Use of LSTM for Sinkhole-Related Anomaly Detection and Classification of InSAR Deformation Time Series

Anurag Kulshrestha ⓘ, *Member, IEEE*, Ling Chang ⓘ, and Alfred Stein ⓘ

Abstract—**Sinkholes exhibit precursory deformation patterns. Such deformation patterns can be studied using InSAR time-series analysis over constantly coherent scatterrers (CCS). In the past we identified Heaviside and Breakpoint changes as two important forms of anomalous behavior. It is challenging to efficiently detect and classify these sudden step and sudden velocity changes in deformation time series, especially in the presence of tens of thousands CCS. To address this challenge, we propose to classify these forms of anomalous behavior with a deep learning-based supervised time series classification. In this study, we used a two-layered bidirectional long short term memory (LSTM) classification model for this purpose. The classified deformation classes were analyzed as well in the context of scattering mechanisms. We implemented this model on a sinkhole affected region spanning $\sim 63 \times 44 \, km^2$ in Ireland, using 104 Sentinel-1 A SAR images acquired between 2015 and 2018. Our results show that the CCS with a linear trend can be correctly classified with a maximum accuracy of $\sim 99\%$, whereas for the CCS categorized as anomalous Heaviside and Breakpoint changes the accuracy drops to a maximum of 62%. Multithreshold-based filtering of samples increased the classification accuracy by as much as 50%. We conclude that the method that we propose is effective in detecting anomalous deformation changes. Future research should investigate how it can be applied to other hazard-related detection and classification problems.**

Index Terms—**Anomaly detection, breakpoint, heaviside, long short term memory (LSTM), sinkholes, time-series classification, time-series InSAR.**

I. INTRODUCTION

LAND subsidence is steadily becoming a major issue in urban and rural areas [1]. Sinkholes, which are characterized by sudden depressions over the ground, are causing major damages to life and property [2]. Therefore, it is important to monitor sinkhole prone areas and detect deformation behaviors which resemble sinkhole-related precursory deformations. Sinkhole prone areas can be monitored by observing spatio–temporal deformation patterns of ground surface, which sometimes show signs of anomalous activity, e.g., cavity migration occurring underneath the ground [3]. Such deformation patterns can be estimated using time-series interferometric synthetic aperture radar (TSInSAR) [4]–[6] with millimeter-level precision. TSInSAR is applied to estimate deformation time series over constantly coherent scatterrers (CCS) [7], e.g., persistent scatterers (PS) [4], or distributed scatterers [8], [9]. Some of the applications of TSInSAR include monitoring of urban infrastructure [10]–[12], dam infrastructure [13], [14], geohazards, such as volcanoes [15], [16] and sinkholes [17], [18]. TSInSAR is by now well established method to study spatio–temporal ground deformation patterns.

The mechanics of ground movement can be studied either spatially, temporally, or jointly in space and time. In this study, we are concerned with modeling the temporal dynamics of deformation time series. The deformations are correlated both in time and space. This allows us to study correlation between deformations event in time at the same location, or at locations spatially distant to each other. Therefore, there are common and distinct trends present in the deformation behavior of scatterers located at separate locations. The reasons for the shared behaviors could be related to the shared geological conditions or common land use types on the top surface, leading to common radar scattering mechanisms (SM). For instance, building corners, lamp posts, and other urban land features cause dihedral scattering, where most of the signal, especially in copolarization mode (HH and VV) is scattered back to the SAR sensor. Vegetation is mostly responsible for volume scattering, where radar waves encounter multipath reflections, and also depolarize the wave. Hence, they register high response in the cross polarization mode (VH and HV). Scatterers, such as water bodies and roads cause very low scattering and, therefore, also register very low signal in the both co-pol and cross-pol channels of SAR data. The reason for common deformation behavior could also be mutual interaction with hydrological phenomena, such as groundwater level variation [19].

Sinkholes may show precursory deformation patterns. Identification and classification of these deformation patterns could be a useful way to identify sinkhole-related CCS. Two of the most reported sinkhole-related precursory deformation patterns are sudden variations in instantaneous velocity [20] and instantaneous acceleration [17]. The former refers to sudden movement in the ground, but does not affect the average deformation velocity of the scatterer, and the latter refers to a change in the average velocity of the scatterer after a certain epoch. These two deformation behaviors can be modeled using Heaviside

step functions and multivelocity Breakpoint models [21], respectively. The importance of distinguishing between these deformation behaviors has been emphasised by [22].

Modeling schemes, such as multiple hypothesis testing (MHT) are well established in using regression to model such patterns. However, a major problem is to classify such patterns, such that the models scale well over large deformation datasets [23]. Classification of these patterns can be done using deep-learning methods which have the advantage of transforming time series in a multidimensional feature space where the samples can be separated using a multidimensional plane efficiently. These transformations are done using a combination of nonlinear transformations performed by activation functions, for e.g., sigmoid [24], hyperbolic tangent [25] used in recurrent neural networks (RNN).

RNNs are used to encode and learn sequences as transformed vectors. There are two main types of tasks for which RNNs are useful, first, sequence–sequence mapping, and second sequence–scalar mapping, e.g., for predicting the next term on a sequence, i.e., time series forecasting, or sequence classification. Recently, there has been a surge in deep learning techniques being used for SAR data analysis [26], [27]. RNN, such as long short term memory (LSTM) [28] have been used for forecasting of deformation time series using TSInSAR derived data [29]. Unsupervised classification of InSAR deformation time series using an LSTM autoencoder model has been attempted by [23]. Furthermore, simulated data has been used to train LSTM networks to identify anomalies in InSAR deformation time series over Italy using Sentinel-1 data [22]. However, supervised classification of deformation time series data using LSTM models and real deformation data has not been studied. Furthermore, the relation between the deformation classes and the SM of the associated CCS has also not been investigated in the context of deformation time series classification. Therefore, the main objective of this study is to classify sinkhole-related deformation time series using supervised LSTM sequence classification and real InSAR derived deformation data. The classified deformation classes are also related to SM classes.

The rest of this article is arranged as follows. Section II introduces the methods for polarimetric decomposition, training data extraction, and LSTM modeling. Section III presents the details of the study area and Sentinel-1 data used. Section IV presents the results, followed by a discussion in Section V. Finally, Section VI concludes this article.

II. METHODS

We consider a multitemporal InSAR (MTInSAR) that is carried out for a stack of SAR images. To detail our methods, we define a set of CCS points $P = P(x_i, y_i, z_i)$, for $i \in [1, b]$, located in a 3-D coordinate system with coordinates x_i, y_i, and z_i, and temporal epochs t_j for $j \in [1, m]$, also referred to as InSAR epochs. The corresponding deformation time series is defined as d_t^P, $(t = \forall t_j)$. Here b and m are the total number of CCS in P and the total number of InSAR epochs, respectively.

Each CCS point in the CCS set P is associated with a SM class that belongs to the SM set denoted by $\mathcal{S}$. The SM classification
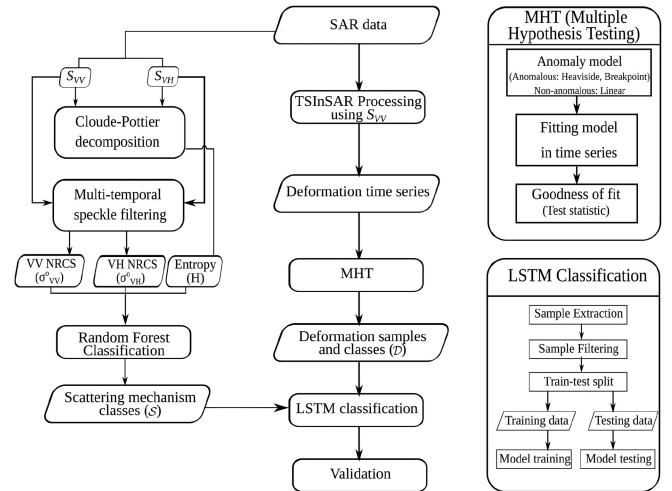

Fig. 1. Methodological flow diagram.

method is presented in Section II-A. Points in P are associated to either nonanomalous (linear) or anomalous deformation classes which are represented by $\mathcal{D}$. The deformation sampling scheme, interclass sample similarity reduction method, and label set preparation method is described in Section II-B. The main LSTM classification model is described in Section II-C. Finally, Section II-D elaborates the choice of hyperparameters, loss function, training-testing strategy and the method for accuracy assessment. These methods are shown in the methodological flow diagram in Fig. 1 as well.

A. Scattering Mechanism Classification

The geolocation of CCS, particularly PS often coincides with double-bounce scattering regions, but sometimes CCS points are also identified over volume and surface scattering regions [30] over which the deformation estimates may be less precise. Therefore, it is important to separate deformations on the basis of SMs. The dominant SM in SAR images can be identified using polarimetric decomposition of SAR datasets [31], [32].

