ORIGINAL ARTICLE

Research of influence of sample size on normal information diffusion based on the Monte Carlo method: risk assessment for natural disasters

Jiafu Liu[1] · Shuai Li[1] · Jin Wu[2] · Xiaojing Liu[1] · Jiquan Zhang[3]

Abstract

Information diffusion techniques and Monte Carlo methods have been widely used in solving all kinds of problems of small samples with incomplete data in the field of natural disaster risk assessment such as environmental resource rating, flood monitoring and temperature changing. Data are not the only thing that matters for natural disaster risk assessment, but with enough data, we can accurately predict the time, place, scale and loss of future disasters. It is important to simulate the complete data scene when there is a minimum sample size. In this paper, we collect temperature data from 3050 meteorological stations in China and use the Monte Carlo simulation method to investigate the effect of sample size on estimating the normal information diffusion. The results show that (1) for the same sample, the information diffusion method is significantly better than the traditional histogram method. (2) Using the hard histogram estimation method, the recommended sample size is 85 or more, which is slightly larger than the traditional threshold value (i.e., 30), while using the information diffusion estimation method, the recommended sample size decreases to 45 or more. (3) Simulation experiments show that, with insufficient samples, both estimation methods, i.e., the information diffusion and the traditional histogram methods become invalid because of its poor correlation, low robustness, high RMSE and variance values. These results indicate that the Monte Carlo simulation method and information diffusion technique have certain practical reference value in the research of natural disaster risk assessment in the case of a small sample.

Keywords Monte Carlo method · Sample size · Information diffusion · Histogram estimation · Correlation coefficient

Introduction

In the process of natural disaster risk analysis and assessment, the sample size is an important part which can affect the precision of assessment results. If the sample size is too small, it will lead to inaccurate and meaningless evaluation results (Kong et al. 2015; Myint et al. 2008; Parsons et al. 2016; Wang et al. 2015a, b). However, in some cases, it is difficult to

extract sufficient information from incomplete data from a few samples (Hao et al. 2014; Li et al. 2013). To solve the problem of lack of information caused by small sample sizes, Professor Chong-fu Huang proposed the information distribution and diffusion method in 1995 (Kazama et al. 2012; Nagata and Shirayama 2012). The method is now widely applied to the risk analysis of a variety of disasters including earthquakes, floods, hailstorms, city fires, loess collapsibility, and rainstorms (Levitan and Wronski 2013; Xue and Gencay 2012; Li et al. 2012; Bai et al. 2014; Mundahl and Hunt 2011). The method is recognized by the industry as the most effective way to manage information incompleteness (Feng and Luo 2008; Lin 2015; Maillé and Saint-Charles 2014).

But what is information incompleteness? The mathematical expressions of information completeness are derived from its rigorous theory (Xu et al. 2013). The theoretical definition of information completeness is too abstract and it is difficult to guide the practical activities based on the theory. A sample size smaller than 30 is statistically considered to cause

✉ Jiafu Liu
liujiafu750506@126.com

1 College of Tourism and Geographical, Sciences, Jilin Normal University, Siping 136000, China

2 Academy of Disaster Reduction and Emergency Management, Beijing Normal University, Beijing 100875, China

3 Department of Environment, Institute of Natural Disaster Research, Northeast Normal University, Renmin Street, Changchun 130024, China

information incompleteness (Sun and Zhang 2008; Huang 2005; Gamboa et al. 2015; Malterud et al. 2015; Perneger et al. 2015). Although 30 is given as the threshold for a small sample, the concept of the small sample itself is very vague (Baio et al. 2015; de Bekker-Grob et al. 2015; Dembkowski et al. 2012; Engblom et al. 2016; Schorr et al. 2014).

In the field of environmental geosciences, when studying the natural factors such as precipitation, temperature and number of earthquakes, the size of the sample is directly related to the accuracy of the final result. Therefore, from a statistical point of view, we attempt to collect data of 3050 weather stations in China as the experimental raw data. The Monte Carlo method is used to resample the original experimental data to build a probability distribution interval with a certain confidence. By comparing the information diffusion method with the traditional hard histogram method, a method can be found that can better reflect the actual state under the condition of small sample, so that an intuitive understanding can be provided when the related scientific workers deal with incomplete information. (Andreotti et al. 2014; Dehghani et al. 2014; Elanique et al. 2012; Haar et al. 2017; Medhat et al. 2017; Newhauser et al. 2007).

