

Land subsidence prediction model based on its influencing factors and machine learning methods

fengkai li (✉ lfkgiser@163.com)

Shandong University of Science and Technology

Guolin Liu

Shandong University of Science and Technology <https://orcid.org/0000-0001-5044-1750>

Qiuxiang Tao

Shandong University of Science and Technology

Min Zhai

Shandong University of Science and Technology

Research Article

Keywords: Land subsidence, InSAR, Pearson correlation coefficient, Land subsidence prediction model, Machine learning methods

Posted Date: August 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1881099/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Natural Hazards on December 29th, 2022. See the published version at <https://doi.org/10.1007/s11069-022-05796-9>.

Abstract

Land subsidence has caused huge economic losses in the Beijing plains (BP) since 1980s. Building land subsidence prediction models that can predict the development of land subsidence is of great significance for improving the safety of cities and reducing economic losses in Eastern Beijing plains. The pattern of evolution of land subsidence is affected by many factors including groundwater level in different aquifers, thicknesses of compressible layers, and static and dynamic loads caused by urban construction. First, we used the small baseline subset Interferometric Synthetic Aperture Radar (SBAS-InSAR) technology on 47 ENVISAT ASAR images and 48 RADARSAT-2 images and used Persistent Scatterers Interferometric Aperture Radar (PS-InSAR) technology on 27 Sentinel-1 images to obtain the land subsidence monitoring results from June 2003 to September 2018. Second, the accuracy of the InSAR monitoring results were validated by using leveling benchmark land subsidence monitoring results. Finally, we built land subsidence rate prediction models and land subsidence gradient prediction models by combining land subsidence influencing factors and four machine learning methods including support vector machine (SVM), Gradient Boosting Decision Tree (GBDT), Random forest (RF) and Extremely Randomized Trees (ERT). The findings show: (1) The InSAR monitoring results revealed that the maximum land subsidence rate reached -110.7 mm/year, -144.4 mm/year and -136.8 mm/year during the 2003–2010, 2011–2015 and 2016–2018 periods, respectively. (2): The InSAR monitoring results agreed well with the leveling benchmark monitoring results with the Pearson correlation coefficients of two monitoring results were 0.97, 0.96 and 0.95 during the 2003–2010, 2011–2015 and 2016–2018 periods, respectively. (3): We found that the land subsidence prediction based on ERT method is the optimal model among four land subsidence prediction models and that the prediction performance of land subsidence prediction model based on ERT method will be greatly improved when apply this prediction model in sub study areas where the land subsidence mechanism is similar owing to the similar hydrogeological parameters.

1. Introduction

With the loss of surface elevation, land subsidence has become one of the most common geological disasters (Castellazzi et al. 2017; Galloway et al. 1998). Land subsidence, especially for uneven land subsidence, could causes huge economic losses and damages infrastructure, including rupturing underground pipelines and sinking foundations (Strozzi et al. 2017; Ng et al. 2012). By the end of 2018, more than 150 countries around the world, including Italy (Samsonov al. 2014), China (Luo et al. 2019; Li et al. 2020), Mexico (Castellazzi et al. 2016; Castellazzi et al. 2017), and United States (Galloway al. 1998), had recorded the occurrence of land subsidence. The cities experienced land subsidence in Chinese mainland are mainly distributed in the North China Plain (NCP) (Zhu et al. 2015; Chen et al. 2016; Chen et al. 2020), the Fenwei Basin (Wan Zhao et al. 2011; Qu et al. 2014) and the Yangtze River Delta (Chen et al. 2003; Hu et al 2014; Ye et al. 2016). Among these areas, the middle and northern regions of the NCP, where Beijing is located, occur most serious land subsidence (Li et al. 2020).

As the capital city of China, Beijing is a megacity and rated as the world's first tier city (Alpha+) by the globalization and World Cities Study Group and network (GAWC) in 2018. Beijing has a Gross Domestic Product (GDP) of over USD 452 billion and total population of over 21 million by the end of 2018. Domestic water and economic development consume a large amount of groundwater (Zhou et al. 2020), which lead to the dramatic decrease of groundwater level and the consolidation of aquifer systems; then causing serious land subsidence (Chen et al. 2020). Beijing Plain (BP) is prone to subsidence owing to the existence of Quaternary loose sediments, especially for the Eastern Beijing Plain where the thickness of Quaternary loose sediments exceed 100 meters and experienced serious land subsidence (Zuo et al. 2019; Li et al. 2020). The static and dynamic loads caused by urban construction have accelerated the process of land subsidence development in the BP (Yang et al. 2018). Previous studies have shown that the land subsidence in BP was mainly affected by groundwater level, the thickness of compressible layers, and urban static and dynamic loads (Yang et al. 2018; Zhou et al. 2019; Li et al. 2020).

As a new quantitative microwave remote sensing technology, Interferometry synthetic aperture radar (InSAR) could obtain land subsidence monitoring information with millimeter accuracy (Berardino et al. 2002). However, its application scope is limited by many factors, including atmospheric delay, spatial and temporal decorrelation, topographic error and thermal noise. In order to solve those obvious deficiencies and expand the application scope of this technology, temporal sequential InSAR (TS-InSAR) technology was proposed. As a typical TS-InSAR technology, Persistent Scatterers Interferometric Aperture Radar (PS-InSAR) could detect points with stable backscattering characteristics as permanent scatterers (PSs) and acquire land subsidence monitoring results with millimeter-scale accuracy (Ferretti et al. 2001). Previous studies have shown that this technology has been widely used in the field of obtaining land subsidence monitoring information (Thapaa et al. 2016). Small baseline subset InSAR (SBAS-InSAR) technology is another expansion and improvement of InSAR technology. Relevant research has proved that this technology can efficiently obtain larger-scale regional land subsidence monitoring results with millimeter-scale accuracy (Chen et al. 2019).

The traditional land subsidence prediction models can be divided into two categories, including physical prediction model and mathematical prediction model. The physical prediction model simulates the process of land subsidence by combining the influencing factors of land subsidence (Nie et al. 2015). This model can explain the mechanism of land subsidence well and obtain reliable prediction results (Zhu et al. 2020). However, physical prediction model need many hydrological parameters that are difficult to obtain, which greatly limits the scope of its application (Zhou et al. 2020). The mathematical prediction model uses mathematical functions to express the statistical characteristics of historical land subsidence (Zhou et al. 2020). The mathematical prediction model have a wider range of applications than physical prediction model. However, mathematical prediction model has a poor prediction performance when the dispersion of land subsidence value is large (Li et al. 2021). This phenomenon should be attributed to the fact that the mathematical prediction model cannot explain the mechanism of land subsidence. The land subsidence prediction model based on machine learning method can integrate the advantages of the physical prediction model and mathematical prediction model and overcome the shortcomings of the physical prediction model and mathematical prediction model. Previous studies

have shown that the land subsidence prediction model based on machine learning method can obtain reliable prediction results (Shi et al. 2020; Li et al. 2021).

This paper was organized as follows. First, we illustrated the geographical and geological conditions of the study area in Section 2. In Section 3, we describe the InSAR technology and machine learning methods in detail. Then, we obtain the land subsidence monitoring results and assess the accuracy of the InSAR monitoring results by using leveling benchmarks in Section 4. We built the land subsidence rate prediction models and land subsidence gradient prediction models by using machine learning methods in Section 5. Finally, we introduced our conclusions in Section 6.

2. Study Area And Datasets

2.1 Study area

Beijing locates in the northwest fringe of NCP and expands from 115°24' E to 117°34' E and 39° 24' N to 41° 03'N (Fig. 1). Geographically, Beijing can be divided into three geographical units, which namely western mountains, northern mountains, and southeastern plains (Chen et al. 2015), respectively. Beijing covers an area approximately 16,400 km², of which plains areas and mountainous areas account for 38% and 62% (Shi et al. 2020), respectively.

The BP lies in the southeast part Beijing city and has typical temperate continental monsoon climate. With an annual average precipitation of 598 mm from 1949 to 2020, BP experienced severe uneven distribution of precipitation across the year, of which 80% concentrated from June to September (Fig. 2). Beijing has a total population of over 21 million and GDP of over USD 452 billion by the end of 2018. Groundwater has been extracted extremely to meet the need of inhabitant survival and economic development, which greatly exceeded its natural replenishment capacity (Gao et al.2018). Excessive groundwater exploitation has led to obvious decline of the groundwater level in BP, which resulted in an increase of effective stress, thus causing consolidation of aquifer systems and serious land subsidence (Galloway et al. 1998; Chaussard et al. 2014).

BP is a typical piedmont alluvial-proluvial plain, which composed by loose sediments carried by Ji Canal, Yongding River, Chaobai River, Daqing River, and Wenyu River (Shi et al. 2020). Different thickness of Quaternary sediments have been deposited in BP during the long geological period, which provide necessary geological conditions for land subsidence (Zhou et al. 2019). According to recharge conditions of groundwater and characteristics of Quaternary loose sediments, the aquifer system in BP can be divided into four groups, including a phreatic aquifer, a first confined aquifer, a second confined aquifer, and a third confined aquifer. The hydrogeological parameters of each aquifer are listed in Table 1.

Previous studies have shown that BP has experienced serious land subsidence since the 1970s (Guo et al. 2019). Five land subsidence funnels have appeared in the Beijing plain, including Shunyi-Pinggezhuang land subsidence funnel,

Shahe-Baxianzhuang land subsidence funnel, Laiguangying land subsidence funnel, Dongbalizhuang-Dajiaoting land subsidence funnel and Yulong-Lixian land subsidence funnel. Dongbalizhuang-Dajiaoting land subsidence funnel is the earliest and most typical among those land subsidence funnels and shows a tendency to merge with Laiguangying land subsidence funnel. Based on those findings, the Laiguangying–Dongbalizhuang–Dajiaoting (LDD) land subsidence funnel was selected as the study area in this study.

Table 1
Hydrogeological parameters of each aquifer in Beijing Plain (BP).

Aquifer	Buried depth of roof	Geological age of strata
The phreatic aquifer	20–40 m	Holocene
The first confined aquifer	80–100 m	Late Pleistocene
The second confined aquifer	100–180 m	Middle Pleistocene
The third confined aquifer	200–300 m	Lower Pleistocene

2.2 Datasets

In this study, three radar images sets, including ENVISAT ASAR (EA), RADARSAT-2 (R2) and Sentinel-1 (S1) were used to obtain land subsidence monitoring results. EA images set contained 47 descending radar images collected from June 2003 to August 2010. R2 images set contained 48 descending radar images obtained from November 2010 to November 2015. S1 images set contained 27 ascending radar images acquired from January 2016 to September 2018. Detailed image parameters for EA images set, R2 images set and S1 images set are listed in Supplementary Table S1, Supplementary Table S2 and Supplementary Table S3, respectively. The EA satellite, R2 satellite and S1 satellite all operate in the 5.6 cm C-band. However, those three satellites have different revisiting time and incident angles. The detailed parameters of each images set are listed in Table 2. Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM) with a 90 m resolution were used to remove the topographic phase in interferograms. SRTM DEM data supplied by the Shuttle Radar Topography Mission covers more than 80% of the land area of the world (detailed information can be found at <http://dds.cr.usgs.gov/srtm/>).

Forty-nine leveling benchmarks—17, 20 and 20 leveling benchmarks which measured precise land subsidence monitoring results during the 2003–2010, 2013–2015 and 2016–2018 periods, respectively—were used to assess the accuracy of the InSAR monitoring results. Their locations can be found in Fig. 1a. The groundwater level in different aquifers during three different periods were used to present the degree of groundwater extraction; the lower the groundwater level indicates more severity of groundwater exploitation. The thickness of Quaternary loose sediments was used to present the strength of

Quaternary sedimentation; the greater the thickness of Quaternary sediments, the more conducive to the development of land subsidence. Building loads were used to present static load information and subway lines, expressways, and urban roads were used to present the dynamic load information.

Table 2
Detailed parameters of each images set

Parameters	ENVISAT ASAR (EA)	RADARSAT-2 (R2)	Sentinel-1 (S1)
Polarization	VV	VV	VV
Orbit direction	Descending	Descending	Ascending
Spatial resolution (m)	30	30	30
Revisiting time (Day)	35	24	12
Incident angle (°)	22.9	27.6	34
Image number	47	48	27
Image acquisition time	June 2003 to August 2010	November 2010 to November 2015	January 2016 to September 2018

3. Methodology And Dataset

Figure 3 displays the flowchart of this study. Firstly, we used the SBAS-InSAR method on EA images set and R2 images set to obtain the land subsidence information during the 2003–2010 and 2013–2015 periods; at the same time, we used PS-InSAR method on S1 images set to obtain the land subsidence information during the 2016–2018 period. Secondly, the accuracy of InSAR land subsidence monitoring results was assess by using the leveling benchmark monitoring results. Thirdly, we chose machine learning methods to build the land subsidence rate prediction models and land subsidence gradient prediction models.

3.1 InSAR Methodology

3.1.1 Persistent Scatterers Interferometric Aperture Radar (PS-InSAR) technology

PS-InSAR technology was proposed firstly in 2000 to eliminate the problem of low correlation of interference phase caused by atmospheric delay, orbit error, terrain error and other factors which greatly limit the application range of InSAR technology (Ferretti et al. 2000a, Ferretti et al. 2000b). Permanent Scatterers (PS) points can maintain stable backscattering characteristics for a long time (Ferretti et al. 2001). These points correspond to urban roads, buildings and exposed rocks in the wild. PS-InSAR technology has been widely used in the field of land subsidence monitoring and has obtained credible results (Shi et al. 2020).

