Identifying Critical Regions in Industry Infrastructure: A Case Study of a Pipeline Network in Kansas, USA

PENG HOU\textsuperscript{1}, XIAOJIAN YI\textsuperscript{1,2}, AND HAIPING DONG\textsuperscript{1}

\textsuperscript{1}Beijing Institute of Technology, Beijing 100081, China
\textsuperscript{2}Academy of Mathematics and Systems Science, Chinese academy of sciences, Beijing 100190, China

Corresponding authors: Xiaojian Yi (yixiaojianbit@sina.cn) and Haiping Dong (donghaipingphd@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71801196.

\begin{abstract}
In the face of the budget cuts and increased size of industry infrastructure, one of the top priorities for industry infrastructure protection is to identify critical regions by vulnerability analysis. Then, limited resources can be allocated to those critical regions. Unfortunately, difficulties can be observed in existing approaches of vulnerability analysis. Some of them are unavailable due to the insufficient data. Others are susceptible to human biases. Here, we propose an approach to overcome these difficulties based on the location data of failure events. The critical geographic regions are determined by the risk ranking of different candidate regions. Risk is calculated by integrating the probability of the failure event occurring (risk uncertainty) and total failure cost (the severity of failure consequences) in each candidate region. By changing the modeled object from the components to the region where the whole industry infrastructure is located, it collects the rarely failure events which are dispersed in different positions of the industry infrastructure to provide sufficient data, then the probability can be obtained by using a Poisson point process and kernel density estimation. Meanwhile, the application of hypothesis testing avoids the susceptibility of the approach to human biases by verifying the correctness of the assumptions used in the approach. Finally, a case study of this approach is performed on a pipeline network in Kansas, USA. In addition to the validation of the feasibility of our approach, risk uncertainty is proven to be less instructive for identifying critical regions than the severity of failure consequences.
\end{abstract}

\begin{IEEEkeywords}
Critical region, critical industry infrastructure, Poisson point process, risk assessment, vulnerability analysis.
\end{IEEEkeywords}

\section{I. INTRODUCTION}
The stable, continuous and safe operation of critical industry infrastructure is significantly affected by different types of hazards. These hazards have caused many failure events in critical industry infrastructures. Due to these failure events, the provision of industry infrastructure services is disrupted. The stability of the economy or national defense will be seriously affected. Therefore, critical industry infrastructure protection has attained great importance in different countries [1]–[4].

As an important part of industry infrastructure protection, the identification of critical regions is to determine the regions where the failure of elements (assets or components) of an industry infrastructure constitutes a grave threat to modern life [5]. However, this is not an easy task in practice. With the increasing size and complexity of industry infrastructure, many types of factors and many elements need to be considered in the task. The identification of critical regions becomes a costly and time-consuming process for a whole industry infrastructure. Therefore, a screening methodology is proposed.

In the screening methodology, the fragilities of an industry infrastructure are modeled with the consideration of different hazards [6]. Widely used modeling approaches include empirical approaches and network-based approaches according to the reference [7]. Based on the description of the research paradigm in figure 1, internal hazards (such as...
corrosion, fatigue and damage) and external hazards (the loading, operating environment and accidents) will act on assets (or components) of an industry infrastructure after those assets are exposed to those hazards. The actions might impact those elements. This triggers the deterioration, fatigue or other failure mechanisms of those elements, which results in failure events with time. Finally, the consequences are an unintentional release of hazardous contents, the loss of production and so on. Clearly, the consequences caused by failures are the effective symptoms of the vulnerability of industry infrastructure and can be measured. Then, critical regions or elements can be identified. Currently, the difference between empirical approaches and network-based approaches is mainly in how to measure the consequences.

In empirical approaches [8]–[11], risk is used to measure the failure consequences of components in industry infrastructure. Risk includes risk uncertainty and severity. Risk uncertainty is measured by the probability of component failure, and severity describes the severity of the consequence caused by the failure of components. Based on the assessed risk of different components, the critical components can be determined in industry infrastructure, and the regions where the failure components are located are the critical regions. For network-based approaches, the industry infrastructure is modeled by a topological network. The nodes of the network denote the elements of the industry infrastructure. Under the assumption that the operation of the industry infrastructure is based on the interdependence of the components (the network connectivity is full), the action of a hazard is the removal of the node that denotes the failure component. The impact is that the removal significantly degrades network connectivity. The consequence is the severity of the disconnection of the topological network. In general, the measure of the disconnection includes the degree of the nodes, the global clustering coefficient, and the average shortest path length [12]–[15]. In this paper, the proposed approach is an empirical approach. Therefore, some recent empirical approaches are reviewed in the following pages.

Despite making a significant contribution to the identification of critical regions, empirical approaches largely apply empirical data, such as expert judgments [9], [11], [16]–[26]. For this reason, empirical approaches deeply rely on the judgment of engineers, which makes these approaches subject to human biases and errors. To overcome this problem, approaches based on machine learning have been proposed [27]–[30]. With the help of machine learning, empirical approaches remove the influence of subjective judgment by refining insights from big data. Unfortunately, insufficient data is a barrier for empirical approaches to apply machine learning. In practice, the number of times that failures happen is small for a component or an asset in many industry infrastructures, which means the amount of data is not enough for each component of an industry infrastructure to build an available machine learning model.

Against this backdrop, an improved empirical approach is proposed. By changing the modeled object from the components of industry infrastructure to the regions in which a whole industry infrastructure is located, the failure data will be the aggregation of the rarely failure events which are dispersed in different positions of the industry infrastructure. Therefore, there are sufficient data to use the proposed approach.

Hypothesis testing is used to verifying the satisfaction of the properties of the model: Poisson point process before modelling, which avoids the susceptibility of the approach to human biases in analysis. Kernel density estimation is used to learn the model parameter with the location data of failure events. Therefore, the empirical data is not used, which effectively reduces the impact of subjective preference in empirical data.

By using the built model, the probability of the failure event occurring is estimated in different geographic regions where the industry infrastructure is located. At the same time, the severity of failure consequences is measured based on the total failure cost. Based on that, risks are assessed in those regions and used to identify critical areas.

The rest of this paper includes four sections. Section II explains the application of the proposed approach. Then, a study case is proposed in section III, which is studied based on performing a group of experiments. Section IV details the results, and the findings are discussed. Finally, the conclusions are provided in section V.

II. METHOD

The proposed approach includes four steps: verification of the availability of the Poisson point process, model selection, model building and risk assessment. The figure 2 describes the process of the proposed approach.

