ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 16, No. 2, 2021, *pp*.148-159

The determination of ionospheric TEC disturbance based on the cross-validation method

Dan Li，Jian-Wei Yang, Peng Lai，Kai Zhao, Ye-Wen Wu
School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing, China.
Correspondence author: Ye-Wen Wu

(Received January 11, 2021, accepted September 30, 2021)

Abstract: The ionosphere has an impact on the radio system, therefore, the determination of the ionospheric state is significant. The total electron content (TEC), as an important ionospheric parameter, can characterize the state of the ionosphere. This paper introduces a new ionospheric disturbed index N13 by correcting the existing index N27, and also proposes the theory foundation for determination the ionospheric state using cross validation method. The N13 is defined as the normalized relative variation of the ionospheric TEC, in which the TEC background value is the sliding median of 27 days. Analyzing the N13 calculated from the TEC data at Taipei station from January 2002 to July 2014, the results show that the two indexes N13 and N27 generally have the same statistical characters against with season and local time, however, they are always different at one time. Based on the probability density function of N13，an optimization model is also constructed to determine the ionospheric disturb proportion by cross validation method. It is found that the proportion is about 25%, when the ionospheric disturbed index range is $N_{13} < -1$ or $N_{13} > 1$.

Keywords: Ionospheric disturbed index, Ionospheric disturbed proportion, Ionospheric disturbed determination, cross validation.

1. Introduction

The ionosphere is the upper atmosphere of the Earth in the altitude range of 60km~1000km above the ground. The presence of a large number of free electrons and ions in the ionosphere can change the speed of radio waves going through it, causing in refraction, reflection and scattering. The electron density varies strongly with many factors. Therefore, some efforts have been made to confirm the state of ionosphere to reduce its impact on the radio system.

In recent years, based on the observational important electron density parameters Total Electron Content (TEC) and frequency of F2-layer(foF2), the ionospheric disturbed index and the criteria for determining disturbed events have been proposed. Bremer [1] proposed an ionospheric activity index AI based on European foF2 measurements , comparing current data with undisturbed historical data, and later revised it with Mielich [2] to an ionospheric activity index that can describe ionospheric storms in mid-latitudes, which is one of the most commonly used index by later authors. Another index proposed by Gulyaeva [3] to describe ionospheric storms is the ionospheric weather index (W index), which is defined by setting the corrected ionospheric parameters to the logarithm with respecting to their reference values of the static day, given when $W = \pm 1$ indicates quiet condition, when $W = \pm 2$ indicates a moderate disturbance, when $W = \pm 3$ indicates a storm, and when $W = \pm 4$ indicates a large storm. Jakowski [4] introduces an interference ionospheric index based on GNSS (Global Navigation Satellite System) measurements to reduce the impact of space weather on GNSS navigation positioning. Nishioka [5] propose the method of standardized index AI, which is denoted as $\hat{P}_{TEC}$, and the index can be independent of local time, season and geographical location.

For the determination criteria of ionospheric disturbed events, Kouris [6] proposed that the relative deviation of ionospheric TEC from the monthly median value in which it occurred for 3 consecutive hours exceeds 0.1 as the basis of ionospheric disturbed events. Lekshmi [7] defined a storm event as one in which $\Delta N_{\max}$ exceeds 25% or is less than -25% and lasts longer than 3h, where the background value is the average value of the 7 calm days before the storm. Matamba [8] determined ±20% and ±40% as the ionospheric static time-varying rate based on the deviation of the observed values of foF2 and TEC data from the monthly median values obtained in the month in which they occurred, respectively. Chinese scholars Huang [9-10] analyzed foF2 data from five Chinese stations and concluded that an ionospheric disturbed event is defined when the variation of foF2 exceeds 15% and is continuous for more than 6 hours. Chen [11] studied an event with $df \geq 0.15$ and a duration of 6h or more as a perturbation event based on the change of the observed value of foF2 data relative to the mid-month value noted as df. Gao [12] analyzed the types and the patterns of distribution of ionospheric disturbances at four mid- and low-latitude stations in the East Asian sector and proposed that an ionospheric storm event is considered to have occurred when the change in foF2 exceeds 15% and lasts for more than 4 hours, where the background value is the 27-day sliding median of the observed value. Liu [13] defined a storm event as one in which the absolute R_{TEC} exceeds 15% and lasts at least 3h. Deng [14] analyzed TEC data from six Chinese observatories and gave the definition of positive (negative) phase storm disturbed events in ionospheric TEC as a continuous period of 6h and more $DI > 0.35(DI \leq -0.3)$ and the period DI not satisfying the value must not exceed 2 h. Here the DI index is the AI index, where the background value is the sliding median of 13 days before and after the corresponding moment of the observed day. Li [15] defined $\Delta foF2$ based on the relative deviation of the observed value of foF2 data from the mid-month value, and defined $\Delta foF2$ greater than or equal to 15% and continuous for more than 6 hours as an ionospheric storm event. Based on the disturbed determination criteria proposed by Gao [12] and Deng [14], Wu [16] considered the daily variation of TEC at storm subtracted from the average daily variation of the static days, and the difference exceeded 25% of the average value of the static days and the duration exceeded 3h as an ionospheric storm event, in which the geomagnetic activity index was used as the criterion for the selection of the static days, and the number of static days was not less than 7 days. Liu [17] defined a storm event as one in which ΔTEC exceeds 25% of the background level and lasts for more than 3 hours, where the average of the seven calm days before the storm is used to represent the background value.

In summary, there is no unified standard for the determination of ionospheric disturbed conditions and disturbed events. In order to better study the physical mechanism of ionospheric disturbances (storms), especially from the application point of view, providing the necessary and accurate ionospheric state information to the relevant equipment/users, performing ionospheric disturbed condition determination is the primary problem to be solved.

2. Data and Methods

2.1 Ionospheric TEC data and disturbed index

The data are ionospheric GPS-TEC data observed at the Chinese Taipei station with a data resolution of 15 min in this paper from January 2002 to July 2014. The cumulative duration is more than 12 years, about one solar activity cycle.

The ionospheric disturbed index is the index after normalizing the relative change of ionospheric TEC. It is as follows: the relative change of ionospheric TEC is defined first (equation 1), and then normalized (equation 2).

$$P_{TEC} = \frac{O_{TEC} - R_{TEC}}{R_{TEC}}, \tag{1}$$

$$N_{TEC} = \frac{P_{TEC} - \mu}{\sigma}, \tag{2}$$

Where O_{TEC} is the ionospheric observed TEC, R_{TEC} is the ionospheric TEC background value, the sliding median of 27 days is usually chosen, N_{TEC} is the ionospheric disturbed index μ is the mean of P_{TEC}, σ is the standard deviation of P_{TEC}.

