

# Thermodynamic State Shift Observed Prior to the 2011 Mw 9 East Japan Earthquake through Data Mining Using Cellular Automaton

Hiroyuki Kikuchi (✉ [hkikuchi@kkh.biglobe.ne.jp](mailto:hkikuchi@kkh.biglobe.ne.jp))

Seismic Lab

---

## Research Article

**Keywords:** megathrust earthquake, cellular automaton, thermodynamic fluctuation, state transition, onset of stress manifestation

**Posted Date:** September 14th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3294988/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

**Version of Record:** A version of this preprint was published at Earth Science Informatics on January 26th, 2024. See the published version at <https://doi.org/10.1007/s12145-024-01225-6>.

# Thermodynamic State Shift Observed Prior to the 2011 Mw 9 East Japan Earthquake through Data Mining Using Cellular Automaton

Dr Hiroyuki Kikuchi<sup>1,1\*</sup>

<sup>1\*</sup>, Seismic Lab,307, 12-1, 2-Chome,Wakamiya,Okegawa,3630022, Saitama,Japan.

Corresponding author(s). E-mail(s): [hkikuchi@kkh.biglobe.ne.jp](mailto:hkikuchi@kkh.biglobe.ne.jp);

## Abstract

In this study, we employed our previously developed data mining method to show that a thermodynamic state shift occurred preceding the 2011 Mw 9 East Japan Earthquake (GEJE), coinciding with the onset of crustal stress manifestations. Our discussion starts with the insights obtained from our prior research, which revealed that small ground vibration fluctuations (GVF) detected near the epicenter of the GEJE are thermodynamically equivalent to a cellular automaton (CA). This equivalence allowed us to consider the "Process to GEJE," defined as the transition from GVF to catastrophic rupture in the GEJE, as analogous to the nonequilibrium-to-equilibrium thermodynamic transition in the CA. Therefore, if we find a thermodynamic state shift in the CA, it is reasonable to assume that a similar thermodynamic state shift may exist in the "Process to GEJE." We successfully derived a thermodynamic state function that exhibits a significant change before the nonequilibrium-to-equilibrium thermodynamic transition in the CA. By evaluating this thermodynamic state function using the GVF data of the "Process to GEJE," we discovered its maximum value occurring 1231 days before the GEJE. This maximum value can be considered as an indication of a thermodynamic state shift prior to the GEJE. It has been found that during this thermodynamic state shift prior to the GEJE, the GVF state approaches thermodynamic equilibrium, and an increase in crustal stress becomes visible for the first time.

**Keywords:** megathrust earthquake, cellular automaton, thermodynamic fluctuation, state transition, onset of stress manifestation

## 1 Introduction

We have dedicated our efforts to extracting physically meaningful information from ground vibration fluctuations (GVF) observed during most of the ground vibration observations, which have weak amplitudes and lack specific waveforms. This is because earthquakes typically last less than a minute, and the dominant state of ground motion is seismically silent. We firmly believe that the GVF contains valuable information about the evolution of ground dynamics before and after the megathrust earthquake (excluding the earthquake itself).

The GVF, as a fluctuating system, can be stochastically treated within the framework of the master equation and characterized by thermodynamic parameters. However, the current understanding of how these thermodynamic characteristics of the GVF are linked to physically and practically meaningful phenomena remains limited.

On the other hand, stochastic cellular automata (CA) can simulate complex phenomena such as traffic jams and identify the conditions that lead to them. If CA can establish a connection between the complex thermodynamic parameters calculated from CA data and physically meaningful information, and if GVF and CA are thermodynamically equivalent, we can extract physically meaningful information from GVF data through the analysis of CA.

Our previous study introduced the Data Mining Method in Seismology by Applying Cellular Automaton Equivalence of Ground Vibration Fluctuations Recorded Near the Epicenter of the 2011 Mw 9 East Japan Earthquake (GEJE) to extract physically meaningful information from the GVF (Kikuchi, 2023). We initially demonstrated the equivalence of the GVF and CA. In the data mining process, we established the relationship between thermodynamic parameters in the CA and stress relaxation, and then evaluated these thermodynamic parameters for the GVF to extract stress relaxation signals in the GVF.

In this paper, we apply the data mining method to discover the thermodynamic state shift prior to the GEJE. We define the “Process to GEJE” as the transition from a long-term ground vibration fluctuation (GVF) to a catastrophic rupture in GEJE. The starting time of the “Process to GEJE” is arbitrarily chosen between the last occurrence of a megathrust earthquake in eastern Japan and the GEJE.

First, we demonstrate the equivalence between the “Process to GEJE” and the thermodynamic transition of a cellular automaton (CA) from non-equilibrium to equilibrium. Next, we derive the thermodynamic state function that yields a significant change prior to the thermodynamic transition in the CA, and then evaluate this thermodynamic state function using the GVF data.

Before delving into the main discussion, let us briefly review the data handling procedures and the key terms explained in our previous paper.

## **2 Review of data handling procedures and key terms (Kikuchi, 2023)**

The ground vibration velocity data, or the source of the GVF, was recorded both before and after the occurrence of the GEJE occurred at the epicenter with latitude 38.06N, longitude 142.51E, and depth 24 km on March 11, 2011, at 14:46 (Japan Standard Time, JST). The survey period spans from 00:00 on January 1, 2006, to 23:59 on August 30, 2018.

The data was obtained from the seismic station KSN and downloaded in chronological order from the F-net website, which is the broadband seismograph network of the National Research Institute for Earth Science and Disaster Resilience (NIED, 2019), using the "BLZ" option. In this context, "B," "L," and "Z" respectively indicate that the data has a sampling interval of 0.05 seconds, the instrument type is a strong motion velocity meter, and the velocity direction is up-down (UD). The seismic station KSN is situated at latitude 38.98N, longitude 141.53E, and altitude 260 m.

The instrument response is corrected by multiplying it by a correction factor,  $A_s$ , which is the product of the transfer function, the normalizing constant, and the ratio of sensitivity to gain (SEED, 2012). For the 'KSN BL' data from 2006 to 2016,  $A_s$  was calculated from the instrument data provided with the 'KSN BL' data and treated as a real constant ( $= 6.168853e-08$  m/s/V for 'KSN\_B LZ'). This is because the imaginary component of  $A_s$  is small in the frequency range of 1-10Hz, where the GVF signal is most relevant to our analysis.

The downloaded data is converted into piecewise deviation, which represents the difference between the velocity and the velocity averaged over the subsequent 10 data points. This piecewise deviation, fluctuating around zero, is referred to as the Ground Vibration Fluctuation (GVF). For the remainder of this study, GVF specifically denotes the GVF recorded at KSN.

To analyze the GVF data, it is divided into blocks of 1024 data points, corresponding to a data acquisition time of 51.2 seconds. The Fourier amplitude of each block is then calculated using the Fast Fourier Transform (FFT) algorithm without any overlap or filtering. The frequency domain ranges from 0.02 Hz (determined by the block size of 1024) to 10 Hz, which is half the data acquisition frequency.

In our previous paper, we reported the discovery of "anomalous noise," which forms a convex curve in the high-frequency region of the Fourier amplitude spectrum of the GVF. We found that this convex curve is qualitatively preserved when transforming the GVF data into a binary sequence, leading us to consider the GVF and the binarized GVF as equivalent. We converted the recorded GVF data into a sequence of "1"s and "0"s and analyzed the binarized GVF instead of the original GVF.

By binarizing the GVF, we were able to define its thermodynamic states, whose time evolution is governed by the master equation. This equation describes the time change of a thermodynamic state vector in terms of the state transition rate matrix and the thermodynamic state vector. Additionally, the state transition rate matrix and the thermodynamic state vector define a scalar quantity called the entropy production rate (EPR), which indicates the thermodynamic stability of the GVF. In this manner, we treat the GVF as a thermodynamic system.

To quantify the "anomalous noise," we introduce the concept of "A-noise," which takes into account whether the spectrum is convex or concave, as well as the magnitude of the convexity.

On the other hand, there exists a mathematical scheme called Cellular Automaton (CA), which generates a binary sequence by following an evolution law known as the "local rule." This binary sequence can be viewed as a sequence of thermodynamic states that evolve according to the master equation. Consequently, the CA can also be treated as a thermodynamic system.

We compared the GVF and CA, two thermodynamic systems, and demonstrated that for a given GVF data set, there exists a corresponding CA data set whose thermodynamic parameters match the thermodynamic parameters calculated for the GVF data set. This result suggests that the GVF is thermodynamically equivalent to the discussed CA.