A. POISSON POINT PROCESS

As a widely used spatial statistical technique, a Poisson point process N can model the data of spatial points. The data of spatial points is called a spatial point pattern X. It is a description of the outcome of a Poisson point process [31]. In theory, the definition of a Poisson point process N is based on a region W ⊂ ℝ². In the region W, we can observe the spatial point pattern \( X = \{x_1, x_2, \ldots, x_n\} \) whose element is the point \( x_i \in ℝ² \). Thus, the regions W is also called the observation window. In our research, the point is the failure event that occurs in an industry infrastructure. The amount of times that a failure event occurs \( n(B_i) \) is a well-defined random variable in a region \( B_i \subset W \), and the region \( B_i \) can be arbitrary shapes and any location in the observation window W. All the elements of a Poisson point process are intuitively illustrated in figure 3.
To define a Poisson point process on the given observation window $W$, the following properties are satisfied [31]:

1) Poisson distribution of point counts: $n(B_i) \sim \text{poi}(\Lambda(B_i))$.

2) Independent scattering: the random variables $n(B_1), \ldots, n(B_m)$ are independent if $m$ regions $B_1, B_2, \ldots, B_m \subset W$ are disjointed.

Here, $\Lambda(B_i)$ denotes the mean number of times that the failure event occurs in region $B_i$ and can be calculated by equation (1). The intensity $\lambda$ or $\lambda(x)$ is the number of times per unit area somewhere. $\Delta(B_i)$ is the area of region $B_i$, and $x \in B_i$. For the homogeneous Poisson point process, its intensity doesn’t vary spatially, and its intensity is often denoted by $\lambda$. For inhomogeneous Poisson point process, its intensity varies spatially, and its intensity is often denoted by $\lambda(x)$.

$$\Lambda(B_i) = \begin{cases} \lambda \Delta(B_i) \\ \int_{B_i} \lambda(x) dx \end{cases} \quad (1)$$

### B. VERIFICATION OF THE AVAILABILITY OF THE POISSON POINT PROCESS

This step is to ensure that a Poisson point process is suitable for modeling a spatial point pattern. A null hypothesis is
proposed according to prior research [32], which states that if the proposed approach is applied to analyze the vulnerability properly, then the properties of the Poisson point process should not be violated. Therefore, two hypothesis tests are applied to determine if it should be accepted or rejected.

1) TESTING THE PROPERTY OF THE POISSON DISTRIBUTION OF POINT COUNTS
To verify the first property, we need to determine that whether there is obvious difference between a Poisson distribution and the samples of \( n(W) \). First of all, select the observation window \( W \) to collect samples in order to guarantee that the records of failure events is sufficient to conduct most statistical tests. Based on the records of failure events, we can obtain lots of \( n(W) \) with the given period \( t \). Then, the frequency of each kind of \( n(W) \) can be counted. The frequency count is shown in figure 4.

In the next place, the significance of the difference between a Poisson distribution and the samples of \( n(W) \) is tested through a chi-square goodness of fit test. The null hypothesis \( H_0 \) states that \( n(W) \) appears to be a Poisson distribution. The alternative hypothesis \( H_1 \) is \( n(W) \) does not appear to be a Poisson distribution. The mean \( \Lambda(W) \) need first to be estimated by referring to equation (2). \( c \) is the number of types of \( n(W) \). With the estimated mean \( \Lambda(W) \), the Poisson probability \( p(m_i) \) can be calculated for each \( m_i \) by equation (3). Next, the theoretical frequency \( f_i^e \) is obtained by equation (4). Then, the \( \chi^2 \) test statistic can be calculated by equation (5). According to the level of significance \( \alpha \) and the degrees of freedom \( c - 1 \), the critical value of \( \chi^2_{c-1} \) can be obtained according to the chi-square distribution table. If \( \chi^2 \geq \chi^2_{c-1} \), we can reject \( H_0 \). If not, we do not reject \( H_0 \).

\[
\Lambda(W) = \frac{\sum_{i=1}^{c+1} m_i f_i^o}{n} = \frac{\sum_{i=1}^{c+1} m_i f_i^o}{\sum_{i=1}^{c+1} f_i^o}
\]

\[
P(n(W) = m_i) = \frac{[\Lambda(W)]^{m_i}}{m_i!} e^{-\Lambda(W)}
\]

\[
f_i^e = n \times P(n(W) = m_i)
\]

\[
\chi^2 = \sum_{i=1}^{c+1} \left( \frac{(f_i^o - f_i^e)^2}{f_i^e} \right)
\]

2) TESTING THE PROPERTY OF INDEPENDENT SCATTERING
In the verification of independent scattering, two sets of counts are required to be tested by a permutation test [33], [34]. The first set is \( n(B_1), \ldots, n(B_m) \). The regions \( B_1, \ldots, B_m \) are the candidate regions of an industry infrastructure, and some of them will be identified as the critical regions. The second set is the counts \( n(D_1), \ldots, n(D_j) \). The regions \( D_1, \ldots, D_j \) are divided from the whole observation window and will be used in model selection.

\[
I = \frac{\sum_{i=1}^{m} [n(B_i) - \bar{n}]^2}{\sum_{i=1}^{m} \sum_{j=1}^{m} w_{i,j}}
\]

\[
\bar{n} = \frac{\sum_{i=1}^{m} n(B_i)}{m}
\]

\[
w_{i,j} = \begin{cases} 0, & d_{i,j} > \sigma \\ d_{i,j}^{-1}, & d_{i,j} \leq \sigma \end{cases}
\]

\[
T = -\frac{1}{m-1} \ln \left( \frac{\sum_{i=1}^{m} n(B_i) - \bar{n}}{m-1} \right)
\]

First of all, an initial Moran’s \( I \) and an expected Moran’s \( I \) \( \bar{T} \) need to be calculated for each set of counts by using equations (6)-(9). In general, if \( I > \bar{T} \), the set of counts are positively autocorrelated in the spatial dimension. If \( I \leq \bar{T} \), it thinks that the set of counts are negative autocorrelation in the spatial dimension. If there is no significant difference between them, it indicates spatial independence in the spatial dimension [35]. To test the significance of the difference between them, a permutation test is applied.

The null hypothesis is that the counts \( n(B_1), \ldots, n(B_m) \) (or \( n(D_1), \ldots, n(D_j) \)) are spatially independent. The negation of the null hypothesis is the alternative hypothesis. Based on the null hypothesis, we rearrange the counts \( n(B_1), \ldots, n(B_m) \) (or \( n(D_1), \ldots, n(D_j) \)) for all different regions of the counts \( B_1, \ldots, B_m \) (or \( D_1, \ldots, D_j \)) and recalculate Moran’s \( I \) by performing an exhaustive rearrangement. After that, many Moran’s \( I \) is obtained, and the amount of them is denoted by \( C \). Then, count the number of the Moran’s \( I \) which is not less than the initial Moran’s \( I \), the number is denoted by \( z \). The P-value is \( z/C \). If \( z/C \) is less than the level of significance in the test, we reject the null hypothesis.

