

Interpretation and Attribution of Coastal Land Subsidence: An InSAR and Machine Learning Perspective

Xiaojun Qiao , Tianxing Chu , Evan Krell , Philippe Tissot , Seneca Holland, Mohamed Ahmed , and Danielle Smilovsky 

Abstract—Subsidence, the downward vertical land motion (VLM), plays a pivotal role in contributing to the risk of coastal flooding. Accurately estimating VLM and identifying its potential features related to subsidence can provide crucial information for stakeholders to make better-informed decisions. This study aimed to estimate large-scale subsidence at the Texas Gulf Coast and identify potential subsidence features using explainable models. Nine potential features were considered for modeling the VLM, ranging from natural terrain variations to anthropogenic activities. These features were used to train a random forest (RF) machine learning model. Explainable artificial intelligence (XAI) techniques including SHapley Additive exPlanations (SHAP) and impurity- and permutation-based feature importance were used to identify the contributions to subsidence. The results demonstrated favorable performance of the RF model, achieving an R^2 value of 0.56 during validation. XAI results underscored the significance of the digital elevation model in explaining subsidence at the Texas Coast. Additionally, XAI analysis highlighted the overall contribution of subsidence from anthropogenic activities, such as hydrocarbon extraction and groundwater withdrawal. Furthermore, the sample-level SHAP map provided detailed and reasonable subsidence-attribution results across the study area, showing potential for automatic and data-driven explanations of the VLM.

Manuscript received 28 September 2023; revised 18 December 2023; accepted 30 January 2024. Date of publication 1 February 2024; date of current version 20 February 2024. This work was supported in part by the U.S. Department of Commerce-National Oceanic and Atmospheric Administration (NOAA) through the University of Southern Mississippi (USM) under the terms of Agreement NA18NOS4000198, in part by the National Science Foundation (NSF) under Award 2131263, and in part by the Open Access Publication Fund provided by the Mary and Jeff Bell Library at Texas A&M University-Corpus Christi (TAMU-CC). (Corresponding author: Tianxing Chu.)

Xiaojun Qiao and Tianxing Chu are with the Department of Computer Science and the Conrad Blucher Institute for Surveying and Science, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412 USA (e-mail: xqiao@islander.tamucc.edu; tianxing.chu@tamucc.edu).

Evan Krell is with the innovation in COrputing REsearch (iCORE) Lab, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412 USA (e-mail: ekrell@islander.tamucc.edu).

Philippe Tissot is with the Conrad Blucher Institute for Surveying and Science and the Department of Physical and Environmental Sciences, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412 USA (e-mail: philippe.tissot@tamucc.edu).

Seneca Holland and Danielle Smilovsky are with the Conrad Blucher Institute for Surveying and Science, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412 USA (e-mail: seneca.holland1@tamucc.edu; danielle.smilovsky@tamucc.edu).

Mohamed Ahmed is with the Department of Physical and Environmental Sciences, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412 USA (e-mail: mohamed.ahmed@tamucc.edu).

Digital Object Identifier 10.1109/JSTARS.2024.3361391

Index Terms—Attribution analysis, coastal land subsidence, explainable artificial intelligence (XAI), machine learning (ML).

NOMENCLATURE

API	American Petroleum Institute.
APS	Atmospheric phase screen.
BBL	One barrel of crude oil.
cGNSS	Continuous global navigation satellite systems.
CNN	Convolutional neural networks.
DEM	Digital elevation model.
DTF	Distance to fault.
GEE	Google Earth Engine.
GIA	Glacial isostatic adjustment.
GNSS	Global navigation satellite system.
GWWD	Groundwater withdrawal.
HPRD	Hydrocarbon production.
HWD	Hydrocarbon well density.
IFI	Impurity feature importance.
IMPV	Imperviousness.
InSAR	Interferometric synthetic aperture radar.
LOS	Line of sight.
MCF	Thousand cubic feet of natural gas.
ML	Machine learning.
NDVI	Normalized difference vegetation index.
PFI	Permutation feature importance.
PPP	Precise point positioning.
PS	Persistent scatterer.
RF	Random forest.
RRC	Railroad Commission of Texas.
SAR	Synthetic aperture radar.
SHAP	SHapley Additive exPlanations.
SRTM	Shuttle Radar Topography Mission.
TWI	Topographical wetness index.
USGS	United States Geological Survey.
VIIRS	Visible Infrared Imaging Radiometer Suite.
VLM	Vertical land motion.
VNL	VIIRS nighttime lights.

I. INTRODUCTION

SUBSIDENCE refers to the sinking, settling, and downward VLM of the Earth's surface [1], usually occurring as a result of natural geological processes or human activities such as the

extraction of underground resources like water, oil, and gas, or due to the compaction of soil and sediment [2]. The VLM process can trigger a cascade of consequences including the loss of surface areas, damage to infrastructures, and challenges in urban planning [3], [4]. Apart from these well-documented impacts, subsidence in densely populated coastal regions can compound a host of additional critical issues including flooding and seawater intrusion [5]. These challenges are primarily associated with complex land–water interactions, especially in the context of the rising trend of the global sea level [6], [7]. Consequently, land subsidence emerges as a pivotal factor in shaping the future of coastal regions [8].

Over the past decades, there has been growing attention to quantifying land subsidence. For example, traditional leveling surveys offer precision but demand substantial time and labor, even for a relatively small area [2]. Extensometers and the GNSS provide accurate measurements for land movement, and cGNSS stations have been widely distributed across vast geographic areas in the world for long-term geodetic positioning [2], [9]. Nonetheless, these approaches rely on localized measurements [10], potentially overlooking the detailed and varying spatial characteristics of subsidence patterns.

InSAR emerges as an accurate and effective spaceborne geodetic technique for generating expansive subsidence maps, leveraging the dense scatterer distribution within the SAR imagery. The effectiveness of InSAR is underpinned by the modern SAR sensors, as exemplified by those aboard Sentinel-1 satellites, which enable the acquisition of substantial imagery across both spatial and temporal dimensions. Meanwhile, advanced multitemporal InSAR data processing strategies, such as PS and small baseline subset (SBAS) methods [11], streamline the handling of extensive image datasets. For instance, the integrated spaceborne approach of InSAR and GNSS has yielded comprehensive national land subsidence maps in Japan [12] and Italy [13] based on Sentinel-1 SAR imagery. In contrast to PS, SBAS typically enhances phase coherence through its multilooking filtering and utilizing redundant interferograms, resulting in fewer pixels being masked induced by coherence outages [14]. On the other hand, PS methods exhibit efficiency in data processing and are capable of yielding large-scale deformation results with full resolution [14].

Comprehending and identifying drivers of land subsidence has long been considered difficult. First, complexity arises from the need to consider various anthropogenic and natural factors that may contribute to land motion. For example, a widely recognized anthropogenic driver is the extraction of fluids, such as oil, gas, and groundwater, which leads to underground compaction and subsequent land surface subsidence [1]. Similarly, activities such as mining and construction can induce land subsidence with increased structure loading or weakened mining support [15], [16], [17]. Also, natural processes can instigate land motion including phenomena such as thawing permafrost [18], shrinkage and swelling of clay soils [19], and GIA [20]. Additionally, the spatial patterns of subsidence velocity can be influenced or controlled by growth faults [21], further complicating the practical analysis of subsidence drivers. Finally, inferring subsurface processes and accurately attributing subsidence to

specific drivers may require subsurface observations as well as related geological and geophysical analysis, such as changes in pore pressure resulting from fluid extraction [22].

Due to these challenges, prior investigations in identifying subsidence drivers have mostly concentrated on qualitative analyses of the VLM. Spatial coincidence between the presence of oil/gas wells and observed subsidence hotspots in the vicinity of the Houston-Galveston Area, Texas, was studied [23]. Prior studies also analyzed the quantitative association between subsidence and HPRD. However, the data of hydrocarbon extraction were aggregated at the county level [24], [25], posing challenges in attribution due to the spatial variability inherent in processes related to both land subsidence and anthropogenic activities. Furthermore, even when accounting for various potential features such as HPRD, GWWD, and geological faults, these features were typically examined in isolation [23], [24], [26], [27].

Integrating various potential features into an ML model may provide valuable insights into the processes driving land subsidence. For example, a range of features encompassing geological, land cover, topographical, and climatic aspects have proven effective in enhancing the accuracy of subsidence detection in deltaic metropolitan areas located in southern China [28]. Furthermore, explainable artificial intelligence (XAI) holds promise in enhancing the understanding of VLM modeling through ML. XAI techniques can be distinguished by interpretable ML and post-hoc methods [29]. Interpretable ML refers to the development of architectures that are designed to expose some information about what the model learned to aid understanding. Post-hoc methods are XAI algorithms that consider the model itself to be a black box and analyze the characteristics of the model based only on the inputs and outputs [30], [31]. Post-hoc methods can be applied to interpretable models and provide information that the interpretable model itself does not.

Leveraging large-scale subsidence estimates derived across the Texas Gulf Coast using the PS InSAR technique [27], this study aimed to interpret land subsidence by identifying potential contributors or correlates of the observed VLM within a quantitative model framework. To achieve this, the study applied ML-based regression to establish relations between observed land subsidence and a range of potential features. It is crucial to highlight that this study did not focus on predicting subsidence at unobserved locations or dates using the ML model. Instead, it aimed to unveil underlying drivers and correlates of subsidence through XAI analysis of a diversely-featured ML model. Key acronyms throughout the article can be found in the Nomenclature.

II. STUDY AREA

Fig. 1 illustrates the extent of Sentinel-1 data coverage, which intersects with a total of 47 counties lining the Texas coastline. The study area is defined by the overlapping region between the spatial reach of Sentinel-1 data and the geographical boundaries of these counties. It encompasses an area of approximately 89 000 km². Over geological time, from the Miocene to the Pleistocene periods, sediments were gradually