This section briefly reviews the thermodynamic parameters, namely *A-noise*, thermodynamic states, the density of states, and *EPR*, along with the CA that is thermodynamically equivalent to the GVF.

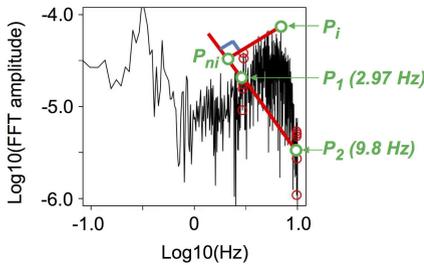
## 2.1 Definition of *A-noise*

In order to quantify the anomalous noise or the convex-curve-like spectrum, we need to define a parameter that takes into account whether the spectrum is convex or concave, as well as the magnitude of the curve. The shape and magnitude of a curve in a flat two-dimensional x-y coordinate system are defined by the 'curvature', which is proportional to the second derivative of the curve. Therefore, a 'convex' curve (i.e., one that curves outward) has negative curvature, and a 'concave' curve has positive curvature.

The curvature of the Fourier amplitude spectrum is defined for a GVF acquired with a sampling time of 51.2 seconds, a sampling frequency of 20 Hz, and a data length of 1024. This is given by the expression  $sign(P_{ni} - P_i)$  times the ratio of  $|P_{ni} - P_i|$  to  $|P_2 - P_1|$ , as illustrated in Fig. 1. Here,  $P_i$  represents a point within the frequency range of 2.97 to 9.8 Hz. The frequency of  $P_1$  and  $P_2$  are 2.97 Hz and 9.8 Hz. The average spectrum value of the nearest 5 points (red circles in Fig. 1) is assigned as the spectrum value for the  $P_1$  and  $P_2$ .  $P_{ni}$  is determined so that the line from  $P_i$  to  $P_{ni}$  is perpendicular to the line connecting  $P_1$  and  $P_2$ . If the spectrum value of  $P_i$  is greater than that of  $P_{ni}$ , then the spectrum is convex, and the curvature is negative. Hence, the definition of curvature presented in Fig. 1 is in line with the standard definition

of curvature in two-dimensional space. Note that no overlap or filtering is applied in the Fourier transform.

The *A-noise* is defined as the product of "−1" and the curvature with the largest absolute value among all the curvatures within the frequency range of 2.97 to 9.8 Hz (Fig. 1). Consequently, an arbitrary GVF signal can be categorized as either positive *A-noise* or non-positive *A-noise* based on this definition.



$$A - noise \equiv -curvature_m;$$

$$|curvature_m| = \max(|curvature_i|)$$

$$curvature_i \equiv sign(P_{ni} - P_i) \frac{|P_{ni} - P_i|}{|P_2 - P_1|}$$

$$2.97 \text{ Hz} < \text{Frequency of } P_i < 9.8 \text{ Hz}$$

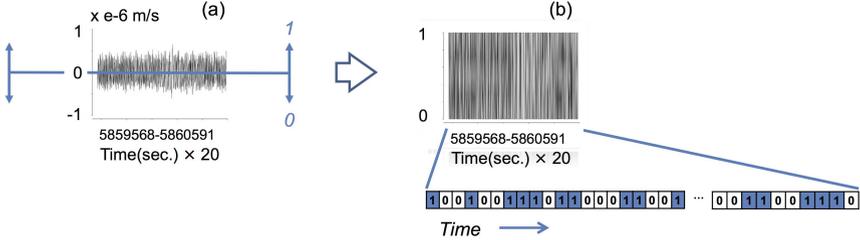
Fig. 1 Definition of spectrum curvature and *A-noise*.

## 2.2 Binarization of GVF

The *A-noise* is calculated based on the binarized GVF. Fig. 2 illustrates the binarization process of the GVF. In Fig. 2(a), if each GVF data is greater than the mean of the dataset, it is converted to "1"; otherwise, it is converted to "0". The resulting binarized GVF data is then expressed as a time sequence of "1" and "0", as shown in Fig. 2(b). The binary sequence of the GVF and the timing to calculate the *A-noise* are defined by Eq. (1)-(2), where the timing and the data length are more explicitly described than before to avoid confusion, as additional parameters are newly introduced in this paper. Eq. (1)-(2) indicate that  $n = 1024(j - 1) + 1$  and  $n = 1, 1025, \text{ and } 2049$  if  $j = 1, 2, \text{ and } 3$ . This indicates that *A-noise* is calculated for every 1024 binarized GVF data points. The  $\Delta t$  is set to  $1/20$  (sec) due to the sampling frequency of the recorded data being 20 Hz.

$$\begin{aligned}
\overline{BS}(t_n) &\equiv \{a(t_1), a(t_2), \dots, a(t_n), \dots, a(t_N)\} : \text{binary sequence} \\
a(t_n) &\in \{0, 1\} \\
t_n &= t_0 + (n - 1)\Delta t : \text{sampling time} \\
t_0 &= 0, \text{ corresponds to 2006-01-01 00:00} \\
\Delta t &= \frac{1}{20} \text{ (sec)}, n = 1, 2, \dots, N \\
N &= 1024 \times 10^4 \times M, M = 1, 2, 3, \dots
\end{aligned} \tag{1}$$

$$\begin{aligned}
\overline{A\text{-noise}} &\equiv \{A\text{-noise}(t''_1), A\text{-noise}(t''_2), \dots, A\text{-noise}(t''_j), \dots, A\text{-noise}(t''_J)\} : \\
&A\text{-noise sequence} \\
A\text{-noise}(t''_j) &: \text{Calculated from the binary sequence } \overline{BS}(t''_j) \\
\overline{BS}(t''_j) &\equiv \{a(t''_j), a(t''_j + \Delta t), \dots, a(t''_j + (1024 - 1)\Delta t)\} \\
t''_j &= t_0 + 1024(j - 1)\Delta t, j = 1, 2, \dots, J; J = N/1024
\end{aligned} \tag{2}$$

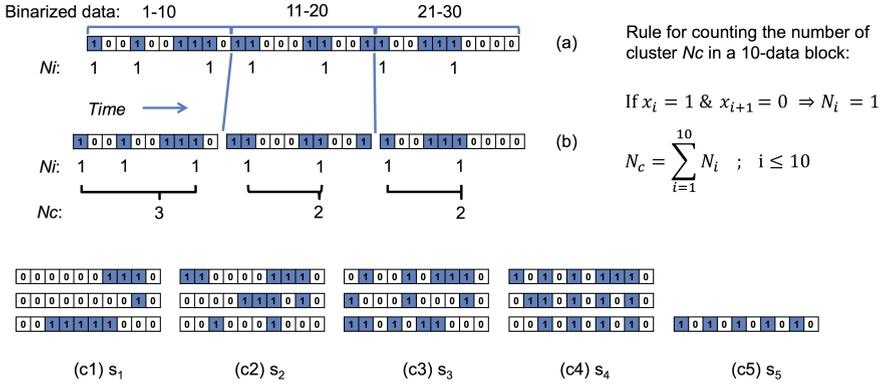


**Fig. 2** Binarization of GVF. The origin of time is 2011-03-03 12:27 (JST). (a) GVF for 51.2 seconds. (b) The binarization result of (a).

### 2.3 Definition of thermodynamic states

To define the thermodynamic state, the binarized GVF sequence (Fig. 3(a)) is divided into blocks of 10 data points, and the number of "1" clusters ( $N_i$ ) is counted in each block. To preserve the total number of clusters, the rule shown in Fig. 3 is applied, where  $x_i \in 0, 1$  is the binary number in the  $i$ th cell in a 10-data block ( $0 \leq i \leq 10$ ). In the 10-data block, we scan the cells from left to right and count one if the sequence of "10" is found. At the end of the scan, at the 10th data point, we count one only if the 10th data point is "1" and the 11th data point is "0" (Fig. 3 (b)). The counting rule restricts the maximum number of clusters in a block to five, and defines five thermodynamic states:  $s_1, s_2, s_3, s_4$ , and  $s_5$ , each containing 1, 2, 3, 4, and 5 clusters, respectively

(Fig. 3(c1)-(c5)). The sequence of the thermodynamic state is defined by Eq. (3), where the timing and the data length are more explicitly described than before to avoid confusion, as additional parameters are newly introduced in this paper. Eq. (3) indicates that the thermodynamic state is defined for every 10 data points of the binary sequence.