C. MODEL SELECTION
In this step, we will answer the question: Which kind of Poisson point process is the suitable model, the inhomogeneous one or the homogeneous one? For this purpose, we need to determine the spatial variability of the intensity in the observation window \( W \). First of all, the division of the observation window \( W \) should be made. After that, parts \( D_1, \ldots, D_j \) (for
example, the division in figure 5) are obtained, and all of the parts must have equal area $a$. Then, we can obtain the counts $n(D_1), \ldots, n(D_j)$. Due to the conclusion in step 1, we know that Poisson point process is the suitable model, then we have no reason to doubt that the counts $n(D_1), \ldots, n(D_j)$ are independent Poisson random variables. If the homogeneous one is appropriated, then $n(D_1), \ldots, n(D_j)$ should follow the Poisson distribution with the mean $\lambda a$. Thus, we have no reason to doubt that there is no spatial variability of intensity. Therefore, we can test whether these counts appear to be a Poisson distribution to determine the spatial variability of intensity. To achieve the aim, we choose the chi-square goodness of fit test due to its availability with the consideration of other tests.

The $\chi^2$ test statistic is calculated by equation (10). $n^c(D_j)$ is the expected count in region $D_j$ and can be calculated by equation (11). By using equation (12), the intensity $\lambda$ is estimated based on the area of the observation window $\Delta(W)$.

Then, determine the critical value of $\chi^2(j-1)$ through the chi-square distribution table. If $\chi^2 \geq \chi^2(j-1)$, the counts didn’t appear to be a Poisson distribution. The inhomogeneous Poisson point process should be selected, and vice versa.

$$\chi^2 = \sum_{j} \frac{(n(D_j) - n^c(D_j))^2}{n^c(D_j)} \tag{10}$$

$$n^c(D_j) = \hat{\lambda} a \tag{11}$$

D. MODEL BUILDING

In the subsection of model selection, the suitable model is determined. In this section, we need to estimate the parameters.

1) KERNEL DENSITY ESTIMATION TECHNOLOGY

Estimating the intensity $\lambda$ or $\lambda(x)$ can intuitively be interpreted as creating a smooth intensity surface over a 2-D

\[
\hat{\lambda} = \frac{\sum_{i=1}^{j} n(D_i)}{\Delta(W)} \tag{12}
\]
geographic space. Because the intensity is a constant, the surface of intensity is actually flat in a homogeneous Poisson point process. Its intensity can be calculated by equation (13). For the intensity of an inhomogeneous Poisson point process, its surface is uneven (as shown in figure 6), and its calculation needs kernel density estimation technology.

\[ \lambda = \frac{n(W)}{\Delta(W)} \]  

(13)

Kernel density estimation is widely used to complete cluster and regression tasks in machine learning [36]–[38]. Based on the locations of failure events \( x_1, \ldots, x_n \in R^2 \), the general form of the kernel density estimator is given by equation (14):

\[ \hat{\lambda}(x) = \frac{1}{nr^2} \sum_{i=1}^{n} \kappa \left( \frac{d_{ix}}{r} \right) \]  

(14)

where \( \lambda(x) \) is the intensity of location \( x \in W \), \( d_{ix} \) is the distance from \( x \) to \( x_i \), and \( r \in R^+ \), which is called the bandwidth. With the help of the bandwidth, only the failure observations within \( r \) are used to estimate \( \lambda(x) \). \( \kappa(\cdot) \) is called the kernel function and allocates the weight to failure observation \( x_i \). By using the kernel function, the estimation of intensity can consider the impact caused by distance decay. Generally, the farther location \( x \) is from failure observation \( x_i \), the less that the failure observation is weighted for calculating the intensity \( \lambda(x) \). In our approach, a Gaussian function is selected as the kernel function, which is given by equation (15):

\[
\begin{align*}
\kappa \left( \frac{d_{ix}}{r} \right) &= \frac{1}{2\pi} \exp\left(-\frac{d_{ix}^2}{2r^2}\right), & \text{when} & \: 0 < d_{ix} \leq r \\
\kappa \left( \frac{d_{ix}}{r} \right) &= 0, & \text{when} & \: d_{ix} > r
\end{align*}
\]

(15)

2) SELECTION OF OPTIMAL BANDWIDTH

The selection of the bandwidth determines the performance of kernel density estimation [39]–[42]. If a small value of \( r \) is selected, a small number of \( x_i \) need to be considered, which will cause higher variance in the estimation. If a large value of \( r \) is selected, the variance will be reduced. However, the bias will be high because unnecessary noise is present. In general, it is determined by experts. In the proposed approach, the optimal bandwidth is obtained by optimization calculation.

The optimal bandwidth is the one that minimize equation (16). MISE[\( \lambda(x) \)] is the mean integrated square error (MISE) of the estimated intensity \( \lambda(x) \), which measures the difference between the estimated intensity \( \hat{\lambda}(x) \) and the target intensity \( \lambda(x) \). Due to its tractability, the MISE is by far the most global measure for the performance of bandwidths [43].

\[ \text{MISE}[\lambda(x)] = E \int_W \{ \lambda(x) - \hat{\lambda}(x) \} dx \]  

(16)

An approximate formula of the MISE is used to build the objective of the optimization [43] and is represented in equation (17):

\[ \text{MISE}[\lambda(x)] \approx \frac{1}{4} h^4 \left[ \int_{w} r^2 \kappa(t) dt \right]^2 \int_{w} \lambda^2(x) dx + n^{-1} h^{-1} \int_{w} \kappa(t)^2 dt \]  

(17)

Its analytical solution is given by equation (18) [43]:

\[ r_{opt} = 1 \int_{w} r^2 \kappa(t) dt \left[ \int_{w} \kappa(t)^2 dt \right]^2 \left[ \int_{w} \lambda^2(x) dx \right]^{-\frac{1}{2}} \]  

(18)

Due to the application of the Gaussian kernel, equation (18) can be replaced by equation (19):

\[ r_{opt} = \frac{1}{\text{opt}} \int_{w} r^2 \kappa(t) dt \left[ \int_{w} \kappa(t)^2 dt \right]^2 \left[ \int_{w} \lambda^2(x) dx \right]^{-\frac{1}{2}} \]  

(19)

where \( n \) is the amount of the failure event that occurred in the observation window \( W \). \( \sigma \) can be calculated by equations (20)-(22). \( w_i \) and \( v_i \) are the geographic coordinates of the failure observation \( x_i \).

\[ \sigma = \sqrt{\frac{\sum_{i=1}^{n} (w_i - \bar{w})^2 \sum_{i=1}^{n} (v_i - \bar{v})^2}{n}} \]  

(20)

\[ \bar{w} = \frac{\sum_{i=1}^{n} w_i}{n} \]  

(21)

\[ \bar{v} = \frac{\sum_{i=1}^{n} v_i}{n} \]  

(22)
TABLE 1. The cost composition of the total cost $C_{B_i}$.