From the above definitions, it is clear that differences in the background values of the ionosphere lead to differences in the final ionospheric disturbed index. In this paper, in order to better characterize the ionospheric disturbances physically, the method of using the median TEC sliding value of the past 27 days as the background value for Nishioka[5] is adjusted to the median sliding value of 13 days (27 days in total) before and after the observed time as the ionospheric background value, and then the ionospheric disturbed index is obtained. For comparison purposes, this paper uses N_{TEC} to denote Nishioka's $\hat{P}_{TEC}$.

2.2 Calculation of ionospheric TEC background values

Missing values may lead to change in ionospheric background values, so it is important to count the number of missing TEC data within the sliding window when obtaining R_{TEC} time series. In this paper, the number of missing TEC values is analyzed for the following calculation of ionospheric background values. The results are shown in Figure 1.

The horizontal axis in Figure 1 is the number of missing values. The vertical axis is the proportion of the number of corresponding missing values in each sliding window (27 O_{TEC} values) and cumulative proportion in Figure 1.a and Figure 1.b, respectively. As shown in Figure 1.a, the sliding window without missing values is about 40%. When the number of missing values increases, the corresponding proportion is decreasing rapidly. When the number of missing values is 7 or more, the corresponding proportion is almost less than 1%. As seen in Figure 1.b, with the gradual increase of the number of missing values, the corresponding proportion increases rapidly at the beginning and then slowly followed. It is almost stable when the number of missing values is more than 7. The proportion has exceeded 80%. Therefore, in order to obtain much data and to get a more accurate perturbation index for improving the confidence of the results, the criterion for calculating the background value R_{TEC} is that there are at most 7 missing values in the sliding window for calculating the value of R_{TEC} at a given moment in this paper. In other words, there are at least 20 valid TEC observations involved in the estimation of the background value.

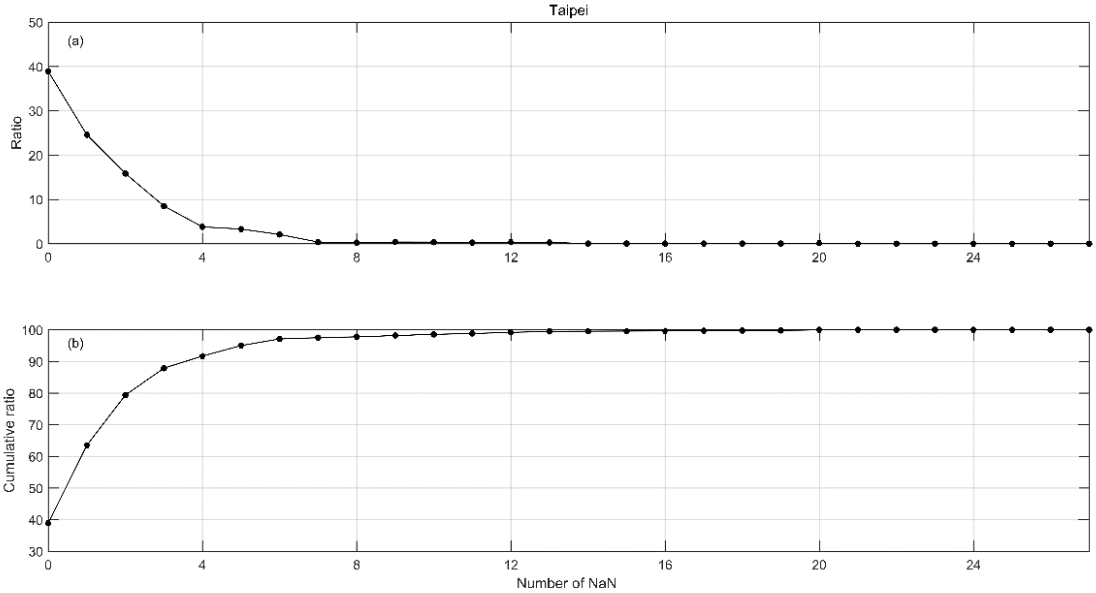


Figure1 The proportion of the number of TEC missing values in the 27-day sliding window

2.3 Cross-validation method

In order to determine the ionospheric disturbed state (ionospheric disturbed index value and the prorotion of the disturbed ionosphere), an optimized model with the idea of cross-validation method is built.

The cross-validation method is a model selected method in statistical learning, which aims to make the learned model have good predictive power for both existing and unknown data, and the predictive power of the learned method for unknown data is usually referred to as generalization power [18]. At the earliest, people trained models based on all data sets and then tested the error estimates of the models within the same data set, but the results of this method were too optimistic and generally failed to yield more accurate estimates, in order to solve this problem, Stone proposed the cross-validation method in 1974. The cross-validation method is a method that can directly estimate the model generalized error without assuming the data distribution in advance, and it is very popular and can be operated easily [19]. The basic idea of the cross-validation method is to repeatedly use the data, slice the given data set, combine the slice data set into a training set and a test set, and on this basis, repeatedly perform training, testing, and model selection to choose the model with the lowest tested error [18]. From the basic idea of the cross-validation method, it is clear that its purpose is to obtain reliable and stable models.

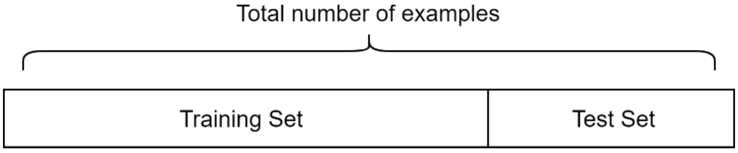


Figure 2 Hold-out verification schematic

Hold-out was proposed by Devroye and Wagner in 1979 [20], and the main idea was to take out a part of the sample set for training the model and the remaining part for testing, as in Figure 2, which was the original form of the cross-validation method. Expressed in mathematical language is that, let I' be a non-empty subset of the set $D_n = \{1, 2, \cdots, n\}$, I'' is its complementary set, use I' as the training set for model training, I'' as the test set for generalization error estimation, The model generated by the training is denoted by $A(D_n)$, the mathematical expression for the generalization error estimate is given by

$$R = \frac{1}{n_k} \sum_{i \in D_n^k} L(A(D_n); \xi_i), \tag{3}$$

where L is the loss function, D_n^k is a test sample, The number of its samples is n_k. Cross-validation methods also include leave-one-out cross-validation [21], leave-one-out P cross-validation [22], v-fold cross-validation [23], and 5 × 2 cross-validation [24].