**Fig. 3** Definition of thermodynamic states. The  $x_i \in \{0, 1\}$  is the binary number of the  $i$ th cell from the left of the 10-data block ( $1 \leq i \leq 10$ ). (a) A sequence of binarized GVF and  $N_i$ , which is the count of “1” clusters. (b) Blocks with 10 data points, divided from the sequence in (a), and the count  $N_i$  for the “1” cluster. (c1)-(c5) Examples of thermodynamic states  $s_1, s_2, s_3, s_4$ , and  $s_5$  which contain 1, 2, 3, 4, and 5 clusters of “1”, respectively.

$$\begin{aligned}
 \bar{S} &\equiv \{s(t'_1), s(t'_2), \dots, s(t'_k), \dots, s(t'_K)\} : \text{sequence of thermodynamic states} \\
 t'_k &= t_0 + 10(k-1)\Delta t, \quad k = 1, 2, \dots, K; \quad K = N/10 \\
 s(t'_k) &: \text{thermodynamic state in } t'_k \leq t_n \leq t'_k + (10-1)\Delta t; \\
 s(t'_k) &\in \{s_1, s_2, s_3, s_4, s_5\} : \text{thermodynamic state of the binary sequence of} \\
 &\quad \{a(t'_k), a(t'_k + \Delta t), \dots, a(t'_k + (10-1)\Delta t)\} \quad (3)
 \end{aligned}$$

## 2.4 Thermodynamics of GVF

The binarized GVF that fluctuates over time with no apparent regularity, implies that the state of the binarized GVF is non-equilibrium or stochastic. It is known that the thermodynamics of fluctuating non-equilibrium systems is described by the master equation defined by Eq. (4), and the *EPR* defined by Eq. (5) (Haitao, Y. & Jiulin, D., 2014, and Ito, S., 2020). Note that for equilibrium systems, which are time-independent, the left-hand side of the master equation is zero and *EPR* is also zero. In contrast, for non-equilibrium systems, *EPR* is positive.

$$\frac{dp_i}{dt} = \sum_{j=1}^n J_{ij}(t) \quad (4)$$

$$EPR(t) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n J_{ij}(t) F_{ij}(t) \quad (5)$$

$$J_{ij}(t) = W_{ij}(t)p_j(t) - W_{ji}(t)p_i(t)$$

$$F_{ij}(t) = \ln \frac{W_{ij}(t)p_j(t)}{W_{ji}(t)p_i(t)}$$

where  $J_{ij}$  and  $F_{ij}$  are called the flow from  $i$ -state to  $j$ -state and thermodynamic force, respectively. “ $t$ ” is time.  $W_{ij}$  is the  $(i, j)$  component of the  $5 \times 5$  transition rate matrix which defines the total number of transitions from  $i$ -state to  $j$ -state. The  $W_{ij}$  is increased by one when a transition occurs from  $i$ -state to  $j$ -state. Here, the term “state” corresponds to the  $s_i$  as depicted in Fig. 3. Fig. 3 (b) shows the three-cluster state, two-cluster state, and two-cluster state in chronological order. The first transition from  $s_3$  to  $s_2$  increases the value of  $W_{32}$  by 1, and the second transition from  $s_2$  to  $s_2$  increases the value of  $W_{22}$  by 1. The  $p_i$ , which corresponds to the state  $s_i$ , is the probability density of states defined in Eq. (6), where the timing and the data length are more explicitly described than before to avoid confusion, as additional parameters are newly introduced in this paper. Eq. (6) indicates that  $p_i$  is calculated from the first 1000 data points extracted from the 1024 data points of the binary sequence, which is the same data used for calculating  $A$ -noise. The initial 1000 data points consist of 100 thermodynamic states, with each state comprising 10 binary digits.

$$\begin{aligned}
 \bar{p}_i &\equiv \{p_i(t''_1), p_i(t''_2), \dots, p_i(t''_j), \dots, p_i(t''_J)\} : \text{sequence of} \\
 &\text{probability density of states} \\
 t''_j &= t_0 + 1024(j-1)\Delta t, \quad j = 1, 2, \dots, J; \quad J = N/1024 \\
 p_i(t''_j) &= \frac{N(s_i(t''_j))}{\sum_{i=1}^5 N(s_i(t''_j))} : \text{probability density of states} \\
 N(s_i(t''_j)) &: \text{number of state } s_i \text{ in } t''_j + t'_1 - t_0 \leq t_n \leq t''_j + t'_{100} - t_0; \\
 &\quad i = 1, 2, 3, 4, 5 \\
 \sum_{i=1}^5 N(s_i(t''_j)) &: \text{total number of states in } t''_j + t'_1 - t_0 \leq t_n \leq t''_j + t'_{100} - t_0 \\
 s_i(t''_j) &\in \{s_1, s_2, s_3, s_4, s_5\} : \text{state } s_i \text{ in } t''_j + t'_1 - t_0 \leq t_n \leq t''_j + t'_{100} - t_0 \\
 \bar{S}(t''_j) &= \{s(t''_j + t'_1 - t_0), s(t''_j + t'_2 - t_0), \dots, s(t''_j + t'_{100} - t_0)\} : \\
 &\quad \text{sequence of states in } t''_j + t'_1 - t_0 \leq t_n \leq t''_j + t'_{100} - t_0 \\
 s(t) &\in \{s_i(t) \mid i = 1, 2, 3, 4, 5\}
 \end{aligned} \tag{6}$$

The thermodynamic parameters in this study are defined as follows: the *EPR*, the *A-noise*, the density of states (which is the ratio of the number of "1" cells to the total number of cells), the five components of the probability density of states, and their combinations and distributions. The *A-noise* is considered a thermodynamic parameter because it corresponds to the pair of *EPR* and the transition rate matrix, as shown in our previous study.

## 2.5 CA review

In this sub-section, we review the CA-184 (or Rule 184) cellular automaton, as well as the  $(d, p)$ -CA184, which is a stochastic extension of CA-184. We then review the data generation protocol that uses the periodic  $(d, p)$ -CA184 (i.e., the  $(d, p)$ -CA184 with periodic boundary conditions), and the equilibrium characteristics of the periodic  $(d, p = 1)$ -CA184.

### 2.5.1 CA-184 (Nishinari & Takahashi, 1999)

CA-184 is commonly used for modeling traffic jams. It is a one-dimensional array of cells, with each cell containing either "1" or "0". The time evolution of each cell, which depends solely on the cells on both sides of it, follows rule-184:

$$\frac{U_{i-1}^t U_i^t U_{i+1}^t}{U_i^{t+1}} = \frac{111}{1}, \frac{110}{0}, \frac{101}{1}, \frac{100}{1}, \frac{011}{1}, \frac{010}{0}, \frac{001}{0}, \frac{000}{0} \tag{7}$$

where  $t$  and  $i$  are natural numbers indicating the discrete time and discrete space, respectively. The denominator  $U_i^{t+1}$  represents the cell value at (time, space) =  $(t+1, x_i)$ , and it is determined by the numerator  $U_{i-1}^t U_i^t U_{i+1}^t$ , which is a sequence of three binary numbers at time  $t$ . Rule 184 states that when the

sequence "10" (one followed by zero) appears at time  $t$ , it becomes "01" (zero followed by one) at time  $t + 1$ . In other words, assuming "0" and "1" denote an empty cell and a cell occupied by "1", respectively, if there is an empty cell to the right of an occupied cell at time  $t$ , the "1" in the occupied cell moves right and occupies the empty cell at time  $t + 1$ .

Therefore, the occupied cells move to the right as time passes. There are eight possible time evolution patterns as shown on the right side of Eq. (7).

### 2.5.2 ( $d, p$ )-CA184 : CA with probabilistic fluctuations

Probabilistic fluctuations need to be introduced in cellular automata to express complex phenomena, such as ground vibration on Earth. By introducing the moving or hopping probability,  $p$  ( $0 < p < 1$ ), Rule 184 (Eq. 7) can be rephrased such that if there is a sequence "10" in the numerator at time  $t$ , then with a probability of  $p = 1$ , the sequence becomes "01" at time  $t + 1$ .

Extending the probability  $p$  to an arbitrary real number, we define the  $p$ -CA184 as follows: if there is a sequence "10" in the numerator at time  $t$ , then with probability  $p$ , the sequence becomes "01" at time  $t + 1$ . In other words, when the  $p$ -CA184 evolves, it evolves with probability  $p$ . The CA-184 with the hopping probability  $p$  is also known as the Asymmetric Simple Exclusion Process (ASEP). However, in this paper, we refer to the ASEP as  $p$ -CA184 because  $p$  is a key parameter associated with thermodynamic characteristics, which are emphasized in this study.