Polarimetric decomposition models multipolarimetric SAR data with respect to various polarization bases, which represent different scattering behaviors [31]. In our study, Cloude–Pottier polarimetric decomposition [33] is recommended to extract polarimetric features because it is effective in decomposing multipolarimetric data into multiple SMs using the division of the alpha angle-entropy plane into as many as nine zones [33, Fig. 4]. In addition to the polarimetric features from the Cloude–Pottier decomposition, the normalized radar cross sections (NRCS) in VV and VH channels are used for SM classification. We use sigma-naughts, i.e., σ_{VV}° and σ_{VH}° as measures of NRCS in VV and VH channels, respectively. Furthermore, in order to reduce the speckle noise in the NRCS features, the SAR images are passed through a multitemporal speckle filter [34]. This filter method is defined as

$$J_k(x, y) = \frac{E[I_k]}{m_1} \sum_{i=1}^{m_1} \frac{I_i(x, y)}{E[I_i]} \quad (1)$$

where $J_k(x, y)$ is the filtered output of SAR image pixel at position (x, y), using image nr. k ($k \in [1, m_1]$), where m_1 is the total number of SAR images. $I_i(x, y)$ represents intensity at (x, y) of ith image, $i \in [1, m_1]$. $E[I_k]$ and $E[I_i]$ represent the local spatial mean of pixel values within a window centered at (x, y).

The filtered SAR intensity images are then converted to sigma-naught images and classified using supervised random forest algorithm [35], [36]. The training datasets for the four classes, i.e., double bounce, high volume, low volume, and surface scatterers are crafted manually. They are referred to as the SM classes $\mathcal{S}$.

B. Deformation Classification and Sample Set Preparation

The deformation time series (d_t^P) can be estimated using single-master TSInSAR processing. Thereafter, standard thresholds to, e.g., the estimated deformation velocities, ensemble coherence, estimated height, and spatio–temporal consistency are applied.

Sample Extraction: The deformation time series is then used to create a sample set ($\mathbf{X}$), which is mapped to target class vector $\mathbf{Y}$ using a bijective mapping. Typical (linear) and anomalous deformation time series are recognized from real data using the overall model test for the linear function. The detection–identification–adaptation procedure [37] is used to calculate the test statistic for Heaviside and Breakpoint functions for each point. Details about test statistic are shown in the Appendix. The functions for Heaviside (M_H) and Breakpoint anomalies (M_{Br}) are separately defined as cf. [21]

$$M_H(\Delta_j) = \Delta_j \mathcal{H}(t - \tau_j), \quad j \in [1, m - 1] \quad (2)$$

$$M_{Br}^k(v_k) = t_j \cdot v_k, \quad j \in [1, \ m - 2]. \quad (3)$$

Here, $\mathcal{H}(t - \tau_j)$, is the Heaviside step function, centered at τ_j, Δ_j is the Heaviside step size, v_k is the deformation velocity before and after the Breakpoint epoch, and k is the total number of breakpoints. For instance, the functions can be defined for fitting the anomaly at only one epoch, i.e., single Heaviside and single Breakpoint function. The model is attempted to fit at all epochs (m) except the extremes, i.e., $m - 1$ for Heaviside and $m - 2$ for Breakpoint function. For a single Breakpoint model, the deformation time series is divided into two parts with two different linear velocity parameters, i.e., v_1 and v_2, at the breakpoint epoch l. The corresponding single Breakpoint model can be written as $M_{Br}^1 = [t_1 \ldots t_l]^T \cdot v_1 + [t_{l+1} \ldots t_m]^T \cdot v_2$. The change in the two velocities is estimated by using $\theta_{t_l} = \arctan((v_1 - v_2)/(1 + v_1 \cdot v_2))$, where θ_{t_l} is the angle between the lines drawn by v_1 and v_2 in the deformation-time space. The test statistic for goodness of fit of functions M_H and M_{Br} at each epoch (t_j) is calculated for fitting these functions for all CCS using (5) in [21].

Filtering of Samples: In order to ensure low interclass similarity (especially between anomalous classes), multiple thresholds can be used to filter the samples further. The first threshold can be applied on the test statistic values for Heaviside function and Breakpoint function. This is done by applying an upper bound on the test statistic value for the class of interest and a lower bound on the test statistic value for all other classes.

For anomalous classes, i.e., Heaviside ($\mathcal{D}_H$) and Breakpoint ($\mathcal{D}_{Br}$), the top nth percentile (p_n) values are chosen, i.e., $X^C = \{d_t^P \,|\, T_P^C \geq p_n(T^C)\}$. Here, X^C are the selected training samples associated to the class of interest C, $C \in \{\mathcal{D}_H, \mathcal{D}_{Br}\}$, T^C is the test statistic for class C for all points, T_P^C is the test statistic for every individual point in the set P, and $p_n(.)$ signifies the application of percentile threshold. This thresholding scheme is chosen because high test statistic value indicates that the tested model strongly rejects the default model, i.e., linear model. Second, a lower bound on the Test statistic values for classes other than the class of interest is applied. This is done using a low percentile threshold p_{n_1} at n_1th percentile. Therefore, $X^C = \{d_t^P \,|\, T_P^{C\sim} \leq p_{n_1}(T^{C\sim})\}$. Note that the tilde $\sim$ is the complement operator. Third, further purification of the samples is done by applying class specific thresholds on the extent of deformation velocity changes (θ) and jumps (Δ) for Breakpoint and Heaviside anomalies, respectively. The samples for class C, $X^C = \{d_t^P \,|\, \mod(\Delta) \in [\Delta_{\min}^C, \Delta_{\max}^C]\}$ and $X^C = \{d_t^P \,|\, \mod(\theta) \in [\theta_{\min}^C, \theta_{\max}^C]\}$. Here, $[\Delta_{\min}^C, \Delta_{\max}^C]$ and $[\theta_{\min}^C, \theta_{\max}^C]$ are the threshold values for both anomalous classes on the jump and deformation velocity change, respectively. For Heaviside class, the threshold for jumps are set at a higher value and that for velocity changes are set at lower values, and vice versa for the Breakpoint class. The $\mod(.)$ function signifies that the thresholds are applied on the absolute values of jumps and velocity change angles.

Application of these thresholds is necessary for reducing the interclass similarity between the deformation classes. Since the deformation estimates are marred by noise (due to, e.g., atmospheric inhomogeneity and incorrect satellite orbit estimation), it becomes difficult to identify unique deformation phenomena for the deformation time series associated to the CCS locations. Our filtering approach increases the chances for the deep learning models to identify unique and distinguishable patterns in the transformed feature space where the samples can be classified accurately.

After the filtering of samples, the samples are accumulated in a training sample set $\mathbf{X} = X^{C_k}; \forall k, k \in [1, N_C]$, where N_C is the total number of classes, and C_k is the kth deformation class.

Label Set Preparation: The classes are defined in two ways, using 1) only deformation classes ($\mathcal{D}$), and 2) deformation classes crossed with SM classes ($\mathcal{S}$), i.e., $C \in \mathcal{S} \otimes \mathcal{D}$. The three deformation classes are linear ($\mathcal{D}_{Lin}$), Heaviside ($\mathcal{D}_H$), and Breakpoint ($\mathcal{D}_{Br}$). Therefore, we take a Kronecker product between the set of SM classes and anomaly classes. The class labels for class C are represented by the symbol Y^C. The labels, $\mathbf{Y} = Y^{C_k}; \forall k, k \in [1, N_C]$, are encoded using the one-hot encoding method. The means that the samples have been assigned hard classification labels.

C. LSTM Model Structure

In traditional RNNs, gates are used to control the influence of new information on the learnt state of the model. A figure of an LSTM unit cell can be seen in Fig. 2. c_t is the cell state, where the information from all time epochs are encoded, and serves as the long term memory for the model, h_t is the hidden state, which is the state made available to the memory. c_{t-1} and h_{t-1}
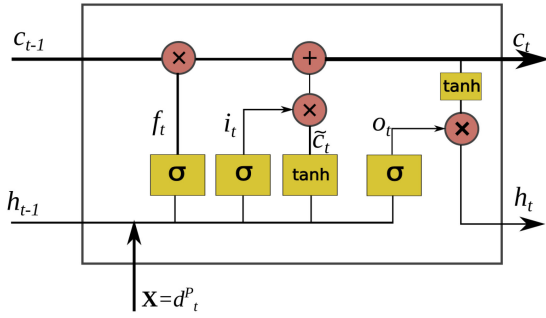
Fig. 2.　LSTM unit cell diagram.