The phase caused by deformation can be separated from the following equation according to the imaging geometric relationship of PS InSAR technology:

$$\phi = \varphi_{deformation} + \varphi_{topographic} + \varphi_{orbit} + \varphi_{atmospheric} + \varphi_{noise}$$

1

where ϕ is differential interferometric phase of two SAR images, $\varphi_{deformation}$ is the deformation phase in line-of-sight (LOS) direction, $\varphi_{topographic}$ is the residual topographic phase, φ_{orbit} is the phase affected by orbit inaccuracy, $\varphi_{atmospheric}$ is the atmospheric delay phase, φ_{noise} is the thermal noise of radar system. After remove other phases, $\varphi_{deformation}$ can be obtained; then the surface deformation rate along the LOS direction can be calculated. Considering that the deformation along horizontal in BP can be ignored (Zhou et al. 2019), land subsidence rate can be estimated by using the following equation:

$$V = \frac{V_{LOS}}{\cos \theta}$$

2

where V is land subsidence rate (deformation rate in the vertical), V_{LOS} is the deformation rate in LOS direction and θ is incidence angle of the S1 radar sensor.

In this study, we applied PS-InSAR technology which has been integrated in SARProz software on S1 images set to obtain the land subsidence monitoring results. According to the principle of minimizing the sum of temporal baseline, spatial baseline and Doppler centers frequency, the image collected in July 2017 was selected as master image. The main image and other slave images form 26 interferograms. Supplementary Fig. 1 shows the network of the spatial and temporal baselines for S1 images set. The points which amplitude stability coefficient greater than 0.7 were selected as PS points.

3.1.2 Small baseline subset Interferometric Synthetic Aperture Radar (SBAS-InSAR) technology

SBAS-InSAR technology was proposed firstly in 2002 (Berardino et al. 2002). SBAS-InSAR technology form interferograms by setting temporal baseline threshold and spatial baseline threshold to further eliminate the influence of temporal and spatial decoherence and atmospheric noise, and obtain more accurate surface deformation information. This technology divides SAR images which spatial baseline is less than the setting threshold into one group, called a subset. According to the setting threshold, all SAR

images are divided into several subsets; SAR images within the same subsets can be form interferograms, and different subsets are connected by common SAR images.

The phase caused by surface deformation can be obtain after remove other phases from interferograms. Previous study have shown that the deformation along horizontal in BP can be ignored (Zhou et al. 2019). Based on the above findings, the land subsidence rate can be estimated by using the Eq. (2).

In this study, SBAS-InSAR technology which has been integrated in GAMMA software were used on EA images set to and R2 images set obtain the land subsidence monitoring results. We generated 267 interferograms for EA images set when temporal baseline threshold and perpendicular baseline threshold were set as 500 days and 500 m, respectively. We generated 163 interferograms for R2 images set when temporal baseline threshold and perpendicular baseline threshold were set as 500 days and 500 m, respectively. Supplementary Fig. 2 shows the network of the spatial and temporal baselines for the EA images set and R2 images set.

3.2 Machine learning methods

Machine learning is an interdisciplinary covering many fields including probability theory, statistics, pattern recognition, approximation theory and artificial intelligence. Machine learning methods simulate the process of human thinking and learning by using computers. Machine learning has experienced four stages of development since it was proposed in the 1950s. Machine learning methods have been used in analyzing of land subsidence influencing factors (Zhou et al. 2019, Li et al. 2020) and building land subsidence prediction model (Shi et al. 2020). In this study, we used SVM method, GBDT method, RF method and ERT method to build land subsidence rate prediction models and land subsidence gradient prediction models. These four algorithms have been programmed in Python. The results of those four prediction models were cross-validated.

3.1.1 Support vector machine

SVM is a machine learning method based on statistics theory. The core of this algorithm is to find the segmentation hyperplane that can correctly divide the training set and has the largest geometric distance (Vapnik et al. 1998). SVM is a supervised learning method, which has been widely used in statistical classification, regression analysis and pattern recognition (Qin et al. 2005, Sun et al. 2002). Owing to excellent classification performance and strong generalization ability, SVM has been widely used in machine learning cases. However, SVM also has some disadvantages, such as low efficiency when used in large size training sample and sensitive to missing data.

3.1.2 Gradient Boosting Decision Tree

GBDT is a effective supervised machine learning method which has been widely used in regression analysis and prediction model cases (Friedman al. 2001). This algorithm obtain the optimal prediction model by adjusting the weight of the basis learner to obtain the minimum value of the loss function. This algorithm were composed by decision tree and gradient boosting. The decision tree method is used to

determine the location of the optimal segmentation by iterative training. Gradient boosting was used to reduce the error of current model in each iteration. The optimal GBDT model can be obtain when the value of the loss function reaches its minimum.

3.1.3 Random Forest

RF is one of most representative algorithm of Bagging in ensemble learning (Breiman al. 1996, Breiman et al. 2001). To improve the accuracy and generalization performance of the whole model, RF obtain the final model result by combining multiple learners and taking the mean value of the combined learners. RF has the many advantages, including high accuracy, processing the training set with high-dimensional features without dimensionality reduction, evaluating the importance of each feature, and insensitive to loss values. RF also has the many disadvantages including large computational burden in feature collection, taking more time in model training, prone to fitting and poor interpretation.

3.1.4 Extremely Randomized Trees

ERT algorithm is an efficient supervised machine learning method expanded from RF algorithm. The ERT model could get more credible than RF mode by using all sample sets in training process. ERT model has stronger robustness owing to its base learner has more representative when compared with RF model. ERT model could obtain a better prediction model and a wider range of applications owing to this algorithm can overcome effectively some shortcomings existing in RF model. ERT model is an efficient and concise method, especially for large datasets which need more time to train the all of sample sets, to get a credible model.

4 Results And Discussion

4.1 Land subsidence monitoring results measured by InSAR method during three time periods

Based on the InSAR land subsidence monitoring results, we obtain three land subsidence rate maps of BP during the 2003–2010, 2011–2015 and 2016–2018 periods. Figure 4 shows that land subsidence rate varied greatly in BP and uneven land subsidence was obvious in BP. The areas where have experienced severe land subsidence mainly located in the east part of Chaoyang District, the northwest part of Tongzhou District, the middle and the south part of Changping District, the west part of Shunyi District and the south part of Daxing District. A total of 272264 pixels were detected in BP from 2003 to 2010, with a density of 50 pixel/km². The maximum land subsidence rate was - 110.7 mm/year during this period. A total of 290197 pixels were detected in BP from 2010 to 2015, with a density of 46 pixel/km². The maximum land subsidence rate reached - 144.4 mm/year during this period. A total of 531087 pixels were detected in BP from 2016 to 2018, with a density of 83 pixel/km². The maximum land subsidence rate was - 136.8 mm/year during this period.

We divide the land subsidence rate maps in the three periods into five categories to better study the evolution law of land subsidence in BP. Figure 5 and Tables 3, 4 and 5 shows the classification results. Table 3 shows that the area with land subsidence rate varied from - 150 to -120 mm/year, -120 to -90 mm/year, -90 to -60 mm/year, -60 to -30 mm/year and - 30 to 15 mm/year account for 0%, 0.3%, 2.6%, 12.2% and 84.9%, respectively. Table 4 shows that the area with land subsidence rate varied from - 150 to -120 mm/year, -120 to -90 mm/year, -90 to -60 mm/year, -60 to -30 mm/year and - 30 to 15 mm/year account for 0.2%, 1.6%, 3.1%, 11.3% and 83.8%, respectively. Table 5 shows that the area with land subsidence rate varied from - 150 to -120 mm/year, -120 to -90 mm/year, -90 to -60 mm/year, -60 to -30 mm/year and - 30 to 15 mm/year account for 0.2%, 1.2%, 1.1%, 9.4% and 88.1%, respectively. We found that the area with land subsidence rate varied from - 150 to -120 mm/year, -120 to -90 mm/year and - 90 to -60 mm/year has increased and that the area with land subsidence rate varied from - 60 to -30 mm/year and - 30 to 15mm/year has reduced by comparing Table 3 and Table 4. Those findings indicate that the development of land subsidence in BP showed an accelerate trend from 2010 to 2015. We found that the area with land subsidence rate varied from - 120 to -90 mm/year, -90 to -60 mm/year and - 60 to -30 mm/year has reduced and that the area with land subsidence rate varied from - 30 to 15mm/year has increased by comparing Table 4 and Table 5. Those findings indicate that the development of land subsidence in BP showed a slowing trend from 2016 to 2018.

Figure 6 shows that land subsidence has experienced a rapid development stage in BP from 2003 to 2018 and the spatial discrepancy of land subsidence has increased gradually. By the end of 2018, the maximum cumulative land subsidence in BP has reached 1716.3 mm; at the same time, the area with cumulative land subsidence exceed 300 mm has reached 1524.9 km², accounting for 23.8% of the total area of the BP. Several land subsidence funnels have formed in BP. LDD subsidence funnels has experienced the most serious land subsidence among them. Therefore, we selected LDD land subsidence funnel as the study area and build land subsidence rate prediction models and land subsidence gradient prediction models for LDD land subsidence funnel by using four machine learning methods.

Table 3
Classification results of land subsidence rate map from June 2003 to August 2010

2003–2010	Land subsidence rete (mm/year)	Number of detected pixels	Proportion (%)
NO 1	-150 ~ -120	0	0
NO 2	-120 ~ -90	1580	0.3
NO 3	-90 ~ -60	9095	2.6
NO 4	-60 ~ -30	36764	12.2
NO 5	-30 ~ 15	224825	84.9

Table 4

Classification results of land subsidence rate map from November 2010 to November 2015

2010–2015	Land subsidence rete (mm/year)	Number of detected pixels	Proportion (%)
NO 1	-150 ~ -120	613	0.2
NO 2	-120 ~ -90	3580	1.6
NO 3	-90 ~ -60	10240	3.1
NO 4	-60 ~ -30	38330	11.3
NO 5	-30 ~ 15	237434	83.8

Table 5

Classification results of land subsidence rate map from January 2016 to September 2018

2016–2018	Land subsidence rete (mm/year)	Number of detected pixels	Proportion (%)
NO 1	-150 ~ -120	932	0.2
NO 2	-120 ~ -90	4587	1.2
NO 3	-90 ~ -60	13624	1.1
NO 4	-60 ~ -30	91330	9.4
NO 5	-30 ~ 15	420614	88.1

4.2 Accuracy verification of InSAR land subsidence monitoring results

A total of 49 leveling benchmarks were used to verify the reliability of InSAR land subsidence monitoring results. We selected leveling benchmarks as datum points and extracted the InSAR monitoring points

within 200m of each benchmark. We compared land subsidence monitoring rate of each benchmark with average land subsidence rate of the extracted points during three time periods. Figure 7 shows that the Pearson correlation coefficient between two land subsidence monitoring results during the 2003–2010, 2011–2015 and 2016–2018 periods were 0.97, 0.96 and 0.95, respectively. In order to further verify the accuracy of InSAR land subsidence results, we counted the maximum error, minimum error and root mean square error (RMSE) between the two land subsidence monitoring results. Table 6 present that the RMSE of two land subsidence monitoring results were less than 10 mm/year in three time periods. Those findings indicated that InSAR land subsidence monitoring results have high accuracy, which provides reliability for subsequent research.

Table 6

Comparison of the InSAR land subsidence monitoring result and leveling benchmark monitoring results in three time periods.

Time period	Maximum error (mm/year)	Minimum error (mm/year)	Root mean square error (mm/year)
2003–2010	-11.4	0.8	5.9
2011–2015	15.2	0.7	6.9
2016–2018	14.7	0.7	5.2

4.3 Land subsidence rate prediction models based on machine learning methods

In this study, 54636 pixels detected by the InSAR technology were selected as research data to build land subsidence rate prediction models by using four machine learning methods including SVM method, GBDT method, RF method and ERT method. Among these data, 70% are selected as training data and the remaining 30% are selected as verification data. Firstly, four land subsidence rate prediction models were built by using four machine learning methods. Secondly, the optimal land subsidence rate prediction model was choose by comparing the accuracy and RSME of the four models; at the same time, the study area was divided into 7 sub study areas according to hydrogeological parameters (Luo al. 2019). Figure 8 shows the classification results of the study area. Table 7 shows the hydrogeological parameters of 7 sub study areas. Finally, the land subsidence rate prediction models was built for each sub study area by using the optimal land subsidence rate prediction model.

Table 7
Hydrogeological parameters of 7 sub study areas.

Sub study area	Thickness of compressible layer (m)	Permeability coefficient (m/day)	Aquifer
A	220–230	< 20	
B	170–190	< 20	
C	150–180	< 20	
D	170–190	20–50	
E	110–130	20–50	,
F	110–130	< 20	,
G	110–130	< 20	

, present the second confined aquifer and the third confined aquifer, respectively.

4.3.1 Comparison of land subsidence rate prediction models based on machine learning methods

We selected the land subsidence rate during the 2011–2015 period as the dependent variable, while groundwater level in the phreatic aquifer, groundwater level in the first confined aquifer, groundwater level in the second confined aquifer, groundwater level in the third confined aquifer, the thickness of compressible layers and additional stress engendered from static and dynamic loads in same time period and land subsidence rate during the 2003–2010 period were the explanatory variables. We built four land subsidence rate prediction models by using SVM method, GBDT method, RF method and ERT method, respectively. Figure 9 shows the results of four land subsidence rate prediction models during the 2011–2015. We found that the land subsidence rate prediction model based on ERT method has the highest accuracy and the lowest RSME among those four prediction models.