Data and methods

Data and processing

Considering the amount of data have a great impact on the experimental results, it requires a large amount of data as the basic data source to test which method can be used to obtain better results in different sample sizes and find out what is the minimum sample size if we want to make the data results good enough. This study selected China as the study area. Its topography is high in the west and low in the east. The climate in China is complex and diverse but distributed in an orderly manner from south to north. China locates in the northern hemisphere, spans six temperature zones. The temperature rises from north to south and also has an obvious change with season. There are many meteorological stations in the country, which are densely populated in the south and sparsely in the north. The overall temperature conditions meet the requirements of the experiment. We collect all the weather data in June 2016 from weather stations across the country in the database of China Weather Network (http://www.weather.com.cn/). First, the weather site data of extreme weather or abnormal temperature was removed through the screening method, and then averaging the remaining data, finally we got a total of 3050 temperature data of the weather stations (Fig. 1).

Regarding the selected temperature data as the parent Ω, the Monte Carlo method is used to resample data in the parent data. Collect n samples ($n = 5, 6, 7,\ldots, 1000$) randomly,

and then collect each of these NTH samples we got just now. In the case of different pseudo-random numbers, repeat the experiment m times ($m = 100$) [11,17]. We count all the sample data generated by re-sampling (we can name it data set A for distinction) using the traditional histogram statistics method (hard histogram) and the improved histogram method under the information allocation (soft histogram) separately to get dataset B. In the next step, we estimate data set B and calculate the expected and standard deviations of the two histogram estimates separately. By comparing the correlation coefficient and the average value (expectation value) of the correlation coefficient between the actual value and the theoretical value of the traditional hard histogram and information diffusion model, the experimental conclusion can be obtained. The core of the experimental design consists of two parts: (1) the histogram estimate and (2) the impact of sample size on normality information diffusion measure design (Fig. 2):

Monte Carlo simulation method

The Monte Carlo (MC) method, also known as the random sampling method, belongs to a branch of computational mathematics. It was developed in the 1940s for the field of atomic energy. The traditional method could not produce results close to the actual values; thus, it was difficult to obtain satisfactory results. However, since the Monte Carlo method is able to simulate the actual physical processes, it is capable of getting reasonable results. Currently, the Monte Carlo method is widely used in various subjects (Andreotti et al 2014; Dehghani et al. 2014; Elanique et al. 2012; Gamboa et al. 2015; Gregory and Graves 2004).

The Monte Carlo (MC) method is very different from general calculation methods. The general calculation method for solving multi-dimensional complex problems is very difficult, unlike the Monte Carlo method. Monte Carlo analysis is based on direct tracing of the particles and thus the physical aspect is relatively clear and easy to understand. It uses the random sampling method that simulates the particle transport process reflecting the statistical fluctuation of the law without the complexity of multidimensional, multi-factor models. The limit is a good way to solve complex particle transport problems with a clear and simple MC program structure. It is easier to obtain intermediate results with MC models and they can be applied with great flexibility to a variety of problems (Del Moral et al. 2011; Haar et al. 2017; Medhat et al. 2017; Tang et al. 2016; Vadapalli et al. 2014).