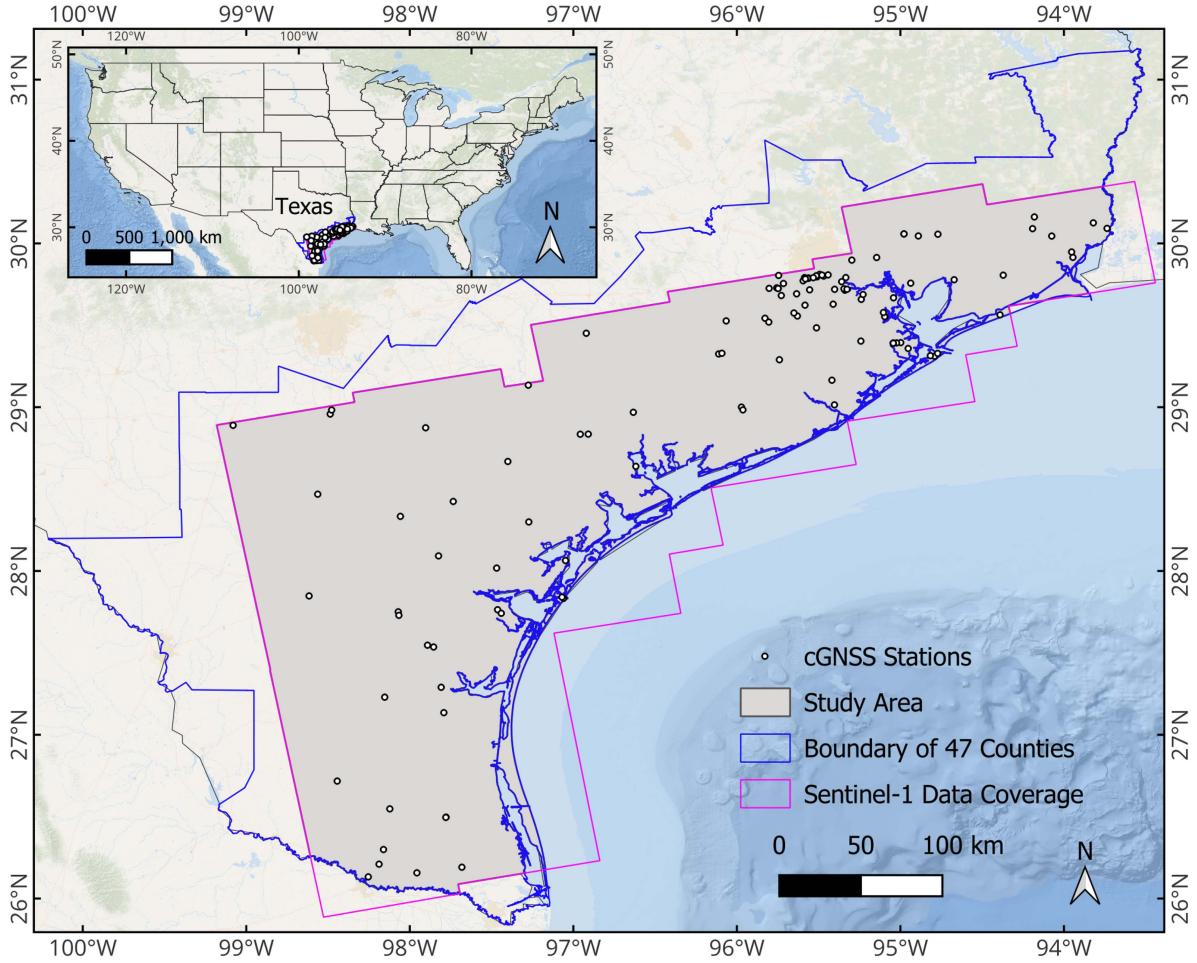


Fig. 1. Study area including a total of 47 coastal counties in Texas, which overlaps the coverage of acquired Sentinel-1 imagery.

deposited along the Texas Gulf Coast. This deposition occurred within a diverse range of environments, including fluvial-deltaic and shallow-marine settings. Within the Coastal Plain, these sediments accumulated on a homoclinal slope gently toward the Gulf of Mexico [32]. Across the Texas Gulf Coast, distinct sedimentary wedges, representing substantial layers of sediment, have been identified. Some of these formations contain deep sandstone reservoirs that have been of benefit for hydrocarbon formation [33]. Additionally, it is worth noting that salt domes are primarily concentrated in the northern region of the Texas Gulf Coast. Additionally, curved faults were formed in parallel with the coastline during the deposition process [32]. The utilization of groundwater resources in this region has been substantial, especially near the Houston-Galveston Area [23] for uses including irrigation, drinking water supply, and industrial needs [1], [32]. Land subsidence is a widespread issue affecting the entire Texas Gulf Coast, primarily attributed to activities related to GWWD and hydrocarbon extraction [34].

III. DATA

A collection of four swaths of Sentinel-1 SAR images was acquired and utilized for the generation of a large-scale

subsidence map. Within the footprint of each swath, the study utilized two or three subswath imagery subsets that overlap the Texas coastlines. A total of 10 subswaths of SAR data were processed independently, and the spatial coverage of these SAR images can be found in Fig. 1. InSAR results were subsequently validated using ground-truth land-motion observations from 115 cGNSS stations (see Fig. 1). Additional details regarding Sentinel-1 SAR and cGNSS data are shown in Table I.

Also, a range of features related to natural and anthropogenic activities was utilized for modeling the VLM including HPRD, HWD, DEM, TWI, GWWD, NDVI, IMPV, VNL, and DTF. The selection of these nine features stemmed from their data accessibility, fine-grained spatial resolution, availability in a numerical format, and their ability to explain or predict subsidence, either as potential drivers or correlative indicators. The detailed relevance of these selected features with land subsidence is summarized as follows.

- 1) HPRD, HWD, and GWWD: Extraction of hydrocarbons (e.g., oil and gas) and groundwater is widely considered to be one of the primary drivers of land subsidence globally [1], [23].
- 2) DEM, TWI, VNL, and DTF: Prior studies utilized these features for mapping subsidence susceptibility, with some

TABLE I
DATA USED FOR ESTIMATING THE VLM AND ANALYZING ITS POTENTIALLY ASSOCIATED FEATURES

Data	Variable	Source	Spatial Detail	Duration (yyyy-mm-dd)	Description
Sentinel-1 SAR	VLM	[42]	5 m × 20 m	• 2016-10-04 – 2022-05-18 • 2017-02-18 – 2022-05-23 • 2017-04-24 – 2022-06-21 • 2017-02-18 – 2022-05-04	• 136_93 (sub-swaths 1 and 2) * • 34_90 (sub-swaths 1, 2, and 3) • 107_88 (sub-swaths 1, 2, and 3) • 107_83 (sub-swaths 1 and 2)
GNSS	VLM	[43]	Point	2017-01-01 – 2021-12-31	The temporal trend of PPP solutions in the vertical direction was estimated using the Hector suite [44] across a dataset comprising 115 cGNSS stations, each of which possessed more than two years of observations between January 1, 2017, and December 31, 2021.
Hydrocarbon	• HPRD • HWD	[45]	Point	1993-01-01 – 2022-12-31	Details can be found in Section III-A.
DEM	• DEM • TWI	[46]	30 m	—	• NASA JPL SRTMGL1 (vers. 3) accessed from GEE [47]. • Refer to details in Section III-B.
Groundwater	GWWD	[48]	Point	1990-01-01 – 2018-12-31	The monthly average of groundwater withdrawal was calculated between 1990 and 2018 for 53,602 distributed wells within the Texas Gulf Coast Aquifer and subsequently extrapolated spatially across the study area.
NDVI	NDVI	[49]	30 m	2017-01-01 – 2022-01-01	Temporal average of $\frac{B_5 - B_4}{B_5 + B_4}$ values while excluding cloudy pixels was calculated in GEE [47] utilizing data from the USGS Landsat 8 satellite, where B_5 and B_4 represent surface reflectance at the near-infrared and red band, respectively.
Imperviousness	IMPV	[50]	30 m	In 2019	The IMPV quantities are in percentages.
Nighttime light	VNL	[51]	Around 500 m	2017-01-01 – 2022-01-01	The average of nighttime light radiance was computed in GEE [47].
Fault	DTF	[52]	polyline	—	Fault data was first cropped spatially to the study area and the DTF was then calculated for each 250 m × 250 m grid cell.

* SAR data were sourced from Path 136 and Frame 93 with sub-swath numbers enclosed in brackets.

being contributing factors to the model prediction [35], [36].

3) IMPV and VNL: Subsidence in Texas may also be influenced by some other natural factors such as fluctuations in soil moisture during dry periods [37], [38] and human activities such as uncontrolled gas flaring and operations related to hydrocarbon extraction, particularly at night in unpopulated areas [27]. IMPV and VNL may provide insights into explaining subsidence.

It should be noted that the 250 m × 250 m spatial grids were utilized to facilitate the calculation and analysis of feature values of both the nine features and the VLM. Besides, the inclusion of around 30-year data of both HPRD and GWWD was a consideration for their enduring influence on land subsidence [39], [40], [41]. Extensive details concerning these nine features are listed in Table I. Additional information regarding obtaining the HPRD/HWD and TWI features is discussed in Sections III-A and III-B, respectively.

A. Hydrocarbon Data

To obtain well-level HPRD, the following three datasets were accessed from the RRC [45].

- 1) For the 47 counties, well layers were accessed to retrieve the geographical coordinates for each unique American Petroleum Institute (API) number. The API-based “bottom well” layers were then spatially combined across the 47 counties to establish the subsurface well locations from which hydrocarbon fluids were extracted.
- 2) For the 47 counties, the API database was obtained that contains the RRC-issued lease number, which is unique to each Oil and Gas Division district.

3) In this study, monthly HPRD data for individual leases in districts with codes 01, 02, 03, and 04, which collectively cover the entire study area, were obtained from the RRC [45]. HPRD is measured in barrels (BBL) for oil and thousand cubic feet (MCF) for gas. It should be noted that HPRD provided by RRC contains the production of oil (BBL) and casinghead (MCF) for oil leases, and gas (MCF) and condensate (BBL) for gas leases.

In summary, these three datasets were used together for calculating HPRD for spatially distributed wells (i.e., API number) through lease-based production data released by the RRC. Specifically, the following equation is used for calculating the HPRD variable on a per API basis:

$$\text{HPRD}(t, \text{api}) = \sum_{j=1}^M \frac{1}{w_j} \left[\frac{P_{\text{cas}}(t, m_j)}{6} + P_{\text{oil}}(t, m_j) \right] + \sum_{i=1}^L \left[\frac{P_{\text{gas}}(t, l_i)}{6} + P_{\text{con}}(t, l_i) \right]. \quad (1)$$

The variables P_{cas} , P_{oil} , P_{gas} , and P_{con} represent HPRD categories, namely casinghead, oil, gas, and condensate, respectively. The variable l_i is drawn from the gas lease set $[l_1, \dots, l_L]$, which is linked to a specific API number. Similarly, m_j belongs to the oil lease set $[m_1, \dots, m_M]$, associated with a particular API number. Time is denoted by the variable t , measured at monthly intervals from January 1993 to December 2022. It is important to note that an oil lease may have been connected to multiple API numbers. To address this, we incorporated a weighting factor $\frac{1}{w_j}$ when allocating the production from oil leases tied to a specific API location, where w_j signifies the total