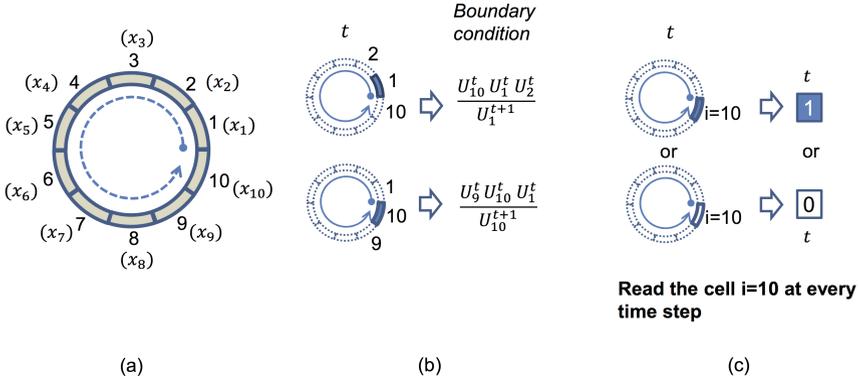
Moreover, since the initial density of states,  $d$ , determines the maximum number of "1" clusters in a state, we define the  $(d, p)$ -CA184 as the  $p$ -CA184 calculated with the initial density of states  $d$ . Here, the density of states  $d$  is the ratio of the number of "1"s to the total number of cells in the cell set under consideration.

### 2.5.3 Periodic ( $d, p$ )-CA184 and data generation

The  $(d, p)$ -CA184 introduced in this paper consists of 10 cells in space, denoted as  $(x_i, 1 \leq i \leq 10)$ , and employs periodic boundary conditions (Fig. 4 (a)-(b)). At the boundaries,  $x_{10}$  and  $x_1$ , the cell value at time  $t + 1$  is determined by referring to the cells on both sides at time  $t$ . The boundary conditions are established by applying the rule-184 to  $(U_{10}^t U_1^t U_2^t)/U_1^{t+1}$  and  $(U_9^t U_{10}^t U_1^t)/U_{10}^{t+1}$  (Fig. 4 (b)). The CA data is generated by reading the cell value of  $x_{10}$  at each time step  $t$  (Fig. 4 (c)).

### 2.5.4 Equilibrium and non-equilibrium CA

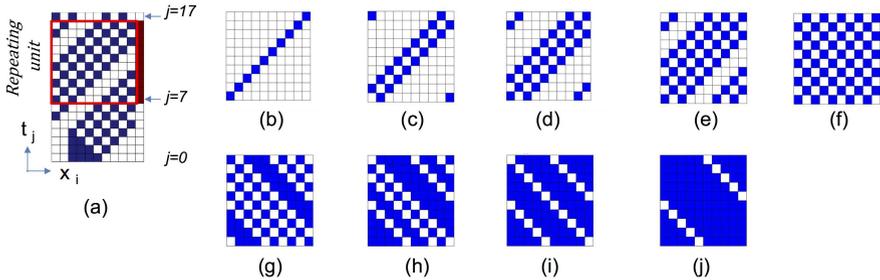
When generating data using periodic  $(d, p)$ -CA184 under the condition of  $p = 1$ , after sufficient time steps, a periodicity in time occurs in the  $(x_i, t_j)$  space, and the same pattern is repeated in the  $x - t$  diagram (Fig. 5). From the  $x - t$  diagram, we can extract a square unit of 10 space-steps  $\times$  10 time-steps, that repeats over time. In this repeating space-time unit, the number of "1" clusters per row, which defines the thermodynamic state, remains constant, implying



**Fig. 4** Periodic  $(d, p)$ -CA184 and data generation. (a) Periodic CA. (b) Periodic boundary condition. (c) Data generation.

that the repeating unit is in equilibrium. Therefore, periodic  $(d, p = 1)$ -CA184 is an equilibrium CA.

It is important to note that the periodic  $(d, p < 1)$ -CA184 is non-equilibrium because the number of “1” clusters per row generally varies over time in the periodic  $(d, p < 1)$ -CA184.



**Fig. 5** Equilibrium CA. The blue and white cells indicate respectively  $U_i^j = 1$  and  $U_i^j = 0$ .  $t_j$  and  $x_i$  are discretized time and discretized space, respectively.  $d$  is the cell density obtained by dividing the number of occupied cells in a row by the total number of cells in a row. (a) A  $x-t$  diagram exhibiting the periodicity in time. Since  $U_x^{17} = U_x^7$ , the blue-white pattern between  $j = 7$  and  $j = 16$  is repeated after  $j = 17$ . (b) Periodic  $(d, p = 1)$ -CA184 with  $d = 0.1$ . (c) Periodic  $(d = 0.2, p = 1)$ -CA184. (d) Periodic  $(d = 0.3, p = 1)$ -CA184. (e) Periodic  $(d = 0.4, p = 1)$ -CA184. (f) Periodic  $(d = 0.5, p = 1)$ -CA184. (g) Periodic  $(d = 0.6, p = 1)$ -CA184. (h) Periodic  $(d = 0.7, p = 1)$ -CA184. (i) Periodic  $(d = 0.8, p = 1)$ -CA184. (j) Periodic  $(d = 0.9, p = 1)$ -CA184.

## 3 Methods

### 3.1 Equivalence between the “Process to GEJE” and CA transition

The commutative diagram in Fig. 6 is introduced to show the equivalence between the “Process to GEJE” and the non-equilibrium-to-equilibrium thermodynamic transition of CA. The “Process to GEJE”, which is the transition from GVF to catastrophic rupture, is denoted by the right-pointing horizontal arrow indicated by “(a)” in Fig. 6. Our objective is to demonstrate that this arrow “(a)” corresponds to the arrow “(f)” which indicates the thermodynamic transition from the periodic ( $d = 0.5, p < 1$ )-CA184 to the periodic ( $d = 0.5, p = 1$ )-CA184.

It is important to note that the CA depicted in Fig. 6 is periodic, although it is not explicitly mentioned.

The arrow “(b1)” indicates that  $EPR(10days)$  is positive in GVF, which will be further explained in a later section. The arrow “(b2)” signifies that a positive  $EPR(10days)$  corresponds to a non-equilibrium state, as per the definition of  $EPR$ . The arrow “(b3)” implies that the non-equilibrium state corresponds to the periodic ( $d = 0.5, p < 1$ )-CA184, which holds true based on the definition of the periodic ( $d = 0.5, p < 1$ )-CA184. In this context, the periodic ( $d = 0.5, p < 1$ )-CA184 is specifically chosen because previous studies have shown that the GVF is equivalent to the periodic ( $d = 0.5, p < 1$ )-CA184 (Kikuchi, 2023), as indicated by the dashed arrow “(b)”.

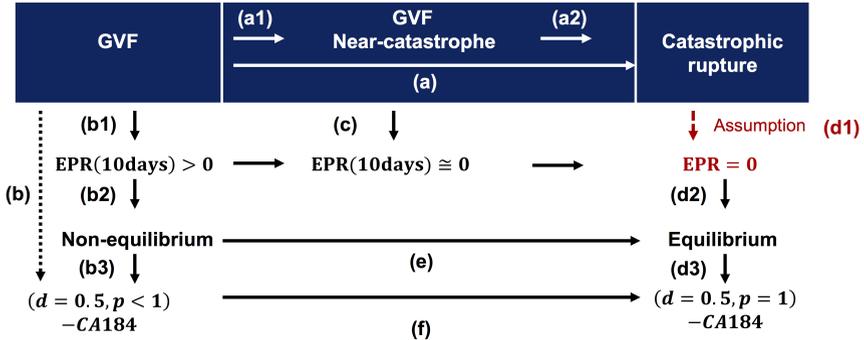
Note that the “Process to GEJE” necessarily includes the sub-processes “(a1)” and “(a2)” via “Near-catastrophe,” as the catastrophe itself cannot be discussed within the framework of the GVF. We attempt to infer the nature of the catastrophe by examining the “Near-catastrophe”. The arrow “(c)” indicates that  $EPR(10days)$  is approximately zero near the catastrophe (This is an observation and will be discussed in a later section). By extension, we assume that  $EPR$  is exactly zero at the “Catastrophe” (dashed arrow “(d1)”). In this context, the data length for the  $EPR$  is not specified, as the time scale of the corresponding “Catastrophic event” is unknown and expected to be much shorter than 10 days.