We selected the land subsidence rate during the 2016–2018 period as the dependent variable, while groundwater level in the phreatic aquifer, groundwater level in the first confined aquifer, groundwater level in the second confined aquifer, groundwater level in the third confined aquifer, the thickness of compressible layers and additional stress engendered from static and dynamic loads in same time period and land subsidence rate during the 2011–2015 period were the explanatory variables. We built four land subsidence rate prediction models by using SVM method, GBDT method, RF method and ERT method,

respectively. Figure 10 shows the results of four land subsidence rate prediction models during the 2016–2018. We also found that the land subsidence rate prediction model based on ERT method has the highest accuracy and the lowest RSME among those four prediction models. Based on the above findings, we choose the land subsidence rate prediction model based on ERT model as the optimal land subsidence prediction model in this study.

4.3.2 Land subsidence rate prediction models based on ERT method in sub study areas.

We built land subsidence rate prediction models by using ERT method during the 2011–2015 and 2016–2018 periods in 7 sub study areas. The parameter settings in this section are the same as those in Section 4.3.1. Figure 11 and Fig. 12 show the results of land subsidence rate prediction models based on ERT method in 7 sub study areas during the 2011–2015 and 2016–2018 periods, respectively. We found that the accuracy of the land subsidence rate prediction model based on ERT method in each sub study area increased and the RSME has decreased when compared with land subsidence rate prediction model based on ERT method applied on the whole study area, except for the sub study area labeled B. This is mainly due to the similar hydrogeological conditions in each sub study area.

4.4 Land subsidence gradient prediction models based on machine learning methods

In this study, we also selected 54636 pixels detected by the InSAR technology as research data to build land subsidence gradient prediction models by using four machine learning methods including SVM method, GBDT method, RF method and ERT method. Of these data, 70% are selected as training data and the remaining 30% are selected as verification data. Firstly, four land subsidence gradient prediction models were built by using four machine learning methods. Secondly, the optimal land subsidence gradient prediction model was choose by comparing the accuracy and RSME of the four models. Finally, the land subsidence gradient prediction models was built for each sub study area by using the optimal land subsidence gradient prediction model.

4.4.1 Comparison of land subsidence gradient prediction models based on machine learning methods

We selected the land subsidence gradient during the 2011–2015 period as the dependent variable, while groundwater level gradient in the phreatic aquifer, groundwater level gradient in the first confined aquifer, groundwater level gradient in the second confined aquifer, groundwater level gradient in the third confined aquifer, the gradient of thickness of compressible layers and the gradient of additional stress in same time period and land subsidence gradient during the 2003–2010 period were the explanatory variables. Four land subsidence gradient prediction models were built by using SVM method, GBDT method, RF method and ERT method, respectively. Figure 13 shows the results of four land subsidence gradient prediction models during the 2011–2015. We found that the land subsidence gradient prediction model

based on ERT method has the highest accuracy and the lowest RSME among those four prediction models.

We selected the land subsidence gradient during the 2016–2018 period as the dependent variable, while groundwater level gradient in the phreatic aquifer, groundwater level gradient in the first confined aquifer, groundwater level gradient in the second confined aquifer, groundwater level gradient in the third confined aquifer, the gradient of thickness of compressible layers and the gradient of additional stress in same time period and land subsidence gradient during the 2011–2015 period were the explanatory variables. Four land subsidence gradient prediction models were built by using SVM method, GBDT method, RF method and ERT method, respectively. Figure 14 shows the results of four land subsidence gradient prediction models during the 2016–2018. We also found that the land subsidence gradient prediction model based on ERT method has the highest accuracy and the lowest RSME among those four prediction models. Based on the above findings, we choose the land subsidence gradient prediction model based on ERT model as the optimal land subsidence prediction model in this study.

4.4.2 Land subsidence gradient prediction models based on ERT method in sub study areas.

We built land subsidence gradient prediction models by using ERT method during the 2011–2015 and 2016–2018 periods in 7 sub study areas. The parameter settings in this section are the same as those in Section 4.4.1. Figure 15 and Fig. 16 show the results of land subsidence gradient prediction models based on ERT method in 7 sub study areas during the 2011–2015 and 2016–2018 periods, respectively. We found that the accuracy of the land subsidence gradient prediction model based on ERT method in each sub study area increased and the RSME has decreased when compared with land subsidence gradient prediction model based on ERT method applied on the whole study area, except for the sub study area labeled B. This is mainly due to the similar hydrogeological conditions in each sub study area.

5 Conclusion

In this research, we obtained the land subsidence monitoring results from June 2003 to September 2018 covering Beijing Plain by using SBAS-InSAR method on 47 EA images and 48 R2 images and using PS-InSAR method on 27 S1 images; at the same time, we assess the accuracy of InSAR land subsidence monitoring results. Based on the land subsidence monitoring results, we analyzed the evolution law of land subsidence in Beijing Plain. Land subsidence rate prediction models and land subsidence gradient models in LDD land subsidence funnel were built by using machine learning methods. The following conclusions can be derived:

1. The Beijing plain experienced serious land subsidence from 2003 to 2018, with maximum cumulative land subsidence reached 1716.3 mm and maximum land subsidence rate reached – 110.7 mm/year, –144.4 mm/year and – 136.8 mm/year during the 2003–2010, 2011–2015 and 2016–2018 periods, respectively. The area with the most serious subsidence occurs in the east part of Chaoyang District and the northwest part of Tongzhou District, where LDD land subsidence funnel

located. The InSAR land subsidence monitoring results agree well with the that of leveling benchmark, and the Pearson correlation coefficient between two land subsidence monitoring results all greater than 0.95 during three study period.

2. With highest accuracy and the lowest RSME, the land subsidence rate prediction model based on ERT method was the optimal prediction model among four land subsidence rate prediction models based on SVM method, SVM method, GBDT method, RF method and ERT method. The accuracy of the land subsidence rate prediction model based on ERT method during the 2011–2015 and 2016–2018 periods was 94% and 93.7%, respectively. Because the land subsidence mechanism is similar in areas with similar hydrogeological parameters, the land subsidence rate prediction models in the sub study area have better prediction performance when compared with the performance of land subsidence rate prediction model based on ERT method applied to the whole study area.
3. We found that land subsidence gradient prediction model based on ERT method was the optimal prediction model among four land subsidence gradient prediction models. The prediction performance of land subsidence gradient prediction model will be greatly improved when land subsidence gradient prediction model based on ERT method was applied to sub study areas with similar hydrological parameters. This finding provides a new method for land subsidence gradient prediction.

Declarations

Acknowledgments

This work was Supported by the National Natural Science Foundation of China (Grant No. 42074009; Grant No. 41404003), and the Shandong Natural Science Foundation (Grant No. ZR2020MD044; Grant No. ZR2020MD043).

Conflict of interest: We do not have any conflict of interest, financial or non-financial that are directly or indirectly related to the work submitted here for publication.

Ethical approval: The study followed the accepted principles of ethical and professional conduct throughout the research work. No animals or human participants are involved in this research.

References

1. Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 1996.
2. Breiman, L. (2001). Random Forests. *Mach. Learn.* 2001, 45, 5–32
3. Berardino P, Fornaro G, Lanari R, Sansosti E (2002). A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans. Geosci.Remote Sens*, 40, 2375–2383. Doi: 10.1109/TGRS.2002.803792.
4. Berardino P, Fornaro G, Lanari R, Sansosti E. (2003) A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on*

- Geoscience & Remote Sensing, 40(11):2375-2383. doi 10.1109/TGRS.2002.803792.
5. Castellazzi P, Arroyodomínguez N, Martel R et al (2016) Land subsidence in major cities of Central Mexico: Interpreting InSAR-derived land subsidence mapping with hydrogeological data. *Int J Appl Earth Obs* 47:102-111. doi:10.1016/j.jag.2015.12.002
 6. Castellazzi P, Garfias J, Martel R, Brouard C, Rivera A (2017) InSAR to support sustainable urbanization over compacting aquifers: The case of Toluca Valley, Mexico. *Int. J. Appl. Earth Obs. Geoinf*, 63, 33–44. doi: 10.1016/j.jag.2017.06.011.
 7. Chaussard E.; Wdowinski S.; Cabral-Cano E.; Amelung F (2014). Land subsidence in central Mexico detected by ALOS InSAR time-series. *Remote Sens. Environ*, 140, 94–106. doi: 10.1016/j.rse.2013.08.038.
 8. Chen B, Gong H, Li X, Lei K, Gao M, Zhou C (2015) Spatial–temporal evolution patterns of land subsidence with different situation of space utilization. *Nat Hazards* 77(3): 1765–1783. doi:10.1007/s11069-015-1674-1.
 9. Chen B, Gong H, Lei K, Li J, Zhou C, Gao M, Guan H, Lv W (2010). Land subsidence lagging quantification in the main exploration aquifer layers in Beijing plain, China. *Int. J. Appl. Earth Obs. Geoinf*. 2019, 75, 54–67.
 10. Chen B, Gong H, Chen Y, Li X, Zhou C, Lei K, Zhu L, Duan L, Zhao X, (2020). Land subsidence and its relation with groundwater aquifers in Beijing Plain of China. *Sci. Total Environ*, 735, 139111
 11. Chen C, Pei S, Jiao J (2003) Land subsidence caused by groundwater exploitation in Suzhou City, China. *Hydrogeol J* 11(2):275-287. doi:10.1007/s10040-002-0225-5
 12. Chen M, Tomás R, Li Z, Motagh M, Li T, Hu L, Gong H, Li X, Yu J, Gong X (2016) Imaging Land Subsidence Induced by Groundwater Extraction in Beijing (China) Using Satellite Radar Interferometry. *Remote Sens-Basel* 8(6):468. doi: 10.3390/rs8060468
 13. Ferretti A , Prati C, (2000) Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry . *IEEE Transactions on Geoscience & Remote Sensing*,38(5):2202-2212..DOI 10.1109/36.868878
 14. Ferretti A, Prati C, Rocca F (2000). Analysis of Permanent Scatterers in SAR interferometry [C]//IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120). IEEE, 2(132):761–763.
 15. Ferretti A, Prati C, Rocca F (2001) Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39(1): 8–20.
 16. Friedman J H., 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189-1232. DOI: 10.1214/aos/1013203451
 17. Galloway D L, Hudnut K W, Ingebritsen S E, Phillips S P, Peltzer G, Rogez F, et al. (1998) Detection of aquifer system compaction and land subsidence using interferometric synthetic aperture radar, Antelope Valley, Mojave Desert, California. *Water. Resour Res* 34(10):2573–2585. doi: 10.1029/98WR01285

18. Gao, M.; Gong, H.; Chen, B.; Li, X.; Zhou, C.; Min, S.; Yuan, S.; Zheng, C.; Duan, G. (2018) Regional Land Subsidence Analysis in Eastern Beijing Plain by InSAR Time Series and Wavelet Transforms. *Remote Sensing*. 10(3):365. doi: 10.3390/rs10030365.
19. Guo L , Gong H , Zhu F , Zhu L, Zhang Z, Zhou C, Gao M, Sun Y. (2019) Analysis of the Spatiotemporal Variation in Land Subsidence on the Beijing Plain, China. *Remote Sensing*, 11(10) doi: 10.3390/rs11101170.
20. Hu R., Yue Z, Wang L, Wang S. (2014) Review on current status and challenging issues of land subsidence in China. *Eng. Geol.* 2004, 76, 65–77.
21. Li F, Gong H, Chen B, Zhou C, Guo L (2020). Analysis of the Contribution Rate of the Influencing Factors to Land Subsidence in the Eastern Beijing Plain, China Based on Extremely Randomized Trees (ERT) Method. *Remote Sensing*, 12(18):2963. doi 10.3390/rs12182963.
22. Li H, Zhu L, Dai Z, Gong H, Guo T, Guo G, Wang J, Teatini, P (2021). Spatiotemporal modeling of land subsidence using a geographically weighted deep learning method based on PS-InSAR. *Sci. Total Environ*, 799, 149244–149257. <https://doi.org/10.1016/j.scitotenv.2021.149244>.
23. Luo Y, Chen B, Lei K, Hu R, Ye C, Cui W (2019). Optimum design of level monitoring points for land subsidence. *Bulletin of Engineering Geology and the Environment*. 78(11). doi:[10.1007/s10064-018-01442-6](https://doi.org/10.1007/s10064-018-01442-6).
24. Ng H M, Ge L, Li X, Abidin H Z, Andreas H, Zhang K (2012) Mapping land subsidence in Jakarta, Indonesia using persistent scatterer interferometry (PSI) technique with ALOS PALSAR. *Int J Appl Earth Obs* 18:232-242. doi: 10.1016/j.jag.2012.01.018
25. Nie L, Wang H, Xu Y, Li Z. 2015. A new prediction model for mining subsidence deformation: the arc tangent function model. *Nat Hazards*. 75(3):2185–2198. doi: 10.1007/s11069-014-1421-z
26. Qin J, He Z, (2005). A SVM face recognition method based on Gabor-featured key points[C]// *Machine Learning and Cybernetics. Proceedings of 2005 International Conference on*. 2005.
27. Qu F, Zhang Q, Lu Z, Zhao C, Yang C, Zhang J (2014) Land subsidence and ground fissures in Xi'an, China 2005–2012 revealed by multi-band InSAR time-series analysis. *Remote Sens Environ* 155:366-376. doi: 10.1016/j.rse.2014.09.008
28. Samsonov S V, D'Oreye N, González P J, Tiampo K F, Ertolahti L, Clague J J (2014) Rapidly accelerating subsidence in the Greater Vancouver region from two decades of ERS-ENVISAT-RADARSAT-2 DInSAR measurements. *Remote Sens Environ* 143(3):180-191. doi: 10.1016/j.rse.2013.12.017
29. Shi L, Gong H, Chen B , Zhou C (2020) Land Subsidence Prediction Induced by Multiple Factors Using Machine Learning Method. *Remote Sensing*, 12(24):4044. doi: 10.3390/rs12244044.
30. Strozzi, T.; Caduff, R.; Wegmüller, U.; Raetzo, H.; Hauser, M. (2017) Widespread surface subsidence measured with satellite SAR interferometry in the Swiss alpine range associated with the construction of the Gotthard BaseTunnel. *Remote Sens. Environ.* 190, 1–12. doi: 10.1016/j.rse.2016.12.007.