Fig. 3.　Model design.

represent the cell state and hidden states from the previous recurrent unit. i_t is the input gate, which controls the extent to which new data values are allowed to change the cell state, f_t is the forget gate introduced by [28], which controls how much of the previous state is transferred onto the next state, o_t is the output gate which controls the part of the learnt state returned by the model, and $\tilde{c}_t$ is the prospective new cell state. The subscript t refers to the temporal epoch corresponding to the input data d_t^P. The symbols $\otimes$ and $\oplus$ represent pointwise multiplication and summation, respectively. The yellow boxes represent the application of sigmoid and hyperbolic tangent functions.

Like in ANNs, the weights associated with the gates are learnt using back propagation through time (BPTT) [28]. There are three sets of weights W, U, and B associated to samples ($\mathbf{X}$), hidden state (h_t), and biases, respectively. The three sets of weights belong to a set of real numbers with dimensions as follows:

$$W \in \mathcal{R}^{H_k \times F}; U \in \mathcal{R}^{H_k \times H_k}; B \in \mathcal{R}^{H_k \times 1} \qquad (4)$$

where H_k refers to the number of neurons in the kth LSTM layer, and F refers to the number of features in the input dataset, which is equal to 1, for the deformation time series, in our case. Each of the three sets of weights are defined for four gates, i.e., input: i, forget: f, prospective state: $\tilde{c}$, and output: o, and $(W_i, W_f, W_c, W_o) \in W$, $(U_i, U_f, U_c, U_o) \in U$, and $(B_i, B_f, B_c, B_o) \in B$.

The samples are defined using the set $\mathbf{X}$, and the associated target labels are represented by $\mathbf{Y}$. In RNNs, the learnt state is calculated as

$$h_t = \tanh(\mathbf{X}_t^T \cdot W + h_t^T \cdot U + B_1) \qquad (5)$$

where T refers to the transpose operator. However, the signal mentioned above is controlled through the gates in LSTMs and the values of the weights associated to the gates are calculated using the following equations:

$$
\begin{aligned}
i_t &= \sigma(U_i \cdot h_{t-1} + W_i \cdot X_t + B_i) \\
f_t &= \sigma(U_f \cdot h_{t-1} + W_f \cdot X_t + B_f) \\
\tilde{c}_t &= \tanh(U_c \cdot h_{t-1} + W_c \cdot X_t + B_c) \\
o_t &= \sigma(U_o \cdot h_{t-1} + W_o \cdot X_t + B_o) \\
c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \\
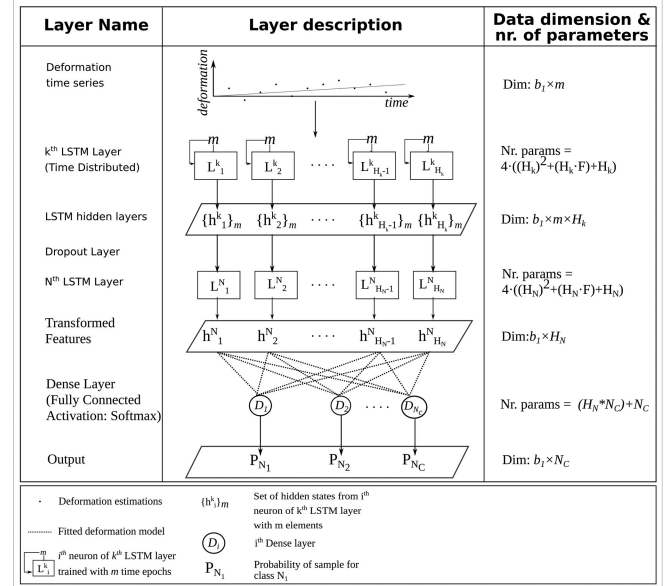h_t &= o_t \cdot \tanh(c_t).
\end{aligned}
\qquad (6)
$$

The sigmoid and hyperbolic tangent functions act as gates and control the flow of information to the cell state c_t due to the variation of its output in the range $[0, 1]$ and $[-1, 1]$, respectively. A detailed description about the working of the LSTM model is given in [38].

LSTM in Our Context: The model used for this work is described in Fig. 3. We use two-layered LSTM model separated by a Dropout layer. The deformation time series samples of dimensions $b_1 \times m$ (b_1 CCS points and m InSAR epochs) is given as an input to the first LSTM layer. The LSTM layer and its individual neurons are represented by L^k, and L_i^k, respectively, where $k \in [1, N]$ and $i \in [1, H_k]$, N is the total number of LSTM layers and H_k is the total number of neurons in the kth layer. The total number of weights for the kth LSTM layer is equal to $4 \cdot ((H_k)^2 + H_k \cdot F + H_k)$. The output of the kth time-distributed LSTM layer of ith neuron is a set of hidden layer values $\{h_i^k\}_m$, where m refers to the size of the set. This output from the kth LSTM layer is given as inputs to the $(k+1)$th layer. This continues until the Nth LSTM layer, and the final output is represented by H_i^N $i \in [1, H_N]$. The dimension of the final output is $b_1 \times H_N$. This shows that input of the model, i.e., samples representing the deformation time series having dimension $b_1 \times m$ have been transformed to a dimension of $b_1 \times H_N$. The outputs of the LSTM layers are then connected to a Dense layer with $(H_N \cdot N_C) + N_C$ parameters, which returns the probability of each of b_1 samples for N_C classes.

The first LSTM layer is chosen to be a bidirectional layer, which is often used to classify sequences with a better classification accuracy [39]. Bidirectional LSTMs use two LSTMs to train one input sequence. It trains the original and a reversed copy of the sample for training the two LSTMs in the same layer. This provides additional context in the model from the same time series. This also results in doubling of the total number of parameters.

D. Model Training and Evaluation

The LSTM neurons (H_k), LSTM iterations, and batch sizes (B_s) are the three hyperparameters which are varied to find the maximum classification accuracy. The LSTM iterations and batch sizes are defined in such a way that their product is always kept equal to the total training samples. The value of H_k is generally chosen to be in the same order as the time length of the samples. The batch size defines the size of chunks of sample which are used to train the model in each LSTM iteration.

Loss Function: Categorical cross entropy is chosen as the loss function for this study. It is defined as follows [40]:

$$ CE = \frac{-1}{b_1 \cdot N_C} \sum_{i=1}^{b_1} \sum_{c=1}^{N_C} y_i^c \cdot \log(\hat{y}_i^c). \tag{7} $$

CE represents categorical cross entropy, b_1; c and N_C are the total number of samples, class index, and total number of classes, respectively; y are the true class labels in one hot-encoding scheme and belong to the label set $\mathbf{Y}$ and $\hat{y}$ are the predicted class scores for all classes. The optimization of the loss can be done using the Adam optimizer [41], which is a commonly used stochastic gradient optimization method.

Training and Testing Strategy: After the sample ($\mathbf{X}$) and corresponding label set ($\mathbf{Y}$) preparation (Section II-B), the samples are divided into training and test sets in the ratio of 2:1. The validation split is set to 0.2, i.e., the 20 percent of the training samples are used for validation. For each combination, the method training and testing are repeated for 15 times, so as to get a statistically significant result. At each iteration, the training and test set samples are randomized to account for correlation between samples, and, therefore, reducing over-fitting.

Accuracy Assessment: The accuracy is assessed by calculating the Kappa coefficient (κ) [42]. The Kappa coefficient is a measure of interrater reliability and takes into account the possibility of correct classification occurring by chance. The accuracy for each class is assessed by calculating the classwise F-scores [43]. The F-score measure is calculated as the harmonic mean of precision and recall. Precision is the ratio of the total number of true positives to the total number of positively labeled samples, whereas recall is the ratio of the total number of true-positives to total number of positively labeled samples.

III. STUDY AREA AND DATASETS

We chose the test site which is near a Gypsum mining site in the county of Monaghan in Northern region of Ireland. This area coincided with a sinkhole event on 25-09-2018 [44]. The location of the study area is shown in Fig. 4(a) in red. Sentinel-1 SAR image and burst coverage are shown in Fig. 4(a) as well. Fig. 4(b) shows the SRTM digital elevation model over the zoomed area around the sinkhole spot, which is shown in red. Fig. 4(c) shows the corresponding hillshade map. It shows a depression immediately south-east of the sinkhole site, which corresponds to the location of the mine.

The deformation estimations of the CCS (i.e., PS) were estimated using 104 Sentinel 1-A ascending repeat pass SAR SLC (single look complex) images covering the sinkhole area. The
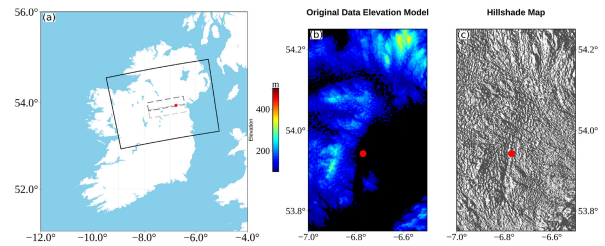

Fig. 4. (a) Sentinel-1 SAR image coverage in black. The two bursts used in the study in the dark and light gray. (b) Digital elevation model and (c) Hillshade map covering the sinkhole area. The red dot in the figures shows the sinkhole location.

images are taken over a time period between 05-09-2015 and 31-12-2018. The repeat pass period for all except 9 images is 12 days. The remaining images have a temporal resolution of 24 days. Fig. 5(a) shows the CCS deformation velocity map superimposed on the amplitude map over the study area and over a subset around the sinkhole location. This map corresponds to the two bursts shown in Fig. 4(a). The extent of the study area spanned $\sim 63 \times 44$ km.

