. Seismol. Res. Lett. 2019, 90, 1650–1659. [CrossRef]
6. Geller, R.J. Earthquake prediction: A critical review. Geophys. J. Int. 1997, 131, 425–450. [CrossRef]
7. Kagan, Y.Y. Are earthquakes predictable? Geophys. J. Int. 1997, 131, 505–525. [CrossRef]
8. Main, I. Is the reliable prediction of individual earthquakes a realistic scientific goal? Nature 1999, 397. [CrossRef]
9. Sichuan Province Earthquake Office. The 1976 Songpan Earthquake; Seismological Press: Chengdu, China, 1979. (In Chinese)
10. Zhu, F.M.; Wu, G. (Eds.) The Haicheng Earthquake of 1975; Seismological Press: Beijing, China, 1982; p. 220. (In Chinese)
11. Zheng, J.Z. Precursors to the Haicheng and Tangshan earthquakes. J. Seismol. Soc. Japan 1981, 34, 43–59. [CrossRef]
12. Anifrani, J.-C.; Le Floch’, C.; Sornette, D.; Souillard, B. Universal Log-periodic correction to renormalization group scaling for rupture stress prediction from acoustic emissions. J. Phys. I 1995, 5, 631–638. [CrossRef]
13. Gluzman, S.; Sornette, D. Classification of Possible Finite-Time Singularities by Functional Renormalization. Phys. Rev. E 2002, 6601, 016134. [CrossRef]
14. Ide, K.; Sornette, D. Oscillatory Finite-Time Singularities in Population, Finance, and Rupture. Phys. A Stat. Mech. Its Appl. 2002, 307, 63–106. [CrossRef]
15. Le Floch’, C.; Sornette, D. Predictive Acoustic Emission: Application on Helium High Pressure Tanks; Instrumentation MesureMetrologie; RS Serie I2M; Hermes Science: Paris, France, 2003; Volume 3, pp. 89–97.
16. Sornette, D. Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes and human birth. Proc. Natl. Acad. Sci. USA 2002, 99 (Suppl. 1), 2522–2529. [CrossRef]
17. Sornette, D.; Sammis, C.G. Complex critical exponents from renormalization group theory of earthquakes: Implications for earthquake predictions. J. Phys. I 1995, 5, 607–619. [CrossRef]
18. Goldenfeld, N. Lectures on Phase Transitions and the Renormalization Group (Frontiers in Physics); CRC Press: Boca Raton, FL, USA, 1992.
19. Sornette, D. Critical Phenomena in Natural Sciences (Chaos, Fractals, Self-Organization and Disorder: Concepts and Tools), 2nd ed.; Springer Series in Synergetics; Springer: Berlin/Heidelberg, Germany, 2004; p. 528.
20. Bowman, D.D.; Ouillon, G.; Sammis, C.G.; Sornette, A.; Sornette, D. An observational test of the critical earthquake concept. J. Geophys. Res. 1998, 103, 24349–24372. [CrossRef]
21. Guilhem, A.; Bürgmann, R.; Freed, A.M.; Ali, S.T. Testing the accelerating moment release (AMR) hypothesis in areas of high stress. Geophys. J. Int. 2013, 195, 785–798. [CrossRef]
22. Mignan, A. Retrospective on the Accelerating Seismic Release (ASR) hypothesis: Controversy and new horizons. Tectonophysics 2011, 505, 1–16. [CrossRef]
23. Johansen, A.; Sornette, D. Evidence of discrete scale invariance by canonical averaging. Int. J. Mod. Phys. C 1998, 9, 433–447. [CrossRef]
24. Johansen, A.; Saleur, H.; Sornette, D. New Evidence of Earthquake Precursory Phenomena in the 17 January 1995 Kobe Earthquake, Japan. Eur. Phys. J. B 2000, 15, 551–555. [CrossRef]
25. Qian, F.-Y.; Zhao, Y.-L.; Lu, J. Georeisistivity precursors to the Tangshan earthquake of 1976. Ann. Geophys. 1997, 40, 251–260.
26. Mearns, E.; Sornette, D. The great 1976 Tangshan Earthquake (Learning from the 1966–1976 Chinese Prediction Program); Cambridge Scholars Publishing: Newcastle, UK, 2021; in press.
27. Mearns, E.; Sornette, D. A transfer fault complex to explain the geodynamics and faulting mechanisms of the M7.8 Tangshan Earthquake. J. Asian Earth Sci. 2021, 213, 104738. [CrossRef]
28. Mei, S. The precursory complexity and regularity of the Tangshan earthquake. J. Phys. Earth 1986, 34, 193–212. [CrossRef]
29. Ouillon, G.; Sornette, D. The Concept of ‘critical earthquakes’ applied to mine rockbursts with time-to-failure analysis. Geophys. J. Int. 2000, 143, 454–468. [CrossRef]
30. Johansen, A.; Sornette, D.; Wakita, H.; Tsunogai, U.; Newman, W.I.; Saleur, H. Discrete scaling in earthquake precursory phenomena: Evidence in the Kobe earthquake, Japan. J. Phys. I 1996, 6, 1391–1402. [CrossRef]
31. Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. J. Am. Stat. Assoc. 1988, 83, 9–27. [CrossRef]
32. Raleigh, B.; Bennett, G.; Craig, H.; Hanks, T.; Molnar, P.; Nur, A.; Savage, J.; Scholz, C.; Turner, R.; Wu, F. Haicheng earthquake study delegation, Prediction of the Haicheng Earthquake. EOS Trans. Amer. Geophys. Union 1977, 58, 236–272.
33. Tufte, E. The Visual Display of Quantitative Information, 2nd ed.; Graphics Press: Cheshire, CT, USA, 2001.