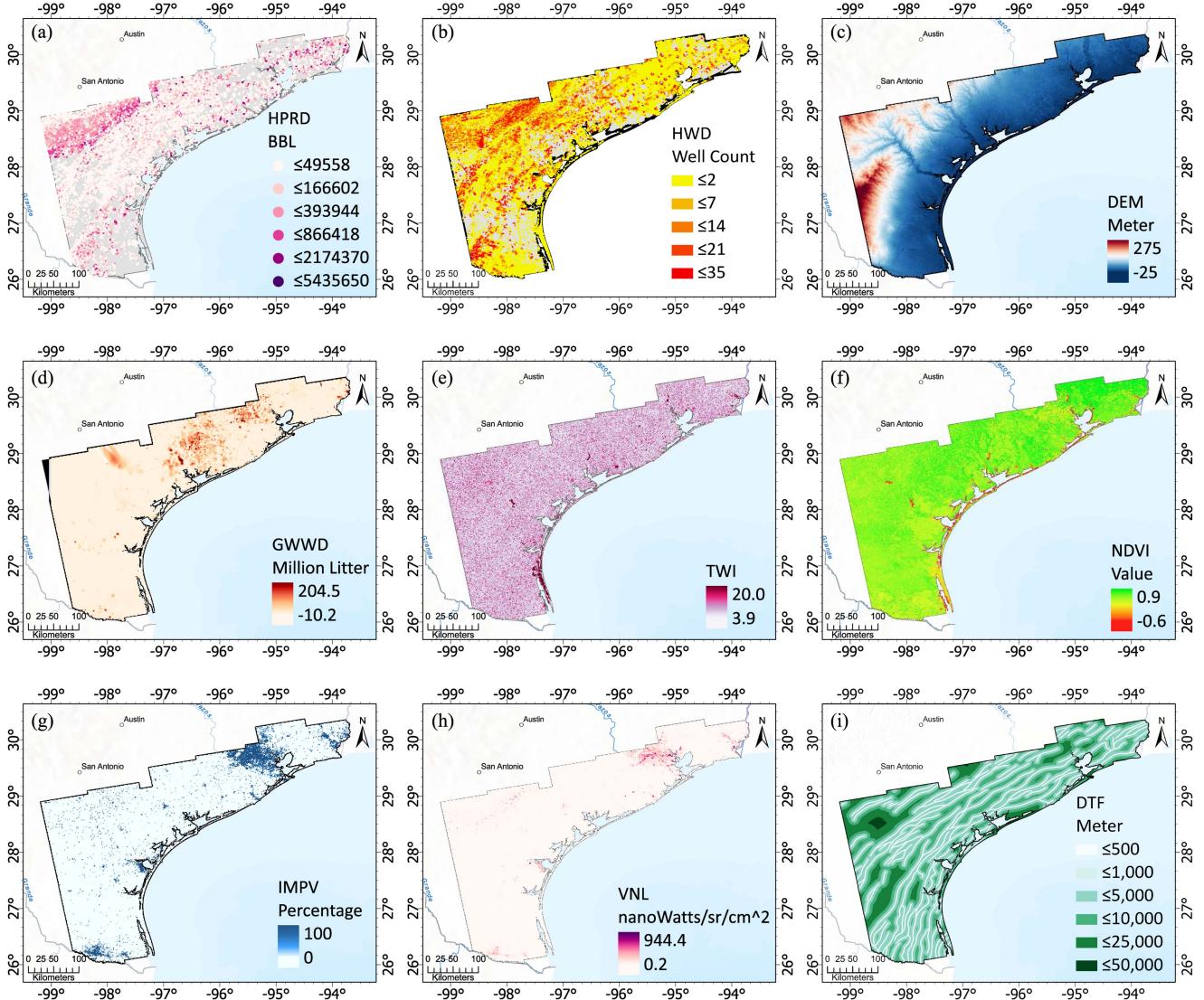


Fig. 2. Features used in the ML methods for analyzing land subsidence attribution at the study area, including (a) HPRD, (b) HWD, (c) DEM, (d) GWWD, (e) TWI, (f) NDVI, (g) IMPV, (h) VNL, and (i) DTF.

count of API numbers linked to a given oil lease in the district. In (1), the coefficient 6 was used to convert MCF units into a BBL equivalent measure, according to the work in [53]. Finally, the monthly time series of HPRD at a specific API number was calculated through (1).

In this study, the HWD and HPRD production were calculated within a 2-km circular buffer zone centered at the 250 m \times 250 m grids to account for potential local influences of hydrocarbon extraction activities. The cumulative HPRD from 1993 to 2022 was calculated for the HPRD variable at each API. As a result, the HPRD and HWD across the 47 counties were spatially cropped to the study area as shown in Fig. 2(a) and (b), corresponding to a total of 108 988 API locations.

B. Topographical Wetness Index

In hydrologic analysis, TWI is an index used to indicate the ability to accumulate water related to local topographical

conditions. This hydrologic feature was included in this study to examine its potential influence on subsidence. TWI is calculated through [54]

$$\text{TWI} = \ln \left(\frac{\alpha}{\tan \beta} \right) \quad (2)$$

where α is the drained area per unit contour length, and β is the local slope angle. Specifically, TWI was computed using the DEM data in this study by abiding by the method proposed in [55].

IV. METHOD

As depicted in Fig. 3, the study approach comprised the following three key stages:

- 1) estimating a land subsidence map utilizing InSAR through the integration of SAR and GNSS data;

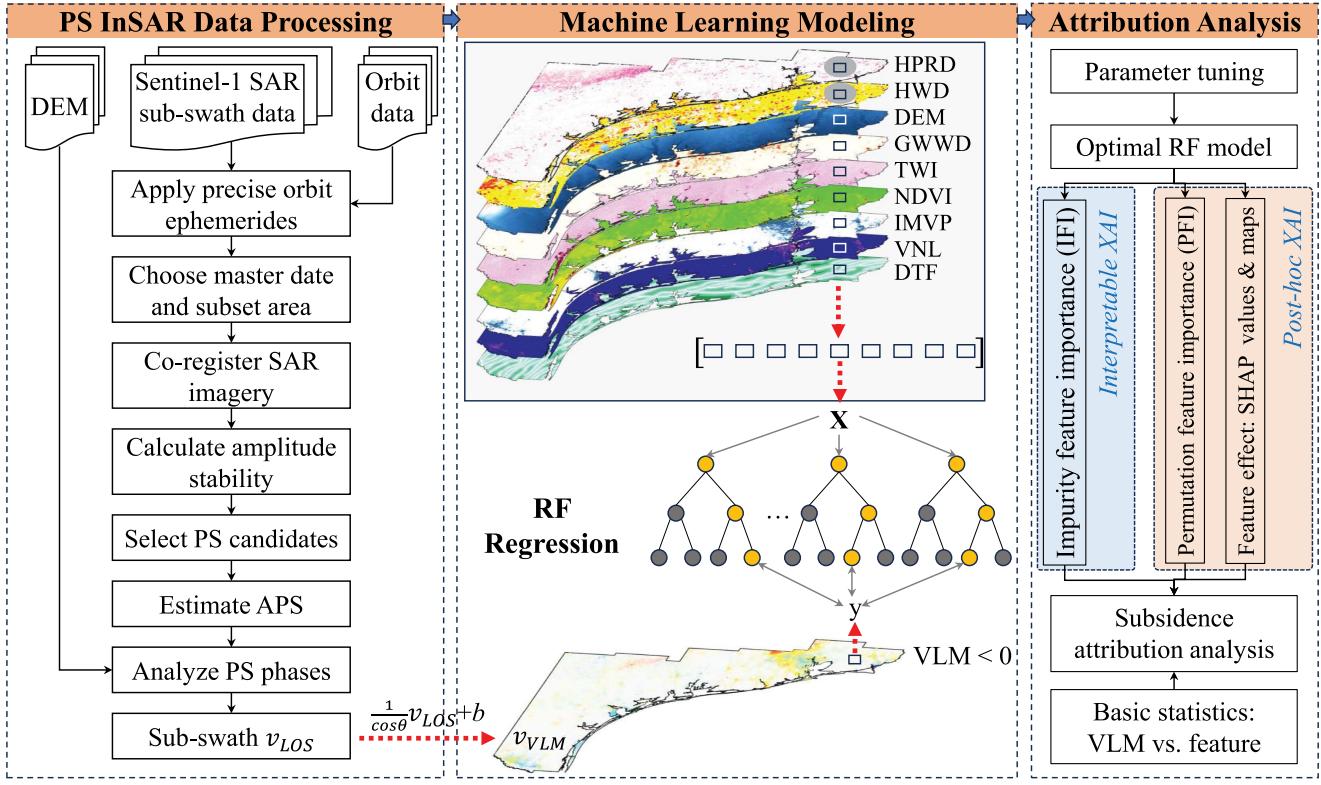


Fig. 3. Research workflow.

- 2) establishing a model that correlates observed land subsidence with potentially contributing features using the ML model;
- 3) performing attribution analysis of subsidence based on feature-importance and feature-effect XAI methods.

A. InSAR Processing

For each subswath dataset within the four sets of Sentinel-1 SAR images, PS InSAR analysis was applied to derive land deformation results along the LOS direction. The essential steps of the PS InSAR processing are detailed in Fig. 3. The PS InSAR processing routines were performed using the SARPROZ software [56]. The processing workflow involved several pre-processing steps including image coregistration, data coverage subsetting, and the selection of a master image. Subsequently, amplitude stability was computed for each pixel. Amplitude stability serves as an indicator of the potential phase-stable targets over time and is defined as $1 - \frac{\sigma_A}{m_A}$, where σ_A and m_A represent the standard deviation and mean statistics of the amplitude time series at a specific pixel [56]. Pixels exhibiting high-amplitude stability were initially identified as the PS candidates. These selected PS candidates were then used to establish a spatial network aimed at isolating the phase contributions associated with atmospheric effects, referred to as the APS. APS manifests as an overlay on interferograms due to its high temporal variability and low spatial variability characteristics in atmospheric effects [14]. Following APS estimation, land deformation was computed on a per-pixel basis by removing

unrelated phase components, primarily stemming from viewing geometry, topography, and the estimated APS [57], [58]. Phase components induced by viewing geometry and topography were corrected using the SAR orbit data and external DEM data [57].

In each subswath dataset, the estimated LOS InSAR velocity, denoted as v_{LOS} , was converted into the vertical direction, represented as v_{VLM} . This conversion is achieved using the formula $v_{VLM} = v_{LOS} \cdot (\cos \theta)^{-1}$ [59], with the contributions in both the North–South and East–West directions neglected because of Sentinel-1’s near-polar orbit motion and the utilization of SAR data from only ascending orbit. The parameter θ represents the average-looking angle of the Sentinel-1 SAR sensor and its values were determined to be 33.9°, 39.3°, and 43.9° for subswath datasets 1, 2, and 3, respectively. This study introduced an additional parameter, denoted as b , to calibrate self-consistent InSAR results obtained from individual subswath datasets. This parameter was determined by comparing the differences between converted InSAR vertical velocities and those derived from two carefully selected cGNSS stations within each subswath coverage of Fig. 3 [27]. Detailed information about PS-InSAR data processing, including the selection of reference points and cGNSS stations for calibration, can be found in [27].

B. ML Modeling

The feature values for HPRD, HWD, and DTF were calculated using 250 m × 250 m grid cells, as explained in Section III. Additionally, the average values were computed

for the 250-m-resolution grids regarding the remaining features and the VLM. The grid-level VLM values are denoted as $\mathbf{y} = [y_1, y_2, \dots, y_i, \dots, y_m]^T$, where m is the sample size of grid cells. For each individual grid as indexed by i , values of nine variables for subsidence attribution analysis were extracted and combined into a vector $\mathbf{x}_i = [x^1, x^2, \dots, x^j, \dots, x^9]$. This process resulted in the creation of the feature matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m]^T$. Subsequently, the ML approach was employed to conduct regression between \mathbf{y} and \mathbf{X} , by satisfying $\mathbf{y} = f(\mathbf{X})$, where $f(\cdot)$ represents the regression function.

The ML modeling was performed using the RF algorithm. This is an ensemble technique consisting of decision trees, which allows for the aggregation of decisions from multiple tree predictors [60]. During the training phase, randomly selecting subsets from both samples and features enhances model diversity. RF has been used to achieve high-performance models in many practical applications [61]. Another advantage of using RF is that it demands minimal data preparation. For example, there is no need for data scaling due to its inheritance from decision trees [61]. Additionally, RF offers the advantage of providing feature importance scores, which indicates the extent to which a specific feature influences the output of the RF model, on average.

C. ML Model Performance

The R -square metric, denoted as R^2 , was utilized for evaluating the RF model during both parameter tuning and result validation. Its definition is provided in [62]

$$R^2 = 1 - \frac{\sum_{i=1}^m [y_i - f(\mathbf{x}_i)]^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (3)$$

where \bar{y} is the mean value of \mathbf{y} .

D. Explainable Artificial Intelligence

In this study, two feature importance methods and one feature effect method were used. Feature importance methods include PFI [60] and IFI [60], [63]. For feature effect, on the other hand, SHAP were used to explain the model's decisions for each sample [30], [64]. The IFI is the interpretable XAI method while both PFI and SHAP are the post-hoc XAI methods. Fig. 4 illustrates the relation of the XAI methods adopted in the study.

1) *Feature Importance Estimated With RF*: The IFI method exploits the interpretable nature of RFs to rank features, as a byproduct of the trained RF model, using the importance scores of the accumulated impurity decrease across all nodes of the trees [60], [61], [63]. The PFI is defined as the decrease of a model score when a single feature value is randomly shuffled [60]. The impurity is usually the variance in the case of regression trees, and the R^2 score is adopted for estimating the PFI. Both the IFI and PFI were utilized in the study to estimate the global feature importance.