In this paper, we choose v-fold cross-validation [23], the basic idea is to divide the sample set equally into v copies, take out v-1 copies of the sample set from v copies of the dataset as the training set each time, and the remaining copy of the dataset as the validation set, repeat the experiment v times, as in Figure 3, and finally average the results of v times as the generalized error estimate.

The v-fold cross-validation method, expressed in mathematical terms, is to have a data set D_n with sample size n, $A_1, \cdots, A_n$ is a subset of dataset D_n, and for any subset A_j we have $M(A_j) \approx n / v$, M is the number of samples in the subset and the final generalization error is estimated as

$$R = \frac{1}{v} \sum_{j=1}^{v} \left[\frac{1}{M(A_j)} \sum_{i \in A_j} L(s(D_n^{-(A_j)}); \xi_i) \right], \tag{4}$$

where L is the loss function, s is the training model, $D_n^{-(A_j)}$ is the remaining sample after removing the subset A_j. From the definition of v-fold cross-validation, it can be seen that this method only needs to train the samples v times, which can reduce the complexity of the computation and is a widely used model selected method in practical applications [19]. For the selection of the number of folds in the cross-validation method is not fixed, when the number of folds is large, the accuracy of the generalized error is better, but the computed time is very long; when the number of folds is small, the computed time and the number of experiments are reduced, but the accuracy of the generalized error is poor.

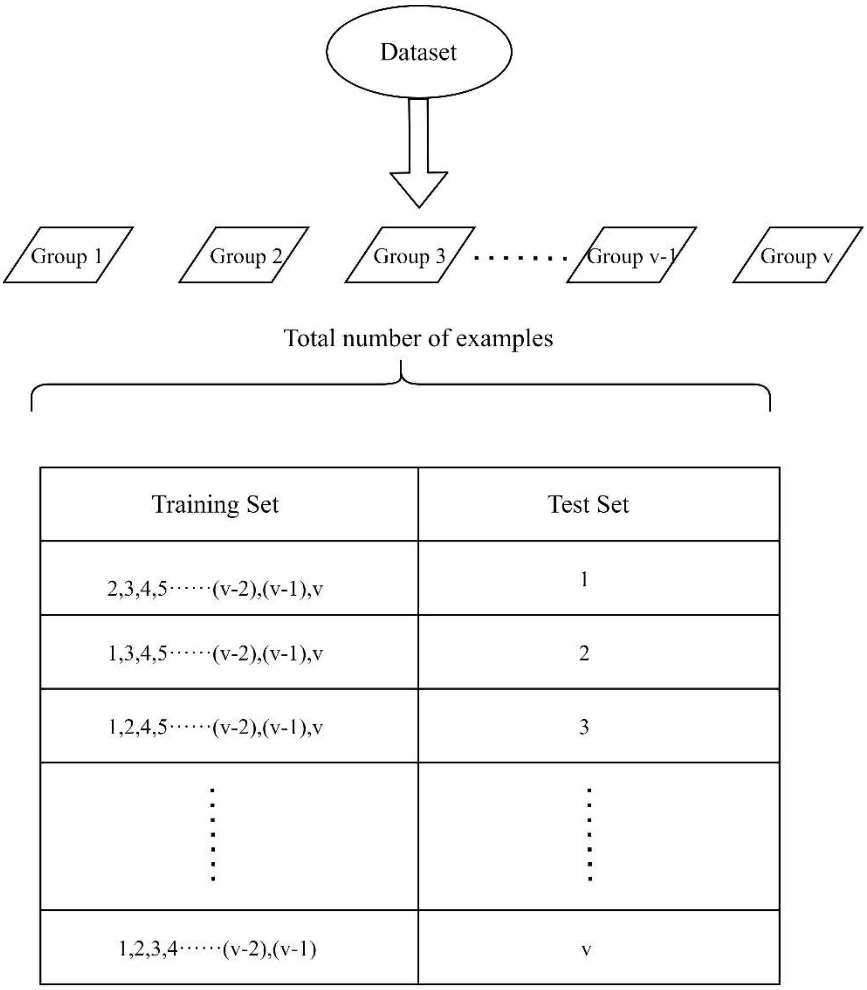
In the above method, the loss function L takes the following specific form.

(1) ME(Mean Error): measures the degree of unbiasedness of the estimate. the accuracy of the valuation, the smaller the value the more accurate, with the following formula.

$$ME = \frac{1}{n} \sum_{i=1}^{n} (\hat{Z}(x_i) - Z(x_i)), \tag{5}$$

(2) RMSE(Root Mean Square Error): measures the closeness of the model estimate to the true value, the smaller the value the closer it is, with the following formula.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{Z}(x_i) - Z(x_i))^2}{n}}, \tag{6}$$



Training Set	Test Set
2,3,4,5······(v-2),(v-1),v	1
1,3,4,5······(v-2),(v-1),v	2
1,2,4,5······(v-2),(v-1),v	3
⋮	⋮
1,2,3,4······(v-2),(v-1)	v

Figure 3 Schematic diagram of v-fold cross-validation

(3) MSE(Mean Standard Error): denotes the mean of the predicted standard error, and the formula is as follows.

$$MSE = \sqrt{\frac{\sum_{i=1}^{n} \sigma^2(x_i)}{n}}\ , \tag{7}$$

(4) RMSSE(Root Mean Standard Square Error): the closer the value is to 1, the more valid the standard error of the prediction is, with the following equation.

$$RMSSE = \sqrt{\frac{\sum_{i=1}^{n} \left[(\hat{Z}(x_i) - Z(x_i)) / \hat{\sigma}(x) \right]^2}{n}}\ , \tag{8}$$

where $\hat{Z}(x_i)$ is the estimated value of sampling point x_i, $Z(x_i)$ is the true value of sampling point x_i, n is the number of sample points.

2.4 Kernel density estimation method

The optimized model based on the cross-validation is constructed from the probability density function of the ionospheric disturbed index N_{13}.

There are two methods of calculating the probability density function: parametric estimation and nonparametric estimation. Parametric estimation is empirically given a specific distribution that the sample set obeys, and nonparametric estimation is to fit the density function from the data itself without assuming that the sample set obeys any specific distribution. The kernel density estimation method is used to estimate the unknown density function in probability statistics, and belongs to one of the nonparametric test methods that can fit complex nonlinear density functions, proposed by Rosenblatt [25] and Parzen [26], also known as the Parzen window. The kernel density estimation method does not use a priori knowledge about the data distribution and does not attach any assumptions to the data distribution; it is a method to study the characteristics of the data distribution from the data sample itself. The kernel density estimation method is expressed in mathematical language as, let $x_1, x_2, \cdots, x_n$ be an independent and identically distributed random variable in R. The formula for the kernel density estimation method is as follows.