Besides the advantages, the MC method has a few shortcomings: one of them is the slow convergence speed problem and error in models of probabilistic nature. Increasing the number of simulations to reduce the sampling error greatly increases the computational intensity of the model (Ableidinger et al. 2017; Dornheim et al. 2015; Newhauser et al. 2007; Wang et al. 2015a, b).
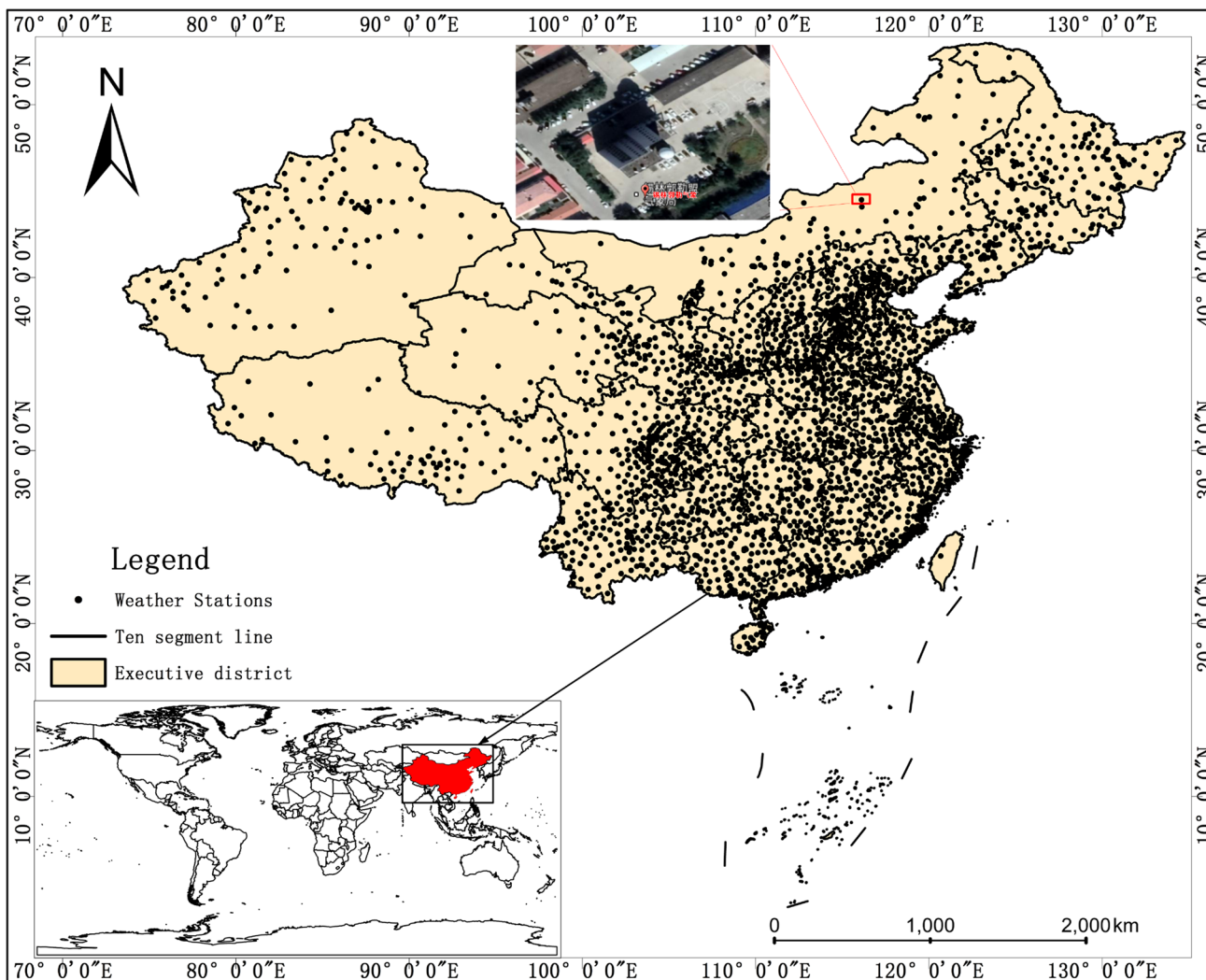
Fig. 1 Study area map

If experimental calculations are performed directly on the basis of existing data, there are some interference conditions that cannot be avoided which can lead to errors in the experimental results. Based on the advantages of the MC method, many limitations can be avoided early in the experiment, and real-world sample data can be simulated more realistically. Especially in the case of a small sample state, the MC method is very suitable to resample the existing data again so that we can get accurate experimental results.

The principle of information diffusion and diffusion estimation

According to the principle of information diffusion, the maternal probability density function estimation is defined as diffusion estimation, $n > 0$, is a constant, so (Bai et al.

2015; Hao et al. 2014; Li et al. 2013; Liang et al. 2012; Lin 2015; Maillé and Saint-Charles 2014; Shin 2009):

$$\hat{f}(y) = \frac{1}{n\Delta_n} \sum_{i=1}^{n} \mu\left(\frac{y - y_i}{\Delta_n}\right). \tag{1}$$

Type (1) is a diffusion estimation of maternal probability density function $f(y)$, $\mu(x)$ is called the diffusion function and n is the window width,

$$x = \frac{y - y_i}{\Delta_n}. \tag{2}$$

The concrete forms of diffusion $\mu(x)$ are the key to the diffusion estimation. For different $\mu(x)$ values, the results can predict different diffusion values. According to the theory of molecular diffusion, the function of normal diffusion can be deduced as:
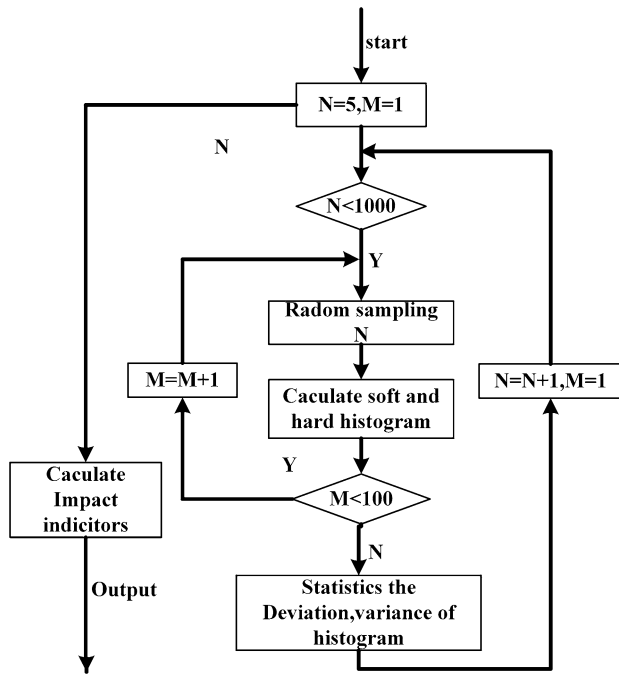
Fig. 2 Monte Carlo simulation flowchart

$$\mu(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{i=1}^{n} \exp\left[-\frac{(y-y_i)^2}{2h^2}\right]. \tag{3}$$