IV. RESULTS

A. InSAR Time Series Extraction

The deformation time series was estimated by implementing single master TSInSAR processing using Delft implementation of persistent scattering interferometry (DePSI) [6]. We took the image acquired on 10-03-2017, as the reference master image. Then the deformation time series was readjusted with respect to the first acquisition date. Therefore, the temporal baseline varied between 0 to 3.65 a. Thereafter, the deformation time series was resampled to a common temporal baseline of 12 days using the bilinear interpolation method. From the TSInSAR derived results, standard postprocessing methods were used to reduce the number of noisy samples. Therefore, the total number of InSAR epochs were augmented to 112. Thereafter, the deformation velocity subset was set between -25 and 5 mm·yr^{-1}. After applying this condition the total number of CCS were reduced from 92 456 to 40 838. The spatial distribution of CCS in radar coordinates can be observed both in Fig. 5(a), and in the cropped scene Fig. 5(b), which corresponds to the black box (denoted as Area-1) in Fig. 5(a). The gray box on the top-right of the image (Area-2) corresponds to a nonsinkhole affected area which is used for comparison of results in Section IV-D.

B. Decomposition and Scattering Mechanism Classification

The false color composite of the filtered Sentinel-1 multi-temporal images in VV and VH channels, as well as the SM classification of a subset of the study area (Area-1) are shown in Fig. 5(c) and (d), respectively. In the polarimetric false color composite [see Fig. 5(c)], the normalized radar cross sections in VV, $\sigma_{VV}^{\circ 2}$, was assigned to the red color channel, the normalized radar cross sections in VH, $\sigma_{VH}^{\circ 2}$, was assigned to the green color channel and $(\sigma_{VV}^{\circ 2} + \sigma_{VH}^{\circ 2})/2$ was assigned to the blue color channel. In this false color composite, the red colored areas
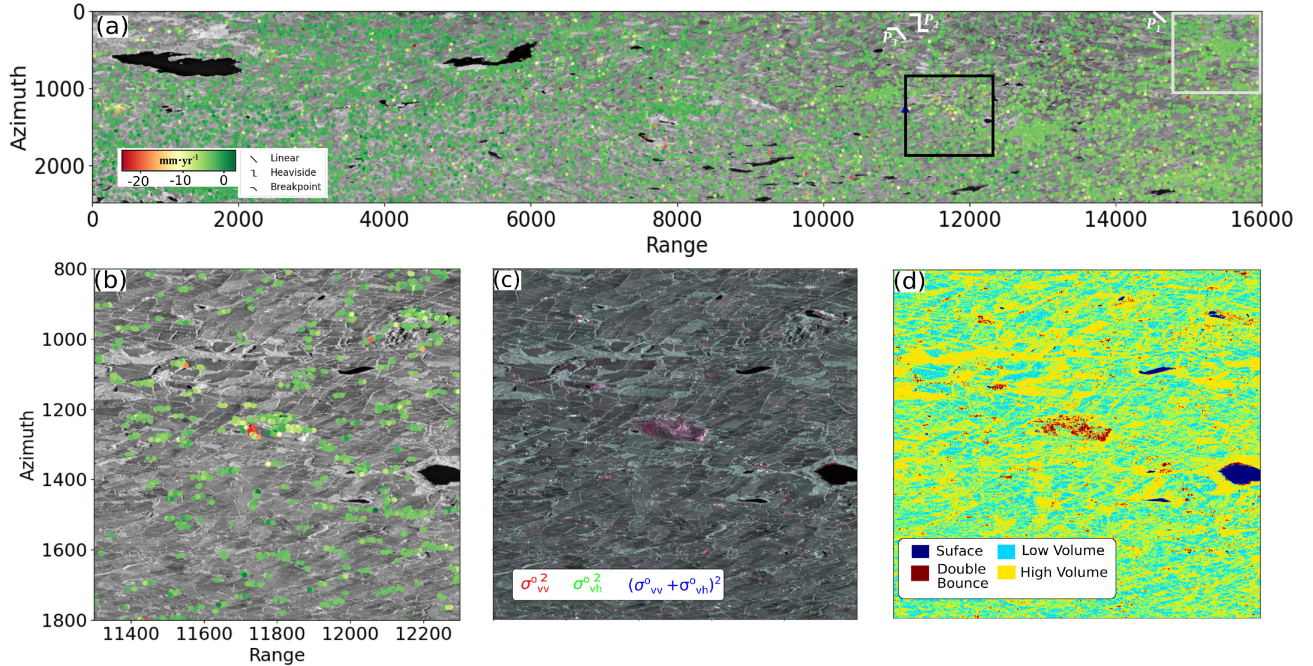
Fig. 5. (a) CCS deformation velocity map overlaid over the filtered VV channel image. (b) Cropped image, (c) polarimetric false color composite, and (d) scattering classification map corresponding to the black box in (a) of size 1000 × 1000 pixels (Area-1). The white markers show the approximate locations of the examples of Linear, Heaviside, and Breakpoint anomaly classes shown in Fig. 6. The gray box represents the alternate nonsinkhole area (Area-2) used for comparison, see Fig. 11(d) and (e).

and bright white pixels show correspondence to double-bounce scattering. Most prominently, this SM can be observed over the step-terraces of the mine [see Fig. 5(d)]. The green areas correspond to volume scattering. The dark patches relate to surface scattering, majorly over water bodies.

The SM classification was done using supervised random forest model. The SM classes were surface, low volume, high volume, and double bounce scattering. Volume scattering was bifurcated into two classes, (low and high) so as to differentiate between high volume scattering features, such as trees and low volume scattering features like farms. The training samples were manually crafted from the image shown in Fig 5(a). The samples for the surface scattering were pixels selected from water bodies and paved roads. The samples for the low and high volume scattering were selected from farmlands and trees, respectively. The samples for the double bounce scattering were selected from urban areas, mostly from buildings. The input features used for this classification were the normalized radar cross sections, σ_{VV}°, σ_{VH}°, in the VV and VH channels, respectively, as well as the Cloude–Pottier entropy calculated using the 2×2 coherency matrix.

For the random forest classification, the total number of decision trees were 500, the maximum number of features used to split the data at the nodes of decision trees was selected as $\sqrt{2}$ times the total number of features, i.e., 2 and Gini index was used to select the feature and feature value for splitting data at each node. The number of samples for the four classes in the same order as described above were 1521, 450, 1750, and 541, respectively. The out-of-bag error for the ensemble of 500 trees was 6.69%.

C. Deformation Class and Sample Extraction

Four deformation classes were extracted using MHT, which gave test statistics for each deformation class. The *a priori* variance used for the stochastic modeling was 5 mm^2. The level of significance (α) is defined using the total number of temporal epochs (m), as $\alpha = 1/(2\ m) = 1/(2 \times 112) \approx 0.45\%$. The power of the test γ was assigned as 50%.

The deformation classes were further separated using thresholds on the estimates of the test statistics, jumps and deformation velocity change. The bounds on the test ratios were applied in the form of p-values. The percentiles for the upper bound for the test statistics of the primary class, i.e., for T^C, $C \in \{\mathcal{D}_H, \mathcal{D}_{Br}\}$ was set to 95%. The percentile threshold for the lower bounds for the secondary class ($T^{C^\sim}$) was set to 30%. There is a tradeoff between the purity of samples and the total number of selected samples. High extremes for choosing the thresholds mentioned above, i.e., high values for T^C and low values for $T^{C^\sim}$, would result in improved likelihood of samples belonging to class C. However, that would reduce the total number of filtered samples, which may lead to inadequate number of samples for training the LSTM model. These values were set after experimenting between various values between [80%, 95%] for T^C, and [5%, 35%] for $T^{C^\sim}$.