31. Sun A, Lim E, Ng W, (2002). Web classification using support vector machine. In Proceedings of the 4th international workshop on Web information and data management.
32. Thapaa S, Chatterjee R, Singhb K, Kumar D (2016). Land Subsidence Monitoring Using Ps-Insar Technique for L-Band Sar Data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 2016, XLI-B7, 995–997.
33. Vapnik, V. *Statistical learning theory* [M]. Springer, 1998
34. Yang Q, Ke Y, Zhang D, Chen B, Gong H, Lv M, Zhu L, Li X (2018). Multi-Scale Analysis of the Relationship between Land Subsidence and Buildings: A Case Study in an Eastern Beijing Urban Area Using the PS-InSAR Technique. *Remote Sens.*10, 1006.
35. Ye S, Luo Y, Wu J, Yan X et al. (2016). Three-dimensional numerical modeling of land subsidence in Shanghai, China. *Hydrogeol J* 24(3): 695-709. doi:10.1007/s10040-016-1382-2.
36. Zhao C, Zhang Q, Yang C, Zou W (2011). Integration of modis data and short baseline subset (sbas) technique for land subsidence monitoring in datong, china. *J Geodyn* 52(1): 16-23. doi:10.1016/j.jog.2010.11.004
37. Zhou C, Gong H, Chen B, Li X, Li J, Wang X , Gao M, Si Y., Guo L, Shi M, (2019) . Quantifying the contribution of multiple factors to land subsidence in the Beijing Plain, China with machine learning technology. *Geomorphology*, 335, 48–61. doi:10.1016/j.geomorph.2019.03.017.
38. Zhou Q, Hu Q, Ai M, Xiong C, Jin H. (2020) An improved GM(1,3) model combining terrain factors and neural network error correction for urban land subsidence prediction, *Geomatics, Natural Hazards and Risk*, 11:1, 212-229, doi:10.1080/19475705.2020.1716860
39. Zhu L, Gong H, Teatini P, Li X, Wang R, Chen B, Dai Z, Teatini P (2015). Land subsidence due to groundwater withdrawal in the northern beijing plain, china. *Eng Geol* 193: 243-255. doi:10.1016/j.enggeo.2015.04.020
40. Zhu L, Franceschini A, Gong H, Ferronato M, Dai Z, Ke Y, Pan Y, Li X, Wang R, Teatini P (2020). The 3-D Facies and Geomechanical Modeling of Land Subsidence in the Chaobai Plain, Beijing. *Water Resources Research*, 56(3).doi 10.1029/2019WR027026.
41. Zuo J, Gong H, Chen B, Liu K, Zhou C, Ke Y (2019) Time-series evolution patterns of landsubsidence in the eastern Beijing Plain, China. *Remote Sens.*11, 539. doi: 10.3390/rs11050539.

Figures

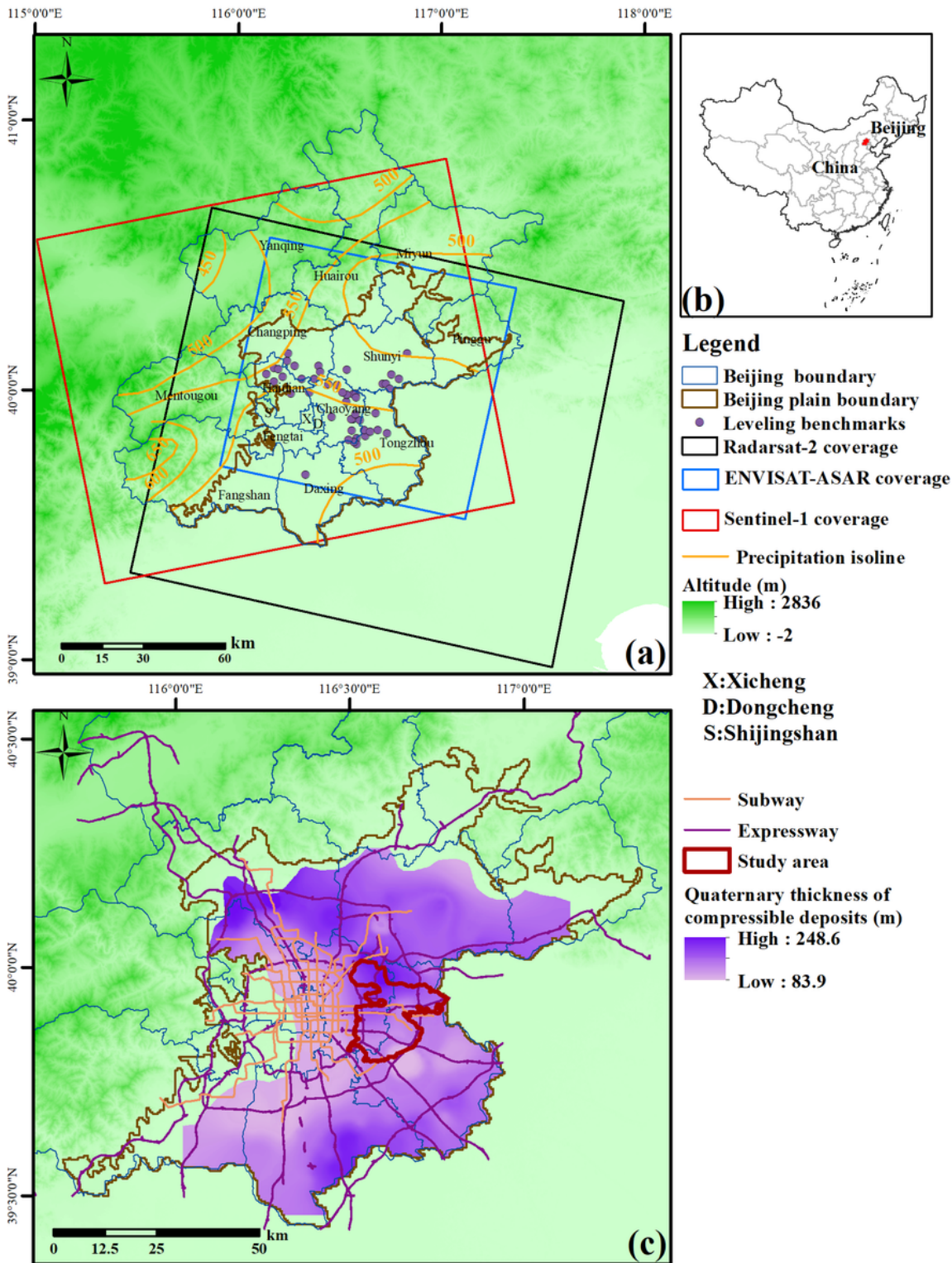


Figure 1

Geographical conditions of the study area. (a): shows the coverage of Radarsat-2, ENVISAT-ASAR and Sentinel-1 and distribution of leveling benchmarks, which were used to compare the deformation results derived by Interferometric Synthetic Aperture Radar (InSAR) technology. (b): shows the location of Beijing. (c): shows the location of study area (the Laiguangying–Dongbalizhuang–Dajiaoting (LDD) land subsidence funnel). (ArcGIS)

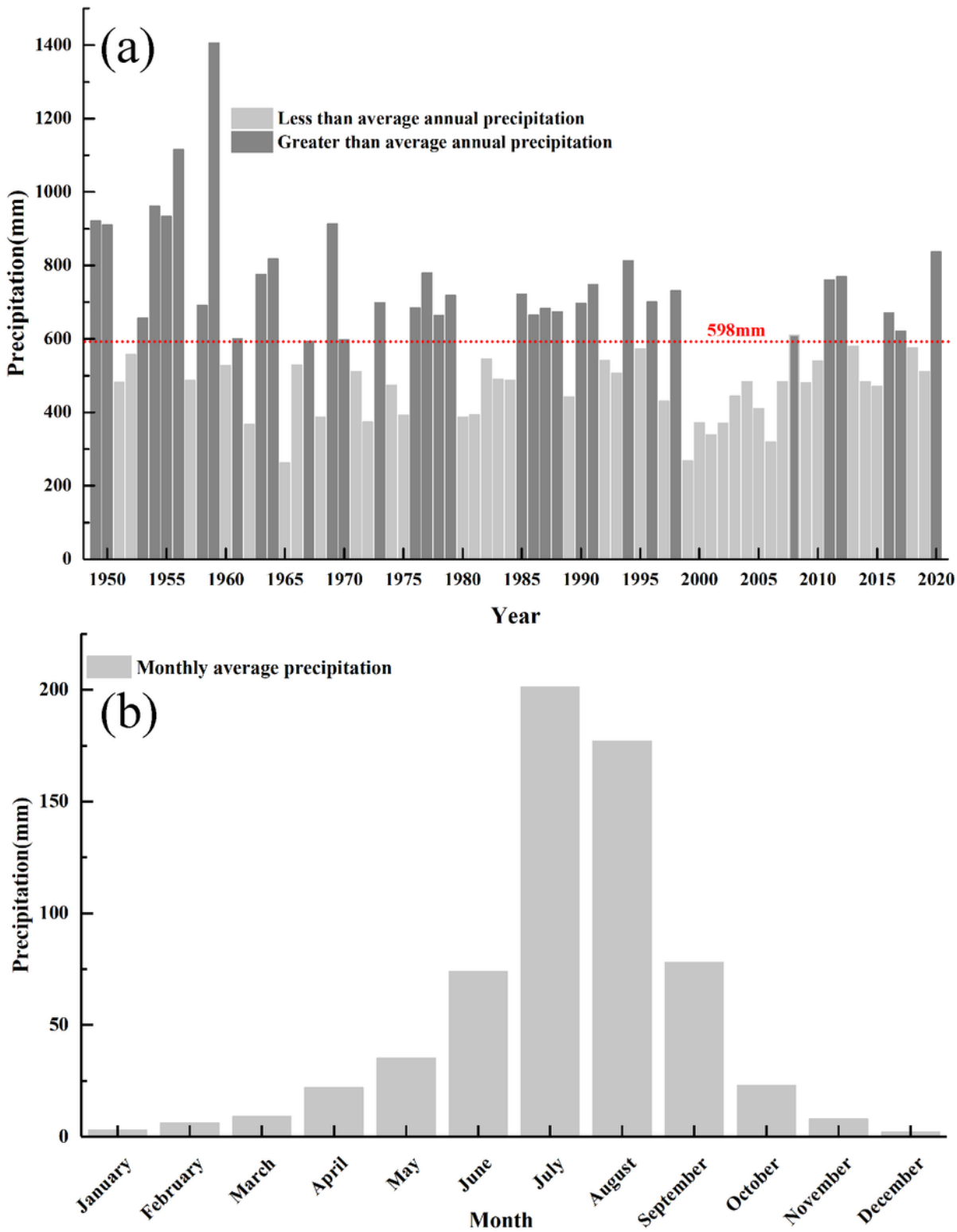


Figure 2

Precipitation in Beijing Plain (BP). (a): shows annual average precipitation in BP from 1949 to 2020. (b): shows average monthly precipitation in BP from 1949 to 2020. (Origin 2016)