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} k(\frac{x_i - x}{h}), x \in R, \quad (9)$$

Where $h > 0$, h is the window width or smooth parameter, n is the total number of samples, $\sum_{i=1}^{n} k(\frac{x_i - x}{h})$ is the kernel function, there are three common kernel functions of the following forms.

(1) Gaussian kernel:

$$k(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}), -\infty < u < \infty, \quad (10)$$

(2) Epanechnikov kernel:

$$k(u) = \frac{3}{4}(1-u^2), |u| \le 1, \quad (11)$$

(3) Bigweight or Quartic kernel:

$$k(u) = \frac{15}{16}(1-u^2), |u| \le 1, \quad (12)$$

Where $u = x_i - x$. In this paper, we choose to use the Gaussian kernel. The probability density function for fitting the Taipei station N_{13} data using the Gaussian kernel density estimation is shown in Figure 4.

Figure 4 shows the probability density distribution of Taipei station N_{13}, the horizontal axis is the value of N_{13}, Figure 4a shows the probability density function of N_{13}, named as $\mathrm{f_{all}}$ in section 3.2. Figure 4b shows the probability distribution function of N_{13}.

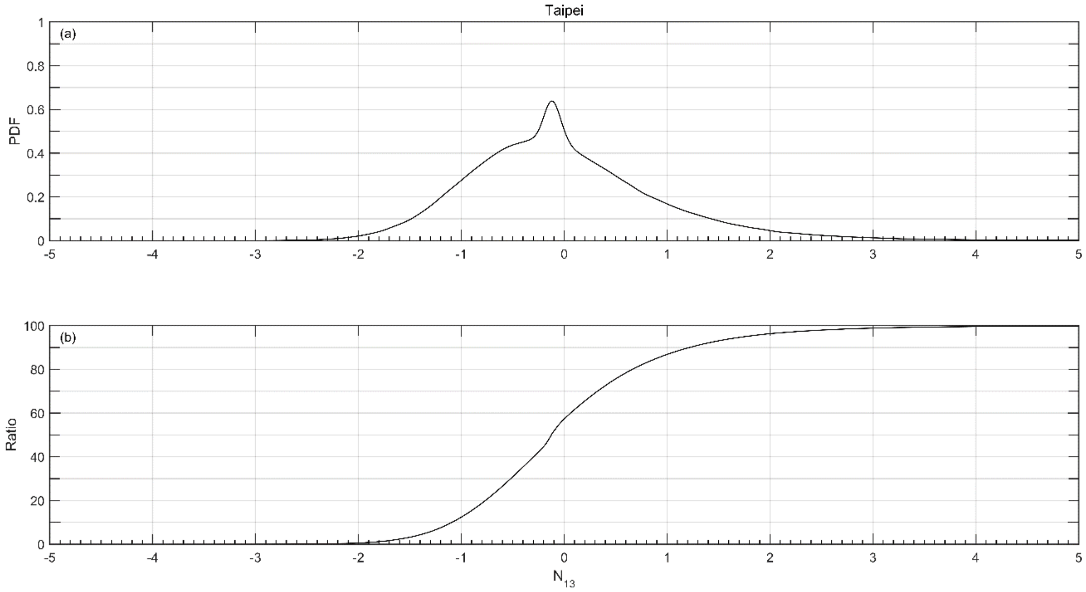


Figure 4 Probability density distribution of N13 at Taipei

3. Results

3.1 The difference between the modified index (N_{13}) and the original index (N_{27})

1) Differences in ionospheric background values

In this paper, the ionospheric disturbed index is corrected by adjusting the ionospheric background value. As mentioned above, different ionospheric background values can lead to the difference of ionospheric disturbed index. Figure 5 shows the variation of the difference of a different method for ionospheric background values with season and local time at Taipei station. The vertical coordinate axis in the figure is the difference between the sliding median value of the past 27 days minus the sliding median value of the 13 days before and after under the same moment. The horizontal axis is LT (Local Time). The red curve with star is the difference LT in summer, and the red line is its mean value. The black, blue and green lines correspond to winter, spring and autumn, respectively. Here, the spring, summer, autumn and winter seasons are formed from 45 days before and after the vernal equinox, summer solstice, autumn equinox and winter solstice, which are labeled as ME (March Equinox season), JS (June Solstice season), SE (September Equinox season) and DS (December Solstice season), respectively. As seen in Figure 5, the difference becomes larger when the local time is from 06:00 to 18:00 LT, and the maximum difference is about 4.4TECu, 4.6TECu, 4.9TECu and 2TECu in spring, summer, autumn and winter, respectively. In addition, the difference is positive in summer and winter, while it is almost negative at all LT in spring and autumn. The mean values of the difference in the four seasons can reach 2.2TECu, 2TECu, 1.3TECu and 1.3TECu, respectively. From the Taipei station, it can be seen that the ionospheric background values vary in both LT and season. In general, the different ionospheric background values vary significantly and are not constant.

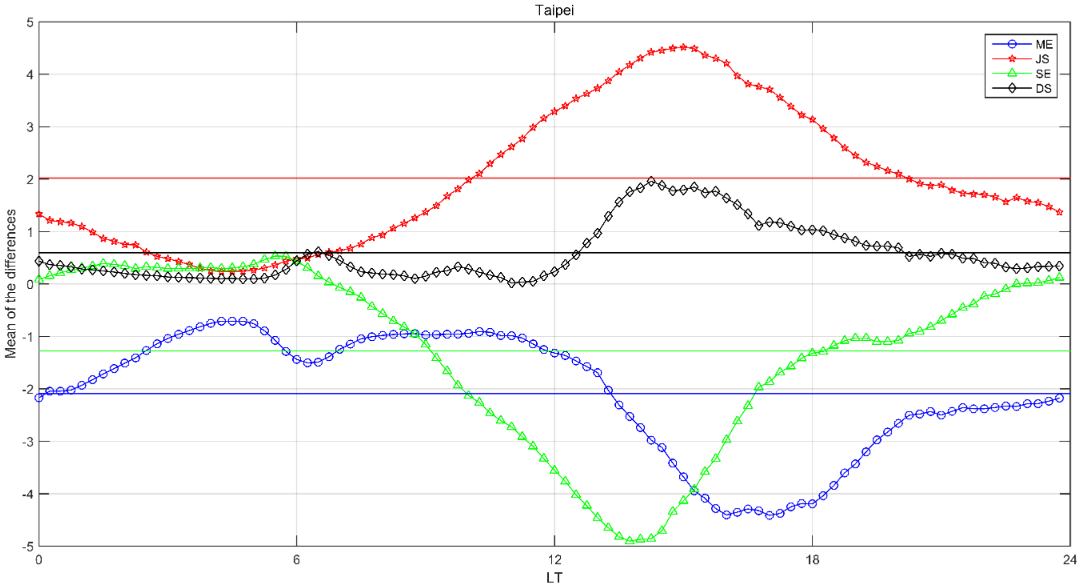


Figure 5 Variation of the difference of different ionospheric background values with LT and season