2) *Feature Effect Estimated With Kernel SHAP*: Drawing inspiration from game theory to quantify the fair contribution of each feature to a prediction, SHAP measures the impact of

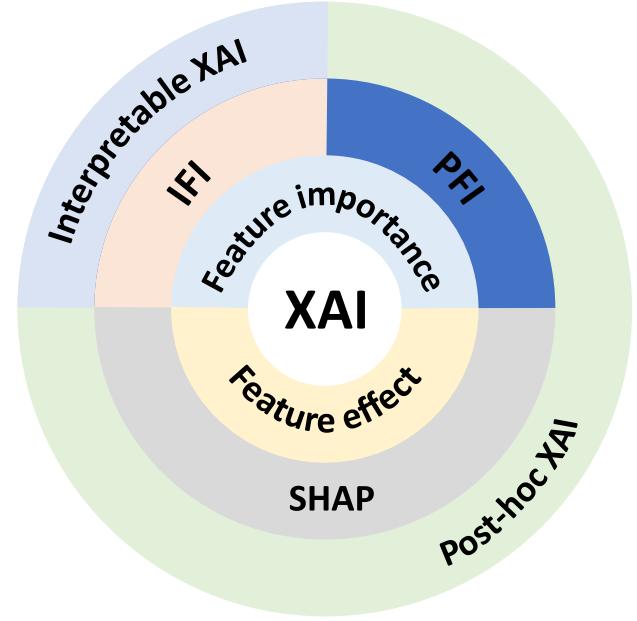


Fig. 4. Relations among different methods related to XAI adopted in this study. For example, the PFI is a post-hoc XAI method to estimate feature importance.

a specific feature value by comparing it to a scenario where the feature value is replaced with baseline values [64]. Under properties of local accuracy, missingness, and consistency, (4) shows the Shapley values, $\phi_i(f, x)$, in the cooperative game theory to calculate the feature effects after removing feature i [65], [66]

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} \left[f_x(z') - f_x(z' \setminus i) \right] \quad (4)$$

where x' is the simplified input of x to approximate the original model $f(\cdot)$. The vector $z' \in \{0, 1\}^M$ of length M controls which features are present in the data using 1 s for present features and 0 s for missing features. $z' \setminus i$ means the feature subset with feature i removed and $|z'|$ is the number of the nonzero entries in z' .

The Kernel SHAP method was utilized to compute Shapley values. This involved optimizing a squared loss function specified in (5), akin to the approach used in the local interpretable model-agnostic explanations algorithm [66], [67]. Algorithm 1 presents a method for determining the impact of individual features on the predicted value for an instance x [68]

$$L(f) = \sum_{z' \in Z} \left\{ f(h_x(z')) - \left[\phi_0 + \sum_{i=1}^M \phi_i z'_i \right] \right\} \cdot \frac{M - 1}{\binom{M}{|z'|} |z'| (M - |z'|)}. \quad (5)$$

Further details on the kernel SHAP method can be explored in [66], [68], and [69]. The kernel SHAP analysis was performed utilizing the SHAP Python software suite [69]. And the baseline values utilized in the $h_x(z'_i)$ function were the medians of training samples.

Algorithm 1: Kernel SHAP to Estimate Shapley Values.

- 1: Sample a series of $z' \in \{0, 1\}^M$.
- 2: Map z' to the original feature space using $h_x(z')$:

$$h_x(z'_i) = \begin{cases} \text{feature values of } x, & \text{if } z'_i = 1 \\ \text{baseline values,} & \text{otherwise.} \end{cases}$$

- 3: Predict for $f(h_x(z'))$.
- 4: Compute weight for z' with the SHAP kernel

$$\pi_{x'}(z') = \frac{M-1}{\binom{M}{|z'|}|z'|(M-|z'|)}.$$

- 5: Fit weighted linear model of $g(z')$ by optimizing the loss function in (5)

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i.$$

- 6: Return Shapley values ϕ_i .

V. RESULTS

A. VLM Estimate and Validation

Land deformation results were acquired for each subswath using the PS InSAR processing steps, as depicted in the left block of Fig. 3. The LOS subswath results were then converted to the vertical direction and merged into one subsidence map across the entire study area. The subsidence results of the PS points were then aggregated within 250 m × 250 m grid cells. As shown in Fig. 5, the mean subsidence velocities were calculated along the Texas Coast in the form of a network of 173 064 grid cells. For validation, InSAR-derived subsidence velocities were compared with that obtained from 115 cGNSS stations. Only the InSAR grid cells within which the cGNSS stations fell were used for the validation. The resulting R^2 value of 0.54 was documented between InSAR grid cells and cGNSS stations regarding their subsidence velocities.

A distribution plot, illustrated in Fig. 5, displays the velocity differences between using GNSS and InSAR techniques. It is worth emphasizing that this distribution was constructed based on velocity difference values that were rounded to the nearest whole numbers. The distribution plot demonstrates an overall favorable agreement between InSAR and GNSS measurements. According to the distribution histogram in Fig. 5, at a total of 75 out of 115 cGNSS stations (approximately 65%), the differences in subsidence rates between cGNSS and InSAR lie within the range of ±1.5 mm/yr, which corroborated InSARs reported accuracy of submillimeter to millimeter per year [14]. Larger differences may arise when comparing InSAR and cGNSS velocities due to several factors, including the distance between their observation locations and differences in their observation directions (i.e., vertical versus LOS). These factors could explain the differences between cGNSS and InSAR at additional 24 cGNSS stations (approximately 21%), where the differences range between ±1.5 mm/yr and ±2.5 mm/yr. Two

conspicuous outliers can be observed from the histogram plot in Fig. 5, corresponding to the cGNSS stations of TXAI (27.73°N, -98.07°W) and TXRF (28.30°N, -97.27°W). These two outliers are attributed to unreliable trend estimates at these two cGNSS stations, stemming from a short temporal observation window, nonlinear processes, and a high level of observational uncertainty [27]. To sum up, these statistics, along with the histogram distribution depicted in Fig. 5, accord well with the expected InSAR accuracy from prior studies [70], [71], [72] and demonstrate the validity of the InSAR results obtained in this study.

B. RF Modeling

Fig. 6 showcases scatter plots, each visualizing the correlation between an individual feature and the VLM across the study area. Besides, Fig. 6 suggests that none of these features exhibit strong individual correlations with land deformation, highlighting the intricate challenge of identifying true drivers and/or correlates of subsidence. This reinforces our study's motivation to leverage ML techniques to address the complex relationships between the VLM and different features.

A total of 143 787 grid cells, where the VLM rate was less than zero, were used for analyzing the relationships between subsidence and potential features. The retained 143 787 grid cells were treated as independent samples for training and validating the RF model, under the regression scenario as mentioned in Section IV-B. A grid search scheme was used to optimize the RF model hyperparameters. The parameter sets included in the search are 1) $\text{tree_size} \in \{50, 100, 300, 500\}$, and 2) the size of training samples, referred to as $\text{training_size} \in \{10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%\}$. Bootstrapping was used for sampling the training dataset in both the sample and feature dimensions. The remaining samples were used to estimate the RF model performance and the R^2 across different parameter combinations is illustrated in Fig. 7(a). It can be observed that the parameter of training size plays a more important role than the tree size in improving R^2 .

Based on the results of the grid search between “`training_size`” and “`tree_size`,” 80% data were sampled for training ML models. Then further grid search tests were conducted by focusing on parameters such as “`max_depth`” and “`max_features`” to introduce more randomness but insignificant improvements of R^2 were documented. Therefore, for both interpretability and computational efficiency purposes, a simpler RF model with “`n_estimators`” set to 300 and all other parameters left at their defaults [73] was trained using 80% data. In Fig. 7(b), a comparison is presented between the VLM values of 28 758 validation samples and their corresponding predictions generated by the RF model. The scatter plot visually demonstrates a good correlation, with R^2 equal to 0.56, indicating a robust model performance. It is notable that this study heavily prioritized comprehending feature interactions and contributions while employing the ML model. Predicting subsidence velocity at unobserved locations and times exceeds the scope of this study due to challenges such as considerable spatiotemporal subsidence variability.

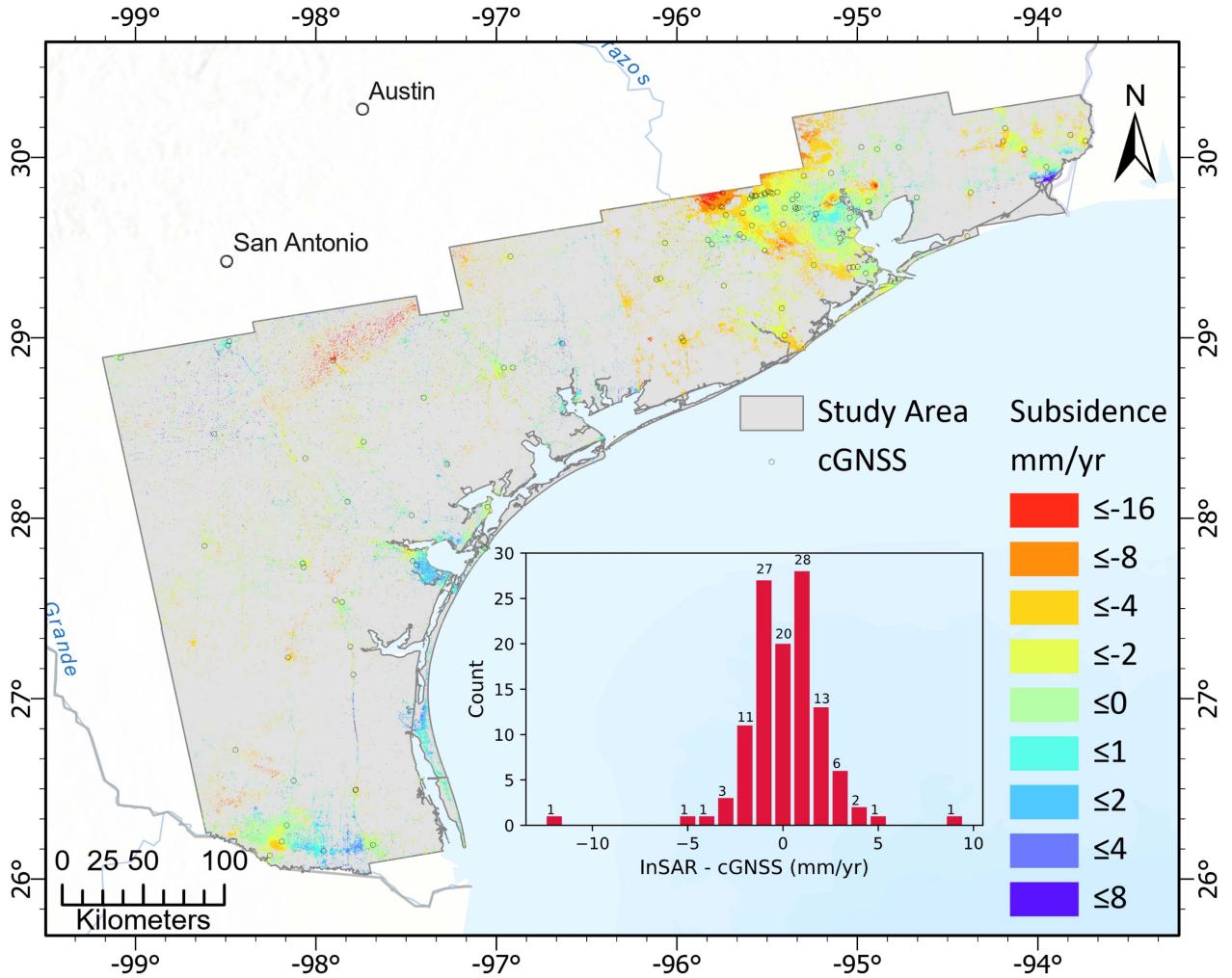


Fig. 5. Land subsidence map represented in 250 m \times 250 m grid cells. The lower right subplot visually portrays the distribution of subsidence velocity differences between results obtained using InSAR versus cGNSS, which have been rounded to the nearest whole numbers.