| No. | Constitution                                                                 |
|-----|-------------------------------------------------------------------------------|
| 1   | The costs caused by damage to the asset of the industry infrastructure         |
| 2   | The costs caused by damage to products                                         |
| 3   | The costs caused by damage to public and private properties                    |
| 4   | The costs caused by emergency response                                        |
| 5   | The costs caused by environmental remediation                                 |
| 6   | Other costs                                                                   |

TABLE 2. Information about each failure event included in the data.

| Attribute | Description                                                                 |
|-----------|----------------------------------------------------------------------------|
| Location  | The location of a failure event that occurred in the pipeline network, which includes the latitude and longitude coordinates |
| Date      | The year and month of a failure event happened                              |
| Cause     | The failure reason                                                          |
| Loss      | The total failure cost of each failure event                                |

E. RISK ASSESSMENT
In this section, we first estimate the probability that the failure event occurs in different candidate regions of the industry infrastructure through the built model. Then, the severity of failure consequences is measured in each candidate region. Based on that, the risks of all the candidate regions are evaluated.

1) ESTIMATING THE PROBABILITY THAT THE FAILURE EVENT OCCURS IN EACH CANDIDATE REGION
Based on the above conclusions and the finite-dimension distribution property [44], the Poisson random variable $n(B_i)$ follows the Poisson distribution with the mean $\Lambda(B_i)$. By using the estimated intensity $\lambda$ (or $\lambda(x)$) and the equation (1), we can calculate the mean $\Lambda(B_i)$. Through the equations (23) and (24), $P_{B_i}$ is obtained, which is the probability that the failure event occurs in region $B_i$.

$$P(n(B_i) = k) = \frac{[\Lambda(B_i)]^k}{k!} e^{-\Lambda(B_i)} \quad (23)$$

$$P_{B_i} = 1 - P(n(B_i) = 0) \quad (24)$$

2) CALCULATING THE SEVERITY OF THE FAILURE CONSEQUENCE IN EACH CANDIDATE REGION
In our approach, the consequence of failure is defined as the total negative collection in the aspects of environment, finance and humanity. The collection should include the deaths caused by the failure events and subsequent events, the damage to the environment, the injury to the public, and the impact on businesses. According to this, the total failure cost $C_{B_i}$ is used to measure the severity of the failure consequence in region $B_i$. Table 1 describes the cost composition of the total failure cost.

3) ASSESSMENT OF THE RISKS OF DIFFERENT CANDIDATE REGIONS AND IDENTIFYING THE CRITICAL REGIONS
Using the estimated probability and calculated total failure cost, risk can be assessed for each candidate region. Our approach suggests using the product of the estimated probability and the calculated total failure cost to measure risks of different candidate regions. Then, the criticality can be given to each of them by sorting the risks of these candidate regions.

III. EXPERIMENTS
A. OBSERVATION WINDOW
The critical industry infrastructure used in the experiments is a pipeline network in Kansas, USA. We choose the whole state of Kansas to be the observation window $W$. It is shown in figure 3. The area of it is 348,078 km$^2$.

B. DATASET
In the dataset, 141 failure events are collected from January 2010 to January 2017. The information included in the dataset is shown in table 2.

C. EXPERIMENTAL DESIGN
1) EXPERIMENTAL DESIGN 1
The feasibility of the proposed approach is first verified in order to prove that our approach has the ability to identify the critical region among four candidate regions. All four candidate regions are different in shape, area and location, as illustrated in figure 7.

In table 3, the frequency of each kind of $n(W)$ is obtained with the period of one month and the samples of $n(W)$. Based on this, the first test of availability verification can be conducted. For the second test, the regions $D_1, D_2, D_3, D_4$ are obtained through the division of the observation window $W$, and all the four regions are shown in figure 5. Both set of counts are listed in table 4. For these candidate regions, the centroid distances between every two regions are shown in figure 8, and the threshold distance is 219,465 m. For regions $D_1, D_2, D_3$, and $D_4$, the centroid distances between every two regions are shown in figure 9, and the threshold distance is 205,360 m. Based on these values, the spatial weight matrices can be obtained. Then, the property of independent scattering is tested.
After that, we need to determine the category of the model. \( n(D_1), n(D_2), n(D_3), \) and \( n(D_4) \) are shown in table 4.

In the procedure of model building, the information of the failure event locations is applied to estimate the intensity and comprises the latitude and longitude. The number of failure events \( n \) is 141. After the calculation of the bandwidth of the kernel density estimator, the intensity is estimated.

Using the estimated intensity in model building, the probability can be estimated for each candidate region. The total failure cost of each candidate region is obtained by summing the total failure cost of every failure event that occurred in the candidate region. The total failure cost of each candidate region can be observed in table 5. Finally, the risks of the candidate regions are calculated, and the criticality of each candidate region is determined.

### EXPERIMENTAL DESIGN 2

In experiment 2, the proposed approach is applied to observe the distribution of risk in the whole state of Kansas. The objective is to identify critical regions under the assumption that there are no candidate regions. The whole state of Kansas is divided into sixteen regions, and each region has an approximately equal area of 21,754.93 km\(^2\). All the divided regions are shown in figure 10. According to the description of the proposed approach, the estimated probability and calculated total failure cost of each candidate region are obtained based on the intensity estimated in experiment 1. Then, the risks of all sixteen regions are obtained. Based on this approach, we can observe the distribution of risk in the whole state of Kansas.

Additionally, the spatial dependency is investigated based on the counts, the total failure costs and the probabilities that a failure event occurred in each region. The test of significance based on Moran’s \( I \) is applied to achieve the objective, which is also used in step 1 of the proposed approach. In experiment 2, the significance test is conducted for three sets of data. The three sets of data are the counts \((n(E_1), \ldots, n(E_{16}))\), the probabilities that a failure event occurred in each region \((P_{E_1}, \ldots, P_{E_{16}})\) and the total costs of the sixteen regions \((C_{E_1}, \ldots, C_{E_{16}})\). The centroid distances between every two regions are provided in figure 11. The threshold distance is 113,158 m. With the results of the tests, the spatial dependency can be reviewed.

### IV. RESULTS AND DISCUSSION

#### A. RESULTS OF EXPERIMENT 1

In the step of the validation of the Poisson distribution of point counts, the calculated \( \chi^2 \) test statistic is 9.97, and the critical
value of $\chi^2_{0.05}(7)$ is 14.067. By comparison, the decision is not to reject $H_0$. $n(W)$ appears to be a Poisson distribution. The first property is not violated.

For the validation of the property of independent scattering, the two spatial weight matrices are determined and are shown in figure 12 and figure 13. The results are listed in table 6 and table 7, and indicate that both sets of counts are not significantly different from the counts with spatial independence. Therefore, there is no reason to doubt that the second property is satisfied in the case.