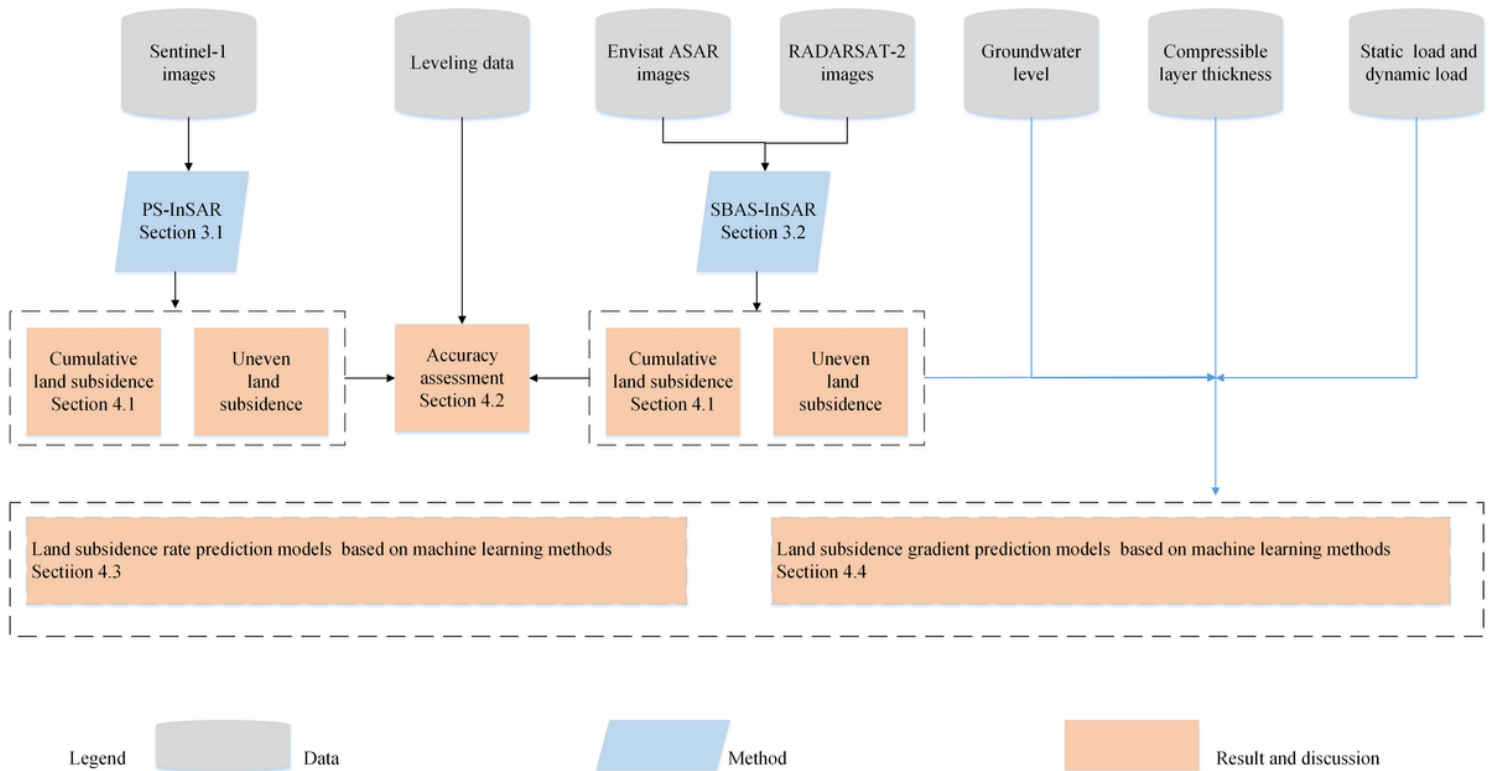


Figure 3

The flowchart of this study. (Visio2013)

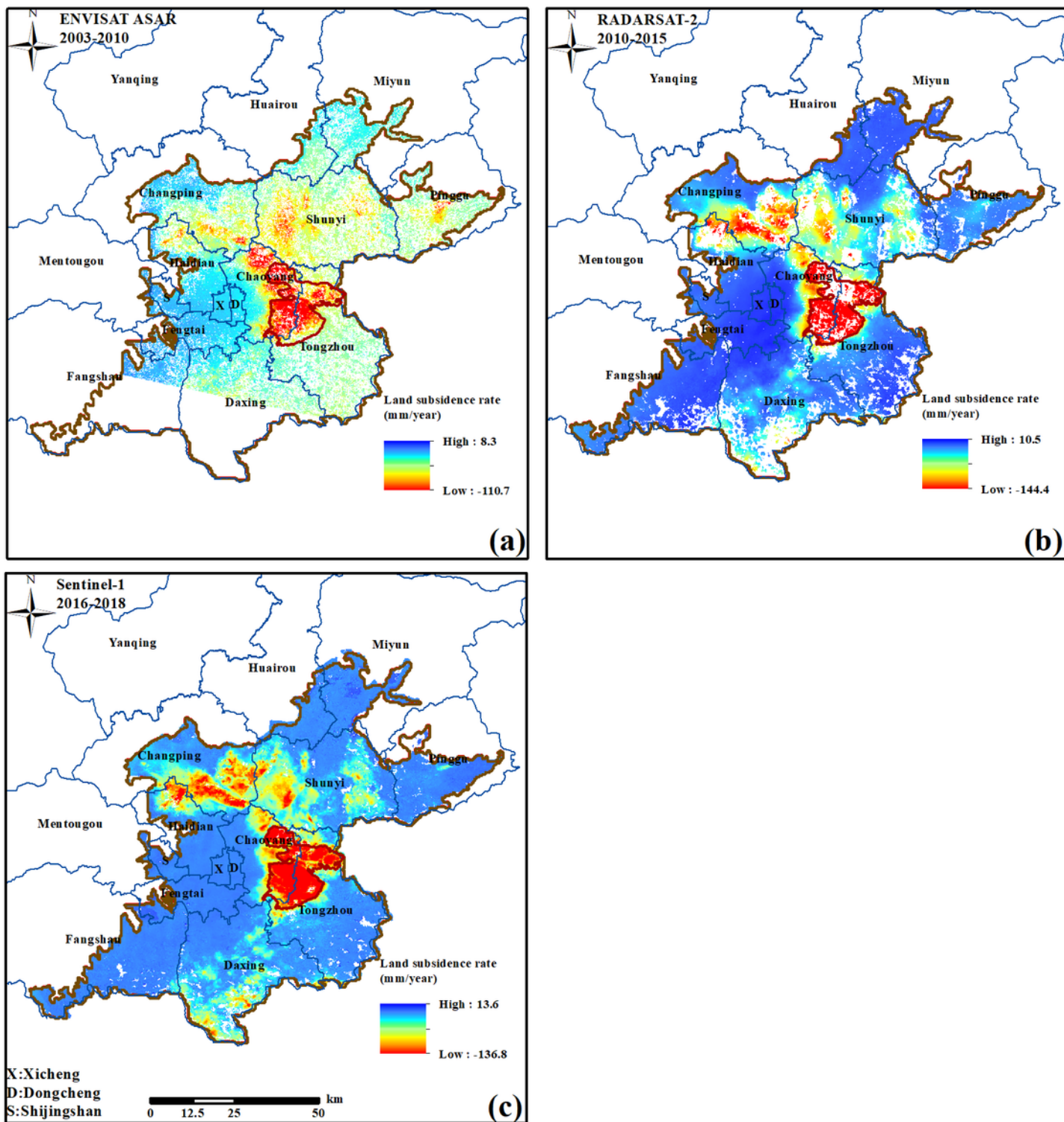


Figure 4

Land subsidence rate maps of BP in three time periods. (a): shows the land subsidence rate map from June 2003 to August 2010. (b): shows the land subsidence rate map from November 2010 to November 2015. (c): shows the land subsidence rate map from January 2016 to September 2018. (ArcGIS)

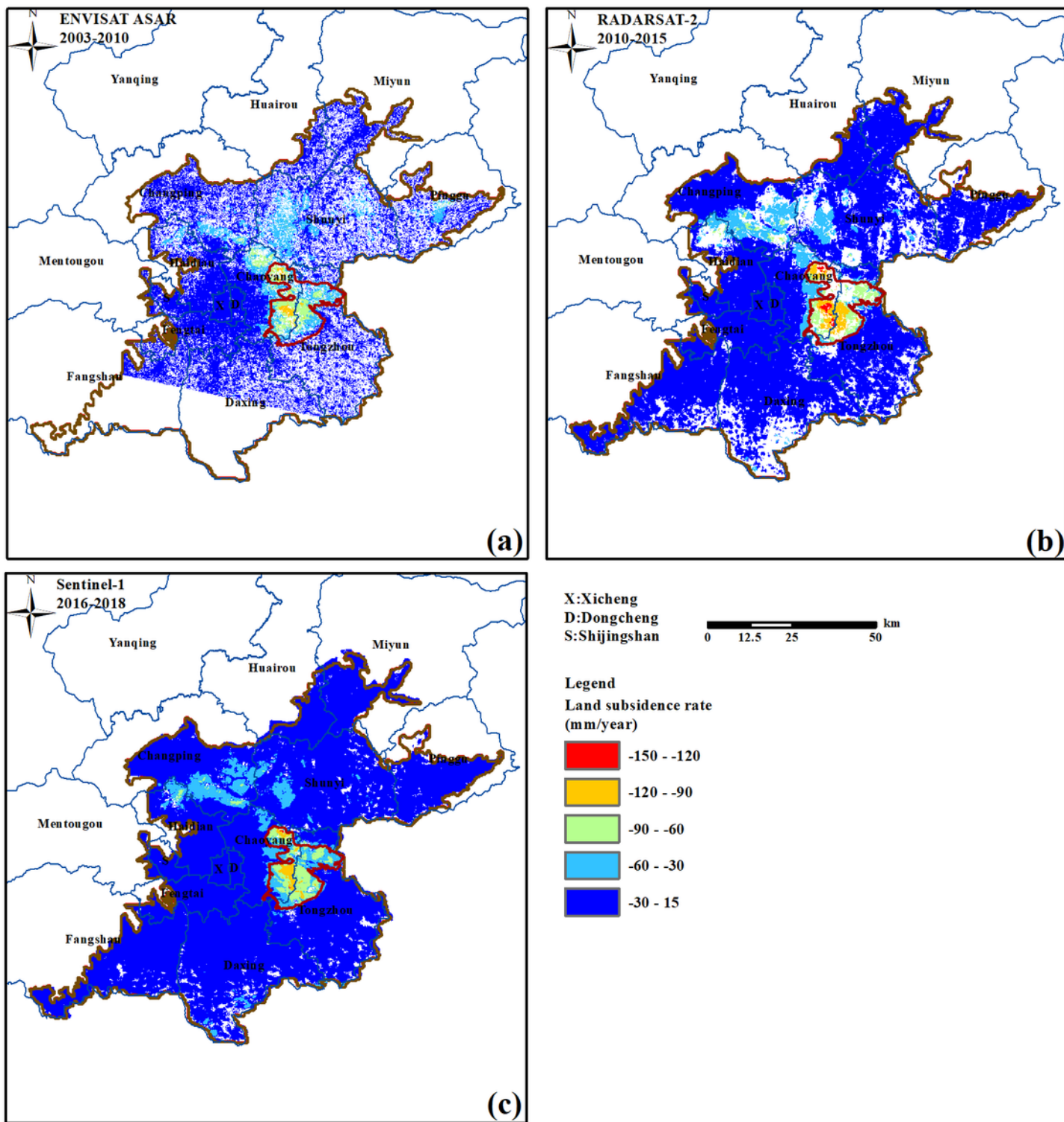


Figure 5

Classification results of subsidence rate maps of BP in time periods. (a): June 2003 to August 2010. (b): November 2010 to November 2015. (c): January 2016 to September 2018. (ArcGIS)

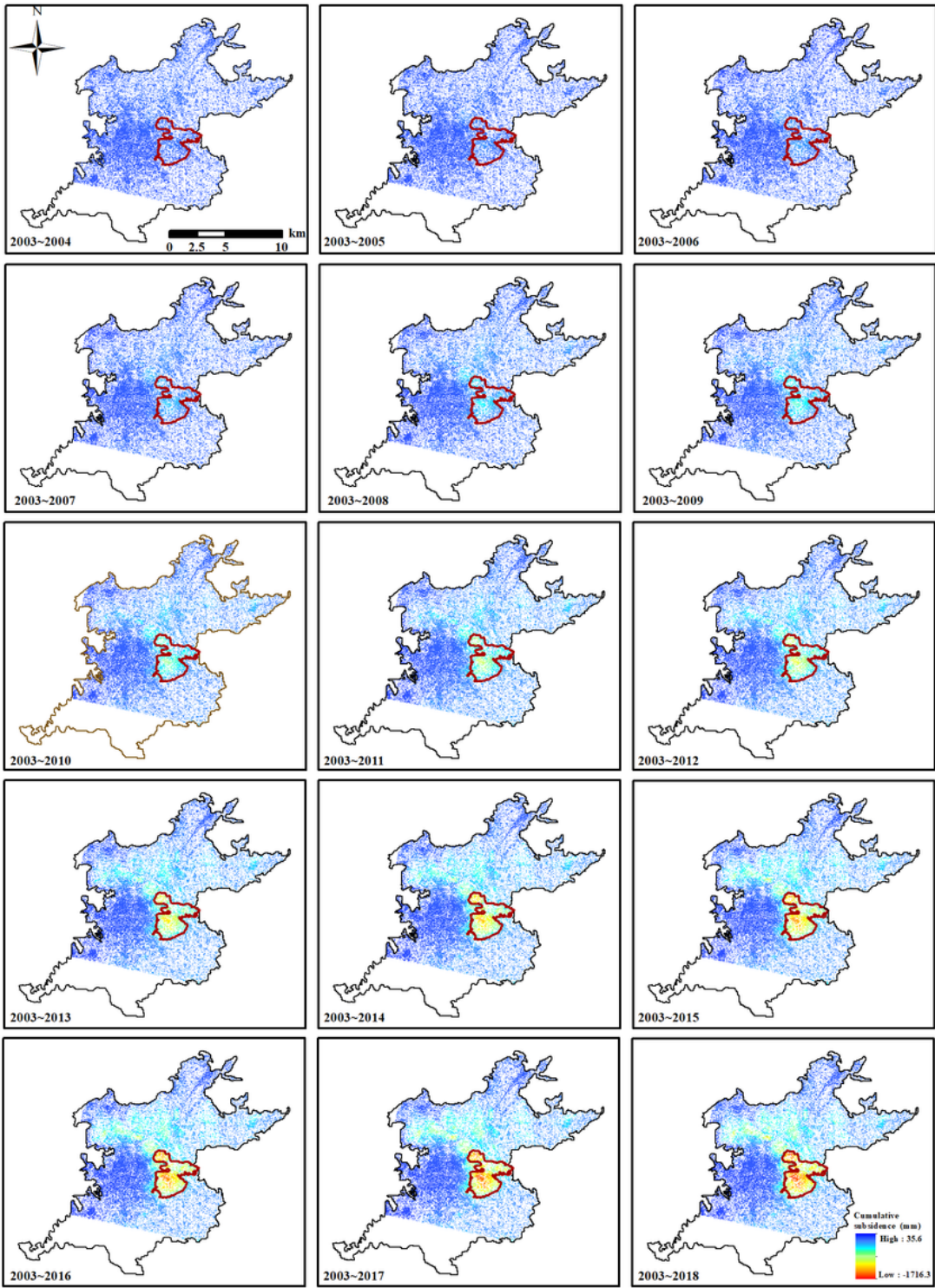


Figure 6

Cumulative land subsidence map of BP from 2003 to 2018 (ArcGIS)