C. Feature Importance and Attribution Analysis

Feature importance results of IFI and PFI, computed using the training set, are displayed in Fig. 8(a) and (b) while the SHAP values derived from the validation set are presented in Fig. 8(c). For a given output, the SHAP value of each feature represents the magnitude to which that feature pushed the model's decision away from a given baseline value. Positive SHAP values indicate that the feature increased the output relative to the baseline, and negative SHAP values indicate a decrease. In this work, the baseline values were the median values for the VLM predictor dataset. Therefore, the SHAP values represent how the features move a specific decision away from the model's mean decision. Overall, the XAI methods agreed on the ranking of global feature importance, as shown in Fig. 8. However, some differences were documented among certain features between the IFI, PFI, and SHAP methods due to different underlying algorithms. Based on the results obtained from individual feature scatter plots (see Fig. 6) and XAI (see Fig. 8), a detailed analysis delved into the following aspects to comprehend the connections between subsidence and individual features.

- 1) First, XAI results of both feature importance and effect consistently identified DEM and HPRD as the primary features influencing the modeling of VLM. The significance of DEM can be rationalized by the observations [see Fig. 6(c)] that the reddish cluster tends to exhibit higher subsidence velocity as elevation values increase. This trend could potentially be attributed to the subsidence due to heightened levels of anthropogenic activities occurring farther away from the coastlines [27], where elevations are approximately at or below sea level. Also, the importance of the HPRD as revealed by feature importance and SHAP methods can also be observed from the spatial coincidence between the HPRD and the VLM at some oil fields, e.g., the HPRD and VLM hotspot near San Antonio, TX, as observed from Figs. 2(a) and 5, respectively.
- 2) The feature of HWD ranked differently in IFI results compared with that in PFI and SHAP. The potential reason for this phenomenon is the cardinal nature of the well density, ranging from 0 to 45. The IFI method can be sensitive to features with high cardinality with many unique

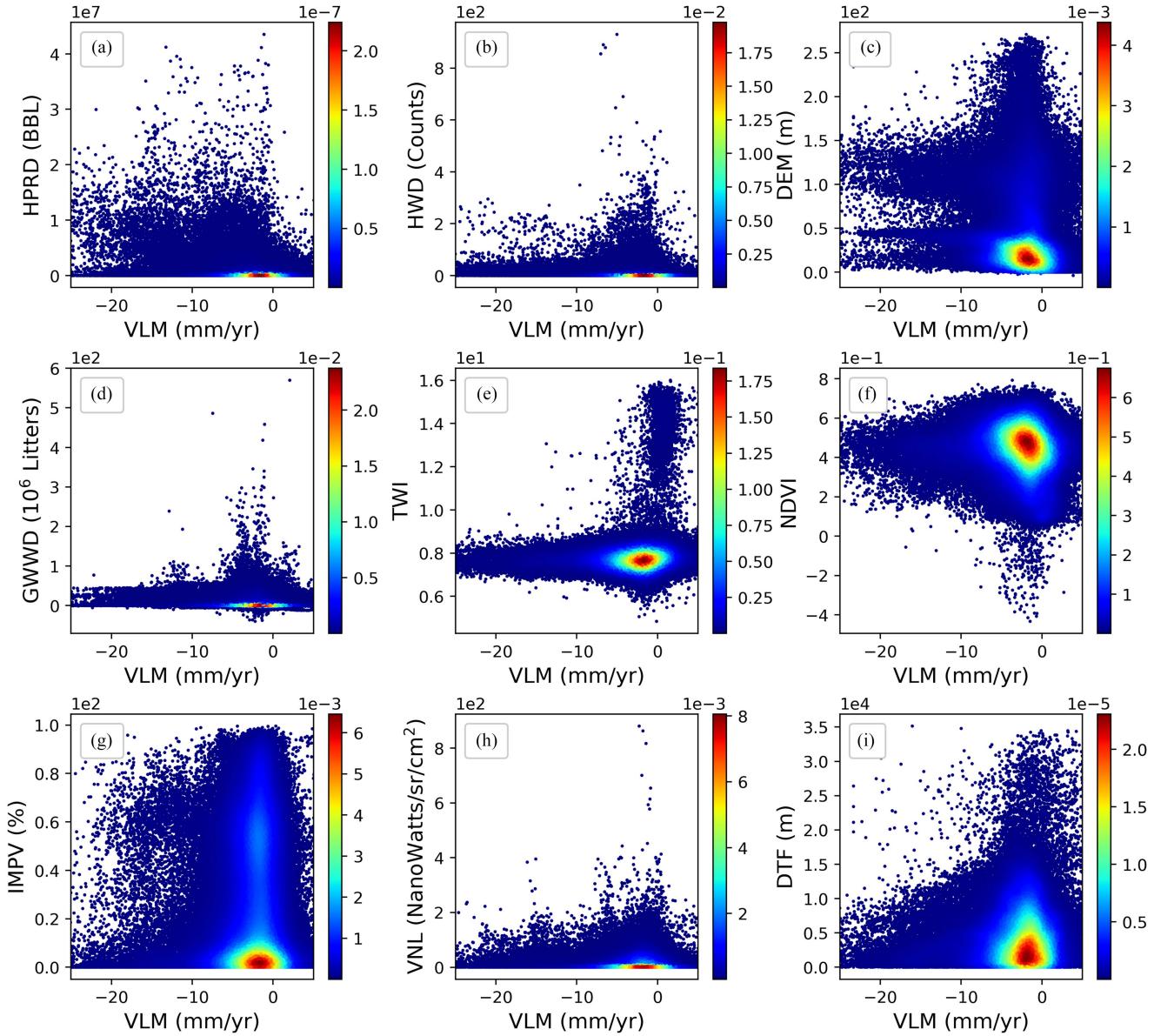


Fig. 6. Scatter plots depicting the relationship between VLM and individual features of (a) HPRD, (b) HWD, (c) DEM, (d) GWWD, (e) TWI, (f) NDVI, (g) IMPV, (h) VNL, and (i) DTF. Reddish hues in each subplot represent high sample density estimated with Gaussian kernel density estimation.

values [74], which might lead to misleading results when including the HWD feature. However, both PFI and SHAP suggest that the HWD was the important feature following the HPRD.

- 3) Results of both feature importance and feature effect methods indicated that VNL tended to be another influential feature affecting the VLM. This may be explained by the overlap between VLM and nighttime light emissions originating from activities such as hydrocarbon extraction (e.g., gas flares and drilling operations), urban areas with a high likelihood of groundwater demand, or less evapotranspiration during dry seasons [27], [37], [38].
- 4) The contribution of GWWD tended to be less important than hydrocarbon extraction activities in the global feature importance of both IFI and PFI. The reasons

may be twofold: 1) While a significant amount of GWWD occurred, particularly in areas outside the Houston metropolitan, no apparent subsidence was documented, which may obscure the contribution assessment performed by feature importance analysis, and 2) land subsidence was attributed to GWWD locally near the Houston-Galveston area. Thus, the sample-level attribution using SHAP maps may be more beneficial for attribution purposes at localized areas.

- 5) The remaining features generally exhibited a more limited impact on VLM, as indicated by both feature importance/effect assessments. For instance, high-VLM grids tended to cluster in proximity to faults [see Fig. 6(i)], and their NDVI values likely hovered around 0.5 [see Fig. 6(f)]. Nevertheless, the relationships observed were

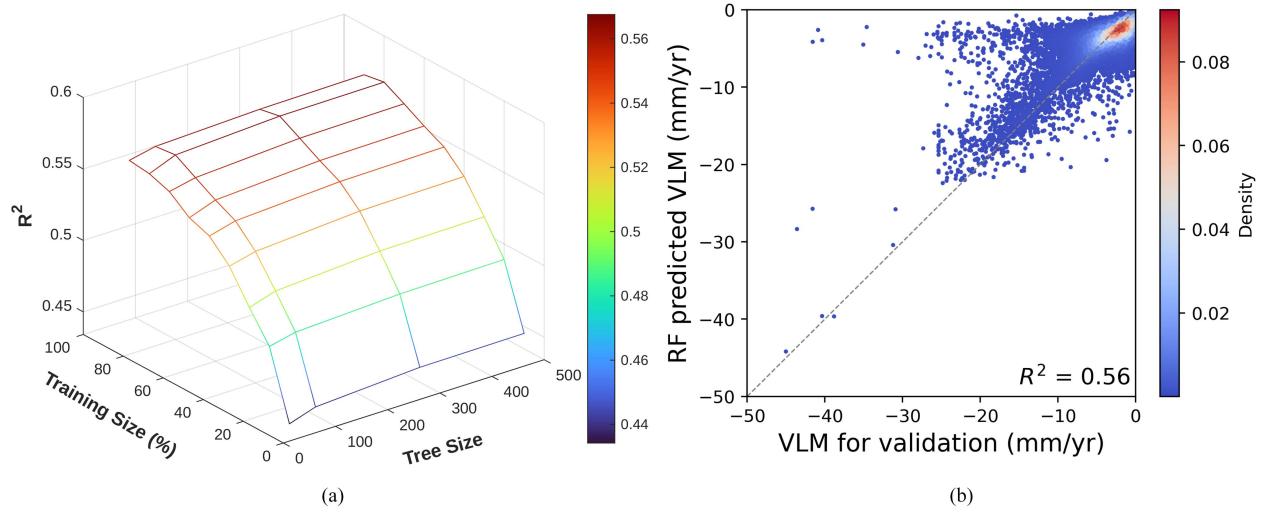


Fig. 7. ML model training and performance validation. (a) Parameter tuning of the RF model. (b) Scatter plot for RF model validation.

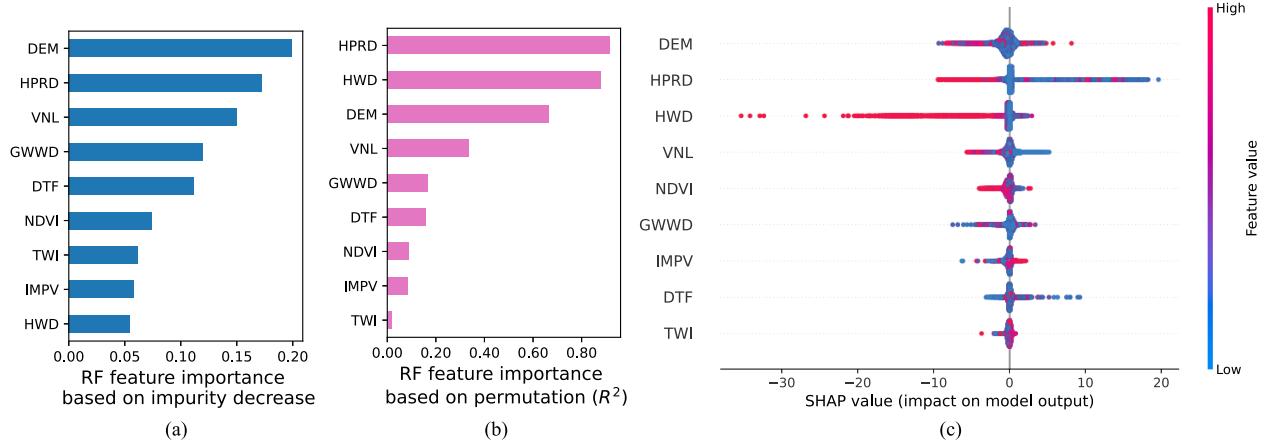


Fig. 8. Feature importance estimated with RF and SHAP. (a) RF feature importance based on impurity decrease. (b) RF feature importance based on feature permutation using score R^2 . (c) SHAP values using the validation set in the RF model.

neither consistent nor definitive in terms of their contribution to VLM. Besides, the limited contribution of TWI may be a result of the overall flat nature of the study area with inconspicuous DEM relief, except for the western region of the study area.