Thereafter, the deformation time series samples were filtered using jumps and deformation velocity change angles. The lower bound for jumps in the Heaviside class was applied based on the limit due to the *a priori* known signal noise ~ 5 mm, and the upper bound was set by keeping the phase unwrapping errors into account. For a C-band sensor, with wavelength $\lambda = 5.6$ cm,
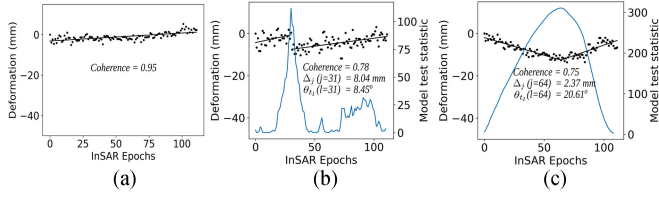
Fig. 6. Samples belonging to (a) Linear, (b) Heaviside (b), and (c) Breakpoint classes marked as P_1, P_2, and P_3, respectively, in Fig 5. The blue curve shows the model test statistics. The dots represent deformation estimations.
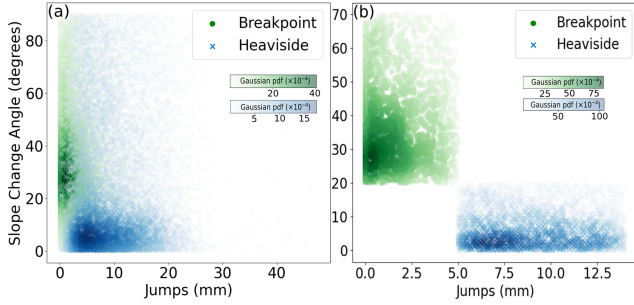

Fig. 7. (a) Density scatterplot between jumps and velocity change angles estimated at the detected anomaly epoch for Heaviside and Breakpoint samples. (b) Result after applying thresholds.

this value would be $\lambda/4 = 14$ mm. Therefore, $\Delta_{min}^{\mathcal{D}_H} = 5$ mm and $\Delta_{max}^{\mathcal{D}_H} = 14$ mm. The maximum jump for the Breakpoint class, i.e., $\Delta_{max}^{\mathcal{D}_{Br}}$ was set to 5 mm. This is because the maximum jumps for the Breakpoint class should be as low as possible so as to distinguish it from the Heaviside class and also realistic. The minimum jump threshold for the same class, i.e., $\Delta_{min}^{\mathcal{D}_{Br}}$ was set as the least possible value, i.e., 0. The minimum and maximum angle for the Breakpoint class, i.e., $\theta_{min}^{\mathcal{D}_{Br}}$, and $\theta_{max}^{\mathcal{D}_{Br}}$ at the velocity change epoch was set to $20°$ and $70°$, respectively. This is set so as to ensure that the velocity change angle for the Breakpoint class should be higher than that for the Heaviside class. Consequently, the maximum angle of the velocity change for the Heaviside class, i.e., $\theta_{max}^{\mathcal{D}_H}$ was set to $20°$.

As mentioned in Section II-B, the application of these thresholds is important to reduce the interclass similarity. The interclass similarity after the application of these thresholds can be assessed visually in Fig. 6, which shows three distinct models, i.e., Linear, Heaviside, and Breakpoint corresponding to CCS P_1, P_2, and P_3, respectively, marked in Fig. 5. Fig. 6(a) shows an example of linear deformation class which has an ensemble coherence of 0.95. The model test statistic (12) is the highest at the epoch of velocity change for the Breakpoint sample [Fig. 6(c)], and also at the epoch of highest jump in the Heaviside sample [Fig. 6(b)]. Fig. 6(b) and (c) also show the ensemble coherence, jumps (Δ) and velocity change angles (θ) for Heaviside and Breakpoint class samples.

The effect of thresholding can be seen in Fig. 7. In Fig. 7(a), the color shows the kernel density scatterplot. Even though the phenomena of jumps and deformation velocity changes are not orthogonal to each other, they have been represented as such just for visualizing the distribution of two key temporal features for Heaviside (jumps) and Breakpoint (deformation velocity change) classes. We observed significant reduction in the
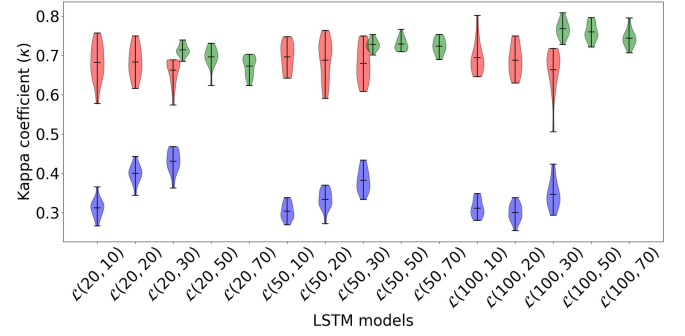

Fig. 8. Violinplot for the κ values calculated after running 15 iterations over each model defined as $\mathcal{L}(H, B_s)$.

number of class samples after applying these strict thresholds, see Fig. 7(b). It can be seen that majority of the jumps correspond to low velocity change angle for the Heaviside samples and vice-versa for the Breakpoint samples.

D. LSTM Modeling and Evaluation

The combination of values for the LSTM neurons and batch size are shown in Fig. 8, where the associated violinplot for kappa values over a set of model training repetitions is also shown. Each combination is represented by $\mathcal{L}(H, B_s)$, where H is the number of neurons in both the LSTM layers and B_s is the batch size. The learning rate was set as 0.001, and exponential decay rates for the first and second moments estimated by the Adam optimizer were 0.9 and 0.999, respectively.

We tested a three values of LSTM neurons $H = \{20, 50, 100\}$, testing from small values to the number of InSAR epochs, and batch sizes were varied in the set $B_s = \{10, 20, 30, 50, 70\}$. The test was done first by using the samples from just the first Sentinel-1 image burst (red and blue violins). Second, the test was performed by using samples from both image bursts (green violins) in order to assess the additional advantage of increasing sample size. The evaluation metrics were calculated over the test dataset. The red violins in the violinplot show the results from the training set after thresholding of samples. The blue ones, however, show the results before the application of these thresholds. The effect of thresholding is clearly visible here where there is $40\%-60\%$ consistent increase in the κ coefficient values, inspite of the larger sample size in the blue case.

The red and green violins show that the accuracy increases after increasing the number of samples by a factor of ~ 2 (2 Sentinel-1 bursts against 1 burst). Furthermore, the variability in the kappa value decreases (see, smaller interquartile ranges in green). The results in Fig. 8 show that lower batch sizes and high number of neurons generally perform well in the deformation classification task. The best combination was found out to be $\mathcal{L}(100, 30)$. This combination was chosen for further experimentation in the case of class crossing. Furthermore, it was observed that the model started overfitting, i.e., the loss values for validation set started rising, after ~ 30 LSTM epochs. The loss values for both training and validation set decreased by $\sim 37.5\%$ during the course of 30 LSTM epochs, as opposed
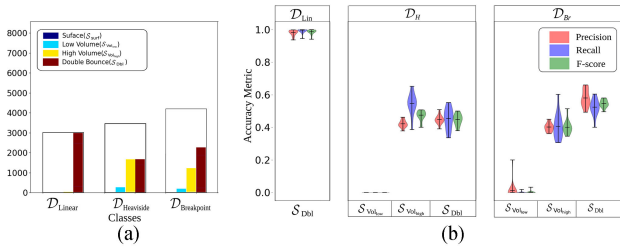
Fig. 9. (a) Sample distribution, and (b) classwise precision, recall and F-score for deformation classes and intersecting scattering mechanism classes.

Fig. 10. (a) Values of state (c_t) learnt by all 100 neurons of the 2nd LSTM layer for three deformation classes crossed with the double bounce scattering class. (b) Values of hidden state (h_t) with respect to InSAR epochs. The gray lines represent the state values for all samples, and black lines show the average values.

to an increase of loss value by $\sim 37.5\%$ for 120 LSTM epochs. Therefore, for further analysis, the number of LSTM epochs was set to 30.

Class Crossing: There was highest relative intersection for all the deformation classes with the double bounce SM class $(\mathcal{S}_{Dbl})$. This was especially the case for the linear deformation class samples $(\mathcal{D}_{Lin})$, see Fig. 9(a). There were no intersecting samples between the surface scatterring class and any of the deformation classes. The low volume scattering class also intersected lesser number of samples than high volume scattering and double bounce scattering classes. This is expected because surface scatterers, such as water bodies, and low volume scatterers, such as farms, do not scatter consistent and high signal back to the radar sensor. Amongst the Heaviside and Breakpoint classes, the Breakpoint class intersected with higher proportion of double bounce scatterers.

The classification accuracy of all crossed samples in the test set is shown in Fig. 9(b). It shows that the accuracy metrics: 1) precision, 2) recall, and 3) F-score for the best chosen LSTM model $\mathcal{L}(100, 30)$. The linear double bounce scatterers were almost completely correctly classified, but the classification accuracy for anomalous classes, i.e., Heaviside and Breakpoint ranged between 0.4 to 0.6. The accuracy for the double bounce scatters $(\mathcal{S}_{Dbl})$ was higher than the high volume scattering class $\mathcal{S}_{Vol_{high}}$ in case of the Breakpoint deformation class. These results are consistent with the sample distribution as seen in Fig. 9(a).