Therefore, substituting type (3) in type (1), the normal diffusion estimation of the maternal probability density function can be deduced as:

$$\hat{f}(y) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^{n} \exp\left[-\frac{(y-y_i)^2}{2h^2}\right], \tag{4}$$

where $h = \sigma\Delta_n$ was considered as the window width of the standard normal diffusion. From type (4), the normal diffusion estimation $f(y)$ of the parent probability density is not only related to the observed values y_i and the numbers of the observed values n, but also related to the window width H of the standard normal diffusion.

When the observations are completed, the observed values and the number of observed values n are known, and the window width h is unknown. For confirming the window width H, a simple calculation method is adapted for checking the width of the window. According to the close principle, the empirical formula of window width h can be deduced as:

$$h = \begin{cases} 0.8146(b-a), & n=5 \\ 0.5690(b-a), & n=6 \\ 0.4560(b-a), & n=7 \\ 0.3860(b-a), & n=8 \\ 0.3362(b-a), & n=9 \\ 0.2986(b-a), & n=10 \\ 2.6851(b-a)/(n-1), & n\geqslant 11 \end{cases},$$

where $b = \max\{x_i\}, a = \min\{x_i\}; \quad i \in (1,n)$.

Histogram estimation

The hard histogram model is a traditional histogram model. To construct a traditional histogram model, it needs to be divided into N intervals of width h, then set the midpoint of each interval and select the control point interval. The soft histogram estimation is based on the traditional histogram model through the distribution of information obtained by improving the histogram model. According to the below equation, we calculate the probability of each random sample point that falls within each section:

$$\mu(x_i) = \frac{1}{h_x\sqrt{2\pi}} \exp\left[-\frac{(u_j-u_i)^2}{2h_x^2}\right]. \tag{5}$$

The following equation calculates the cumulative probability:

$$q_j = \sum_{i=1}^{n} \frac{1}{h_x\sqrt{2\pi}} \exp\left[-\frac{(u_j-u_i)^2}{2h_x^2}\right], \tag{6}$$

where x_i is the ITH sample point, $h_x = 0.5$, u_j is the midpoint of the interval for J.

Normalization of different interval probabilities is presented in the following equation:

$$q_j = q_j / \sum q_i. \tag{7}$$

To investigate the influence of the sample size on information diffusion, this experiment uses two indicators: the correlation coefficient and root mean square error [Eqs. (8), (9)]. The correlation coefficient (r) is a measure of the degree of correlation between variables. Its value lies between -1 and 1. For $|r|$ close to 1, the correlation between two variables is stronger.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}, \tag{8}$$

where $\bar{x}, \bar{y}$ are the mean values of the variables.

The root mean square error is a measure of the difference between two variables and is defined by:

$$rmse = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2 / (n-1)}. \qquad (9)$$

Conclusions

To make the experimental procedure simply and select the subsequent impact indicators conveniently in the experiment, first of all, based on the understanding of normal probability density function, this paper divides this curve into nine intervals on [0,54], the midpoint of normal distribution interval is [3,9,15,21,27,33,39,45,51], and then the function is integrated in each interval to get the normal distribution probability under each interval, corresponding to the midpoint of the normal distribution one by one, which is [0.0 401,0.0655,0.121,0.1747,0.1974,0.1747,0.121,0.0655,0.04 01]. The results are shown in Table 1. Then the resulting

Table 1 Normal probability density function value in the cumulative probability under each section

Extent	Midpoint	Probability
[0, 6]	3	0.0401
(6,12]	9	0.0655
(12,18]	15	0.1210
(18,24]	21	0.1747
(24,30]	27	0.1974
(30,36]	33	0.1747
(36,42]	39	0.1210
(42,48]	45	0.0655
(48, 54]	51	0.0401

probability density values are subjected to a hard histogram density estimation operation (Table 1).

According to the above experimental scheme, the data are compiled and implemented using Python. Using Monte Carlo simulation results, the influence of the sample size on the diffusion of normal information is shown in Figs. 3 and 4, respectively.