Then, the arrow “(d2)” is generated based on the definition of  $EPR$ . It is worth noting that the “Process to GEJE” can be viewed as the transition from the non-equilibrium state ( $EPR(10days)$ ) to equilibrium ( $EPR = 0$ ), as indicated by the arrow “(e)”. The arrow “(d3)” is generated with reference to the equilibrium properties of periodic ( $d = 0.5, p = 1$ )-CA184. Finally, the arrow “(f)” is generated in parallel to the arrow “(e)” since the transition from periodic ( $d = 0.5, p < 1$ )-CA184 to periodic ( $d = 0.5, p = 1$ )-CA184 represents the transition from a non-equilibrium state to an equilibrium state, a transition already indicated by the arrow “(e)”.

Here, the periodic ( $d = 0.5, p = 1$ )-CA184 is specifically chosen as the destination of the non-equilibrium-to-equilibrium transition from the periodic

( $d = 0.5, p < 1$ )-CA184 since the periodic ( $d = 0.5, p < 1$ )-CA184 exists and fluctuates very close to the periodic ( $d = 0.5, p = 1$ )-CA184.

So far, we have demonstrated that the “Process to GEJE” (“(a)”) is equivalent to the transition from the periodic ( $d = 0.5, p < 1$ )-CA184 to the periodic ( $d = 0.5, p = 1$ )-CA184 (“(f)”), assuming  $EPR = 0$  at the “Catastrophic rupture” (“(d1)”) and postponing the discussion on “(b1)” and “(c)”. Now, we need to further explain the thermodynamics of GVF, specifically focusing on “(b1)” and “(c)”.



**Fig. 6** Commutative diagram to discuss equivalence between “Process to GEJE” and CA transition. Although not described, the CAs are periodic.

### 3.1.1 Thermodynamics of the GVF (Fig. 6 (b1) and (c))

This subsection completes the deferred discussion on Fig. 6 (b1) and (c). The  $EPR(10days)$ , which represents the  $EPR$  calculated every 10 days, and the logarithmic  $A$ -noise are extracted from the previous paper (Kikuchi, 2023) and presented in Fig. 7 along with earthquake events. The elapsed time has been extended by 20 months from the original plot, covering the period from 00:00 on January 1, 2006, to 23:59 on August 30, 2018. Fig. 7 (a) illustrates the location of the KSN (white-filled circle), where the ground vibration data was recorded, the measurement point of the seismic intensity (blue-filled square), and the epicenters of the earthquakes that recorded a seismic intensity of 5 or greater (red plus-circle). These earthquakes are identified as  $A, B, C,$  and  $D$ , with earthquake  $B$  being the GEJE of magnitude 9. The table in Fig. 7 includes the identifiers, dates, magnitudes, seismic intensities, and epicenters of the earthquakes, retrieved from the website of the Japan Meteorological Agency of the Ministry of Land, Infrastructure, Transport and Tourism (JMA, 2019). The search conditions used are a seismic intensity greater than 4, the location of seismic intensity observation, and the search period. For the GVF of the UD velocity data acquired every 0.05 seconds at KSN,  $EPR$  is calculated every 10 days, and the results are shown in Fig. 7 (b). Fig. 7 (c) illustrates the time evolution of positive  $A$ -noise in a log-scale.

Fig. 7 (b) shows that  $EPR(10days)$  is consistently positive, confirming the validity of the arrow "(b1)" in Fig. 6, which indicates that  $EPR(10days)$  is positive in the GVF.

From the beginning of 2008 to the end of 2014 (731 to 3285 days),  $EPR(10days)$  occasionally decreases to a minimum of 0.03 (Fig. 7 (b)). The timing of the  $EPR(10days)$  decrease is similar to the occurrence of a strong peak of positive  $A-noise$  (Fig. 7 (c)). For instance, near the timing of the GEJE, marked by the green vertical line "B",  $EPR(10days)$  drops to around 0.04 (red solid circle), while  $\log_{10}(A-noise)$  sharply increases and reaches approximately 0.5, the maximum axis value (black dot). The sharp decrease in  $EPR(10days)$  near the GEJE suggests that "Near-catastrophe" in GVF corresponds to  $EPR(10days) \approx 0$  and validates the arrow (c) in Fig. 6. Consequently, by assuming  $EPR = 0$  at the rupture, we have demonstrated that the "Process to GEJE" is equivalent to the transition from the periodic ( $d = 0.5, p < 1$ )-CA184 to the periodic ( $d = 0.5, p = 1$ )-CA184.

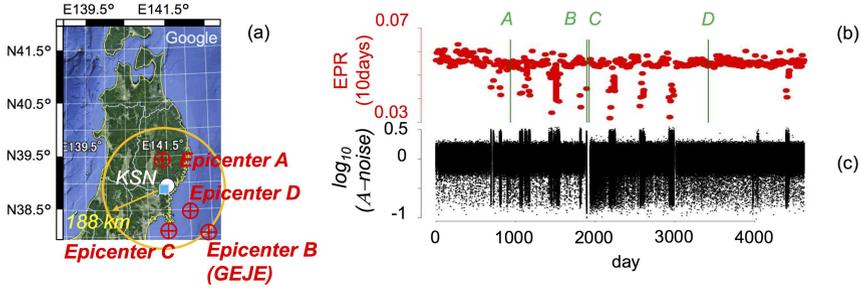
Since the state of  $EPR(10days) = 0$  corresponds to the equilibrium state in which no state transition occurs, the decrease in  $EPR(10days)$  implies a change from a non-equilibrium state towards an equilibrium state. Therefore, the arrow (e) in Fig. 6 is confirmed to be correct.

The decrease in  $EPR(10days)$  does not coincide with the timing of earthquakes A, C, and D, which have magnitudes less than or equal to 7.2. However, the timing of the GEJE (earthquake B of magnitude 9) is close to the sharp decrease in  $EPR(10days)$ . This implies that a state change from non-equilibrium toward equilibrium is a phenomenon associated only with earthquakes of magnitude greater than 7.2.

### 3.2 Derivation of thermodynamic state functions in CA

We are searching for thermodynamic state functions that fulfill two conditions: their significant change should occur only once prior to the GEJE, and there should be a clear thermodynamic state shift associated with it. The first condition, the uniqueness of the significant change of the thermodynamic state function, is based on the prior observation that the onset of the crustal stress manifestation occurs only once before the GEJE, and we believe this manifestation likely corresponds to a thermodynamic state shift or a significant change in the state function. Earthquakes, positive peaks in  $A-noise$ , and sharp decreases in  $EPR$  do not uniquely appear before the GEJE and are unsuitable as thermodynamic state functions. We expect the thermodynamic state functions to be weaker than these signals and to manifest during the seismically silent period. To detect these weak signals, we focus on the distribution of thermodynamic parameters rather than the thermodynamic parameters themselves. One potential candidate is the  $A-noise$  distribution.

We are interested in analyzing the large positive  $A-noise$ , so we employ Extreme Value Analysis (EVA), which statistically evaluates the positive maximum values of the subsets within a given sample of  $A-noise$ . In EVA, the  $r$  stochastic variables are divided into  $s$  subsets, and the maximum value

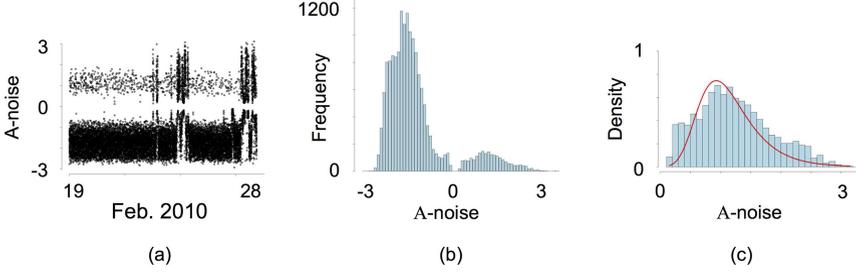


ID	Date	Magnitude	Seismic intensity	Latitude	Longitude	Depth(km)
A	Jul. 24, 2008	6.8	5	39.43N	141.38E	108
B	Mar. 11, 2011	9	6	38.06N	142.51E	24
C	Apr. 07, 2011	7.2	5	38.12N	141.55E	66
D	May 13, 2015	6.8	5	38.52N	142.09E	46

**Fig. 7** The  $EPR(10days)$  and  $A\text{-noise}$  at KSN during the period from 2006-01-01 00:00 to 2018-08-30 23:59, and the major earthquakes detected near KSN during the same period. For the earthquakes with a seismic intensity greater than 4, the identifier (uppercase letters A, B, C, and D), time, magnitude, seismic intensity, and epicenter location are summarized in the table. (a) The location of the KSN (white filled circle), the earthquake epicenter (red plus-circle), and the point of seismic intensity measurement (blue-filled square). The seismic intensity is recorded at Kesenuma-city, approximately 10 km from KSN, which is 188 km from the GEJE epicenter. (b)  $EPR(10days)$ . The green line indicates the timing of the earthquake identified by the uppercase letter at the top of the line. (c) Positive  $A\text{-noise}$ .