In model selection, the calculated $\chi^2$ test statistic is $46.887$ with the estimated intensity $\hat{\lambda} = 4.0508E$-10 times/m². The critical value of $\chi^2_{0.05}(3)$ is 7.815. By comparison, the decision is to reject $H_0$. Therefore, we choose the inhomogeneous Poisson point process.

In model building, the calculated bandwidth is 24.905 km. In figure 6, we can observe the estimated intensity of any location in the observation window.

After that, the probability that the failure event occurred is estimated in each candidate region. Then, the risk of each candidate region is assessed with the total failure cost. All of the results are shown in table 8. According to the assessed risk, region $B_2$ is the most critical among the regions.

B. THE RESULTS OF EXPERIMENT 2

According to the requirements of experiment 2, all the results are obtained, which are shown in table 9. Figures 14, 15 and 16 are used to illustrate the distributions of the probability, the total failure cost and the risk in the spatial dimension.

Based on the centroid distances between every two regions in experiment 2, the spatial weight matrix is determined,

| FIGURE 11. The centroid distances between every two regions in experiment 2 (m). |
| --- |
| $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ | $E_6$ | $E_7$ | $E_8$ | $E_9$ | $E_{10}$ |
| $19519.03$ | $195867.03$ | $3400111.38$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ |
| $19551.28$ | $3400111.38$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ |
| $20166.96$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ |
| $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ |
| $35960.32$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ | $1134.67$ |
| $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ |
| $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ |
| $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ | $262016.96$ | $1134.67$ | $295470.70$ | $35960.32$ | $597095.32$ |

| TABLE 7. The results of the significance test for regions $B_1$, $B_2$, $B_3$, and $B_4$. |
| --- |
| $B_1$ | $B_2$ | $B_3$ | $B_4$ |
| 0 | 0.51E-06 | 0 | 0 |
| 0 | 0.51E-06 | 0 | 4.55E-06 |
| 0 | 0 | 4.55E-06 | 0 |

| FIGURE 12. The spatial weight matrix of the candidate regions. |
| --- |
| $D_1$ | $D_2$ | $D_3$ | $D_4$ |
| 0 | 4.86E-06 | 0 | 4.90E-06 |
| 0 | 4.86E-06 | 0 | 4.90E-06 |
| 0 | 4.86E-06 | 0 | 4.90E-06 |

| TABLE 8. The results of risk assessment for every candidate region. |
| --- |
| Region | Mean | Probability | Total cost ($) | Risk |
| $B_1$ | 29.99 | 0.9999999999999906 | 1385231 | 1385231 |
| $B_2$ | 24 | 0.999999999622249 | 4180171 | 4180171 |
| $B_3$ | 14 | 0.999999168471281 | 1578163 | 1578161.688 |
| $B_4$ | 13 | 0.999997739670593 | 439909 | 439908.0057 |

| FIGURE 14. Illustration of the probability of occurrence of a failure event in experiment 2. |
| --- |

In model selection, the calculated $\chi^2$ test statistic is 46.887 with the estimated intensity $\hat{\lambda} = 4.0508E$-10 times/m². The critical value of $\chi^2_{0.05}(3)$ is 7.815. By comparison, the decision is to reject $H_0$. Therefore, we choose the inhomogeneous Poisson point process.

In model building, the calculated bandwidth is 24.905 km. In figure 6, we can observe the estimated intensity of any location in the observation window.

After that, the probability that the failure event occurred is estimated in each candidate region. Then, the risk of each candidate region is assessed with the total failure cost. All of the results are shown in table 8. According to the assessed risk, region $B_2$ is the most critical among the regions.

B. THE RESULTS OF EXPERIMENT 2

According to the requirements of experiment 2, all the results are obtained, which are shown in table 9. Figures 14, 15 and 16 are used to illustrate the distributions of the probability, the total failure cost and the risk in the spatial dimension.

Based on the centroid distances between every two regions in experiment 2, the spatial weight matrix is determined,
which is shown in figure 17. Then, significance tests are conducted based on the counts (the counts can be observed in table 9), the total failure costs and the probabilities. All the results are listed in table 10.

C. DISCUSSION

According to the estimated intensity in experiment 1, some hotspots can be observed in figure 6 in the southeast part of Kansas. These hotspots intuitively reveal that the areas where they are located have higher concentrations of failure events due to the vulnerability of the pipeline network in those areas. However, the clusters of failure events within a limited geographical region are not statistically significant. Although a comparison between the calculated Moran’s index and the expected Moran’s index (0.089246 > -0.066667) provides evidence to support these hotspots, the high P-value (0.333988) indicates that the hotspots are not statistically significant. Therefore, these regions of hotspots are useless to target for stakeholders’ efforts knowing that they will achieve a positive effect with limited resources.

By observing the spatial distribution of the probability of the failure event in experiment 2, we think that the probability of the failure event occurring in each region is very high, and the variation of those probabilities is very small in the spatial dimension. According to the probability results shown in table 9, thirteen of the sixteen values are larger than 0.95, while the probabilities of regions $E_1$, $E_2$, and $E_9$ are less

![FIGURE 17. The spatial weight matrix in experiment 2.](image)

![FIGURE 15. Illustration of the total failure cost in experiment 2.](image)

![FIGURE 16. Illustration of the risk in experiment 2.](image)

| Region | Count | Probability | Total failure cost ($) | Risk |
|--------|-------|-------------|------------------------|------|
| $E_1$  | 1     | 0.632120558828558 | 3888                   | 2457.68473 |
| $E_2$  | 2     | 0.864646716763387 | 35034                  | 30292.6636 |
| $E_3$  | 3     | 0.950212931632136 | 124000                 | 117826.403 |
| $E_4$  | 3     | 0.950212931632136 | 814800                 | 774233.496 |
| $E_5$  | 12    | 0.999993853636796 | 1242889                | 1242881.36 |
| $E_6$  | 5     | 0.993264383927303 | 816821                 | 811319.207 |
| $E_7$  | 12    | 0.999993853636796 | 416643                 | 416640.44 |
| $E_8$  | 3     | 0.950212931632136 | 2837840                | 2697407.45 |
| $E_9$  | 0     | 0            | 0                      | 0     |
| $E_{10}$ | 9   | 0.999876632888319 | 194733                 | 194708.976 |
| $E_{11}$| 47   | 1.000000000000000 | 4653061                | 4653061 |
| $E_{12}$| 6    | 0.997522105323231 | 1408753                | 1405262.25 |
| $E_{13}$| 8    | 0.999664537372097 | 3373355                | 3372223.36 |
| $E_{14}$| 13   | 0.999979741290628 | 878987                 | 878985.014 |
| $E_{15}$| 12   | 0.999993853636796 | 1538302                | 1538292.54 |
| $E_{16}$| 5    | 0.99326205300915  | 40401                  | 40128.7802 |

| Statistic | Result based on the number of failure events | Result based on the probability of occurrence of the failure event | Result based on the total failure cost |
|-----------|----------------------------------------------|---------------------------------------------------------------|--------------------------------------|
| Calculated Moran’s I | 0.089246 | -0.108406 | -0.1277220 |
| Expected Moran’s I | -0.066667 | -0.066667 | -0.066667 |
| P-value | 0.333988 | 0.784655 | 0.813529 |