Figure 3 shows the results of 100 Monte Carlo simulations for the soft histogram coefficient of the correlation between the theoretical and simulated value and hard histogram coefficient of the correlation between the theoretical and simulated value of the mean (expected). It can be obviously seen that the correlation coefficient of the theoretical value of the soft histogram estimation is much better than that of the traditional hard histogram, which shows that the soft histogram estimation method can get better experimental results.

Figure 4 shows the results of 100 Monte Carlo simulations for the soft RMSE and hard histogram coefficient of the correlation between the theoretical and simulated value of the mean (expected). The conclusions can be drawn from the Fig. 3 and Fig. 4 that (1) with the increase in sample size, for both the traditional and soft histogram estimation method. This is in line with the statistics of the large numbers theorem; hard and soft histogram estimation methods are more likely to converge compared to the histogram method. (2) With the increase in the sample size, for both soft and hard estimation methods, the sample variance converges. (3) To achieve the same RMSE or correlation coefficient value, the required sample size is far less when the soft approach is used (45 or more) as the sample estimation method than when the hard approach is used (85 or more).
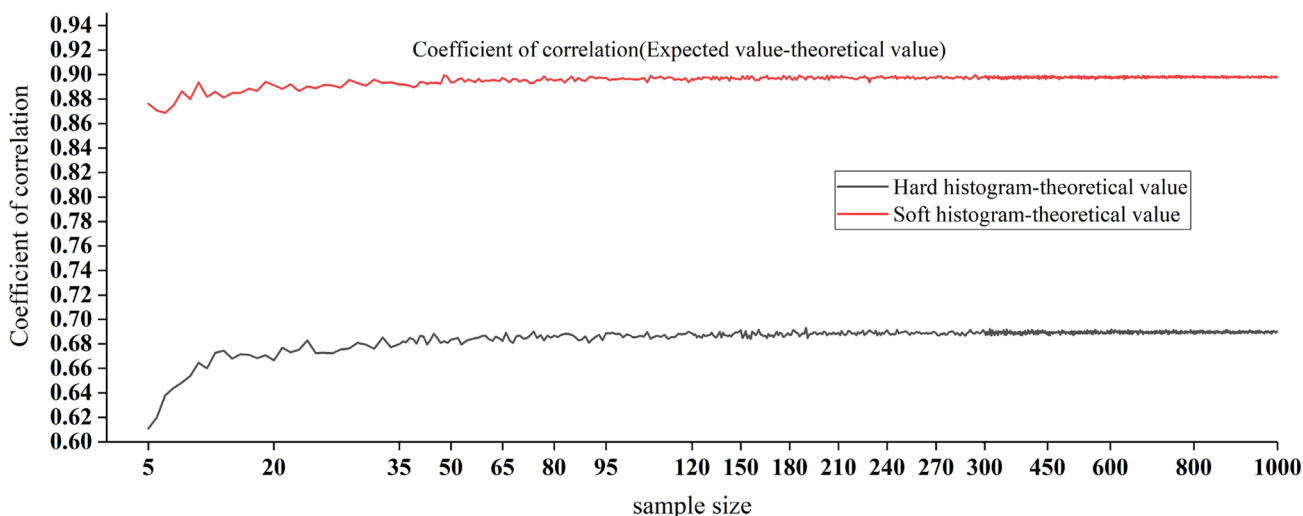

Fig. 3 Sample size estimation for normal diffusion (coefficient of correlation)
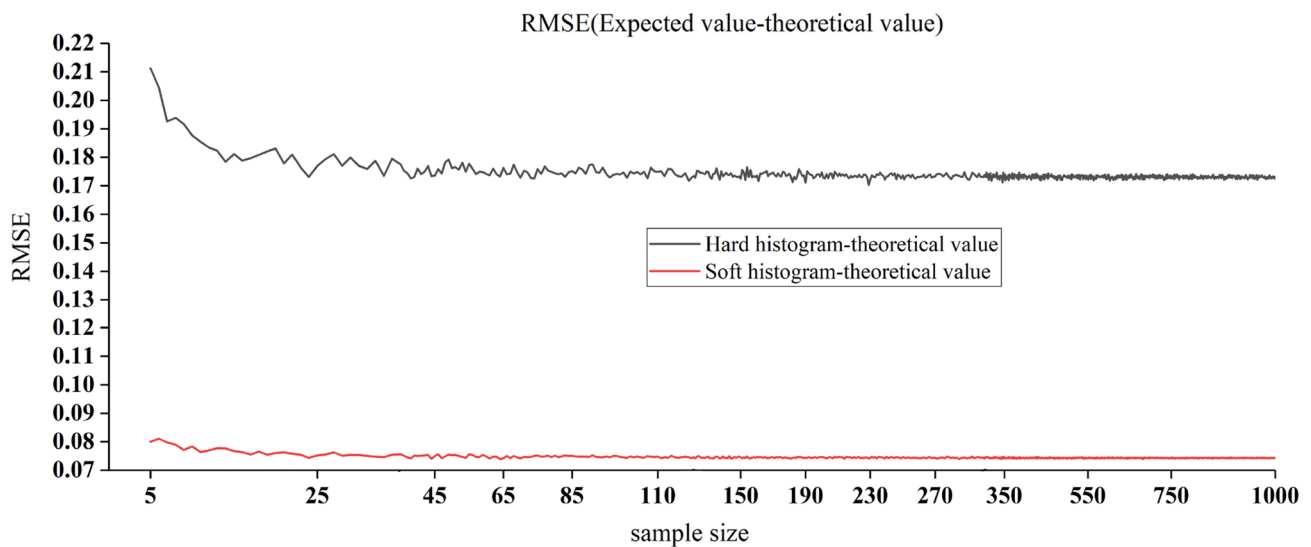
Fig. 4 Effect of sample size on normal diffusion estimation (RMSE)