of each subset is extracted to create a new set  $M_s$  comprising the positive maximum values. This new set undergoes statistical operations. According to the Fisher-Tippett-Gnedenko theorem (Charras-Garrido & Lezaud, 2013), after the proper normalization,  $M_s$  converges in distribution to one of three distributions: the Gumbel distribution, the Fréchet distribution, or the Reversed-Weibull distribution (Eq. (8)). In the Gumbel distribution in Eq. (8), for the limit of  $\gamma \rightarrow 0$ , the term  $1 + \gamma z$  is approximated as  $1 + \gamma z \approx \sum_{n=0}^{\infty} (\gamma z)^n / n! = e^{\gamma z}$ . We opt for the Gumbel distribution as it fits well with the frequency distribution of the positive  $A\text{-noise}$  data, as shown in Fig. 8.

$$\begin{aligned}
 G_{\gamma}(z) &= \exp\left(- (1 + \gamma z)^{-1/\gamma}\right), \quad 1 + \gamma z \geq 0, \quad \gamma, \beta, x \in \mathbb{R}, \quad \gamma > 0 \\
 \text{Gumbel : } G_0(x) &= \exp(-e^{-x}) \\
 \text{Fréchet : } G_{1/\beta}(\beta(x-1)) &= \exp(-x^{-\beta}) \\
 \text{Reversed - Weibull : } G_{-1/\beta}(\beta(x+1)) &= \exp(-(-x)^{\beta}). \quad (8)
 \end{aligned}$$



**Fig. 8** Distribution of *A-noise*. Recorded at KSN from 2010-02-19 00:00 to 2010-02-28 23:59. (a) Time evolution of *A-noise*. (b) Frequency distribution of *A-noise*. (c) Frequency distribution of positive *A-noise* and its Gumbel approximation (red curve).

### 3.2.1 Protocol of the EVA with Gumbel distribution

Extreme Value Analysis (EVA) is conducted on the *A-noise* sequence calculated from either binarized GVF or the binary sequence generated by CA. For the binarized GVF, the *A-noise* is calculated as described in Eq. (2). When applying Eq. (2) to the binary sequence generated by CA, we need to set  $\Delta t \equiv 1$  and  $t_0 \equiv 1$ . Consequently,  $t_j''$  becomes  $t_j'' = 1 + 1024(j - 1)$ .

In EVA, for both of the binarized GVF and CA, the  $r$  sets of *A-noise* ( $=\{A-noise(t_j''), A-noise(t_{j+1}'') \cdots, A-noise(t_{j+r-1}'')\}$ ) is divided into 20 subsets to create  $M_{30-s}$ ;  $s = 1, 2, \dots, 20$ , with each subset containing  $r/20$  *A-noise* values.  $M_{30-s}$ , and  $\omega$ , which contains 160 sets of *A-noise* intervals, are defined as (Eq. (9)). Eq. (9) indicates that for  $(s, j1) = (1, 1)$ ,  $M_{30-s}(\tau_{j1}) = M_{30-1}(\tau_1) = \{A-noise(t_j'') \mid j = j2 = 1, 2, \dots, r/20\} = \{A-noise(t_1''), A-noise(t_2''), \dots, A-noise(t_{r/20}'')\}$ , which is the first  $r/20$  terms of the *A-noise* sequence in Eq. (2). For  $(s = 1, 2, \dots, 20, j1 = 1)$ ,  $M_{30-s=1,2,\dots,20}(\tau_1) = \{A-noise(t_1''), A-noise(t_2''), \dots, A-noise(t_r'')\}$ , which is the first  $r$  terms of the *A-noise* sequence in Eq. (2), and on which the first EVA is conducted.

$$\begin{aligned}
 M_{30-s} &\equiv M_{30-s}(\tau_{j1}) \equiv M_{30-s}(t_{(j1-1)r+1}'') = M_{30-s}(t_0 + 1024(j1 - 1)r\Delta t) \\
 &\equiv \{A-noise(t_j'') \mid j = (j1 - 1)r + (s - 1)r/20 + j2\} \\
 &\quad j1 = 1, 2, \dots, N/1024/r \\
 &\quad j2 = 1, 2, \dots, r/20 \\
 \omega &= \{\omega_{j3} \equiv [-4.0 + (j3 - 1) \times 0.05, -4.0 + j3 \times 0.05]\} \\
 &\quad j3 = 1, 2, \dots, 160.
 \end{aligned} \tag{9}$$

For each  $\tau_{j1}$ , the histogram of  $M_{30-s}$  with class interval  $\omega$  is calculated, and then the maximum  $\omega_j$  with non-zero frequency is extracted and defined as *A-noise*<sub>s</sub>. A new set of  $M_{20-j1}$  consisting of 20 sets of *A-noise*<sub>s</sub> is formed to approximate the histogram of  $M_{20-j1}$  with a Gumbel distribution using class

interval  $\omega$ . In the Gumbel approximation, the input data  $x_j \equiv (\omega_j + \omega_{j+1})/2$  represents the mean of the class interval  $\omega_j$ , and the input data  $y_j$  represents the frequency of occurrence for the  $\omega_j$  in the histogram for  $M_{20-j1}$ . Gumbel fitting is applied in the domain of  $x_j > 0$ .

For the Gumbel fitting,  $G_0(x)$  in Eq. (8) is rewritten as follows:

$$G_0(x) = \exp\left(-\exp\left(-\frac{x-\lambda}{\eta}\right)\right)$$

$$\Rightarrow -\log(-\log(G_0(x))) = \frac{1}{\eta}x - \frac{\lambda}{\eta}, \quad -4.0 \leq x < 4.0 \quad (10)$$

where  $-\log(-\log(G_0(x)))$  is expressed as the linear function of  $x$ . The cumulative distribution,  $G_0(x_j)$ , is calculated from  $y_j$  as shown in Table 1, where  $y_{sj}$  denotes the cumulative frequency,  $G_0(x_j) \equiv y_{sj}/(y_{total} + 1)$  is the cumulative distribution, and  $y_{total} \equiv \sum_{i=1}^{160} y_i$ . The 20 pairs of  $(x_j, G_0(x_j))$  are fitted to a Gumbel distribution using linear regression, and the gradient, y-intercept, and Pearson's coefficient in Eq. (11) are calculated.

**Table 1** A detailed example of calculating the cumulative distribution  $G_0(x_j)$ .

Class(j)	$\omega_j$	$x_j$	$y_j$	$\sum_{i=1}^j y_i \equiv y_{sj}$	$G_0(x_j) \equiv y_{sj}/(y_{total} + 1)$
1	[0, 0.05)	0.025	2	2	$2/(y_{total} + 1)$
2	[0.05, 0.1)	0.075	3	5	$5/(y_{total} + 1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
160	[3.95, 4.0)	3.975	1	$y_{total}$	$y_{total}/(y_{total} + 1)$

*Gumbel fitting parameters :*

$$\text{gradient} = \frac{1}{\eta} \text{ in Eq. (10)}$$

$$\text{Intercept} = -\frac{\lambda}{\eta} \text{ in Eq. (10)}$$

$$R2 = \text{Pearson's correlation coefficient} \quad (11)$$

### 3.2.2 Thermodynamic state function derived with the periodic ( $d = 0.5, p$ )-CA184

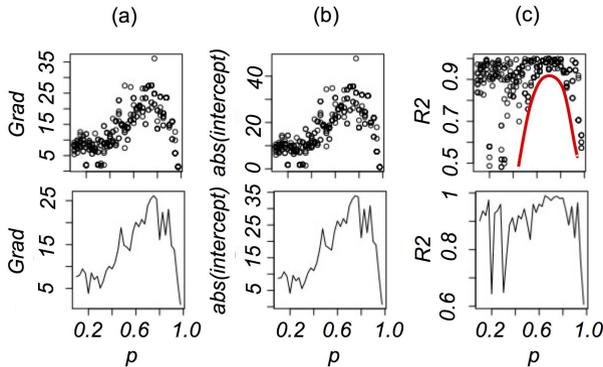
The Gumbel fitting parameters defined by Eq. (11) are calculated for the periodic ( $d = 0.5, p$ )-CA184 and plotted with respect to the hop probability  $p$  as shown in Fig. 9. The bottom graph in Fig. 9 shows the top graph averaged for each  $p$ . The *Grad*, *abs(intercept)*, and *R2* represent the gradient, absolute value of intercept, and Pearson's correlation coefficient, respectively (Fig. 9 (a)-(c)). It should be noted that the *A-noise* in CA is notationally identical to

the definition in Fig. 1, but is calculated from the data generated by CA. For each of the 36 cases of  $p = \{0.1, 0.125, \dots, 0.95, 0.975\}$ , the fitting parameter is calculated for 9 initial conditions. Each fitting parameter is calculated from a binary sequence of length  $2.15 \times 10^7$ , corresponding to  $r=20,996$  *A-noise* values, resulting in  $M_{30-s}$  consisting of  $r/20 = 1049$  *A-noise* values.