By observing the spatial distribution of the probability of the failure event in experiment 2, we think that the probability of the failure event occurring in each region is very high, and the variation of those probabilities is very small in the spatial dimension. According to the probability results shown in table 9, thirteen of the sixteen values are larger than 0.95, while the probabilities of regions $E_1$, $E_2$, and $E_9$ are less
than 0.95. The reason for the low-probability regions may be incorrect division of the observation window. Due to the unknown deployment locations of the pipeline, our division of regions may make the divided regions cover more areas that have no deployment of the pipeline, which will achieve a negative effect in reducing the probability of a failure event occurring in each candidate region. Based on this limitation, we think that the thirteen high-probability regions have the objective estimation of the probability, and the mean of their probabilities is 0.918187. For the industry infrastructure, the mean implies that the likelihood of occurrence of a failure event is very high. Additionally, the coefficient of variation is calculated based on the thirteen high probabilities. The result is 0.022236. For the total failure cost, the coefficient of variation is 0.955012. There is obvious difference between the two coefficients of variation. By comparison, the variability of risk uncertainty is very small in spatial dimension. Moreover, small variability can also be observed in the spatial dimension. According to the results of the probabilities shown in table 10, there is not sufficient evidence to support that there is a cluster of high or low values of the probability with reference to the average in the spatial dimension. Therefore, we conclude that the probability of occurrence of a failure event is very high in the regions where the pipeline is located, however the spatial variability is too small. This implies that risk uncertainty is less instructive for identifying critical regions than the severity of failure consequences. It also indicates that the total failure cost is more affected by location.

According to the investigation of the spatial dependencies of the total costs and the probabilities of occurrence of a failure event in the sixteen regions in experiment 2, there is not enough evidence to prove that geospatial interdependency exists in industry infrastructure. As an interdependency type of industry infrastructure, geospatial interdependency supposes that the proneness to failure of adjacent components is similar because of the proximity of infrastructure components in the spatial dimension [45]. Geospatial interdependency is widely used for industry infrastructure as an assumption in vulnerability analysis [14], [46]–[48]. However, the investigation of the spatial dependency reveals that the total failure costs and the probabilities of occurrence of the failure event in the sixteen regions are both spatially independent, which can be reviewed in table 10. This means that the preferences of failures of components are not similar even if components of the industry infrastructure are located in adjacent regions.

D. APPROACH COMPARISONS

To help researchers observe the differences between our approach and other existing methods, a comparison is made in five aspects.

In the first aspect, the comparison is made with the consideration of the difficulty of acquiring the required input data or information. For the locations and loss of failure events which are used in our approach, they are easily obtained through news. For other approaches, the required input data or information includes the constitution of the system and the design file of components, which is not only important for stakeholders but also for homeland Security. Therefore, our approach has an advantage over other approaches in this aspect.

The second aspect is the size of the data which is need for modelling. By comparison with network-based approaches, our approach is not good. However, our approach doesn’t need large amounts of data by comparison with the empirical approach, in particular to the approach which applies other machine learning technologies. Therefore, we think our approach is not bad in this aspect.

The third aspect is the ability of approaches to consider the system characteristics in modelling. In this aspect, our approach is weak. In our approach, geographic coordinates are used to be the variables of risk, which means there is no system characteristics that are considered in risk assessment. Therefore, our approach has no ability to consider the system characteristics in modelling.

For the fourth aspect, the comparison is made in computational cost. the approaches which are used in comparison is limited to the empirical approaches that also use machine learning technologies [49]. In computational complexity, kernel density estimation is one of the machine learning technologies which have the very low computational complexity. In addition, the small amount of input data reveals that our approach doesn’t need large computational cost.

The fifth aspect is the falsifiability of the approach. In the aspect, our approach is excellent. Because it lacks a common standard for evaluating the accuracy of vulnerability analysis, the results of almost all approaches can’t be falsified. For this reason, our approach applies hypothesis testing before modelling, which effectively improves the falsifiability.

V. CONCLUSION

This paper proposes an improved empirical approach for the vulnerability analysis of industry infrastructure to identify critical regions. It is accomplished by determining a risk value for any region where an industry infrastructure is deployed. This risk value is characterized by risk uncertainty and the severity of consequences caused by failures. The risk uncertainty is measured by the probability of the failure event occurring in a geographic region, and the total failure cost is used to measure the severity of the consequences. With the assessed risk, a ranking of the different geographic regions is obtained. Then, the criticalities of the different geographic regions are determined.

A. THE CONTRIBUTIONS OF THE PROPOSED APPROACH

The first contribution is categorized as the contribution in application. Being one of the spatial statistical techniques, the Poisson point process is innovatively applied for industry infrastructures to analyze their vulnerability. By overcoming the difficulty of the insufficient data and the impact of human biases, it effectively and objectively estimates the probability.
that the failure event occurs in different geographic regions. On top of that, the following benefits are obtained:

1) By comparison with existing approaches, the uncommonly used location information in the records of failure events is used for modeling, which means that the records of failure events are deeply uncovered by using our approach.

2) The difficulty of insufficient data, which most approaches are facing, is overcome by the transformation of the modeled object. In essence, the transformation is achieved by the Poisson point process. It changes the modeled object from the components to the region where the whole industry infrastructure is located, all the infrequent failures of different components are aggregated to form a large amount of data. Therefore, there are sufficient data to use the proposed approach.

The second is the contribution in methodology. A hypothesis test is applied to test the assumptions used in our approach, which increases the falsifiability of the results. In current research on vulnerability analysis, it lacks a common standard for evaluating the correctness of the analysis for the existing approaches. It means that there is no foundation to make comparison among different approaches. Due to this reason, we explore to evaluate the correctness of the analysis based the approach itself. By using hypothesis testing, the correctness of the applied assumptions and hypotheses can be determined, which effectively improves the falsifiability of the proposed approach.

In theory, the proposed approach has two contributions. The first one is the findings of risk uncertainty. We find that the variability of risk uncertainty is very little in different geographic regions where an industry infrastructure is located. The diversity of risk is mainly caused by the severity of failure consequences. At the same time, the significant variability of the severity of failure consequences in spatial dimension indicates that the severity of failure consequences is more sensitive to the geographical position of failure events.