Discussion

Figures 3 and 4 reveal the unique advantages of the soft estimation methods in solving the problem of small sample sizes (incomplete samples). Small samples can be used with the traditional histogram method for the large sample estimation of the effect of the same law; this is also the main reason why the current method mentioned in the paper is for complete samples under non-extensive applications.

After processing the air temperature data, we found that if the traditional hard histogram estimation method is used, the ideal convergence state will be achieved when the sample size is above 85, while the soft histogram estimation method, can obtain good results when the sample size reaches 45; so, to get the best sample capture, use a sample size of at least 45. Although the soft histogram method requires fewer samples than the traditional hard histogram method, the data obtained are far superior to the traditional hard histogram method. However, with very few sample data, both the traditional hard histogram estimation method and the soft histogram estimation method cannot get good results, resulting in a large final result error. The sample size is still a key factor.

The experimental results show that when the sample size reaches the best capture value, the histogram will achieve the best convergence. After that, the trend is stable and the fluctuation is very slow. Even if the sample volume is increased, there will be no obvious change. Therefore, to reduce the consumption of calculation time, when the sample size reaches the best capture, that is to say, when the sample size of the soft histogram method is 45, stop the sample size increase and complete the simulation of the sample size.

For small samples, neither soft nor hard estimation methods can effectively solve the problem of incomplete information from the process under the inversion of reality. In current research efforts, scientists mostly use soft estimation methods for the analysis and evaluation of natural disaster risks. They use smaller samples to analyze the potential uncertainties associated with these analyses and to determine whether they could satisfy users' production practice needs. This could be a future research topic.

The traditional sample size for the community to determine whether the information is complete is 30, which is considered unreasonable. Since the sample size is small, the points are relatively large and, therefore, they may be blurry between the complete and incomplete information regions. Therefore, the soft estimation method is in fact not a good solution to the problem of incomplete information. A better solution to this problem is needed to increase the model accuracy.

Compliance with ethical standards

Conflict of interest The authors confirm that this study has no conflicts of interest.

References

Ableidinger M, Buckwar E, Thalhammer A (2017) An importance sampling technique in Monte Carlo methods for SDEs with a.s. stable and mean-square unstable equilibrium. J Comput Appl Math 316:3–14. https://doi.org/10.1016/j.cam.2016.08.043. (**Accepted 9 August 2016**)

Andreotti E, Hult M, Marissens G, Lutter G, Garfagnini A, Hemmer S, von Sturm K (2014) Determination of dead-layer variation in HPGe detectors. Applied radiation isotopes 87:331–335. https://doi.org/10.1016/j.apradiso.2013.11.046. (**Accepted 18 November 2013**)

Bai C-Z, Zhang R, Hong M, Qian L-x, Wang Z (2015) A new information diffusion modelling technique based on vibrating string equation and its application in natural disaster risk assessment. Int J Gen Syst 44:601–614

Baio G, Copas A, Ambler G, Hargreaves J, Beard E, Omar RZ (2015) Sample size calculation for a stepped wedge trial. Trials 16:1–15. https://doi.org/10.1080/03081079.2014.980242. (**Accepted 21 Oct 2014**)

Bekker-Grob EWD, Donkers B, Jonker MF, Stolk EA (2015) Sample size requirements for discrete-choice experiments in healthcare: a practical guide. The patient. Patient Cent Outcomes Res 8:373–384. https://doi.org/10.1007/s40271-015-0118-z