Looking Into the Box: The classification results were also analyzed in the context of learnt state values for different classes. The weights of the trained model, $\mathcal{L}(100, 30)$, see Fig. 8, were used to calculate the learnt state value for all neurons for three deformation class samples crossed with the double bounce scattering class. The results of the average learnt state values (c_t), which are shown in Fig. 10(a), show that certain neurons transform the samples for the $\mathcal{D}_{Lin}$ to a much higher absolute value than that for the samples for the $\mathcal{D}_H$ and $\mathcal{D}_{Br}$ classes. The extent of values is proportional to the number of InSAR epochs because the values are iteratively added to the learnt state c_t of the LSTM neuron, see Fig. 2. Fig. 10(b) shows the variation of the hidden state value (h_t) with respect to InSAR epochs. It can be seen that the values for the linear class are significantly distant from the Heaviside and Breakpoint classes. Furthermore, the separation between the anomalous and nonanomalous classes increases as the InSAR epochs increase.

From both results shown above, i.e., Fig. 10(a) and (b), the contrast between samples for the two anomalous classes is much
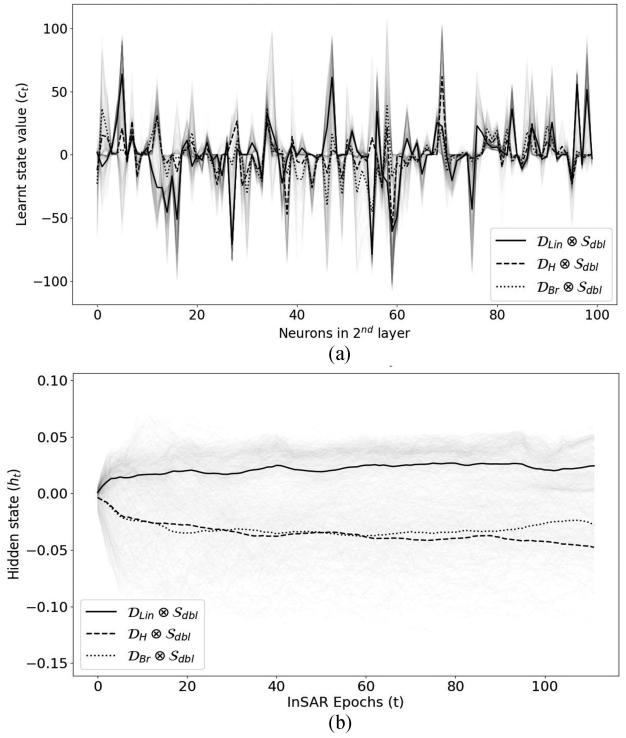
lower, which explains the reason for the lack of distinction between the two anomalous classes in the classification results, see Fig. 9. However, the distance between the two anomalous classes increases with the increase in InSAR epochs.

Anomaly Analysis: The deformation class map over the sinkhole are is shown in Fig. 11(a). The majority of scatterers belongs to the double bounce scattering class (red). We note a higher density of Heaviside and Breakpoint anomalies detected near the sinkhole site as compared to the linear deformation class. Out of the 202 points in Fig. 11(a), $\sim70\%$ are classified as one of these anomalies, and the remaining ones as linear scatterers. Furthermore, the anomalies occur very close to the sinkhole epoch, which was at the 90th epoch, see Fig. 11(c). However, the extent of jumps and velocity changes increases as we move away from the sinkhole spot, see Fig. 11(b). Fig. 11(d) and (e) show the deformation classes corresponding to the gray box in Fig. 5(a). Fig. 11(d) shows the deformation class overlaid with the SM class and Fig. 11(e) shows the jump and velocity change angles of Heaviside and Breakpoint classes, respectively. These figures show substantially less anomalies (false positives) suggesting that this area is less prone to sinkhole-related deformation behavior. In fact, 229 of the 386 points, are classified as linear. Furthermore, any large clusters of anomalous deformations are absent here, in contrast to the mining area in the sinkhole region, cf. the black box in Fig. 11(f). When comparing the Heaviside jump and Breakpoint velocity change maps to the deformation velocity maps, we note that the area in the 200-m vicinity of the sinkhole subsides with a velocity of -20 to -25 mm·yr^{-1},
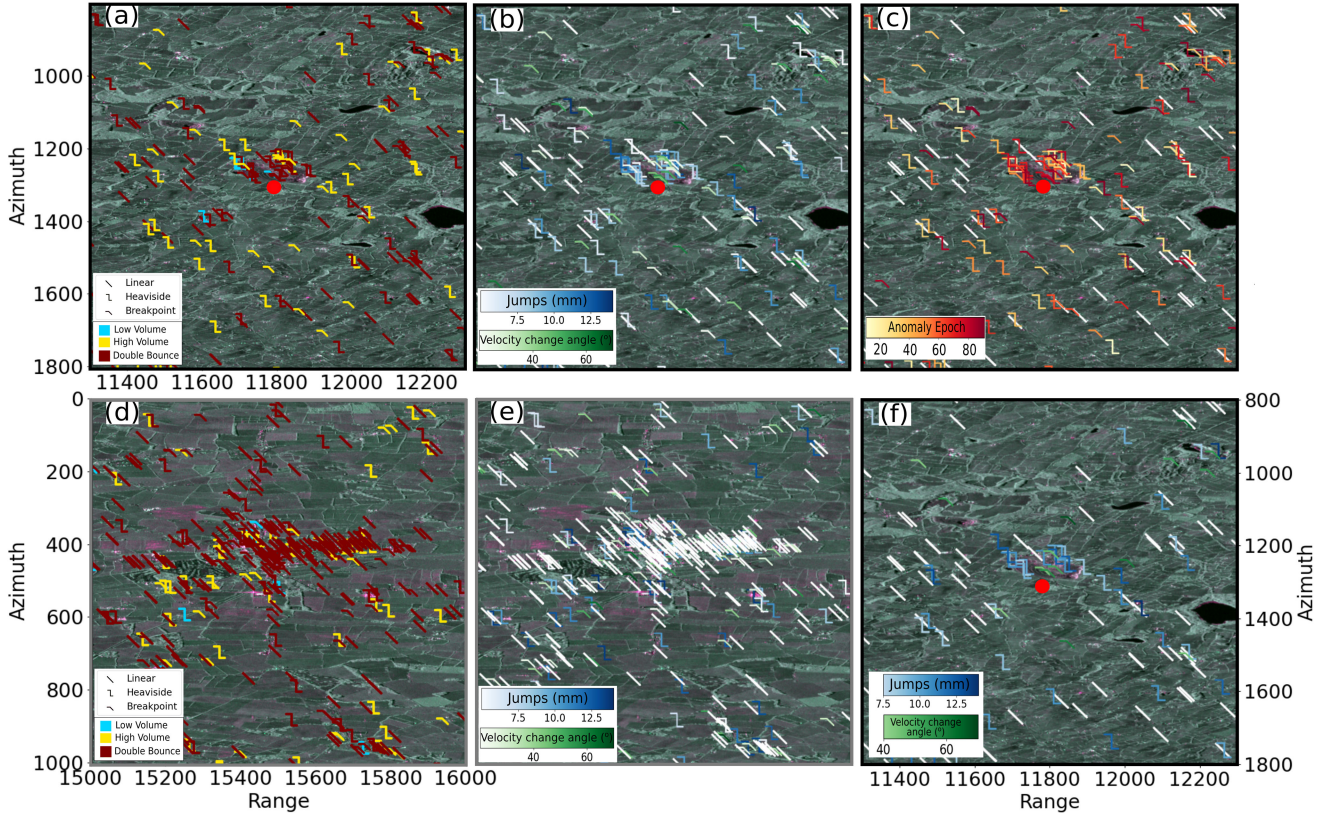
Fig. 11. (a) and (d) Map of linear, Heaviside and Breakpoint deformation classes crossed with scattering mechanism classes. (b) and (e) Heaviside jumps and Breakpoint velocity change angles; and (c) detected anomaly epochs identified over the sinkhole area. (f) Stricter threshold applied to map in (b). (d) and (e) correspond to Area-2 (gray box), whereas others correspond to Area-1 (black box) in Fig. 5(a), see the borders of subfigures. The sinkhole spot is shown as a red dot.

while the area in the gray box region is stable, i.e., deforming with a velocity ranging between 0 and -5 mm·yr^{-1}. False positives can be reduced by applying stricter thresholds on the jump and velocity angle changes, when contextual information is available. This can be noted for instance in Fig. 11(f), showing a significant reduction of false positives by more than 50%.

V. DISCUSSION

The estimated deformation velocity showed clusters of significantly high subsidence velocities, ~ -20 to -25 mm ·yr^{-1} over the mining area near the sinkhole site. This is also similar to what was observed in [18]. Over the study area, the CCS spatial density was found to be low because of nature of the study area which is predominantly rural. This was also observed in the SM classification map, which showed a higher presence of volume scattering as compared to double bounce scattering. For the SM classification, the volume scattering class was split into two subclasses due to the bimodal nature of σ°_{VH}. Upon analysis of the polarimetric features, the entropy of the 2×2 coherency matrix was observed to highly contrast between the double bounce and volume scattering pixels, and, therefore, was used as a feature in the SM classification using the random forest method. This method was effective in classifying the SAR image, with a low out-of-bag error of $\sim 6\%$.