In addition to assessing subsidence drivers or correlates on an individual feature basis as aforementioned, a more automated approach is to identify the features that correspond to the high-extent SHAP values for each grid cell. This can result in a map indicating the top contributor associated with land subsidence. In this study, the features with the highest absolute SHAP value were considered to have a significant impact on the VLM. Fig. 9(a) and (b) depicts the primary and secondary contributing features to subsidence for each grid cell.

To facilitate quantitative analysis by delving deeper into the SHAP map findings, five key subsidence zones were visually identified [see Fig. 9(c)] with cyan polygons. Previous investigations have linked localized subsidence near Polygons 1, 2,

and 4 to hydrocarbon extraction activities [23], [24], [27], [34] (see Fig. 10). Polygon 3 was chosen as a prime candidate due to documented subsidence in the vicinity of Katy, TX, which was potentially attributed to GWWD [23], [26], [34], [75]. Polygon 5 serves as an example that showcases the valuable insights we can glean from SHAP results about the contributing features. The individual-feature effect was calculated as mean SHAP values across samples within each polygon (see Fig. 10). Findings in each of the five polygons are summarized as follows.

- 1) Figs. 9(a) and (b) and 10(a) reveal that HPRD and HWD had a significant influence on the model's predictions within Polygon 1. This aligned with the drivers attributed in previous studies [27], [34].
- 2) Polygon 2 reveals a compelling interplay among features, with both DEM and HPRD significantly influencing the model's predictions [see Fig. 10(b)]. HPRD's role as a key driver in Polygon 2 aligns well with its proximity to the Eagle Ford Shale [see Fig. 10(c)]. DEMs contribution [see

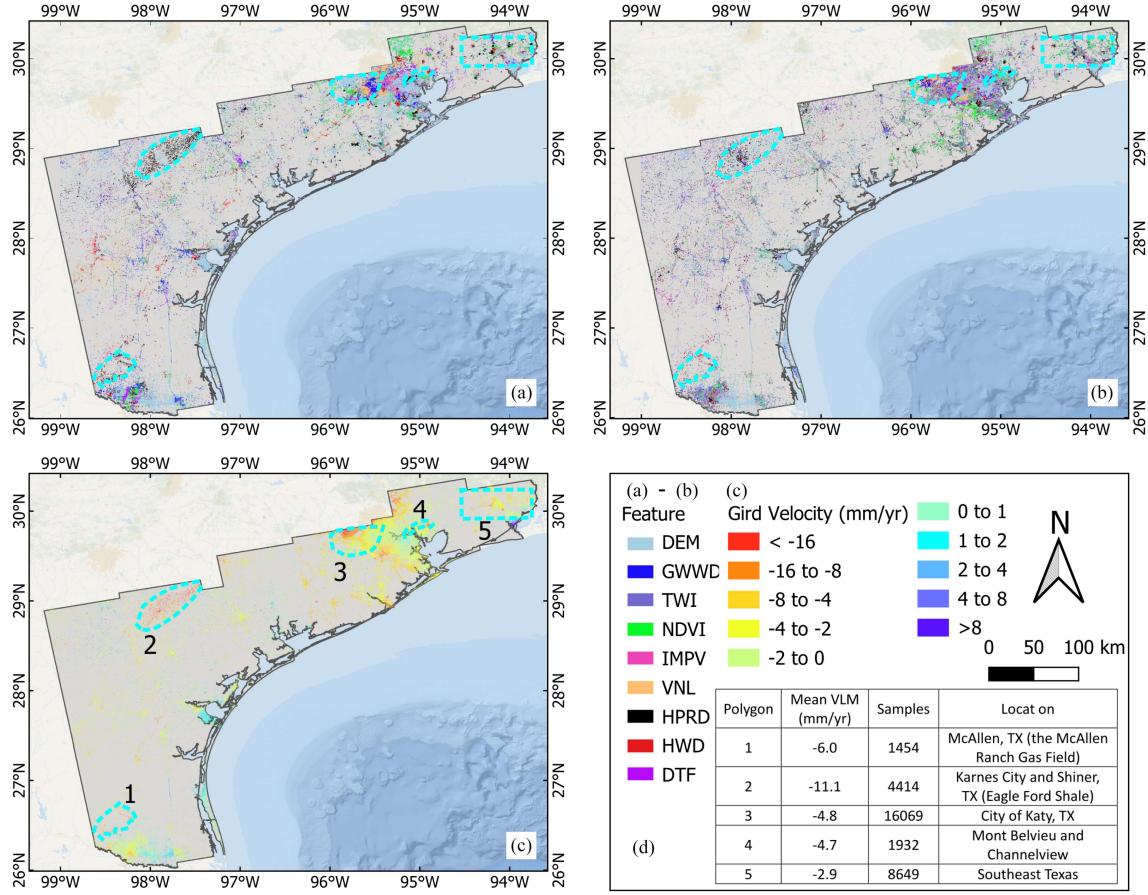


Fig. 9. SHAP maps and comparison with the VLM. (a) Attribution map illustrating the primary SHAP values. (b) Attribution map displaying secondary SHAP values. (c) InSAR-derived VLM map with the cyan-colored dashed polygons highlighting selected VLM hotspots. (d) Legend with the table showing details associated with each polygon.

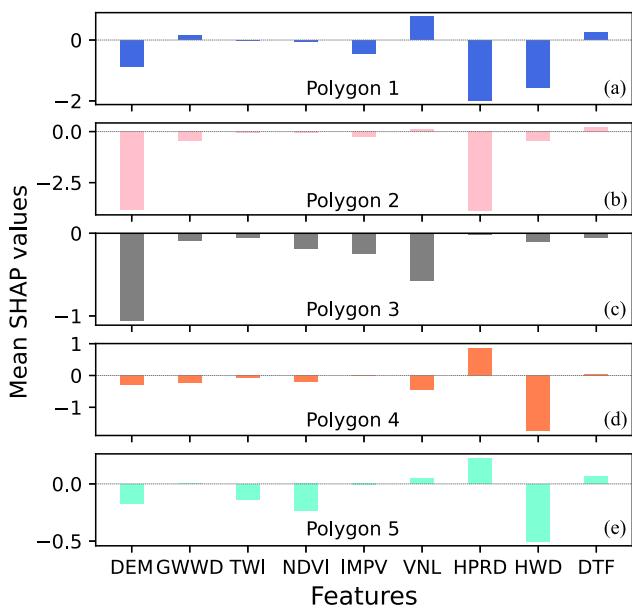


Fig. 10. Average SHAP values representing feature contribution to the model's prediction, across samples within each of five selected polygons as depicted in Fig. 9, visualized in subplots (a)–(e).

Fig. 10(b)], possibly related to the high elevation gradient in this region, warrants further investigation.

- Polygon 3, an area including part of the Houston metropolitan region, has emerged as a documented epicenter of subsidence linked to GWWD, as supported by previous studies [1], [23], [75]. While GWWD emerges as the dominant driver at certain locations [see Fig. 9(a) and (b)], the SHAP values in Fig. 10(c) reveal a different story, with DEM and VNL taking center stage as key contributors. The discrepancy could be attributed, at least in part, to the potential limitations of interpolated GWWD data in capturing localized changes. Expanding the study area to include the entire local subsidence hotspot may enhance our ability to analyze the importance of GWWD. However, this falls outside the scope of our study, which was focused on interpreting subsidence near the Texas coastlines.
- Polygon 4, encompassing Mont Belvieu and Channelview, TX, exhibits intriguing “bowel-shaped” subsidence patterns, a spatial signature consistent with the documented linkage between subsidence and hydrocarbon extraction activities in the region [23], [27]. This pattern aligns with SHAP attribution results in Fig. 10(d).

5) Fig. 10(e) leans toward hydrocarbon extraction as the dominant driver of subsidence across Polygon 5, echoing similar findings in prior studies [27]. However, the frequent spatial variations in subsidence within this area necessitate further investigation to refine this attribution and potentially uncover additional contributing factors.

It should be noted that Fig. 10 only shows the statistics of average SHAP values over the entire individual polygons. Looking into sample-level SHAP explanations is also important as detailed in SHAP maps [see Fig. 9(a) and (b)]. Overall, SHAP values and maps provided reasonable attributions among selected polygons in this study, providing a valuable tool for identifying potential causes and features related to land subsidence.

VI. DISCUSSION

The study assumed that the observed subsidence could be explained and modeled using certain potential features within a data-driven modeling framework. However, it is important to note that the actual causes of subsidence are more intricate. First, several natural processes can contribute to subsidence, such as drought, tectonic movements, and GIA [76], [77]. For example, prior research has shown a strong correlation between peatland subsidence and drought conditions driven by climate factors [37], and the exceptional drought between 2011 and 2016 has caused widespread impacts in Texas [38], [78]. Moreover, the occurrence and magnitude of subsidence are contingent on subsurface geological settings, including factors like the compressibility of stratification layers. These complexities may help explain the challenges encountered in achieving a perfect model fit for VLM using RF in this study, as evidenced by the outliers distributed away from the 1:1 line in Fig. 7(b).

It should be noted that the ML model may be affected, to some extent, by the utilized data and parameter settings. First, while this study investigated the long-term impacts of underground fluid extraction (i.e., GWWD, HPRD, and HWD) on land subsidence, the temporal difference in availability between these datasets and the SAR data may introduce variations in the results. Second, the adopted RF model with mostly default parameters achieved an R^2 value close to that of the optimally tuned model. However, such a parameter choice may introduce slight variations in the results. Furthermore, the XAI results in this study necessitate further investigation and validation to fully understand their implications. One specific example is the hypothesis that the importance of DEM in the RF model arises from increased anthropogenic activities away from the shoreline (i.e., higher elevation values) [27]. This hypothesis requires further exploration through the integration of additional data sources. Additionally, the SHAP map results need practical validation at the sample level or in localized areas to confirm their findings and assess their applicability to specific scenarios.

Although this study demonstrates the potential of utilizing ML and XAI for interpreting and attributing land subsidence, it is considered essential to combine refined InSAR data, cutting-edge ML techniques, and additional informative features. This is crucial for mapping a high-resolution subsidence map and

deepening our understanding in the attribution analysis. Further development is necessary in several key aspects, which are as follows.

- 1) Integrating data from both ascending and descending tracks of SAR data from multiple missions can lead to more accurate VLM estimates with better spatial coverage.
- 2) Advanced InSAR processing methods that consider the low coherence nature near the Texas coastlines, influenced by factors like vegetation coverage and land–water interaction dynamics, should be studied [79]. It is supposed to enhance InSAR results by expanding spatial coverage and mitigating phase unwrapping errors, which may result in the striping effect between subswath images, particularly over densely vegetated areas [27].
- 3) It is believed that it is necessary to introduce additional features to enhance the model’s performance. In particular, prioritizing efforts to collect subsurface observations is of vital importance, including data related to compressibility, clay thickness, and pressure changes associated with fluid extraction [22].
- 4) It would be advantageous to leverage the capabilities of other ML models, such as neural networks (NNs) to explore the complex relationships between subsidence and potential attributes and to effectively integrate categorical data like geological maps and land use types. This would enable the inclusion of a broader range of features in the model, potentially leading to improved performance.
- 5) More complex ML approaches such as CNN may yield additional advantages in regard to providing attribution results by analyzing patterns existing in both the feature and VLM layers from an image perspective, rather than individual samplewise attribution as conducted in this study. To facilitate image analysis using CNN, it is essential to utilize SBAS InSAR results and employ necessary spatial interpolation to ensure the continuity of subsidence results across spatial areas.
- 6) Creating a more detailed and large-scale groundwater change map is believed to be helpful for comprehending the causes of subsidence. This may involve the separation of GWWDs related to hydrocarbon extraction from those associated with irrigation and municipal water demands. Moreover, obtaining spatially detailed GWWD or groundwater level data is crucial for building reliable ML models and identifying accurate contributors.