Dehghani M, Saghafian B, Nasiri Saleh F, Farokhnia A, Noori R (2014) Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. Int J Climatol 34:1169–1180, https://www.researchgate.net/publication/258395784

Del Moral P, Doucet A, Jasra A (2011) An adaptive sequential Monte Carlo method for approximate Bayesian computation. Stat Comput 22:1009–1020, https://doi.org/10.1007/s11222-011-9271-y. (**Accepted 7 July 2017**)

Dembkowski DJ, Willis DW, Wuellner MR (2012) Comparison of four types of sampling gears for estimating age-0 yellow perch density. J Freshw Ecol 27:587–598. https://doi.org/10.1080/02705060.2012.680932

Dornheim T, Groth S, Filinov A, Bonitz M (2015) Permutation blocking path integral Monte Carlo: a highly efficient approach to the simulation of strongly degenerate non-ideal fermions. New J Phys arXiv:1504.03859

Elanique A, Marzocchi O, Leone D, Hegenbart L, Breustedt B, Oufni L (2012) Dead layer thickness characterization of an HPGe detector by measurements and Monte Carlo simulations. Applied radiation and isotopes: including data, instrumentation and methods for use in agriculture. Ind Med 70:538–542. https://doi.org/10.1016/j.apradiso.2011.11.014. (**Accepted 4 November 2011**)

Engblom H, Heiberg E, Erlinge D, Jensen SE, Nordrehaug JE, Dubois-Rande JL, Halvorsen S, Hoffmann P, Koul S, Carlsson M et al (2016) Sample size in clinical cardioprotection trials using myocardial salvage index, infarct size, or biochemical markers as endpoint. J Am Heart Assoc 5:e002708. https://doi.org/10.1161/JAHA.115.002708

Feng L, Luo G (2008) Flood risk analysis based on information diffusion theory. Hum Ecol Risk Assess Int J 14:1330–1337. https://doi.org/10.1080/10807030802494691. (**Accepted 02 Feb 2008**)

Gamboa F, Janon A, Klein T, Lagnoux A, Prieur C (2015) Statistical inference for Sobol pick-freeze Monte Carlo method. Statistics 50:881–902. https://doi.org/10.1080/02331888.2015.1105803. (**Accepted 05 Oct 2015**)

Gregory A, Graves (2004) A Monte Carlo simulation of biotic metrices as a function of subsample size. J Freshw Ecol 19(2):195–201. https://doi.org/10.1080/02705060.2004.9664532. (**Accepted 12 Nov 2003**)

Haar T, Kamleh W, Zanotti J, Nakamura Y (2017) Applying polynomial filtering to mass preconditioned hybrid Monte Carlo. Comput Phys Commun 215:113–127, https://doi.org/10.1016/j.cpc.2017.02.020 (**Get rights and content. Accepted 19 February 2017**)

Hao L, Yang LZ, Gao JM (2014) The application of information diffusion technique in probabilistic analysis to grassland biological disasters risk. Ecol Model 272:264–270. https://doi.org/10.1016/j.ecolmodel.2013.10.014. Accepted 13 October 2013

Huang CF (2005) Risk assessment of natural disaster (theory and practice). Science Press, Beijing

Kazama K, Imada M, Kashiwagi K (2012) Characteristics of information diffusion in blogs, in relation to information source type. Neurocomputing 76:84–92. https://doi.org/10.1016/j.neucom.2011.04.036

Kong LY, Tao YC, Liu J, Wu AH (2015) Study on integrated spatial database for composite urban disaster risk assessment. Appl Mech Mater 738–739:285–288. https://doi.org/10.4028/www.scientific.net/AMM.738-739.285

Levitan L, Wronski J (2013) Risk prediction model of LNG terminal station based on information diffusion theory. Procedia Engineering 52:60–66. https://doi.org/10.1016/j.proeng.2013.02.106

Li Q, Zhou JZ, Liu DH, Jiang XW (2012) Research on flood risk analysis and evaluation method based on variable fuzzy sets and information diffusion. Saf Sci 50:1275–1283. https://doi.org/10.1016/j.ssci.2012.01.007. (**Accepted 14 January 2012**)

Li ZR, Lu TJ, Shi WH, Zhang XH (2013) Predicting the scale of information diffusion in social network services. J China Univ Posts Telecommun 20:100–104. https://doi.org/10.1016/S1005-8885(13)60239-3

Liang Z, Xu BY, Jia Y, Zhou B (2012) Mining evolutionary link strength for information diffusion modeling in online social networks. Appl Mech Mater 157–158:567–572

Lin S (2015) information diffusion and momentum in laboratory markets. J Behav Finance 16:357–372. https://doi.org/10.1080/15427560.2015.1095757