The deformation classes were chosen in order to distinguish between anomalous and nonanomalous classes. In this research,

the nonanomalous class was chosen to be a default linear model. Anomalous classes were defined with respect to sudden variation in deformation time series and in deformation velocity of CCS, i.e., showing Heaviside and Breakpoint anomalies, respectively. These deformation classes were identified by fitting anomalous deformation models to samples for which the linear model residuals are large. This method is illustrated in [21, Fig. 2]. Anomalies are defined only at single epochs to limit the complexity in the class samples and demonstrate the application of the LSTM model in classifying deformation time series classes. In the future, multiepoch anomalies can be defined and tested. Classification of multiepoch anomalies can be done by increasing the number of layers in LSTM model.

Filtering of samples in anomalous classes was an essential step in increasing the classification accuracy. Deformation jumps and velocity change angles were identified as two key features which represented the Heaviside and Breakpoint classes, respectively. In this way, the distance between the samples of the two classes in the jump-velocity change angle feature space increased. This resulted in approximately a fivefold reduction in the total number of samples, and in an increase in the classification accuracy by an order of two. In this filtering step, the manual thresholds were set to test statistics, jumps, and velocity change angles. The values of these objective thresholds can be applied to other cases with only minor tuning.

For the LSTM classification, the training dataset is crafted by using only the deformation classes and then using a crossed

configuration between the deformation and SM classes. In this research. one-hot encoding was used to assign hard classes to the samples. In the future, soft classes can be assigned using class-wise likelihoods for all samples. Two model hyperparameters, LSTM neurons, and batch size, were first tuned by conducting experiments on only the deformation class samples. The range of parameter values was based upon two rule-of-thumb notions: first, the number of LSTM neurons should be close to the size of the time dimension of the time series data, and second, the batch sizes should be small. The results suggested that this was true, and the best results were achieved when the number of LSTM neurons was approximately equal to the time length of samples, i.e., ~ 100. Furthermore, it was observed that increasing the number of neurons and reducing the batch size increased the classification accuracy provided that the number of training samples were in the order of magnitude of 10^3.

Post tuning, the model was tested on the crossed class configuration, where the deformation classes were crossed with the SM classes. The linear class samples intersected almost completely with double bounce scatterers. The Heaviside and Breakpoint class samples, however, intersected with both the volume and double bounce scattering classes. Interestingly, class crossing acted as an additional sample filtering method, where the precision for double-bounce scattering samples was higher than that for the high-volume scattering class, in the case of both Heaviside and Breakpoint classes. The F-score for the Breakpoint-double bounce class was significantly higher than that for the Heaviside-double bounce class, which may have been due to higher number of samples in the former class.

The results of the LSTM classification were further analyzed by assessing the learnt and hidden states of the second LSTM layer. This layer encompassed the state values of the first layer as well. The values of the states showed high separation between the linear and anomalous classes. Separation between the two anomalous classes, however, was relatively low, but the interanomalous class similarity reduced with the increase in InSAR epochs. This suggested that increasing the number of InSAR epochs further could improve the distinction between Heaviside and Breakpoint classes.

Finally, the deformation classification map was analyzed spatially, which revealed that the density of anomalies with high jumps and velocity changes, was significantly higher near the sinkhole area than in other areas. This suggested that there may have been a relation between the sinkhole occurrence and anomalous deformation activity. The anomaly epoch was closer to the sinkhole epoch as we moved closer to the sinkhole spot. The extent of jumps ($\sim$10–12 mm) and velocity change angles (~ 50–$60°$) were also found to be high over the mining area, which is situated just across a road to the sinkhole site.

The design and use of the LSTM model in deformation time series classification is the main novelty of this study. The application of this LSTM model is generic and in the future we aim to continue this research to other areas and to compare our method with other methods, e.g., [45]. This model can also be used for deformation time series derived from other regions and SAR satellite images. For the SAR datasets with the same number of images and temporal resolution as used in this research, this trained model could be used as it is after minor tuning of model weights. For other datasets, the same model structure can be directly used, but the model requests training with the newly selected training datasets using MHT.

A major advantage of our model is that it scales well to large areas, and, therefore, is well suited for classification of deformation trends. Finally, any deformation classification map should be analyzed along with the deformation velocity map, the SM map, and the jump and velocity change angle maps in order to identify critical areas of sinkhole-related precursory deformation patterns. Accurate prediction of sinkhole-related anomalies remains a challenging task because of underlying geological and hydrological factors. These factors can either be taken into account explicitly, or their effects can be learnt from the data. A spatio–temporal sequence learning model would then be useful.

VI. CONCLUSION

This study focuses on identifying sinkholes that are associated with Heaviside and Breakpoint anomalies. Our proposed LSTM method effectively differentiates hazard-related anomalous deformation time series from linear ones. Sample filtering on the basis of the different velocity changes is recommended to improve the separation between the two anomaly-related classes as it increases the classification accuracy. SM classes are useful in filtering the samples.

Our results showed that the accuracy for distinguishing between the two anomalies can be increased by using more deformation time series samples and by increasing the number of InSAR epochs. Increasing this number may better separate the two classes. More features like multiepoch jumps, multivelocity changes, and classification on the basis of anomaly epochs, are expected to result in a further increase of the classification accuracy. Alternatively, addressing multiple bursts may also lead to more samples, while a higher number of LSTM layers should be considered if the complexity in the class samples increases, such as the presence of multiple anomalous epochs. False positives, i.e., nonsinkhole-related anomalies, can be reduced by using the parameter information of, e.g., the extent of jumps, velocity change angles and anomalous epochs. The correlation between these parameters could also be used in the analysis.

APPENDIX
TEST STATISTIC COMPUTATION

Herein, the method for calculating the test statistic values for Heaviside and Breakpoint deformation functions are shown. This formulation has been adapted from [46].

We assume that m (deformation) observations $\underline{y}$ and its stochastic model Q_{yy} are predesigned. The null hypothesis H_0 includes the default model, a linear function of time for our case, see (8). Against H_0, the alternative hypothesis H_a can be established by introducing additional parameter(s) ∇ and specification matrix C, see (9).

$$H_0 : E\{\underset{m\times 1}{\underline{y}}\} = \underset{m\times n}{A}\ \underset{n\times 1}{x}; D\{\underset{m\times 1}{\underline{y}}\} = \underset{m\times m}{Q_{yy}} \tag{8}$$

$$H_a : E\{\underset{m\times 1}{\underline{y}}\} = \underset{m\times n}{A}\ \underset{n\times 1}{x} + \underset{m\times q}{C}\ \underset{q\times 1}{\nabla}; D\{\underset{m\times 1}{\underline{y}}\} = \underset{m\times m}{Q_{yy}}. \tag{9}$$

Here $E\{.\}$ and $D\{.\}$ are the expectation operator and dispersion operator, respectively. A is the design matrix, x is the unknown parameter, and n is the number of unknowns. For instance, for a linear function of time, A is the temporal baseline, and x is the constant linear velocity. In this study, (2) and (3) are the examples of $C\nabla$. As q is the number of new additional parameters, and single Breakpoint and Heaviside both have only one additional parameter, then $q = 1$.

The residual between observations and model, in H_0, is defined by

$$\hat{e}_0 = \underline{\hat{y}} - A\hat{x}. \tag{10}$$

Then the test statistic is expressed as

$$T_q = \hat{\nabla}^T C^T Q_{yy}^{-1} Q_{\hat{e}_0 \hat{e}_0}^{-1} Q_{yy}^{-1} C^T \hat{\nabla}^T \tag{11}$$

where $Q_{\hat{e}_0 \hat{e}_0} = Q_{yy} - Q_{\hat{y}\hat{y}}$.

Then (11) is reformulated as

$$T_q = \hat{e}_0^T Q_{yy}^{-1} C (C^T Q_{yy}^{-1} Q_{\hat{e}_0 \hat{e}_0} Q_{yy}^{-1} C)^{-1} C^T Q_{yy}^{-1} \hat{e}_0. \tag{12}$$
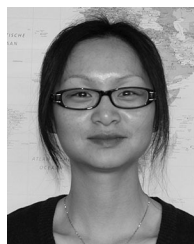
ACKNOWLEDGMENT

REFERENCES

[1] B. G. Herrera-García *et al.*, "Mapping the global threat of land subsidence," *Science*, vol. 371, no. 6524, pp. 34–36, 2021.

[2] R. G. Hermosilla, "The Guatemala city sinkhole collapses," *Carbonates Evaporites*, vol. 27, pp. 103–107, 2012.