VII. CONCLUSION

This study estimated land subsidence across 47 coastal counties in Texas using the PS InSAR technique, and the VLM results were calibrated and validated with cGNSS stations over the study area. The differences in VLM rates between ground-truth cGNSS and co-located InSAR results were found to be within ± 1.5 mm/yr and ± 2.5 mm/yr across 65% and 86% of all validation data used, respectively. This firmly confirmed the validity of the obtained InSAR results. Geospatial data of nine different features potentially contributing to subsidence were accessed, ranging from natural terrain variations (i.e.,

DEM, NDVI, TWI, and DTF) to anthropogenic activities (i.e., HPRD, HWD, GWWD, VNL, and IMPV). The estimated VLM was modeled with the nine features under a data-driven RF framework, and the R^2 value of the optimal RF model reached 0.56, demonstrating the strong performance of the RF model. The subsidence attribution was further analyzed based on XAI provided by RF-based feature importance and SHAP-based feature effect results. Both the RF-based and SHAP-related feature importance results indicated the significance of DEM in explaining VLM variations, as well as the overall contribution of subsidence from anthropogenic activities, such as hydrocarbon extraction and GWWD. Additionally, the subsidence attribution map, based on individual sample-level SHAP analysis, identified reasonable features at various VLM hotspots within the study area. Overall, this study showcased the feasibility of identifying features related to subsidence using ML and XAI approaches. Future work will involve improving the performance of InSAR estimates, incorporating additional potential features in attribution analysis, and employing advanced ML models, such as CNN.

ACKNOWLEDGMENT

The authors also thank Jian Qing for obtaining the lease production data of oil and gas wells from the RRC.

REFERENCES

- [1] D. L. Galloway, D. R. Jones, and S. E. Ingebritsen, *Land Subsidence in the United States*, vol. 1182. Reston, VA, USA: US Geological Surv., 1999.
- [2] M. Shirzaei, J. Freymueller, T. E. Törnqvist, D. L. Galloway, T. Dura, and P. S. Minderhoud, “Measuring, modelling and projecting coastal land subsidence,” *Nature Rev. Earth Environ.*, vol. 2, no. 1, pp. 40–58, 2021.
- [3] D. Sarah et al., “Mapping the environmental impacts from land subsidence hazard in Pekalongan city and its correlation with the subsurface condition,” *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1201, no. 1, 2023, Art. no. 012044.
- [4] L. C. Fowler, “Economic consequences of land surface subsidence,” *J. Irrigation Drainage Division*, vol. 107, no. 2, pp. 151–159, 1981.
- [5] H. Z. Abidin, H. Andreas, I. Gumilar, T. P. Sidiq, and M. Gamal, “Environmental impacts of land subsidence in urban areas of Indonesia,” in *Proc. FIG Work. Week*, 2015, pp. 1–12.
- [6] S. Y. Avoroyo et al., “A scoping review of coastal vulnerability, subsidence and sea level rise in Ghana: Assessments, knowledge gaps and management implications,” *Quaternary Sci. Adv.*, vol. 12, 2023, Art. no. 100108.
- [7] X. Qiao, T. Chu, P. Tissot, I. Ali, and M. Ahmed, “Vertical land motion monitored with satellite radar altimetry and tide gauge along the Texas coastline, USA, between 1993 and 2020,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 117, 2023, Art. no. 103222.
- [8] G. Wang, P. Li, Z. Li, C. Liang, and H. Wang, “Coastal subsidence detection and characterization caused by brine mining over the Yellow River Delta using time series InSAR and PCA,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 114, 2022, Art. no. 103077.
- [9] J. Yu and G. Wang, “Introduction to the GNSS geodetic infrastructure in the Gulf of Mexico Region,” *Surv. Rev.*, vol. 49, no. 352, pp. 51–65, 2017.
- [10] X. Qiao, T. Chu, P. Tissot, J. Louis, and I. Ali, “Land subsidence estimation with tide gauge and satellite radar altimetry measurements along the Texas Gulf Coast, USA,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jul. 2022, Art. no. 3007905.
- [11] A. Hooper, D. Bekaert, K. Spaans, and M. Arikhan, “Recent advances in SAR interferometry time series analysis for measuring crustal deformation,” *Tectonophysics*, vol. 514, pp. 1–13, 2012.
- [12] Y. Morishita, “Nationwide urban ground deformation monitoring in Japan using Sentinel-1 LiCSAR products and LiCSBAS,” *Prog. Earth Planet. Sci.*, vol. 8, no. 1, pp. 1–23, 2021.
- [13] D. Festa et al., “Nation-wide mapping and classification of ground deformation phenomena through the spatial clustering of P-SBAS InSAR measurements: Italy case study,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 189, pp. 1–22, 2022.
- [14] A. Ferretti, *Satellite InSAR Data: Reservoir Monitoring From Space (EET 9)*, European Association of Geoscientists & Engineers, 2014. [Online]. Available: <https://www.earthdoc.org/content/books/9789462820036>
- [15] P. Ma, Y. Zheng, Z. Zhang, Z. Wu, and C. Yu, “Building risk monitoring and prediction using integrated multi-temporal InSAR and numerical modeling techniques,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 114, 2022, Art. no. 103076.
- [16] G. Modeste, C. Doubre, and F. Masson, “Time evolution of mining-related residual subsidence monitored over a 24-year period using InSAR in southern Alsace, France,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 102, 2021, Art. no. 102392.
- [17] N. P. Prokopovich, “Genetic classification of land subsidence,” in *Evaluation and Prediction of Subsidence*. Reston, VA, USA: ASCE, 1979, pp. 389–399.
- [18] R. Li et al., “Monitoring surface deformation of permafrost in Wudaoliang Region, Qinghai–Tibet Plateau with ENVISAT ASAR data,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 104, 2021, Art. no. 102527.
- [19] A. Burnol, M. Foumelis, S. Gourdier, J. Deparis, and D. Raucoules, “Monitoring of expansive clays over drought–rewetting cycles using satellite remote sensing,” *Atmosphere*, vol. 12, no. 10, 2021, Art. no. 1262.
- [20] P. L. Whitehouse, “Glacial isostatic adjustment modelling: Historical perspectives, recent advances, and future directions,” *Earth Surf. Dyn.*, vol. 6, no. 2, pp. 401–429, 2018.
- [21] L. Hu et al., “Land subsidence in Beijing and its relationship with geological faults revealed by Sentinel-1 InSAR observations,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 82, 2019, Art. no. 101886.
- [22] G. Gambolati, P. Teatini, and M. Ferronato, “Anthropogenic land subsidence,” *Encycl. Hydrological Sci.*, 2006.
- [23] F. Qu et al., “Mapping ground deformation over Houston–Galveston, Texas using multi-temporal InSAR,” *Remote Sens. Environ.*, vol. 169, pp. 290–306, 2015.
- [24] M. Haley, M. Ahmed, E. Gebremichael, D. Murgulet, and M. Starek, “Land subsidence in the Texas coastal bend: Locations, rates, triggers, and consequences,” *Remote Sens.*, vol. 14, no. 1, 2022, Art. no. 192.
- [25] M. Younas, S. Khan, O. Tirmizi, and Y. Hamed, “Geospatial analytics of driving mechanism of land subsidence in Gulf Coast of Texas, United States,” *Sci. Total Environ.*, vol. 902, 2023, Art. no. 166102.
- [26] S. D. Khan, O. C. A. Gadea, A. T. Alvarado, and O. A. Tirmizi, “Surface deformation analysis of the Houston area using time series interferometry and emerging hot spot analysis,” *Remote Sens.*, vol. 14, no. 15, 2022, Art. no. 3831. [Online]. Available: <https://www.mdpi.com/2072-4292/14/15/3831>
- [27] X. Qiao, T. Chu, P. Tissot, and S. Holland, “Sentinel-1 InSAR-derived land subsidence assessment along the Texas Gulf Coast,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 125, 2023, Art. no. 103544.
- [28] Z. Wu, P. Ma, Y. Zheng, F. Gu, L. Liu, and H. Lin, “Automatic detection and classification of land subsidence in deltaic metropolitan areas using distributed scatterer InSAR and oriented R-CNN,” *Remote Sens. Environ.*, vol. 290, 2023, Art. no. 113545.
- [29] E. M. Kenny, C. Ford, M. Quinn, and M. T. Keane, “Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies,” *Artif. Intell.*, vol. 294, 2021, Art. no. 103459.
- [30] E. Krell, H. Kamangir, W. Collins, S. A. King, and P. Tissot, “Aggregation strategies to improve XAI for geoscience models that use correlated, high-dimensional rasters,” *Environ. Data Sci.*, vol. 2, 2023, Art. no. e45.
- [31] C. M. Gevaert, “Explainable AI for Earth observation: A review including societal and regulatory perspectives,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, 2022, Art. no. 102869.
- [32] A. H. Chowdhury and M. J. Turco, “Geology of the Gulf Coast Aquifer, Texas,” in *Aquifers of the Gulf Coast of Texas: Texas Water Development Board Report*. R. E. Mace, S. C. Davidson, E. S. Angle, and W. F. Mullican, Eds. Austin, TX, USA: Texas Water Development Board, 2006, pp. 23–50.
- [33] R. G. Loucks, “Controls on porosity and permeability of hydrocarbon reservoirs in lower tertiary sandstones along the Texas Gulf Coast,” Bureau Econ. Geol., The Univ. Texas, Austin, Texas, Rep. No. 149, Jan. 1986.
- [34] F. Qu, Z. Lu, J. Kim, and M. J. Turco, “Mapping and characterizing land deformation during 2007–2011 over the Gulf Coast by L-band InSAR,” *Remote Sens. Environ.*, vol. 284, 2023, Art. no. 113342.

[35] W. L. Hakim, M. F. Fadhillah, S. Park, B. Pradhan, J.-S. Won, and C.-W. Lee, "InSAR time-series analysis and susceptibility mapping for land subsidence in Semarang, Indonesia using convolutional neural network and support vector regression," *Remote Sens. Environ.*, vol. 287, 2023, Art. no. 113453.

[36] O. Rahmati, A. Golkarian, T. Biggs, S. Keesstra, F. Mohammadi, and I. N. Daliakopoulos, "Land subsidence hazard modeling: Machine learning to identify predictors and the role of human activities," *J. Environ. Manage.*, vol. 236, pp. 466–480, 2019.

[37] M. Y. N. Khakim, A. A. Bama, I. Yustian, P. Poerwono, T. Tsuji, and T. Matsuoka, "Peatland subsidence and vegetation cover degradation as impacts of the 2015 el niño event revealed by Sentinel-1 A SAR data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 84, 2020, Art. no. 101953.

[38] B. R. Scanlon, I. Duncan, and R. C. Reedy, "Drought and the water–energy Nexus in Texas," *Environ. Res. Lett.*, vol. 8, no. 4, 2013, Art. no. 045033.

[39] T. J. Kearns, G. Wang, Y. Bao, J. Jiang, and D. Lee, "Current land subsidence and groundwater level changes in the Houston Metropolitan area (2005–2012)," *J. Surv. Eng.*, vol. 141, no. 4, 2015, Art. no. 05015002.