Maillé M-È, Saint-Charles J (2014) Fuelling an environmental conflict through information diffusion strategies. Environ Commun 8:305–325. https://doi.org/10.1080/17524032.2013.851099

Malterud K, Siersma VD, Guassora AD (2015) Sample size in qualitative interview studies: guided by information power. Qual Health Res 26:1753–1760, https://doi.org/10.1177/1049732315617444

Medhat ME, Abdel-hafiez A, Singh VP (2017) Optimization of fast neutron flux in an irradiator assembly using Monte Carlo simulations. Vacuum 138:105–110. https://doi.org/10.1016/j.vacuum.2017.01.026. (**Accepted 24 January 2017**)

Mundahl ND, Hunt AM (2011) Recovery of stream invertebrates after catastrophic flooding in southeastern Minnesota, USA. J Freshw Ecol 26:445–457. https://doi.org/10.1080/02705060.2011.596657. (Accepted 12 May 2011)

Myint SW, Yuan M, Cerveny RS, Giri4 C (2008) Categorizing natural disaster damage assessment using satellite-based geospatial techniques. Nat Hazards Earth Syst Sci 8:707–719. https://doi.org/10.5194/nhess-8-707-2008

Nagata K, Shirayama S (2012) Method of analyzing the influence of network structure on information diffusion. Phys A 391:3783–3791. https://doi.org/10.1016/j.physa.2012.02.031

Newhauser W, Fontenot J, Zheng Y, Polf J, Titt U, Koch N, Zhang X, Mohan R (2007) Monte Carlo simulations for configuring and testing an analytical proton dose-calculation algorithm. Phys Med Biol 52:4569–4584, https://doi.org/10.1088/0031-9155/52/15/014/meta

Parsons M, Glavac S, Hastings P, Marshall G, McGregor J, McNeill J, Morley P, Reeve I, Stayner R (2016) Top-down assessment of disaster resilience: a conceptual framework using coping and

adaptive capacities. Int J Disaster Risk Reduct 19:1–11,. https://doi.org/10.1016/j.ijdrr.2016.07.005. (**Accepted 19 July 2016**)

Perneger TV, Courvoisier DS, Hudelson PM, Gayet-Ageron A (2015) Sample size for pre-tests of questionnaires. Qual Life Res 24:147–151, https://doi.org/10.1007/2Fs11136-014-0752-2

Schorr RA, Ellison LE, Lukacs PM (2014) Estimating sample size for landscape-scale mark-recapture studies of North American migratory tree bats. Acta Chiropterologica 16:231–239. https://doi.org/10.3161/150811014X683426

Shin JK (2009) Information accumulation system by inheritance and diffusion. Phys A 388:3593–3599. https://doi.org/10.1016/j.physa.2009.05.032

Sun CZ, Zhang X (2008) Risk assessment of agricultural drought in Liaoning Province based on information diffusion. Syst Sci Compr Stud 24:507–510

Tang D-S, Hua Y-C, Cao B-Y (2016) Thermal wave propagation through nanofilms in ballistic-diffusive regime by Monte Carlo simulations. Int J Therm Sci 109:81–89. https://doi.org/10.1016/j.ijthermalsci.2016.05.030. (**Accepted 28 May 2016**)

Vadapalli U, Srivastava RP, Vedanti N, Dimri VP (2014) Estimation of permeability of a sandstone reservoir by a fractal and Monte Carlo simulation approach: a case study. Nonlinear Process Geophys 21:9–18. https://doi.org/10.5194/npg-21-9-2014. (**Accepted: 12 Nov 2013**)

Wang D, Sun S, Tse PW (2015a) A general sequential Monte Carlo method based optimal wavelet filter: a Bayesian approach for extracting bearing fault features. Mech Syst Signal Process 52–53:293–308. https://doi.org/10.1016/j.ymssp.2014.07.005. (**Accepted 7 July 2014**)

Wang Y, Zhang J, Guo E, Sun Z (2015b) Fuzzy comprehensive evaluation-based disaster risk assessment of desertification in Horqin Sand Land, China. Int J Environ Res Public Health 12:1703–1725. https://doi.org/10.3390/ijerph120201703. (**Accepted: 22 January 2015**)

Xu LF, Xu XG, Meng XW (2013) Risk assessment of soil erosion in different rainfall scenarios by RUSLE model coupled with information diffusion model: a case study of Bohai Rim. China Catena 100:74–82. https://doi.org/10.1016/j.catena.2012.08.012. (**Accepted 24 August 2012**)

Xue Y, Gencay R (2012) Hierarchical information and the rate of information diffusion. J Econ Dyn Control 36:1372–1401. https://doi.org/10.1016/j.jedc.2012.03.001. (**Accepted 1 March 2012**)