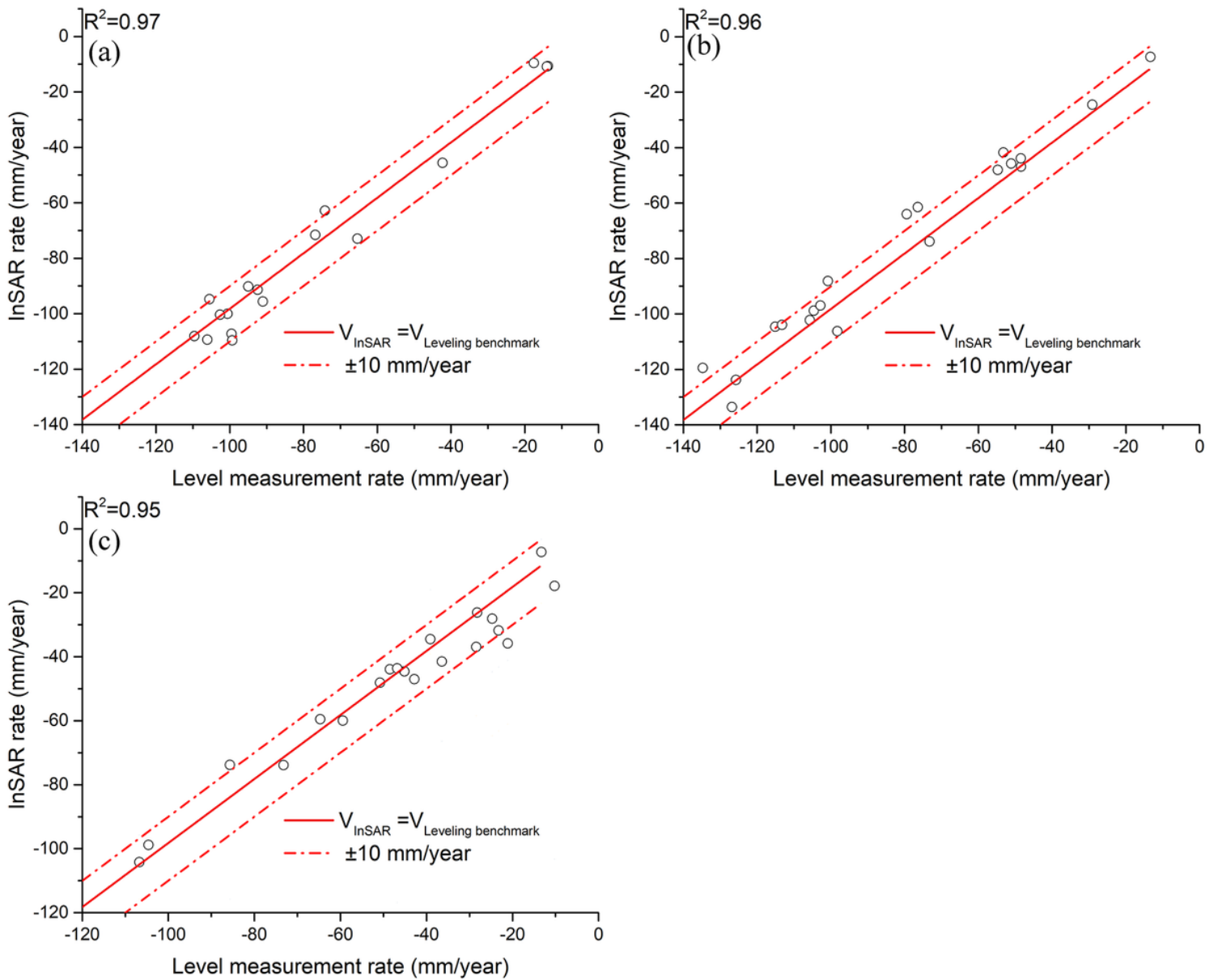


Figure 7

Comparison of the InSAR land subsidence monitoring results and leveling benchmark monitoring results in three time periods. (a): shows the comparison result during the 2003–2010 period. (b): shows the comparison result during the 2011–2015 period. (c): shows the comparison result during the 2016–2018 period. (Origin 2016)

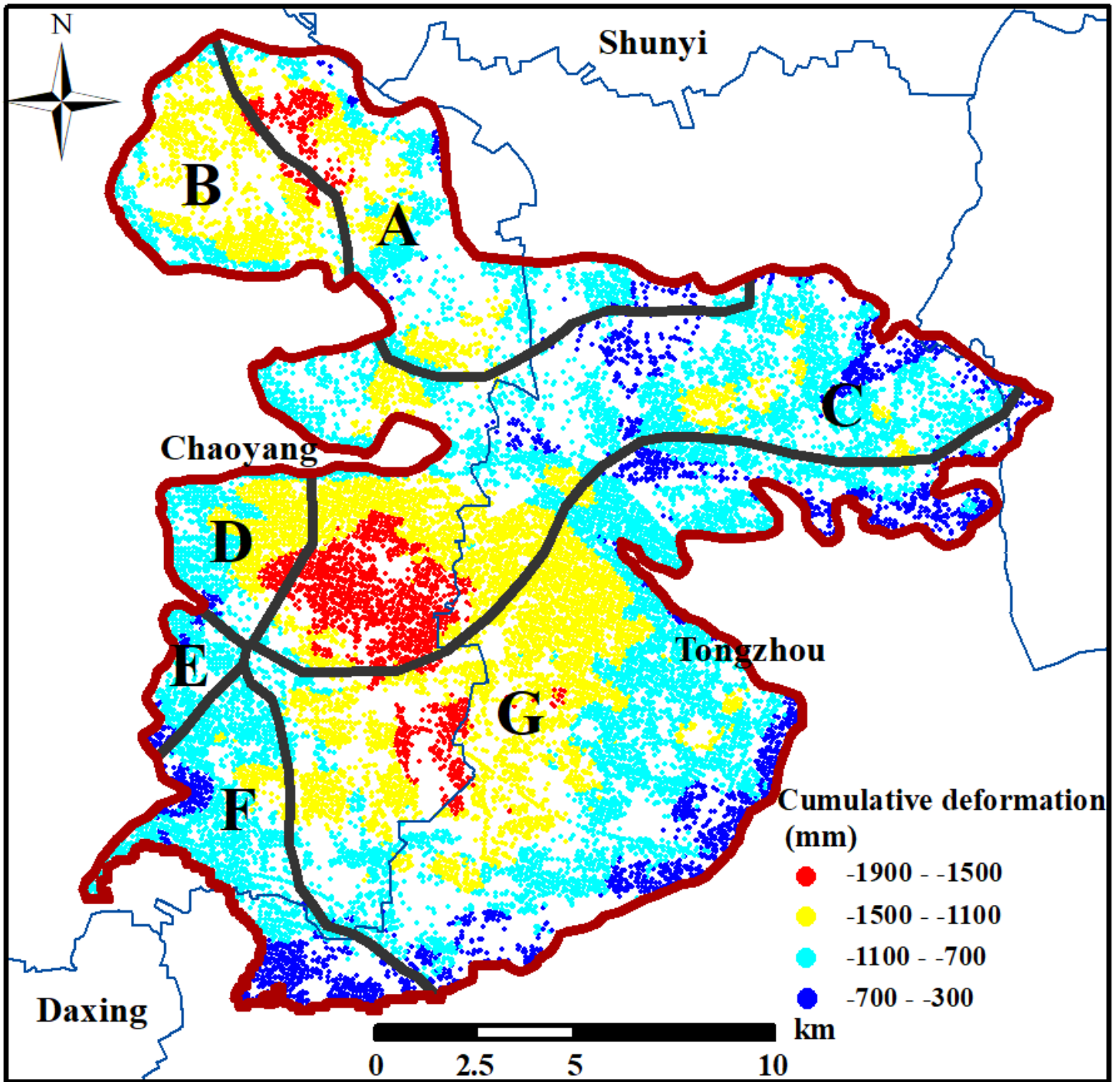


Figure 8

Classification results of the study area. (ArcGIS)

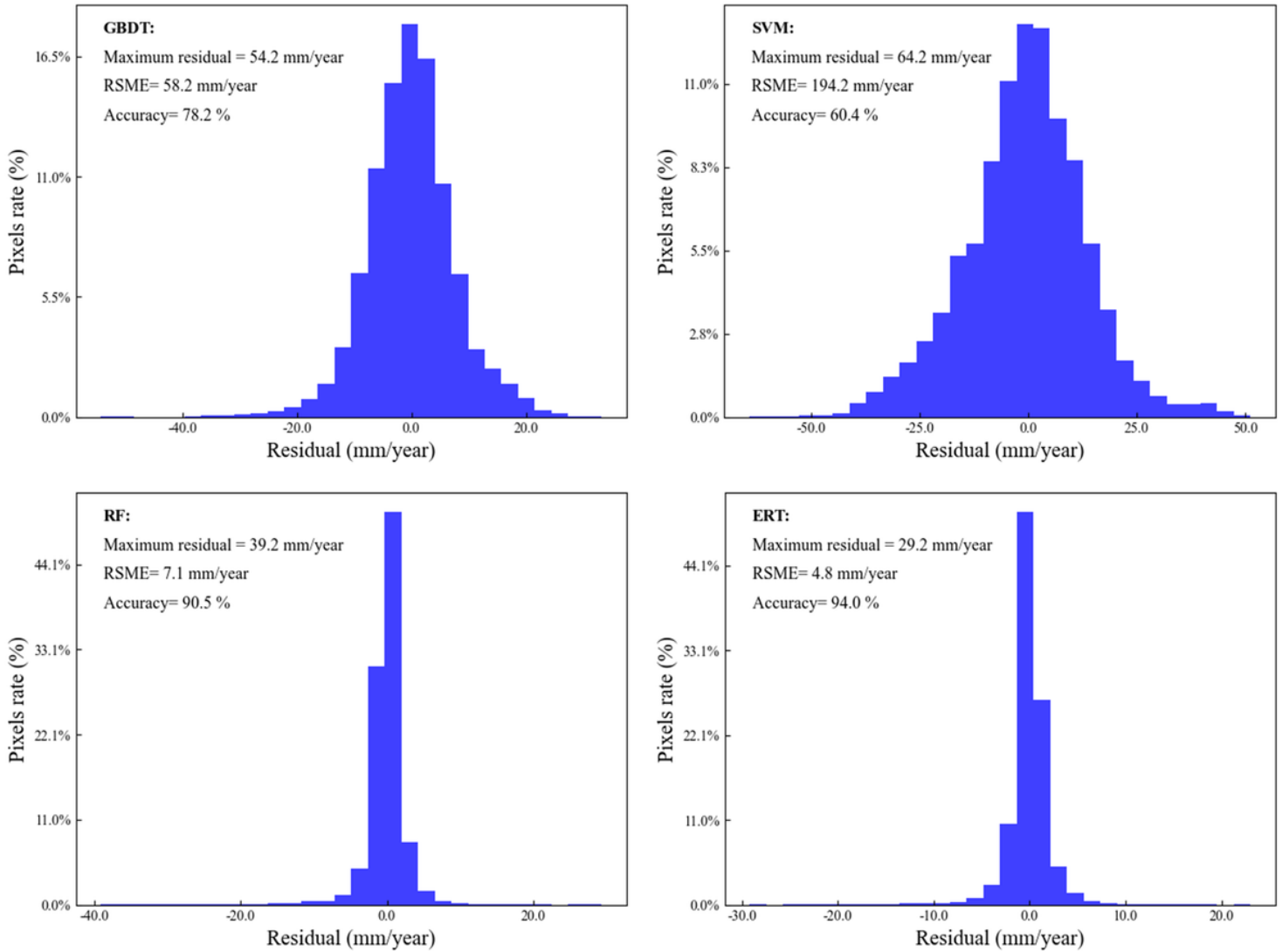


Figure 9

Four land subsidence rate prediction models during the 2011–2015. (Python)