2) N_{13} index characteristics

In this paper, N_{13} is chosen as an index to determine the ionospheric state. The variation of N_{13} with local time and season is analyzed. Figure 6 shows the distribution of N_{13} with local time at each disturb level (the criteria are referred to Nishioka[5]). The horizontal axis represents LT, and the vertical axis is the proportion of the occurrence in the corresponding level. As seen from the figure, at the quiet level, the percentage of quiet periods at each moment is about 80%, which basically does not change with local time. At the positive and moderate negative disturbances, the percentage of each moment is about 5%, which varies somewhat, but not obviously, with local time. At the strong negative storm, the percentage of each moment varies more obviously with LT, which gets the smallest value around midnight and obtains larger values mainly at sunrise and in the afternoon.

Figure 7 shows the variation of Taipei station N_{13} with seasons. Subplots a-e represent five classes: strong negative disturbance, moderate negative disturbance, quiet period, moderate positive disturbance, and strong positive disturbance, respectively. In one subfigure, the horizontal axis represents the four seasons, and the vertical axis represents the occurrence proportion of the corresponding ionospheric state.

As can be seen from Figure 7, the proportion of quiet period in each season is basically unchanged, which is about 80%. It is almost the same in 4 seasons. While in the other disturb levels, the ratio varies with season slightly.

In general, when N_{13} is used as the ionospheric disturbed index, the determination of the ionospheric quiet state basically does not vary with local time and season, which is the same as N_{27} in Nishioka[5]

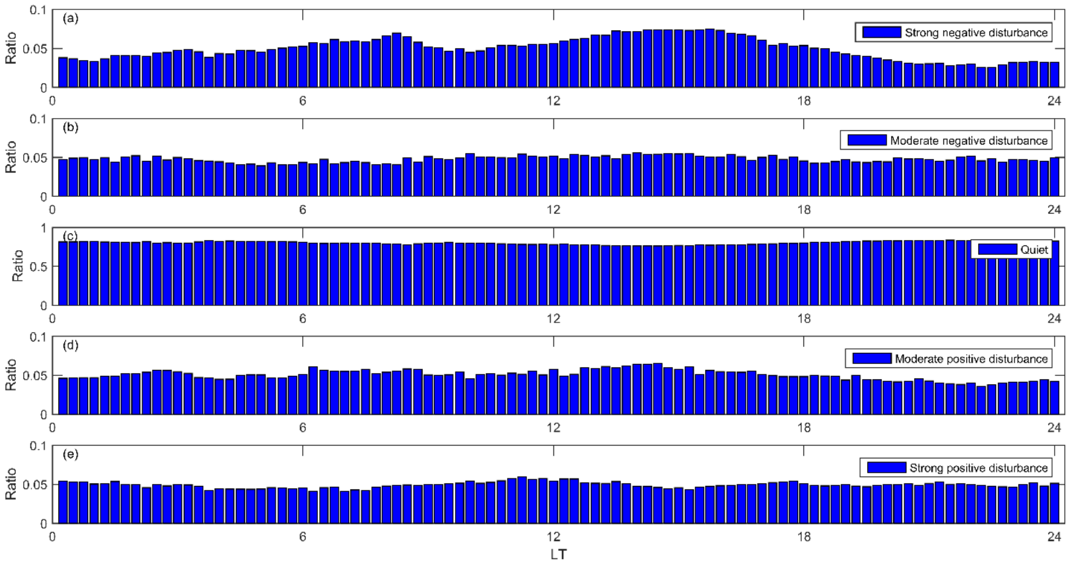


Figure 6 Variations of N_{13} with LT in different disturb level at Taipei station

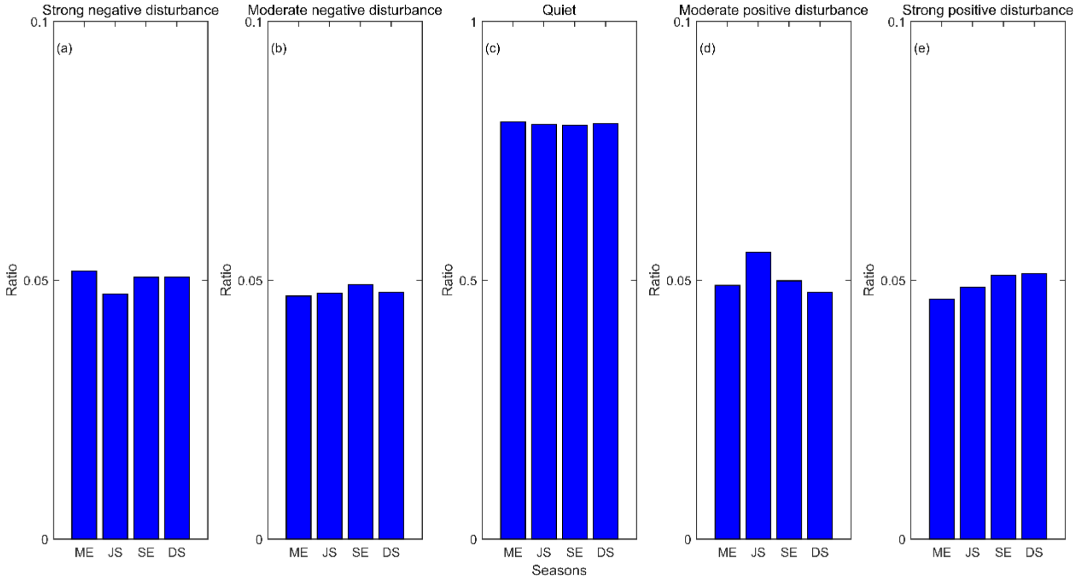


Figure 7 Variation of N_{13} with season in different disturb level at Taipei station

3）Disturbance level difference

According to the disturbed class classification method of Nishioka [5], the ionospheric disturbance is classified into five classes according to the value of N_{TEC} : strong negative disturbance when $N_{TEC} \leq 2$, moderate negative disturbance when $-2 < N_{TEC} \leq -1$, quiet period when $-1 < N_{TEC} \leq 1$, moderate positive disturbance when $1 < N_{TEC} \leq 3$, and strong positive disturbance when $N_{TEC} > 3$. In this section, the class difference is studied between the index of N_{13} and N_{27}. The results are shown in Figure 8.