Large values are observed around  $p = 0.7$  in the averaged graphs of gradient, absolute intercept, and  $R2$  (lower graphs in Fig. 9 (a)-(c)). Additionally, for  $R2$ , the data fluctuation is significantly reduced near  $p = 0.7$ , and the bottom-line envelope of  $R2$  forms a convex curve (red curve in the upper graph in Fig. 9 (c)). To approximate this convex curve, the 100 moving minimum of  $R2$  is defined as  $\min R2(\tau_{j1}) \equiv \min(R2(\tau_{j1}), R2(\tau_{j1+1}), \dots, R2(\tau_{j1+99}))$ .

If we have a system in which the hop probability increases over time, for example, if  $R2(p)$  corresponds to  $R2(t)$ , we would expect to see large values of  $R2$  and large values of  $\min R2$  before the hop probability reaches 1. Therefore,  $R2$  and  $\min R2$  can be appropriate thermodynamic state functions for the transition of the CA from periodic ( $d = 0.5, p < 1$ )-CA184 to periodic ( $d = 0.5, p = 1$ )-CA184.

Assuming that the hop probability increases with time, we propose a 500-Simple-Moving-Average of  $R2$  and a 100-moving minimum of  $R2$  (denoted as  $\min R2$ ) as thermodynamic state functions for the transition of the CA.



**Fig. 9** Curve fitting parameters evaluated in the periodic ( $d = 0.5, p$ )-CA184. The top graph consists of 324 points, with 9 points per each  $p$  value, while the bottom graph displays the average for each  $p$  value. (a) Gradient calculated in the Gumbel approximation. (b) The absolute value of the Intercept. (c) Pearson's correlation coefficient between the fit and the source data.

## 4 Results

### 4.1 Thermodynamic state shift prior to GEJE

In this section, we evaluate the thermodynamic state functions, the 500-Simple-Moving-Average- $R2$  and  $\min R2$ , on the GVF data to discover a

thermodynamic state shift prior to the GEJE and rationalize the assumption in Fig. 6.

Note that the thermodynamic state functions derived in CA are applicable to GVF. We assume the maximum possible hop probability,  $p = 1$ , at the timing of GEJE, causing the hop probability to increase with time in the period leading up to GEJE. This condition satisfies the precondition for deriving the thermodynamic state functions in CA, where it is considered that the hop probability increases with time.

#### 4.1.1 Evaluation of thermodynamic state functions in GVF

In each EVA calculation on the GVF data, the total number of *A-noise* is defined as 600 ( $r = 600$ ) and  $\Delta t$  is  $1/20$  as defined in Eq. (1). Each EVA is performed on the binary sequence of  $\overline{BS}(t''_{600(j1-1)+1} \leq t_n < t''_{600 j1+1}; \Delta t = 1/20)$ .

For the GVF data (source data for Fig. 7 (b)-(c)) recorded at KSN from 2006-01-01 00:00 to 2018-08-30 23:59, the thermodynamic state function, the *minR2*, is evaluated and plotted in Fig. 10 (a). The two prominent peaks labeled "State shift" are visible in Fig. 10 (a) around  $t=664$  days, which is 1231 days before the GEJE. The GEJE occurred on  $t=1895$  days, indicated by the short green vertical line "B". The short green vertical lines labeled "A", "B", "C", and "D" correspond to the earthquake timings shown in the table in Fig. 7. The dashed red stepped line in Fig. 10 (a) is defined at time point  $t$  as the maximum value of *minR2* prior to and including time point  $t$ . Given that the red dashed line remains unaltered following the peaks labeled as "State shift," these two peaks stand out as the highest within the elapsed time range, making them both unique and significant. Consequently, they can be regarded as the sought-after indicators of the thermodynamic state shift. Fig. 10 (b) shows that the 500-Simple-Moving-Average-R2 also reaches its maximum at a similar timing to the *minR2* state shift, suggesting that this maximum is another indicator of the thermodynamic state shift prior to the GEJE.

After the timing of the thermodynamic state shift prior to GEJE, indicated by the vertical dotted line "STM", the probability density of states  $p_1$  and  $p_2$  of the GVF are clearly separated, and the value of  $p_1$  becomes smaller than that of  $p_2$  (Fig. 10 (c)). The fluctuation in the lower level of  $p_1$  than  $p_2$  continues even after the timing of the thermodynamic state shift prior to GEJE, extending up to the GEJE occurrence. This suggests that the GVF state undergoes a state shift towards the GEJE state at the timing of the "State shift" prior to GEJE. In this context, it is important to remember that The GEJE, the "Catastrophic rupture" is assumed to correspond to the periodic ( $d = 0.5, p = 1$ )-CA184 as shown in Fig. 6 (d1)-(d3). Furthermore, it should be noted that the thermodynamic state of the periodic ( $d = 0.5, p = 1$ )-CA184 is independent of time, consistently comprising five clusters of "1," which is the potential maximum number of the cluster in a given state (as indicated in Fig. 5 (f)). Therefore, the GEJE is assumed to be the state  $s_5$  with the probability density of  $\{p_i \mid p_5 = 1.0 \text{ and } p_i = 0; i = 1, 2, 3, 4\}$ . Following the

“State shift” prior to GEJE, in a GVF state, a decrease in the  $p_1$  component, representing the state with a single “1” cluster, leads to an increase in the other components,  $p_i$ ;  $i = 2, 3, 4, 5$ , corresponding to states with more than one “1” cluster. Consequently, the total count of “1” clusters in the GVF state increases after the “State shift”, causing the GVF state to move closer to the GEJE state, characterized by the maximum number of “1” clusters.

Fig. 10 (d) illustrates the temporal evolution of the  $R+A\text{-noise}(10\text{days})$ , which is defined as the ratio of the number of positive  $A\text{-noise}$  values to the total number of  $A\text{-noise}$  values within a 10-day data block of  $A\text{-noise}$ . The peak in  $R+A\text{-noise}(10\text{days})$  in Fig. 10 (d) aligns with the peak in  $A\text{-noise}$  in Fig. 10 (e), as indicated by the vertical red line shown between these figures. Previous research (Kikuchi, 2023) has demonstrated that the peaks in  $A\text{-noise}$  are linked to the stress peaks induced by seasonal storm surges. Furthermore, it has been shown that  $R+A\text{-noise}(10\text{days})$  corresponds to the stress development underground: a large  $R+A\text{-noise}(10\text{days})$  corresponds to a high-stress level.

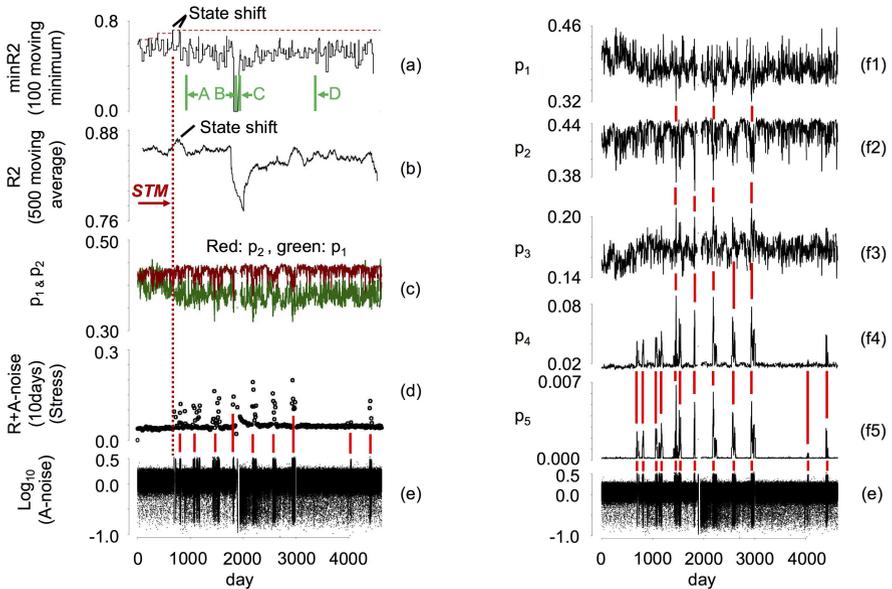
The stress peak, or the  $R+A\text{-noise}(10\text{days})$  peak, becomes first visible at the “State shift” timing (STM) during the elapsed time (Fig. 10 (d)), implying that a change in the underground-stress state undergoes at the “State shift” timing.