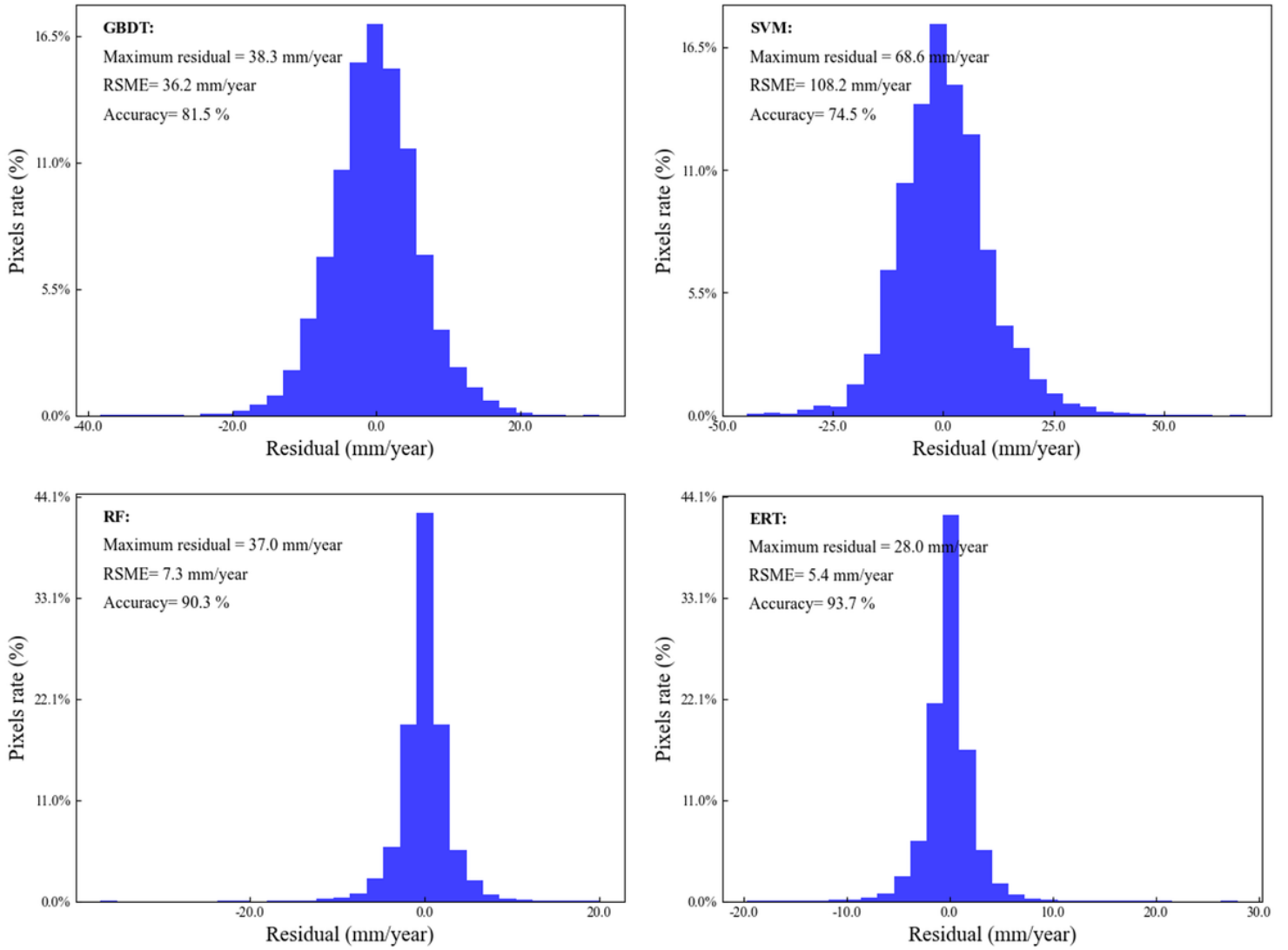


Figure 10

Four land subsidence rate prediction models during the 2016–2018. (Python)

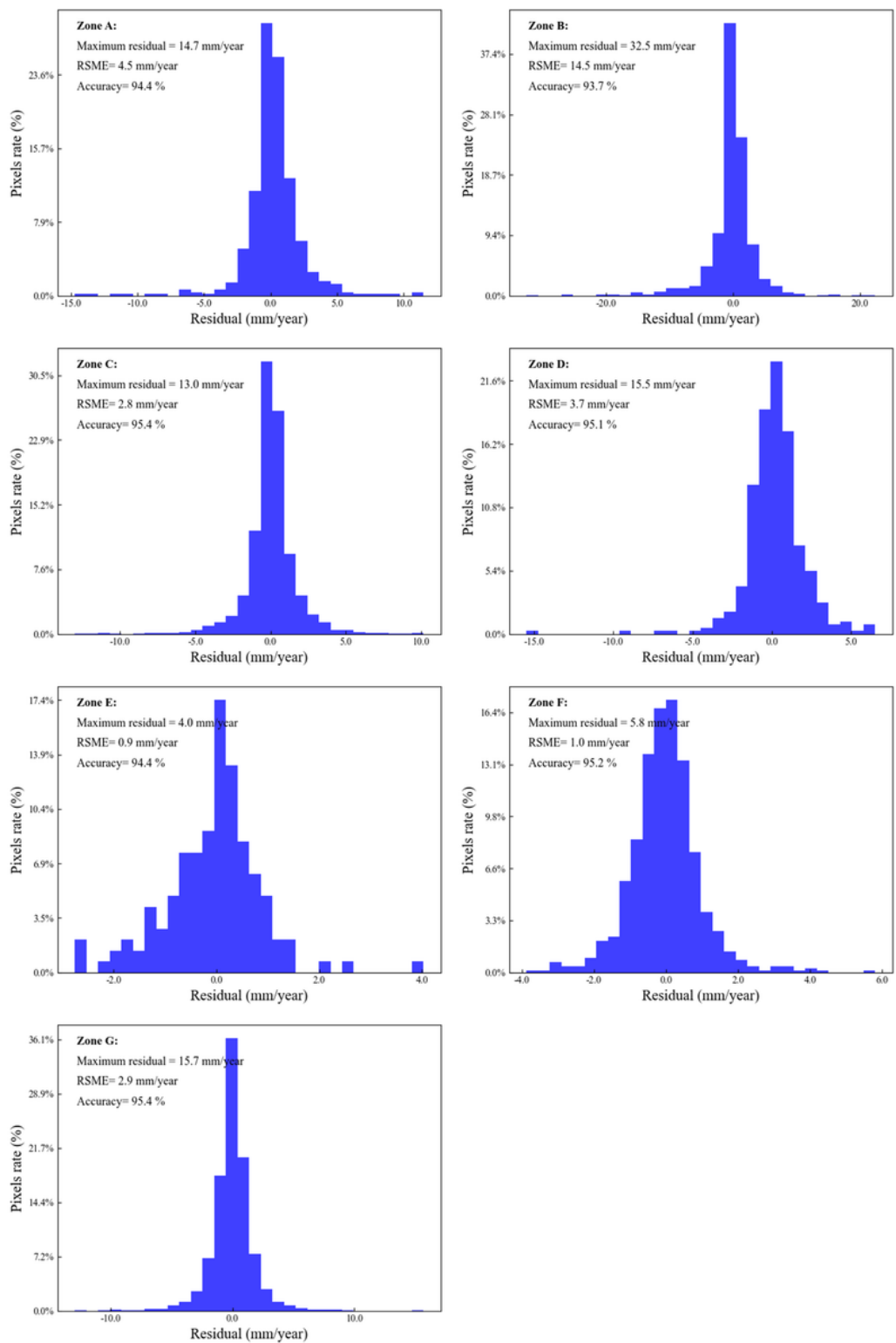


Figure 11

Results of land subsidence rate prediction models based on ERT method in 7 sub study areas during the 2011–2015 period. (Python)

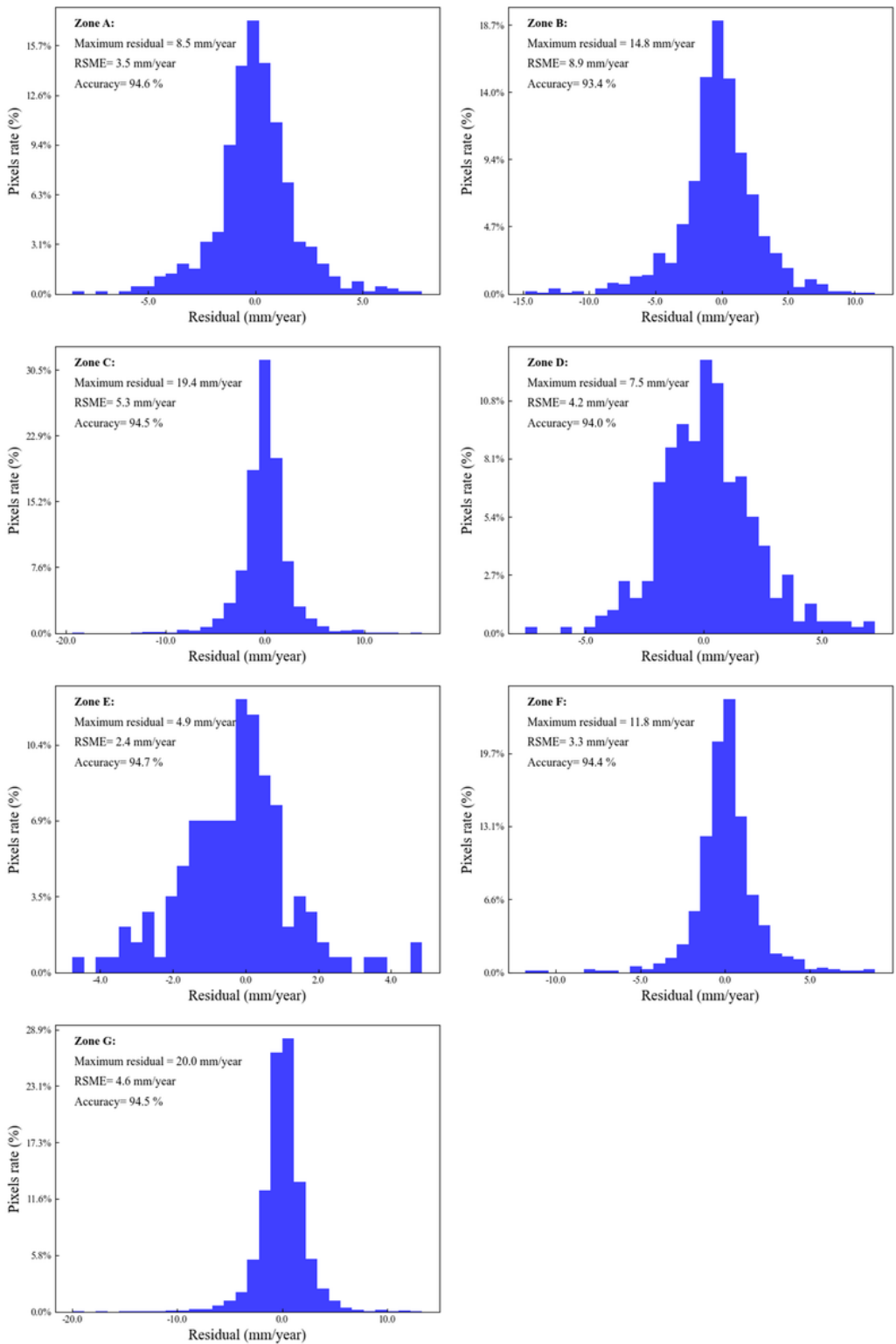


Figure 12

Results of land subsidence rate prediction models based on ERT method in 7 sub study areas during the 2016–2018 period. (Python)

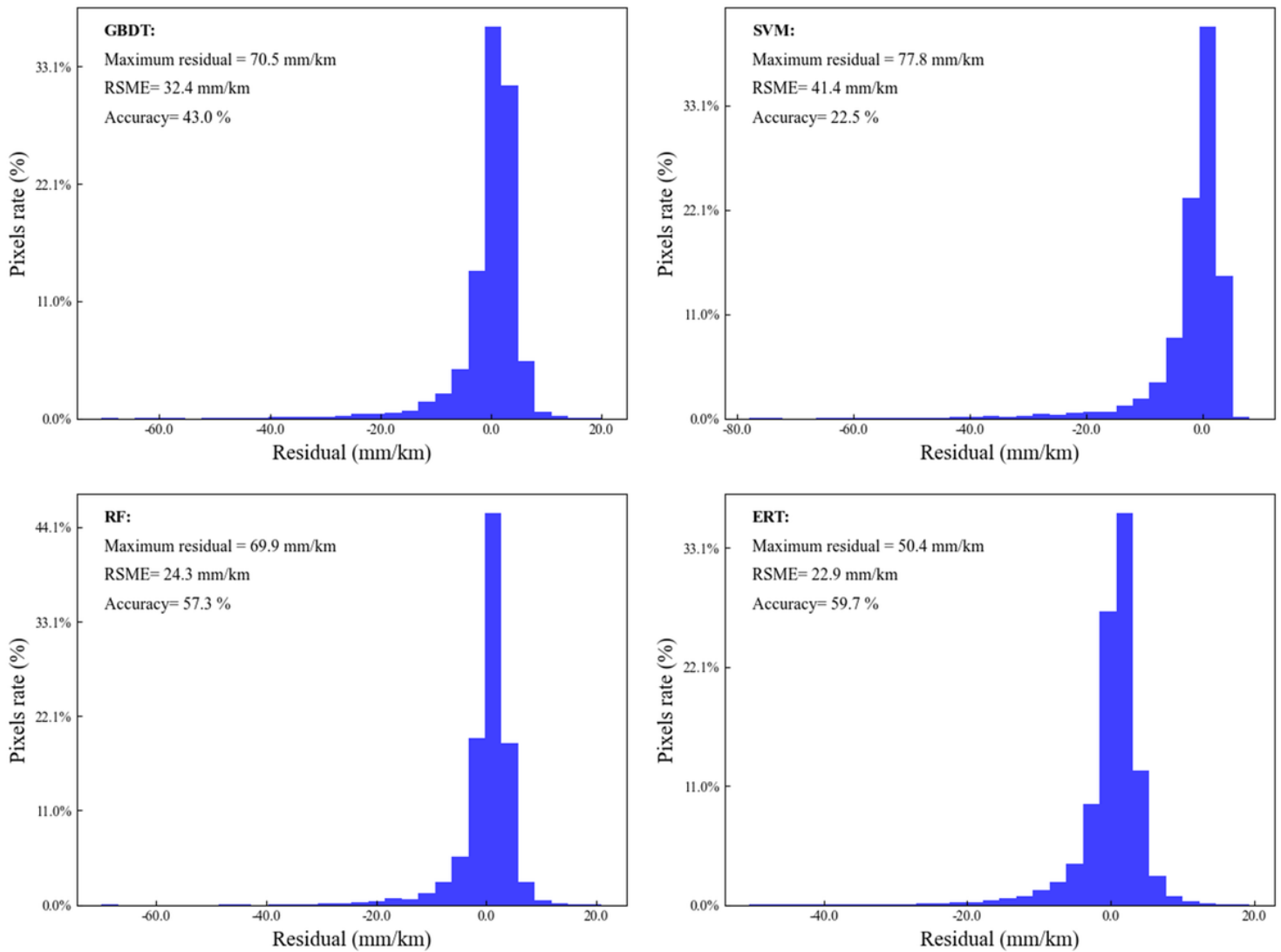


Figure 13

Four land subsidence gradient prediction models during the 2011–2015. (Python)