In figure 8, the ionospheric state is first divided into 5 classes as strong negative disturbance, moderate negative disturbance, quiet period, moderate positive disturbance and strong positive disturbance according to N_{27}, listed in sub-figure a-e. the data amounts of each class are 1103, 47484, 298897, 49312 and 4915, respectively. Then, the ionospheric disturbance index N_{13} is used in each class to judge the ionospheric state based on the same standard above. The same ionospheric state is represented by the same colour. As shown in Figure 8, the strong negative disturbance, moderate negative disturbance, quiet period, moderate positive disturbance and strong positive disturbance are dark blue, pink, yellow, green and light blue, respectively. It can be seen clearly that, in each subfigure, there are always the different ionospheric state based on the different N_{TEC}. The digital details are shown on the top of the color histograms. The numbers are the ratio of the difference, namely the amount of ionospheric state according to N_{27} divided by the amounts of the ionospheric state based on N_{13} in each subfigure. The percentages of the same judgements are 86.13%, 72.15%, 90.71%, 65.1%, and 59.39% for strong negative disturbance, moderate negative disturbance, quiet period, moderate positive disturbance, and strong positive disturbance, respectively. A more obvious difference is shown in subfigure e. It can be seen that, in the period of strong positive disturbance based on N_{27}, there are about 38.51% classified as moderate positive disturbance, even 2.1% classified as quiet period based on N_{13}. It exists in the other classes shown in subfigure a-d.

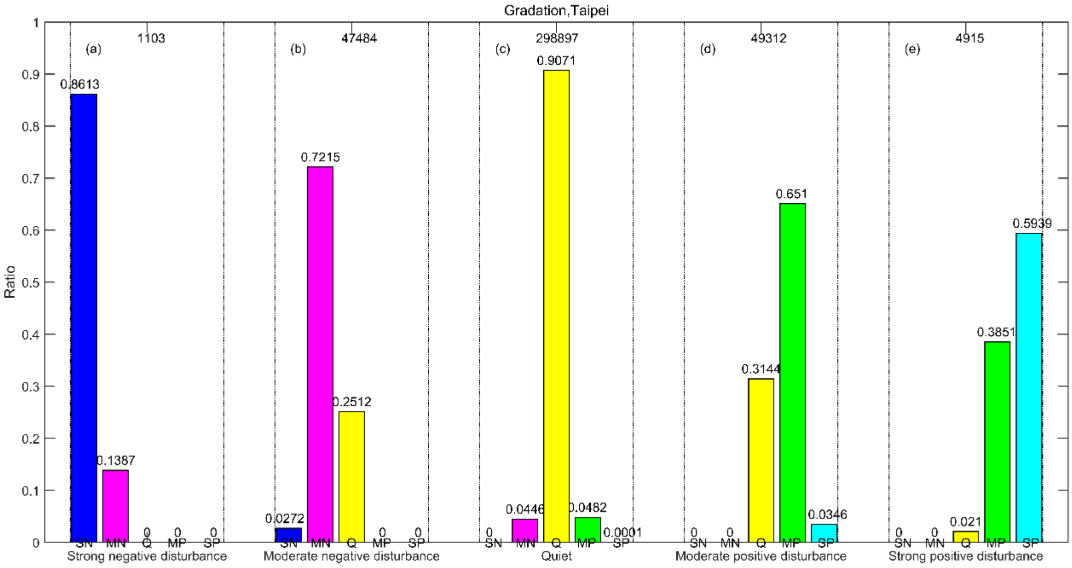


Figure 8 The class difference is studied between the index of N_{13} and N_{27} at Taipei station

3.2 The determination of Ionospheric disturbance

Based on the parameter N_{TEC}, the ionospheric TEC condition can be judged subjectively, but there is no clear basis for the determination criteria of quiet period or disturbed condition. In this paper, based on the N_{13} index, the cross-validation method is used to construct the optimized model to find out the ionospheric disturbed proportion and the criteria of ionospheric disturbance. The specific approaches are as follows.

(1) The critical index value for determining the ionospheric disturbed condition and the quiet period is defined as the $\alpha/2$ quantile of the distribution function (Note: α is a undetermined variable). Based on the distribution function obtained from all data (Figure 4b in Section 2.3), the N_{13} value corresponding to the upper and lower $\alpha/2$ quantile is obtained as the critical value for determining the ionospheric disturbed condition or the quiet period. Naturally, the ionospheric disturbed proportion is α.

(2) The original data are divided into multiple data sets, consisting of training set and test set, for the cross-validation method. Firstly, the data of 2002 to 2014 is divided into 13 copies according to the year. Secondly, the first copy of the 13 ones is selected as the test set, and the left 12 copies are used as the training set, which forms the first group dataset. By analogy, there are still another 12 groups dataset formed. The datasets are named as the year of the test set.

Based on the kernel density estimation method in Section 2.3, the probability density functions of training sets, named f_m, are obtained as shown in Figure 9. In general, the probability density function curves of each group of data have similar trends and are approximately symmetrically distributed, and they have a maximum value of about 0.65, when N_{13} is less than 0.