[40] X. Qiao, T. Chu, and P. Tissot, "Coastal subsidence analysis between Houston and Galveston, Texas, USA," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 4139–4142.

[41] X. Qiao, T. Chu, P. Tissot, and S. Holland, "Long-term vertical-land-motion investigation with space and terrestrial geodetic techniques near San Leon, Texas, USA," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 125, 2023, Art. no. 103580.

[42] Alaska Satellite Facility (ASF), "ASF data search vertex," Copernicus Sentinel Data, Retrieved from ASF DAAC, processed by ESA, 2022. Accessed: Jan. 7, 2022. [Online]. Available: <https://ASF.alaska.edu/>

[43] NGL, "Precise point positioning (PPP) product of GNSS observations by Nevada Geodetic Lab (NGL)," 2023. Accessed: Mar. 12, 2023. [Online]. Available: <http://geodesy.unr.edu/>

[44] M. Bos, R. Fernandes, S. Williams, and L. Bastos, "Fast error analysis of continuous GNSS observations with missing data," *J. Geodesy*, vol. 87, no. 4, pp. 351–360, 2013.

[45] RRC, "The Railroad Commission of Texas," 2023. Accessed: Jan. 1, 2023. [Online]. Available: <https://www.rrc.texas.gov/>

[46] T. G. Farr et al., "The shuttle radar topography mission," *Rev. Geophys.*, vol. 45, no. 2, 2007, Art. no. RG2004.

[47] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017, doi: [10.1016/j.rse.2017.06.031](https://doi.org/10.1016/j.rse.2017.06.031).

[48] N. Houston, D. Wallace, A. Robinson, and L. Foster, "Groundwater-withdrawal data from the coastal lowlands aquifer system in parts of Texas, Louisiana, Mississippi, Alabama, and Florida, 1925–2018," 2021. Accessed: Aug. 20, 2023.

[49] E. Vermote, C. Justice, M. Claverie, and B. Franch, "Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product," *Remote Sens. Environ.*, vol. 185, pp. 46–56, 2016.

[50] J. Dewitz, "National Land Cover Database (NLCD) 2019 products. 2021 National Land Cover Database (NLCD) 2019 products (Ver. 2.0, Jun. 2021)," 2021. Accessed: Aug. 15, 2023.

[51] C. D. Elvidge, K. Baugh, M. ZhiZhin, F. C. Hsu, and T. Ghosh, "VIIRS night-time lights," *Int. J. Remote Sens.*, vol. 38, no. 21, pp. 5860–5879, 2017.

[52] United States Geological Survey, "Faults in the Gulf Coast [gcfaultsg]," 2004. Accessed: Aug. 15, 2023. doi: [10.5066/P9C33AT5](https://doi.org/10.5066/P9C33AT5).

[53] USGS World Energy Assessment Team, "U.S. Geological Survey World Petroleum Assessment 2000—Description and results," 2000, Accessed: Aug. 23, 2023. [Online]. Available: <https://pubs.usgs.gov/dds/dds-060/index.html#TOP>

[54] K. J. Beven and M. J. Kirkby, "A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'Appel variable de l'hydrologie du Bassin versant," *Hydrol. Sci. J.*, vol. 24, no. 1, pp. 43–69, 1979.

[55] M. S. Chowdhury, "Modelling hydrological factors from DEM using GIS," *MethodsX*, vol. 10, 2023, Art. no. 102062.

[56] SARPROZ, "The SAR PROcessing tool by periZ," 2023. Accessed: Jan. 1, 2023. [Online]. Available: <https://www.sarproz.com>

[57] D. Dzurisin, D. Dzurisin, and Z. Lu, "Interferometric synthetic-aperture radar (InSAR)," in *Volcano Deformation: Geodetic Monitoring Techniques*. Berlin, Germany: Springer, 2007, pp. 153–194.

[58] C. Prati, A. Ferretti, and D. Perissin, "Recent advances on surface ground deformation measurement by means of repeated space-borne SAR observations," *J. Geodynamics*, vol. 49, no. 3, pp. 161–170, 2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0264370709001446>

[59] C. Tay et al., "Sea-level rise from land subsidence in major coastal cities," *Nature Sustainability*, vol. 5, pp. 1049–1057, 2022.

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

[61] A. Géron, *Hands-On Machine Learning With Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2022.

[62] A. Akossou and R. Palm, "Impact of data structure on the estimators R-square and adjusted R-square in linear regression," *Int. J. Math. Comput.*, vol. 20, no. 3, pp. 84–93, 2013.

[63] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 431–439.

[64] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 2522–2539, 2020.

[65] S. Lipovetsky and M. Conklin, "Analysis of regression in game theory approach," *Appl. Stochastic Models Bus. Ind.*, vol. 17, no. 4, pp. 319–330, 2001.

[66] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[67] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 1135–1144.

[68] C. Molnar, "Interpreting machine learning models with SHAP," 2023. [Online]. Available: [Lulu.com](https://lulu.com)

[69] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Neural Inf. Process. Syst.*, vol. 30, pp. 4768–4777, Dec. 2017. [Online]. Available: <https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>

[70] C. Liang, Z. Liu, E. J. Fielding, and R. Bürgmann, "InSAR time series analysis of L-band wide-swath SAR data acquired by ALOS-2," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4492–4506, Aug. 2018.

[71] K. Wang and J. Chen, "Accurate persistent scatterer identification based on phase similarity of radar pixels," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5118513.

[72] M. de Michele, D. Raucoules, C. Rault, B. Aunay, and M. Foumelis, "Potential of Sentinel-1 interferometric synthetic aperture radar (InSAR) and offset tracking in monitoring post-cyclonic landslide activities on Réunion," *Earth Surf. Dyn.*, vol. 11, no. 3, pp. 451–460, 2023.

[73] Scikit Learn, "Random forest regressor," 2024. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

[74] Scikit Learn, "Feature importances with a forest of trees," 2024. [Online]. Available: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

[75] Y. Liu, G. Wang, X. Yu, and K. Wang, "Sentinel-1 InSAR and GPS-integrated long-term and seasonal subsidence monitoring in Houston, Texas, USA," *Remote Sens.*, vol. 14, no. 23, 2022, Art. no. 6184.

[76] M. M. Miller, C. E. Jones, S. S. Sangha, and D. P. Bekaert, "Rapid drought-induced land subsidence and its impact on the California aqueduct," *Remote Sens. Environ.*, vol. 251, 2020, Art. no. 112063.

[77] X. Zhou, G. Wang, K. Wang, H. Liu, H. Lyu, and M. J. Turco, "Rates of natural subsidence along the Texas coast derived from GPS and tide gauge measurements (1904–2020)," *J. Surv. Eng.*, vol. 147, no. 4, 2021, Art. no. 04021020.

[78] T. Chaudhary et al., "East Texas forests show strong resilience to exceptional drought," *Forestry*, vol. 96, no. 3, pp. 326–339, 2023.

[79] D. H. T. Minh and Y.-N. Ngo, "Compressed SAR interferometry in the Big Data Era," *Remote Sens.*, vol. 14, no. 2, 2022, Art. no. 390.



Xiaojun Qiao received the B.S. degree in surveying and mapping and the M.S. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Beijing, China, in 2014 and 2017, respectively, and the Ph.D. degree in geospatial computer science from Texas A&M University-Corpus Christi, Corpus Christi, TX, USA, in 2023.

His research interest includes the realm of utilizing spaceborne geodetic techniques and machine learning techniques to assess coastal land deformation and sea-level changes.



Tianxing Chu received the Ph.D. degree in photogrammetry and remote sensing from Peking University, Beijing, China, in 2012.

He is currently an Assistant Professor with Texas A&M University-Corpus Christi (TAMU-CC), Corpus Christi, TX, USA. Prior to his faculty role, he was with TAMU-CC as a Postdoctoral Research Associate and an Assistant Research Scientist from 2014 to 2018. His current research interests include emergent remote sensing and geomatics techniques, ubiquitous navigation and positioning, indoor simultaneous localization and mapping, and machine learning applied to geospatial applications.



Seneca Holland received the B.S. degree in geography from Florida State University, Tallahassee, FL, USA, in 2003 and the master of arts degree in geography from the University of Georgia, Athens, GA, USA, in 2007.

From 2009 to 2012, she was a Geospatial Research Associate with Texas A&M University-Corpus Christi (TAMU-CC), Corpus Christi, TX, USA, and from 2012 to 2017, she taught in the Geographic Information Science (GIS) program with the School of Engineering and Computing Sciences,

TAMU-CC. Before returning to TAMU-CC in 2022, she was the Senior GIS Developer for Katy Independent School District, TX.



Evan Krell received the B.S. and M.S. degrees in computer science in 2016 and 2018, respectively, from Texas A&M University-Corpus Christi (TAMU-CC), Corpus Christi, TX, USA, where he is currently working toward the Ph.D. degree within the Geospatial Computer Science program.

He is currently a member of the National Science Foundation (NSF) AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES). His research interests include artificial intelligence with a focus on developing explainable AI techniques for geospatial models and marine robotics including path planning and navigation for autonomous vessels.



Mohamed Ahmed received the Ph.D. degree in geosciences from Western Michigan University, Kalamazoo, MI, USA, in 2012.

He is currently an Associate Professor with Texas A&M University-Corpus Christi (TAMU-CC), Corpus Christi, TX, USA. He employs numerical modeling techniques, specifically rainfall-runoff models, and harnesses the power of machine learning methods, such as deep neural networks, gradient boosting machines, generalized linear models, and distributed random forests. By integrating these multidisciplinary techniques, he aims to gain comprehensive insights into these complex problems and contribute to a deeper understanding of the intricate interactions between Earth's systems.

His current research endeavors revolve around the utilization of geophysical data from spaceborne, airborne, and ground-based sources (including GRACE, gravity, magnetic, ground penetrating radar, and resistivity data), as well as remote sensing data encompassing optical, microwave, and thermal sensors



Philippe Tissot received the Diploma degree in physics engineering from the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 1987 and the Ph.D. degree in nuclear engineering from Texas A&M University, College Station, TX, USA, in 1994.

He is currently the Conrad Blucher Institute Chair for Coastal Artificial Intelligence with Texas A&M University-Corpus Christi (TAMU-CC), Corpus Christi, TX, USA, and a co-Principal Investigator (Co-PI) for the National Science Foundation (NSF) AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES). Prior to joining TAMU-CC in 1999, he was a Research Scientist with a private research institution working on R&D projects including developing prototypes for companies in the areas of oil exploration, environmental assessment, semiconductor materials analysis, and treatment of engineered wood materials. He has authored or coauthored more than 50 peer-reviewed articles, more than 270 proceedings, abstracts, and technical presentations, a Physical Science textbook for future K-12 teachers and has two U.S. patents. Since his arrival at CBI in 1999, his research has focused on the development of artificial intelligence methods for the analysis and predictions of environmental systems with a focus on coastal physical processes. His research projects include developing operational predictions for coastal inundation, navigation, and emergency management. Other research interests include studying relative sea level rise, the spatial variability of vertical land motion, and their societal impacts.



Danielle Smilovsky received the undergraduate degree in ecosystem sustainability and the master of advanced study degree in geographic information science (GIS) from Arizona State University, Tempe, AZ, USA, and the Ph.D. degree in GIS from Texas State University, San Marcos, TX, USA, in 2023.

She has more than ten years of experience in solving real-world problems with a wide spectrum of geospatial and geomatics techniques. She specializes in performing geotechnical land deformation investigations using InSAR, SAR, light detection and ranging, and water data analysis to study mines, landslides, subsidence, recharge, Earth fissuring, flood control structures, and various types of facilities and transportation structures.