B. THE LIMITATIONS OF THE PROPOSED APPROACH
The first limitation is the poor modeling capability. The variables, reflecting the characteristics of industry infrastructures, is replaced by geographic coordinates. Therefore, the proposed approach can’t consider many system characteristics in vulnerability analysis. The second limitation is the strict constraints of the application of the proposed approach. Due to the requirement for the properties of the Poisson point process to be validated before modelling, the vulnerability analysis may not be made when the properties are not hold. Therefore, the application of the proposed approach is limited.

APPENDIX
The dataset can be acquired online at: https://www.kaggle.com/usdot/pipeline-accidents/metadata. It was collected and published by United States Department of Transportation’s Pipeline and Hazardous Materials Safety Administration.

REFERENCES
[1] G. W. Bush GW, “Homeland security presidential directives-3 (HSPD-3),” Office Federal Register, Washington, DC, USA, Tech. Rep. HSPD-3, 2002.
[2] G. W. Bush, “Homeland security presidential directives-7 (HSPD-7),” Office Federal Register, Washington, DC, USA, Tech. Rep. HSPD-7, 2003.
[3] W. Clinton, “Presidential decision directive PDD-63, protecting America’s critical infrastructures,” Clinton Presidential Library Museum, Washington, DC, USA, Tech. Rep. PDD-63, 1998.
[4] T. G. Lewis, Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation Hoboken, NJ, USA: Wiley, 2006.
[5] J. D. Mottef, C. Copeland, and J. W. Fischer, “Critical Infrastructures: What Makes an Infrastructure Critical?” Congressional Res. Service Rep., Library Congr., Congressional Res. Service, Washington, DC, USA, Tech. Rep. RL31556, 2002.
[6] E. Zio, “Challenges in the vulnerability and risk analysis of critical infrastructures,” Rel. Eng. Syst. Saf., vol. 152, pp. 137–150, Aug. 2016.
[7] M. Ouyang, “Review on modeling and simulation of interdependent critical infrastructure systems,” Rel. Eng. Syst. Saf., vol. 121, pp. 43–60, Jan. 2014.
[8] C. M. J. Tampère, J. Stada, B. Immens, E. Peertmans, and K. Organe, “Methodology for identifying vulnerable sections in a national road network,” Transp. Res. Record, J. Transp. Res. Board, vol. 2012, no. 1, pp. 1–10, Jan. 2007.
[9] E. Jelenski, “Network structure and travel patterns: Explaining the geographical disparities of road network vulnerability,” J. Transp. Geography, vol. 17, no. 3, pp. 234–244, May 2009.
[10] J. Johansson and H. Hassel, “An approach for modelling interdependent infrastructures in the context of vulnerability analysis,” Rel. Eng. Syst. Saf., vol. 95, no. 12, pp. 1335–1344, Dec. 2010.
[11] Z. Wang, A. P. C. Chan, J. Yuan, B. Xia, M. Skitmore, and Q. Li, “Recent advances in modeling the vulnerability of transportation networks,” J. Infrastruct. Syst., vol. 21, no. 2, Jun. 2015, Art. no. 06014002.
[12] S. A. Patterson and G. E. Apostolakis, “Identification of critical locations across multiple infrastructures for terrorist actions,” Rel. Eng. Syst. Saf., vol. 92, no. 9, pp. 1183–1203, Sep. 2007.
[13] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, “Catastrophic cascade of failures in interdependent networks,” Nature, vol. 464, no. 7291, pp. 1025–1028, Apr. 2010.
[14] E. E. Lee, II, J. E. Mitchell, and W. A. Wallace, “Restoration of services in interdependent infrastructure systems: A network flows approach,” IEEE Trans. Syst., Man, Cybern. C. Appl. Rev., vol. 37, no. 6, pp. 1303–1317, Nov. 2007.
[15] Z. Wang, A. P. C. Chan, J. Yuan, B. Xia, M. Skitmore, and Q. Li, “Recent advances in modeling the vulnerability of transportation networks,” J. Infrastruct. Syst., vol. 21, no. 2, Jun. 2015, Art. no. 06014002.
[16] T. Li, L. Rong, and K. Yan, “Vulnerability analysis and critical area identification of public transport system: A case of high-speed rail and air transport coupling system in China,” Transp. Res. Part A, Policy Pract., vol. 127, pp. 55–70, Sep. 2019.
[17] D. R. Marlow, D. J. Beale, and J. S. Mashford, “Risk-based prioritization and its application to inspection of valves in the water sector,” Rel. Eng. Syst. Saf., vol. 100, pp. 67–74, Apr. 2012.
[18] A. J. Brito and A. T. de Almeida, “Multi-attribute risk assessment for risk ranking of natural gas pipelines,” Rel. Eng. Syst. Saf., vol. 94, no. 2, pp. 187–198, Feb. 2009.
[19] D. Fuchs-Hansch, F. Friedl, M. Møderl, W. Sprung, H. Plihal, F. Kretschmer, and T. Ertl, “Risk and performance oriented sewer inspection prioritization,” in Proc. World Environ. Water Resour. Congr., Albuquerque, NM, USA, May 2012, pp. 3711–3723.
[20] M. A. Hahn, R. N. Palmer, and S. M. Merrill, “Prioritizing sewer line inspection with an expert system,” in Proc. WRPMD, Tempe, AZ, USA, 1999, pp. 1–9.
[21] M. A. Hahn, R. N. Palmer, M. S. Merritt, and A. B. Lukas, “Expert system for prioritizing the inspection of sewers: Knowledge base formulation and evaluation,” J. Water Resour. Planning Manage., vol. 128, no. 2, pp. 121–129, Mar. 2002.
[22] P. K. Dey, “Analytic hierarchy process analyzes risk of operating cross-country petroleum pipelines in India,” Natural Hazards Rev., vol. 4, no. 4, pp. 213–221, Nov. 2003.
[23] P. K. Dey, S. O. Ogulana, and S. Naksuksakul, “Risk-based maintenance model for offshore oil and gas pipelines: A case study,” J. Qual. Maintain. Eng., vol. 10, no. 3, pp. 169–183, Sep. 2004.

[24] M. J. Anvari, M. Tabesh, and A. Roozbahani, “Risk assessment model to prioritize sewer pipes inspection in wastewater collection networks,” J. Environ. Manage., vol. 190, pp. 91–101, Apr. 2017.

[25] S. Kaplan and J. B. Garrick, “On the quantitative definition of risk,” Risk Anal., vol. 1, no. 1, pp. 11–27, Mar. 1981.

[26] A. Mancuso, M. Compare, A. Salo, E. Zio, and T. Laakso, “Risk-based optimization of pipe inspections in large underground networks with imprecise information,” Rel. Eng. Syst. Saf., vol. 152, pp. 228–238, Aug. 2016.

[27] A. Cervellera and D. Maccio, “An extreme learning machine approach,” J. Perform. Constructed Facilities, vol. 30, no. 1, Feb. 2016, Art. no. 04014202.