Fig. 10 (f1)-(f5) shows the time evolution of the state probability density  $p_i$ ;  $i = 1, 2, 3, 4, 5$ . The  $p_5$  graph exhibits prominent upward peaks (Fig. 10 (f5)) that align with the peaks in the  $A\text{-noise}$  peaks (Fig. 10 (e)), as indicated by the red vertical line between these figures, suggesting that  $p_5$  corresponds to the stressed state. Similarly, the graph for  $p_4$  also displays pronounced upward peaks corresponding to the  $A\text{-noise}$  peaks (Fig. 10 (f4)).  $p_3$  behaves similarly to  $p_4$ , but with fewer upward peaks (Fig. 10 (f3)). On the contrary,  $p_2$  exhibits downward peaks that correspond to the upward peaks of  $p_3$  (Fig. 10 (f2)). Similarly,  $p_1$  contains downward peaks that correspond to the three downward peaks of  $p_2$  (Fig. 10 (f1)).

Hence, increasing stress results in higher probabilities for states  $p_3$ ,  $p_4$ , and  $p_5$ , while decreasing probabilities for states  $p_1$ , and  $p_2$ . It’s conceivable that a high stress level corresponds to the state with a higher  $p_5$  component. Considering the intuitive assumption that the maximum stress is observed at the GEJE timing, it’s plausible that the state near GEJE closely resembles the  $p_5$  state, or the  $s_5$  state depicted in Fig. 3 (c5), generated by  $(d = 0.5, p = 1)$ -CA184 (Fig. 5 (f)). This finding supports the initial assumption ( $d1$ ) and the deductions  $d2$  and  $d3$  presented in Fig. 6, indicating that the catastrophic rupture (GEJE) is indeed associated with the  $(d = 0.5, p = 1)$ -CA184.

## 5 Conclusions

Assuming  $EPR = 0$  or equilibrium at the catastrophic rupture in the “Process to GEJE”, we applied the data mining method to extract the thermodynamic



**Fig. 10** Evolution of thermodynamic state function and GVF dynamics. The elapsed time starts and ends at 2006-01-01 00:00 and 2018-08-30 23:59. (a) The thermodynamic state function  $minR2$  is the 100 moving minimum of  $R2$ . The short green vertical lines labeled “A”, “B”, “C”, and “D” indicate the earthquake timings shown in the table in Fig.7. The red dashed line is the maximum value of  $minR2$  before time  $t$  including  $t$ . The vertical dotted line “STM” indicates the “State shift” timing. (b) Pearson correlation coefficient  $R2$  (500-Simple-Moving-Average). (c) Probability density of state,  $p_1$  (green line) and  $p_2$  (red line). (d)  $R+A$ -noise, calculated every 10days and corresponding to subsurface stress. The vertical red line in the bottom shows the correspondence between the  $R+A$ -noise and  $A$ -noise peaks. (e)  $A$ -noise in log10 scale. (f1)-(f5) State probability density  $p_i$ ;  $i = 1, 2, 3, 4, 5$ . Red vertical lines connect the prominent peaks between (e) and (f5), (f5) and (f4), (f4) and (f3), (f3) and (f2), and (f2) and (f1).

state shift prior to the GEJE from the GVF recorded at the seismic station KSN. We first demonstrated the thermodynamic equivalence between the “Process to GEJE”, defined as the transition from the GVF to catastrophic rupture, and the transition in CA from periodic ( $d = 0.5, p < 1$ )-CA184 to periodic ( $d = 0.5, p = 1$ )-CA184. Subsequently, we derived thermodynamic state functions that yield a thermodynamic state shift in the CA transition and evaluated these thermodynamic state functions in the GVF to discover the thermodynamic state shift prior to GEJE.

The thermodynamic state function of hop probability  $p$ , denoted as  $R2 \equiv R2(p)$ , was derived as the Pearson’s correlation coefficient in the curve fitting using the Gambel approximation in the EVA applied to the  $A$ -noise calculated from periodic ( $d = 0.5, p < 1$ )-CA184. The  $R2$  showed a significant increase around the hop probability  $p = 0.7$ . Considering a system in which time elapses as  $p$  increases, the  $R2 \equiv R2(p) \sim R2(t)$  can be considered a thermodynamic state function of the transition from periodic ( $d = 0.5, p < 1$ )-CA184 to periodic ( $d = 0.5, p = 1$ )-CA184. Similarly, the 100-point moving minimum of

$R2$ , denoted as  $minR2$ , which approximates the lower envelope of  $R2(t)$ , is considered another thermodynamic state function of the CA transition.

The thermodynamic state shift prior to GEJE were discovered by evaluating the  $R2$  for the  $A$ -noise calculated from the GVF recorded at KSN during the period from January 2006-01-01 00:00 to 2018-08-30 23:59. Two significantly large peaks of the 500-Simple-Moving-Average- $R2$  appeared side by side only once, 1231 days before the GEJE. A large  $minR2$  peak also appeared at similar times to the  $R2$  peaks. Therefore, the large values of  $R2$  and  $minR2$  can be regarded as thermodynamic state shift prior to GEJE.

It is important to note that the thermodynamic state functions derived in CA are suitable for applying to GVF. At the timing of GEJE, we assumed the maximum possible hop probability  $p = 1$ , so the hop probability increases with time in the period before GEJE, which is coherent with the precondition for deriving the thermodynamic state functions in CA: the hop probability increases with time.

The assumption that the hop probability  $p = 1$  at the GEJE (Fig. 6 (d3)), or equivalently that the GEJE corresponds to  $ERR = 0$  (Fig. 6 (d1)), is rationalized as follows. It has been observed that the high-stress level corresponds to the state with a higher  $p_5$  component. Considering the intuitive assumption that the maximum stress is observed at the GEJE timing, it's plausible that the state near GEJE closely resembles the  $p_5$  state, or the  $s_5$  state depicted in Fig. 3 (c5), generated by ( $d = 0.5, p = 1$ )-CA184 (Fig. 5 (f)).

The scope of this research is limited to the specific earthquake GEJE and the data recorded at a specific seismic station KSN. Further investigation is required for generalization.

## Declarations

### Author's Contribution

Hiroyuki Kikuchi contributed to all works on this manuscript.

### Availability of Data and Materials

All data used in this study are publicly available through the: National Research Institute for Earth Science and Disaster Resilience, National Research and Development Corporation under Ministry of Education, Culture, Sports, Science and the Technology. F-Net (Broadband seismograph network) database. <http://www.fnet.bosai.go.jp/top.php?LANG=en>.

### Funding

No funding was obtained for this study.

## Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

## References

- Nishinari, K. & Takahashi, D. (1999). Analytical properties of ultradiscrete Burgers equation and rule-184 cellular automaton. *Journal of Physics A: Mathematical and General*. 31. 5439. DOI: 10.1088/0305-4470/31/24/006
- SEED (2012). Standard for the Exchange of Earthquake Data, SEED Format Version 2.4. Published by the Incorporated Research Institutions for Seismology, United States Geological Survey
- NIED (2019). National Research Institute for Earth Science and Disaster Resilience, National Research and Development Corporation under Ministry of Education, Culture, Sports, Science and Technology. F-Net (Broadband seismograph network) data base. DOI: 10.17598/nied.0005.  
<http://www.fnet.bosai.go.jp/top.php?LANG=en>
- Charras-Garrido, M., Lezard, P. (2013). Extreme Value Analysis : an Introduction. *Journal de la Société Française de Statistique*, 154, 2, 66-97.  
<https://hal-enac.archives-ouvertes.fr/hal-00917995>
- JMA (2019). Japan Meteorological Agency under the Ministry of Land, Infrastructure, Transport and Tourism. Seismic intensity database search.  
<https://www.data.jma.go.jp/svd/eqdb/data/shindo/index.php>
- Ito, S. (2020). Special Feature: Physics and Information Geometry. *Mathematical Sciences*, No. 689, 38-45 (November 2020). ISSN 0386-2240. Printed in Japan
- Kikuchi, H. (2023). Data mining method in seismology by applying cellular automaton equivalence of ground vibration fluctuations recorded near the epicenter of the 2011 Mw 9 East Japan earthquake. *Earth Science Informatics* 16, 2615?2633. <https://doi.org/10.1007/s12145-023-01054-z>