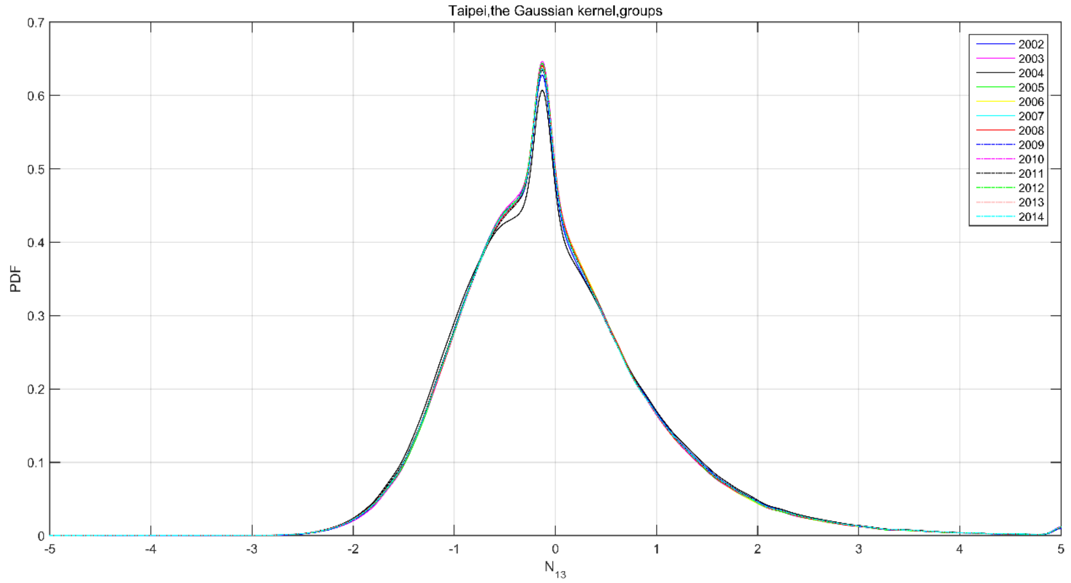


Figure 9 Probability density function of each group of N_{13} data at Taipei station

(3) According to the definition in approach (1), some N13 data in each test set are chosen for a given α in the function of $\mathrm{f}_{\mathrm{all}}$ as disturb state. These data are denoted as $x_m^{1'}, x_m^{2'}, \cdots, x_m^{m'}$. Taking these m' data into formula (13), the sum of squares of errors for the jth dataset is gotten.

$$\overset{\wedge}{\sigma}{}^{2}_{\alpha j} = \frac{1}{m'}\sum_{i'=1}^{m'}(f_{all}(x_m^{i'}) - f_m(x_m^{i'}))^2 , \tag{13}$$

Where $\alpha = 1, 2, \cdots, 40$, $j = 1, 2, \cdots, n$, n represents the number of groups, $\hat{\sigma}^2_{\alpha j}$ is the error sum of squares, namely the average variance.

Figure 10 shows the variation of $\hat{\sigma}^2_{\alpha j}$ with the disturbance proportion α. The horizontal axis is the ionospheric disturbed proportion α, and the vertical axis is the $\hat{\sigma}^2_{\alpha j}$ of each group of data. Generally speaking, each $\hat{\sigma}^2_{\alpha j}$ increase with disturbed proportion α until to about 12% clearly except 2004. In order to get more accurate result, the average situation is considered in the next step.

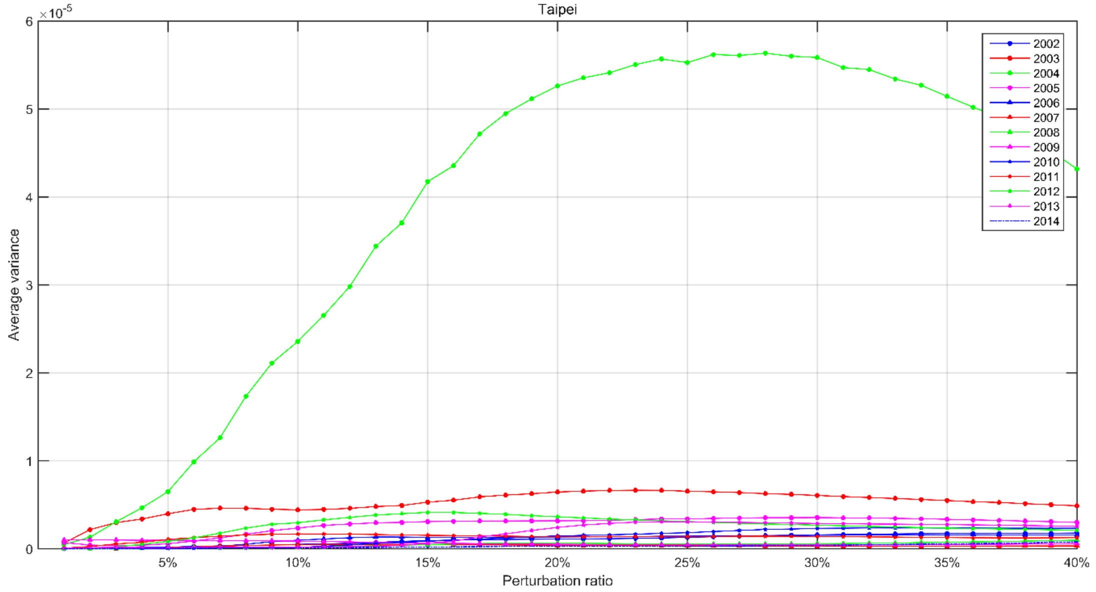


Figure 10 Variation of $\hat{\sigma}^2_{\alpha j}$ with the proportion of disturbance α

(4) According to the cross-validation method based on the probability density function, the optimal ionospheric TEC disturbed proportion α should get the maximum of $\hat{\sigma}_{\alpha}'$ in formula (14).

$$\hat{\sigma}_{\alpha}' = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\hat{\sigma}_{\alpha j}^{2}} \ , \tag{14}$$

Where n is the number of dataset.

Figure 11 shows the variation of Taipei station $\hat{\sigma}_{\alpha}$ ' with the disturbed proportion α . Considering the special variation of 2004 (group 3) in Figure 10, all 13 groups and 12 groups except group3 are checked, which are shown in Fig. 10a and Fig. 10b, respectively. It can be seen both of the $\hat{\sigma}_{\alpha}$ ' increase with α gradually and achieves a maximum value when α is about 25% . Then, the $\hat{\sigma}_{\alpha}$ ' is basically stable and no longer changes significantly. It means that the optimal ionospheric TEC disturbed proportion α is 25%. By checking the f_{all} above, the corresponding N_{13} should be about $N_{TEC} \leq -1$ or $N_{TEC} > 1$.