[28] M. Layouni, M. S. Hamdi, and S. Tahar, “Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning,” Appl. Soft Comput., vol. 52, pp. 247–261, Mar. 2017.

[29] M. B. Abdulla and R. Herzallah, “Probabilistic multiple model neural network based leak detection system: Experimental study,” J. Loss Prevention Process Ind., vol. 36, pp. 30–38, Jul. 2015.

[30] S. Dietrich and A. Penttinen, “Recent applications of point process methods in forestry statistics,” Stat. Sci., vol. 15, pp. 61–78, Jan. 2000.

[31] J. Bailey and A. C. Gatrell, “A survey of recent advances in CNN-based image segmentation,” Image Anal. Stereol., vol. 37, no. 3, pp. 341–342, Aug. 1995.

[32] H. Li, C. A. Calder, and N. Cressie, “Beyond Moran’s I: Testing for spatial dependence based on the spatial autoregressive model,” Geographical Anal., vol. 39, no. 4, pp. 357–375, Oct. 2007.

[33] V. A. Sindagi and V. M. Patel, “A survey of recent advances in CNN-based single image crowd counting and density estimation,” Pattern Recognit. Lett., vol. 107, pp. 3–16, May 2018.

[34] C. Cervellera and D. Maccio, “An extreme learning machine approach to density estimation problems,” IEEE Trans. Cybern., vol. 47, no. 10, pp. 2123–2127, Oct. 2017.

[35] T. C. Bailey and A. C. Gatell, Interactive Spatial Data Analysis. Essex, U.K.: Longman, 1995.

[36] D. O’Sullivan and D. J. Unwin, Geographic Information Analysis, Hoboken, NJ, USA: Wiley, 2002.

[37] D. O’Sullivan and D. W. S. Wong, “A surface-based approach to measuring spatial segregation,” Geographical Anal., vol. 39, no. 2, pp. 147–168, Apr. 2007.

[38] O. Schabenberger and C. Gotway, “Mapped point patterns,” in Statistical Methods for Spatial Data Analysis. New York, NY, USA: Chapman & Hall/CRC, 2005, pp. 97–120.

[39] B. W. Silverman, “Density estimation in action,” in Density Estimation for Reliability Engineering & System Safety, and Quality and Reliability Engineering International.

PENG HOU was born in Taiyuan, Shanxi, China, in 1986. He received the B.S. degree in physics from the College of Jinchong, Shanxi, China, in 2010, and the M.S. degree (Master of Law) from the Shaxi University of Finance and Economics, Taiyuan, Shanxi, in 2013. He is currently pursuing the Ph.D. degree in reliability engineering with the Beijing Institute of Technology, Beijing, China.

From 2012 to 2016, he was an Executive Vice President with Shaxi Youth Business Investment Corporation.

XIAOJIAN YI was born in 1987. He received the B.S. degree in control technology from the North University of China, in 2010, and the M.S. and Ph.D. degrees in reliability engineering from the Beijing Institute of Technology, in 2012 and 2016, respectively.

From 2015 to 2016, he was a Joint Training Ph.D. with the University of Ottawa to study on robot reliability and maintenance. From 2016 to 2018, he was an Associate Professor with the China North Vehicle Research Institute and the Beijing Institute of Technology. He is the author of one book and more than 80 articles. He holds nine patents. His research interests include system reliability analysis, system reliability assessment, optimization design, maintainability engineering, fault diagnosis, point process theory, and statistical learning theory.

Dr. Yi was a member of ASME and SAE-China. He is a member of Technical Committee in IEEE PHM 2018. He was a recipient of the Science and Technology Progress of NORINCO, in 2015 and 2018. He is a Program Chair of PHM-Chongqing 2018 and SDPC 2018. He is a General Chair of SDPC 2019. He is a Reviewer of ten international journals and conferences, such as the IEEE TRANSACTIONS ON RELIABILITY, Reliability Engineering & System Safety, and Quality and Reliability Engineering International.

HAIPING DONG was born in Yongkang, Zhejiang, China, in 1969. He received the B.S. degree in electronic engineering from the Huazhong University of Science and Technology, Wuhan, in 1991, the M.S. degree in automatic control from the Beijing Institute of Technology, Beijing, in 1994, and the Ph.D. degree in gas engineering from the Research Institute of Petroleum Exploitation and Development, Beijing, in 2001.

From 1994 to 1997, he was an Engineer with the Beijing Institute of Information and Control. Since 2001, he has been working with the School of Mechatronical Engineering, Beijing Institute of Technology. From 2001 to 2009, he was a Lecturer with the Beijing Institute of Technology. Since 2009, he has been an Associate Professor with the Beijing Institute of Technology. He is the author of two books, two national military standards, and more than 50 articles. He holds nine patents. His research interests include the reliability design and analysis for complex systems, reliability evaluation method with small samples for initiating devices, life assessment, accelerated life test, and accelerated degradation test.

[40] A. Rachman and R. M. C. Ratnayake, “Machine learning approach for risk-based inspection screening assessment,” Rel. Eng. Syst. Saf., vol. 185, pp. 518–532, May 2019.

[41] A. Mancuso, M. Compare, A. Salo, E. Zio, and T. Laakso, “Risk-based optimization of pipe inspections in large underground networks with imprecise information,” Rel. Eng. Syst. Saf., vol. 152, pp. 228–238, Aug. 2016.

[42] J. Bailey and A. C. Gatrell, “A survey of recent advances in CNN-based image segmentation,” Image Anal. Stereol., vol. 37, no. 3, pp. 341–342, Aug. 1995.

[43] C. Cervellera and D. Maccio, “An extreme learning machine approach to density estimation problems,” IEEE Trans. Cybern., vol. 47, no. 10, pp. 3254–3265, Oct. 2017.

[44] D. J. Daley and D. Vere-Jones, An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. New York, NY, USA: Springer, 2008, pp. 275–276.

[45] P. Pederson, D. Dudenhoeffer, S. Hartley, and M. Perrmann, “Critical infrastructure interdependency modeling: A survey of US and international research,” Idaho Nat. Lab., Idaho Falls, ID, USA, Tech. Rep. INL/EXT-06-11464, Aug. 2006.

[46] S. M. Kimbalad, J. P. Peerenboom, and T. K. Kelly, “Identifying, understanding, and analyzing critical infrastructure interdependencies,” IEEE Control Syst. Mag., vol. 21, no. 6, pp. 11–25, Dec. 2001.

[47] R. Zimmerman, “Social implications of infrastructure network interactions,” J. Urban Technol., vol. 8, no. 3, pp. 97–119, Dec. 2001.

[48] D. Dudenhoeffer, M. Perrmann, and M. Manic, “CIMS: A framework for infrastructure interdependency modeling and analysis,” in Proc. Winter Simul. Conf., Monterey, CA, USA, Dec. 2006, pp. 478–485.