[3] A. Theron and J. Engelbrecht, "The role of Earth observation, with a focus on SAR interferometry, for sinkhole hazard assessment," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1506.

[4] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatters in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, Jan. 2001.

[5] R. F. Hanssen, Radar Interferometry: Data Interpretation and Error Analysis. Dordrecht, Netherlands: Springer, 2001.

[6] F. van Leijen, "Persistent scatterer interferometry based on geodetic estimation theory," Ph.D. dissertation, Dept. Geosci. Remote Sens., Delft Univ. Technol., Delft, Netherlands, 2014.

[7] L. Chang and A. Stein, "Exploring PAZ co-polarimetric SAR data for surface movement mapping and scattering characterization," *Int. J. Appl. Earth Observation Geoinf.*, vol. 96, no. Jul. 2020, 2021, Art. no. 102280.

[8] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2375–2383, Nov. 2002.

[9] S. Samiei-Esfahany, J. E. Martins, F. Van Leijen, and R. F. Hanssen, "Phase estimation for distributed scatterers in InSAR stacks using integer least squares estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5671–5687, Oct. 2016.

[10] L. Chang, R. P. B. J. Dollevoet, and R. F. Hanssen, "Nationwide railway monitoring using satellite SAR interferometry," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 596–604, Feb. 2017.

[11] L. Chang, R. P. B. J. Dollevoet, and R. F. Hanssen, "Monitoring line-infrastructure with multisensor SAR interferometry: Products and performance assessment metrics," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1593–1605, May 2018.

[12] P. Castellazzi *et al.*, "Land subsidence in major cities of central Mexico: Interpreting InSAR-derived land subsidence mapping with hydrogeological data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 47, pp. 102–111, 2016.

[13] P. Milillo *et al.*, "Monitoring dam structural health from space: Insights from novel InSAR techniques and multi-parametric modeling applied to the Pertusillo dam Basilicata, Italy," *Int. J. Appl. Earth Observation Geoinf.*, vol. 52, pp. 221–229, Oct. 2016.

[14] T. Wang, D. Perissin, F. Rocca, and M.-S. Liao, "Three gorges dam stability monitoring with time-series InSAR image analysis," *Sci. China Earth Sci.*, vol. 54, pp. 720–732, 2011.

[15] A. Hooper, H. Zebker, P. Segall, and B. Kampes, "A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers," *Geophysical Res. Lett.*, vol. 31, no. 23, pp. 1–5, 2004.

[16] M. Lazecký *et al.*, "LiCSAR: An automatic InSAR tool for measuring and monitoring tectonic and volcanic activity," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2430.

[17] L. Chang and R. F. Hanssen, "Detection of cavity migration and sinkhole risk using radar interferometric time series," *Remote Sens. Environ.*, vol. 147, pp. 56–64, 2014.

[18] A. Kulshrestha, L. Chang, and A. Stein, "Sinkhole scanner: A new method to detect sinkhole-related spatio-temporal patterns in InSAR deformation time series," *Remote Sens.*, vol. 13, no. 15, 2021, Art. no. 2906.

[19] P. Castellazzi and W. Schmid, "Interpreting C-band InSAR ground deformation data for large-scale groundwater management in Australia," *J. Hydrol.: Regional Stud.*, vol. 34, 2021, Art. no. 100774.

[20] M. E. Martinotti *et al.*, "Landslides, floods and sinkholes in a karst environment: The 1-6 Sep. 2014 Gargano event, southern Italy," *Natural Hazards Earth Syst. Sci.*, vol. 17, pp. 467–480.

[21] L. Chang and R. F. Hanssen, "A probabilistic approach for InSAR time-series postprocessing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 421–430, Jan. 2016.

[22] M. Bianchi, C. Bischof, A. Rucci, M. Basilico, and F. Lattari, "Machine learning methods for SAR-derived time series trend change detection," TRE-Altamira, Milan, Italy, Tech. Rep. D100.6, 2020. [Online]. Available: site.tre-altamira.com/mattch/

[23] H. Ansari, M. Rußwurm, M. Ali, S. Montazeri, A. Parizzi, and X. X. Zhu, "InSAR displacement time series mining: A machine learning approach," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 3301–3304.

[24] S. Narayan, "The generalized sigmoid activation function: Competitive supervised learning," *Inf. Sci.*, vol. 99, no. 1, pp. 69–82, 1997.

[25] A. Farzad, H. Mashayekhi, and H. Hassanpour, "A comparative performance analysis of different activation functions in LSTM networks for classification," *Neural Comput. Appl.*, vol. 31, pp. 2507–2521, 2019.

[26] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[27] X. Zhu *et al.*, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.

[28] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. Int. Conf. Artif. Neural Netw.*, vol. 2, 1999, pp. 850–855.

[29] P. Ma, F. Zhang, and H. Lin, "Prediction of InSAR time-series deformation using deep convolutional neural networks," *Remote Sens. Lett.*, vol. 11, no. 2, pp. 137–145, 2020.

[30] A. G. Mullissa, D. Perissin, V. A. Tolpekin, and A. Stein, "Polarimetry-based distributed scatterer processing method for PSI applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3371–3382, Jun. 2018.

[31] A. Freeman and S. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 963–973, May 1998.

[32] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1699–1706, Aug. 2005.

[33] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.

[34] S. Quegan and J. J. Yu, "Filtering of multichannel SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2373–2379, Nov. 2001.

[35] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[36] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, 2016.

[37] P. J. G. Teunissen, "Quality control in integrated navigation systems," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 5, no. 7, pp. 35–41, Jul. 1990.

[38] C. Olah, "Understanding LSTM Networks," 2015. Accessed: Jun. 9, 2022. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs

[39] T. D. Pham, "Time–frequency time–space LSTM for robust classification of physiological signals," *Sci. Rep.*, vol. 11, 2021, Art. no. 6936.

[40] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–13.

[42] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[43] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Advances in Information Retrieval (ECIR Lecture Notes in Computer Science)*, vol 3408, D. E. Losada and J. M. Fernández-Luna, Eds. Berlin, Germany: Springer, 2005.

[44] BBC, "Sinkhole-hit GAA club in county Monaghan plans new pitches," 2019. Accessed: Jun. 9, 2022. [Online]. Available: https://www.bbc.com/news/world-europe-46748729

[45] A. Vaswani *et al.*, "Attention is all you need," in *Adv. Neural Inf. Process. Syst.*, vol. 30, Red Hook, NY, USA, 2017.

[46] P. J. G. Teunissen, D. G. Simons, and C. C. J. M. Tiberius, *Probability and Observation Theory*. Delft, Netherlands: Delft Univ. Technol., 2005.

Ling Chang received the M.S.E. degree in geodesy and survey engineering from Tongji University, Shanghai, China, and the Ph.D. degree in monitoring civil infrastructure using satellite radar interferometry from the Delft University of Technology, Delft, The Netherlands, in 2010 and 2015, respectively.

She was an Assistant Professor of Microwave Remote Sensing with the ITC, University of Twente, Enschede, The Netherlands. Her research interests include statistical hypothesis testing, time series modeling, and change detection, using satellite-borne remote sensing technology.

Anurag Kulshrestha (Member, IEEE) received the M.Sc. degree in physics from the Indian Institute of Science, Education and Research (IISER), Mohali, Punjab, India, in 2014, and the M.Sc. degree in geoinformation science and earth observation, with a specialization in geoinformatics, in 2018, from the Faculty of Geoinformation Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands, where he is currently working toward the Ph.D. degree in time series insar analysis using deep learning for detection of geological hazards.

His research interests include interferometric and polarimetric SAR signal processing, classification of spatio–temporal remotely sensed data using machine learning and deep learning, and natural hazard monitoring using remote sensing.

Alfred Stein received the M.Sc. degree in mathematics and information science (with a specialization in applied statistics) from the Eindhoven University of Technology, Eindhoven, The Netherlands, and the Ph.D. degree in spatial statistics from Wageningen University, Wageningen, The Netherlands, in 1983 and 1991, respectively.

He is currently a Professor of Spatial Statistics and Image Analysis with the Faculty of Geoinformation Science and Earth Observation, University of Twente, Enschede, The Netherlands, and 10 Ph.D. students are working under his supervision. He holds honorary positions with the University of Pretoria, Hatfield, Pretoria, South Africa, and the University of Cape Town, South Africa. His research interests include the statistical aspects of spatial and spatiotemporal data in the widest sense, such as monitoring data, optimal sampling, image analysis, spatial statistics, including the use of prior information, issues of data quality, fuzzy techniques, and random sets in a Bayesian setting. His research interests also include landuse, hazards and disasters, spatial geohealth, and spatial epidemiology.

Dr. Stein has been the Editor-in-Chief for the *Spatial Statistics Journal*, since 2011, which is the new leading platform in the field of spatial statistics.