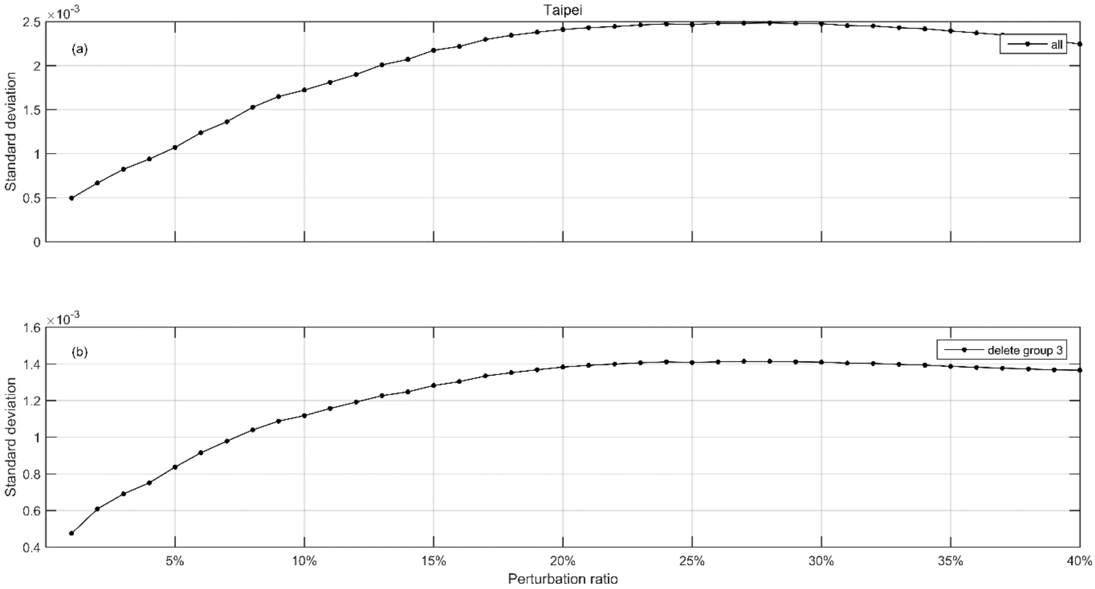


Figure 11 Variation of Taipei station $\hat{\sigma}_{\alpha}$ ' with the proportion of disturbance α

4. Conclusion

Based on the ionospheric TEC data obtained from Taipei station, the ionospheric TEC disturbed characteristics are studied by a modified N_{TEC} index firstly, where the TEC background value is replaced with the 27days centered at the very moment. Then the Cross-validation method is introduced to judge the ionospheric state. The main results are as follows.

1）The new index almost keeps the same statistical characters against with season and local time, however, they are always different at one time. It is suggested that the ionospheric state may be decided by the index.

2）the Cross-validation method is effective method as theory foundation to judge the ionospheric state, which is always defined subjectively. Based on the modified NTEC index, the results of Cross-validation method show that the proportion of the disturbed ionospheric state is about 25%, and the corresponding NTEC range is less than -1 or larger than 1.

References:

1. Bremer J , Cander L R , Mielich J , et al. Derivation and test of ionospheric activity indices from real-time ionosonde observations in the European region. Journal of Atmospheric and Solar-Terrestrial Physics, 2006, 68(18), 2075-2090.
2. Mielich J , Bremer J . A modified index for the description of the ionospheric short- and long-term activity. Annales Geophysicae, 2010, 28(12), 2227-2236.
3. Gulyaeva T L, Stanislwska I, Tomasik M. Ionospheric weather: Cloning missed foF2 observations for derivation of variability index. Annales Geophysicae, 2008, 26(2), 315-321.
4. Jakowski N , Stankov S M , Schlueter S , et al. On developing a new ionospheric perturbation index for space weather operations. Advances in Space Research, 2006, 38(11), 2596-2600.
5. Nishioka M , Tsugawa T , Jin H , et al. A new ionospheric storm scale based on TEC and foF2 statistics. Space

Weather, 2017, 15(1), 228-239.

6. Kouris S S . Thresholds of TEC variability describing the plasmaspheric disturbed state. Acta Geophysica, 2008, 56(2), 408-416.
7. Lekshmi D V, Balan N, Ram S T, et al. Statistics of geomagnetic storms and ionospheric storms at low and mid latitudes in two solar cycles. Journal of Geophysical Research Space Physics, 2011, 116, A11328, doi: 10.1029/2011JA017042.
8. Matamba T M, Habarulema J B, Mckinnell L A. Statistical analysis of the ionospheric response during geomagnetic storm conditions over South Africa using ionosonde and GPS data. Space Weather, 2015, 13(9), 536-547.
9. Huang. The negative phase ionospheric disturbance and its solar-terrestrial correlation. Chinese Journal of Space Science, 1985(4), 303-307.
10. Huang. The research on normal phase ionospheric disturbance. Chinese Journal of Space Science , 1990(2), 52-56.
11. Chen D J ,Wu J ,Wang X Y. A technology study of foF2 forecasting during the ionospheric disturbance. Chinese J . Geophys, (in Chinese) , 2007, 50 (1), 18-23.
12. Qin G , Libo L , Biqiang Z , et al. Statistical study of the storm effects in middle and low latitude ionosphere in the East-Asian sector. Chinese J, Geophys, 2008, 51(3), 626-634.
13. Liu J , Zhao B , Liu L . Time delay and duration of ionospheric total electron content responses to geomagnetic disturbances. Annales Geophysicae, 2010, 28(3), 795-805.
14. Deng Z X, Liu R Y, Zhen W M. Study on the ionospheric TEC storms over China. Geophys, (in Chinese), 2012, 55(7), 2177-2184.
15. Li C B. Statistical study of ionosphere mid/low-latitude anomalous disturbances. Xidian University, 2012.
16. Wu J S, Xu L. Statistical study of the ionospheric storms over 5 latitude zones in the European sector. Geophys. (in Chinese), 58(2), 349-361.
17. Liu W, Liang X, Chao X, et al. The ionospheric storms in the American sector and their longitudinal dependence at the northern middle latitudes. Advances in Space Research, 2016, 59(2), 603-613.
18. Li H. Statistical learning methods. Tsinghua University Press, 2012.
19. Fan Y D. A summary of cross-validation in model selection. Shanxi University, 2013.
20. Devroye L P, Wagner T J. Distribution-free performance bounds for potential function rules. IEEE Transactions on Information Theory, 1979, 25(5), 601-604.
21. Geisser S. A predictive approach to the random effect model. Biometrika, 1974, 61(1): 101-107.
22. Shao J. Linear Model Selection by Cross-validation. Journal of the American Statistical Association, 1993, 88(422), 486-494.
23. Geisser S. The predictive sample reuse method with application. Journal of the American Statistical Association, 1975, 70(350), 320-328.
24. Dietterich T. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation, 1998, 10(7), 1895-1923.
25. Rosenblatt M. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics, 1956, 27(3), 832-837.
26. Parzen E. On estimation of a probability density function and mode. Annals of Mathematical Statistics, 1962, 33(3), 1065-1076.