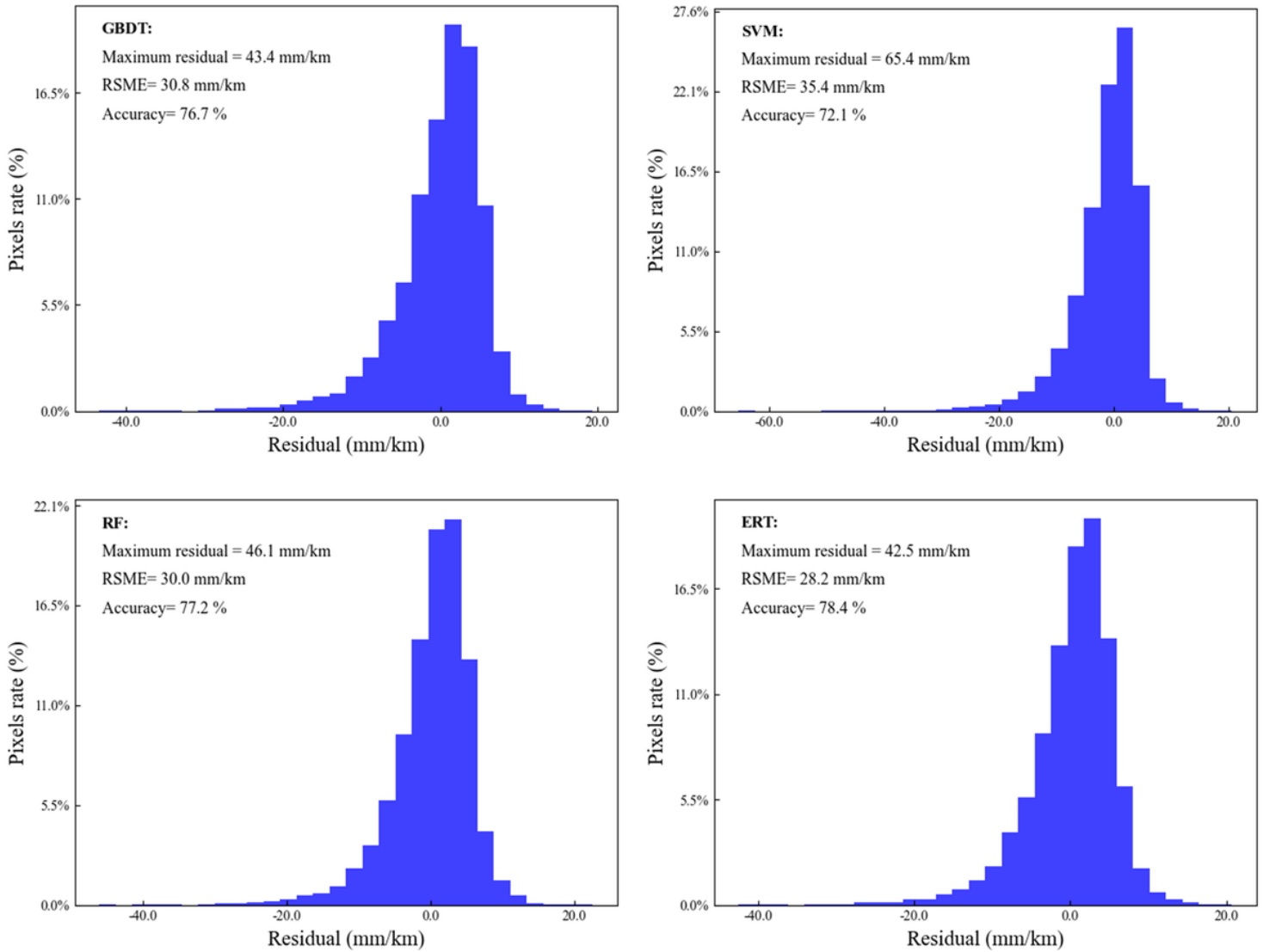


Figure 14

Four land subsidence gradient prediction models during the 2016–2018. (Python)

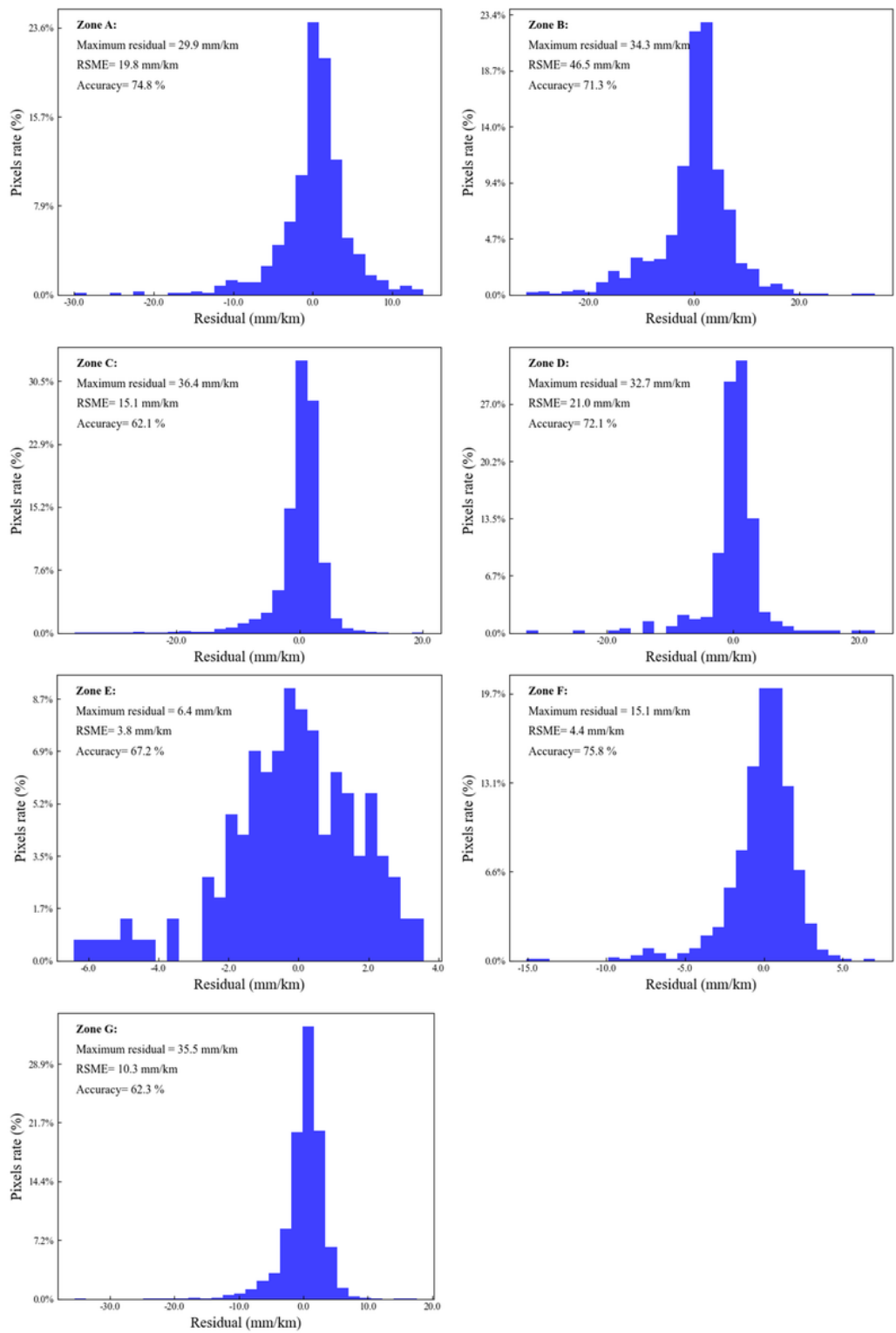


Figure 15

Results of land subsidence gradient prediction models based on ERT method in 7 sub study areas during the 2011–2015 period. (Python)

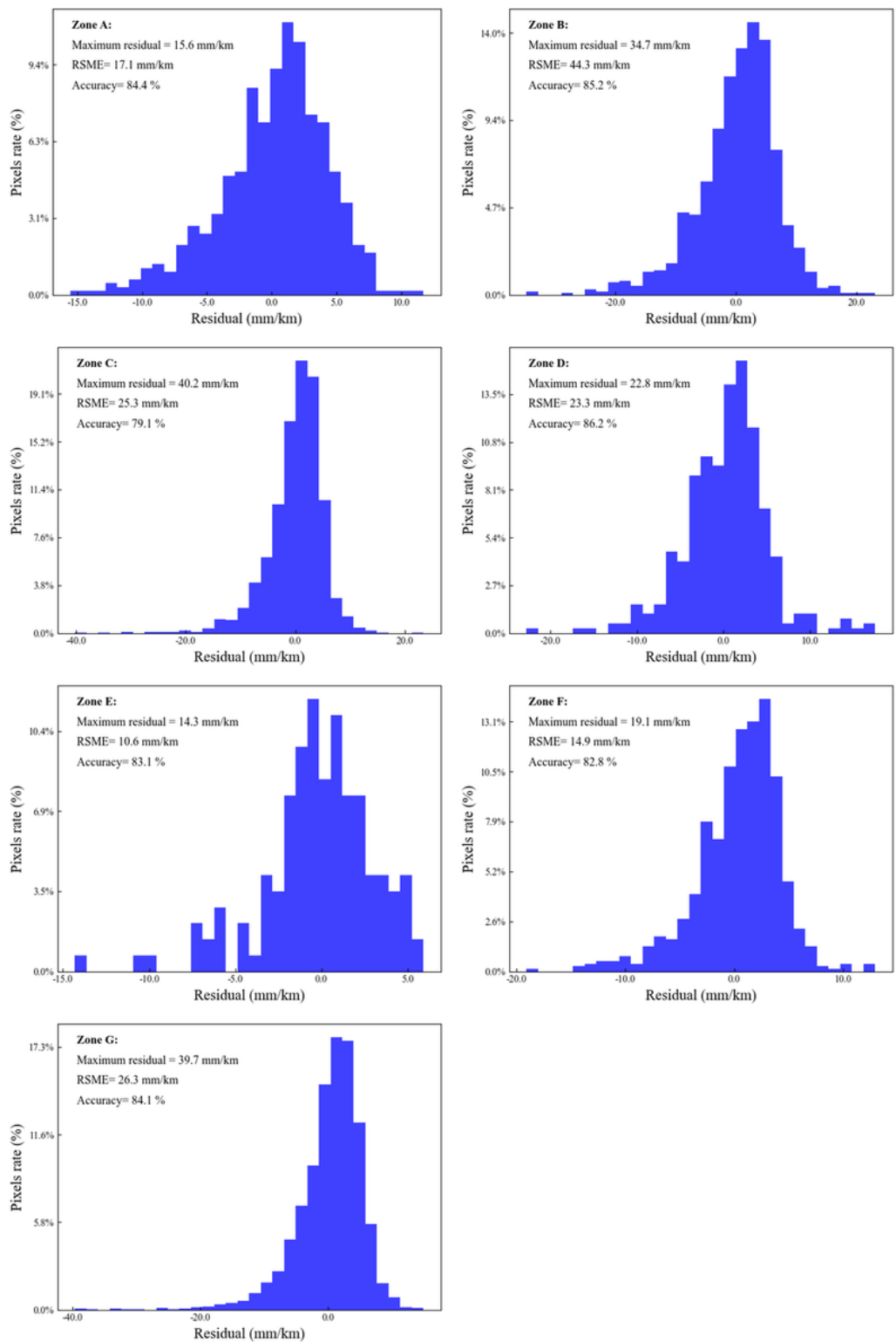


Figure 16

Results of land subsidence gradient prediction models based on ERT method in 7 sub study areas during the 2016–2018 period. (Python)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFig.1.png](#)
- [SupplementaryFig.2.png](#)
- [SupplementaryTables.docx](#)