nature communications

Article

https://doi.org/10.1038/s41467-023-41933-z

Global land subsidence mapping reveals widespread loss of aquifer storage capacity

Received: 11 July 2023

Accepted: 22 September 2023

Published online: 04 October 2023

Check for updates

Md Fahim Hasan [1] ✉, Ryan Smith[1], Sanaz Vajedian[2], Rahel Pommerenke[1] & Sayantan Majumdar [3]

Groundwater overdraft gives rise to multiple adverse impacts including land subsidence and permanent groundwater storage loss. Existing methods are unable to characterize groundwater storage loss at the global scale with sufficient resolution to be relevant for local studies. Here we explore the interrelation between groundwater stress, aquifer depletion, and land subsidence using remote sensing and model-based datasets with a machine learning approach. The developed model predicts global land subsidence magnitude at high spatial resolution (~2 km), provides a first-order estimate of aquifer storage loss due to consolidation of ~17 km^3/year globally, and quantifies key drivers of subsidence. Roughly 73% of the mapped subsidence occurs over cropland and urban areas, highlighting the need for sustainable groundwater management practices over these areas. The results of this study aid in assessing the spatial extents of subsidence in known subsiding areas, and in locating unknown groundwater stressed regions.

Excessive groundwater pumping can cause depletion, loss of aquifer storage capacity, arsenic contamination, saltwater intrusion, and infrastructure damage[1–3]. Despite its importance, many regions of the world with intensive groundwater withdrawals and storage loss are poorly monitored. In absence of spatially dense monitoring networks, publicly available in situ data, and uniform monitoring strategies, it is challenging to quantify groundwater storage loss. To address such data gaps, remote sensing techniques have been used to develop global scale datasets that measure proxies of or drivers for groundwater storage change. However, no current remotely sensed dataset provides a direct estimate of available groundwater storage and storage loss.

One of the most visible and harmful effects of groundwater depletion is land subsidence, which is caused by compaction of aquifer materials following the loss of pore pressure[4] and can cause irreversible loss of aquifer storage capacity[5]. Estimating the amount of subsidence can be used to quantify storage loss in unconsolidated confined aquifer systems[1]. In-situ measurement methods for quantifying subsidence exist; however, they are spatially far too sparse to be used in accurate subsidence estimation at regional to global scales. Interferometric Synthetic Aperture Radar (InSAR) observations have been shown to be a reliable source of subsidence data, providing ~1 cm accuracy at a fine spatial resolution of ~100 m[6], and have been used to monitor groundwater storage depletion in many aquifer systems[7–9]. Despite that, processing InSAR data is computationally expensive and can be challenging to interpret in the presence of tropospheric or ionospheric noise[10,11]; therefore, InSAR-based groundwater studies have been limited to the local or regional level. This hampers our ability to understand the state of subsidence and loss of aquifer storage capacity in regions outside the scope of these studies.

Process-based models provide another method for developing global estimates of groundwater availability and storage change[12,13] Nevertheless, a global model of land subsidence has not been produced to date. Such an effort would require extensive geomechanical and hydrogeologic datasets, and knowledge of the temporal evolution of head changes driving non-linear subsidence processes[1,5], which are not available at the global scale. However, remote sensing and global model-based datasets offer estimates of some of the drivers of subsidence and can be useful for predicting subsidence with statistical or data science approaches. While some studies have attempted to

[1]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA. [2]Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA. [3]Division of Hydrologic Sciences, Desert Research Institute, Reno, NV 89512, USA. ✉e-mail: Fahim.Hasan@colostate.edu

predict loss of aquifer storage capacity from subsidence at regional scales[1,14], and subsidence susceptibility globally[15], no existing study has quantified the magnitude of subsidence and associated groundwater storage loss globally.

In this study, we present a machine learning method to map pumping-induced land subsidence at a high spatial resolution (~2 km) on a global scale, using remote sensing and model-based hydrologic, land use, climatic, and geologic datasets. We trained the model[16] with an extensive InSAR and Global Navigation Satellite System (GNSS)–based land subsidence dataset. Our method produces subsidence estimates in <1 cm/year, 1–5 cm/year, and >5 cm/year classes. This study is a global-scale endeavor to map subsidence across ranges of magnitude and related groundwater storage loss at high (~2 km) resolution, in addition to exploring drivers of land subsidence. Such a study is crucial from the perspective of climate change, population growth, and relative sea level rise, which threaten to increase water scarcity, coastal flooding, and saltwater intrusion. The resulting global subsidence map can be interpreted as a first-order map of global aquifer storage loss due to loss of porosity, which is the primary mechanism for groundwater storage loss in confined alluvial basin aquifers[1]. In addition, the machine learning model[16] provides a global subsidence probability map (Supplementary Figs. 2 and 3, Supplementary Discussion 4) which is critical in identifying regions that are likely to experience subsidence. Using the model results, we are able to generate country level statistics of loss of aquifer storage capacity caused by subsidence, identifying countries with aquifers under the highest threat and putting groundwater stress in a global context.

Results and discussion

Global subsidence map

A random forest algorithm-based machine learning approach has been used in this study to generate a high-resolution (~2 km) global map of land subsidence. The global subsidence model[16] was designed to predict subsidence in three classes: <1 cm/year, 1–5 cm/year, and >5 cm/year, with the <1 cm/year class considered as the nominal or zero subsidence class. It was trained with InSAR-based subsidence datasets from 47 regions and a GNSS-based coastal subsidence data[17] of the world; using hydrologic, land use, and geologic datasets as input variables/predictors that are estimates and proxies of principal drivers of land subsidence. Figure 1 (see Supplementary Fig. 1. for the whole map) shows the global map of subsidence, focused on regions with high subsidence signatures, mapped by our model. It should be noted that the model developed in this study is designed to only estimate subsidence related to aquifer system compaction from groundwater pumping; therefore, the total subsidence estimates over some regions, which are undergoing subsidence from other sources, may not match.

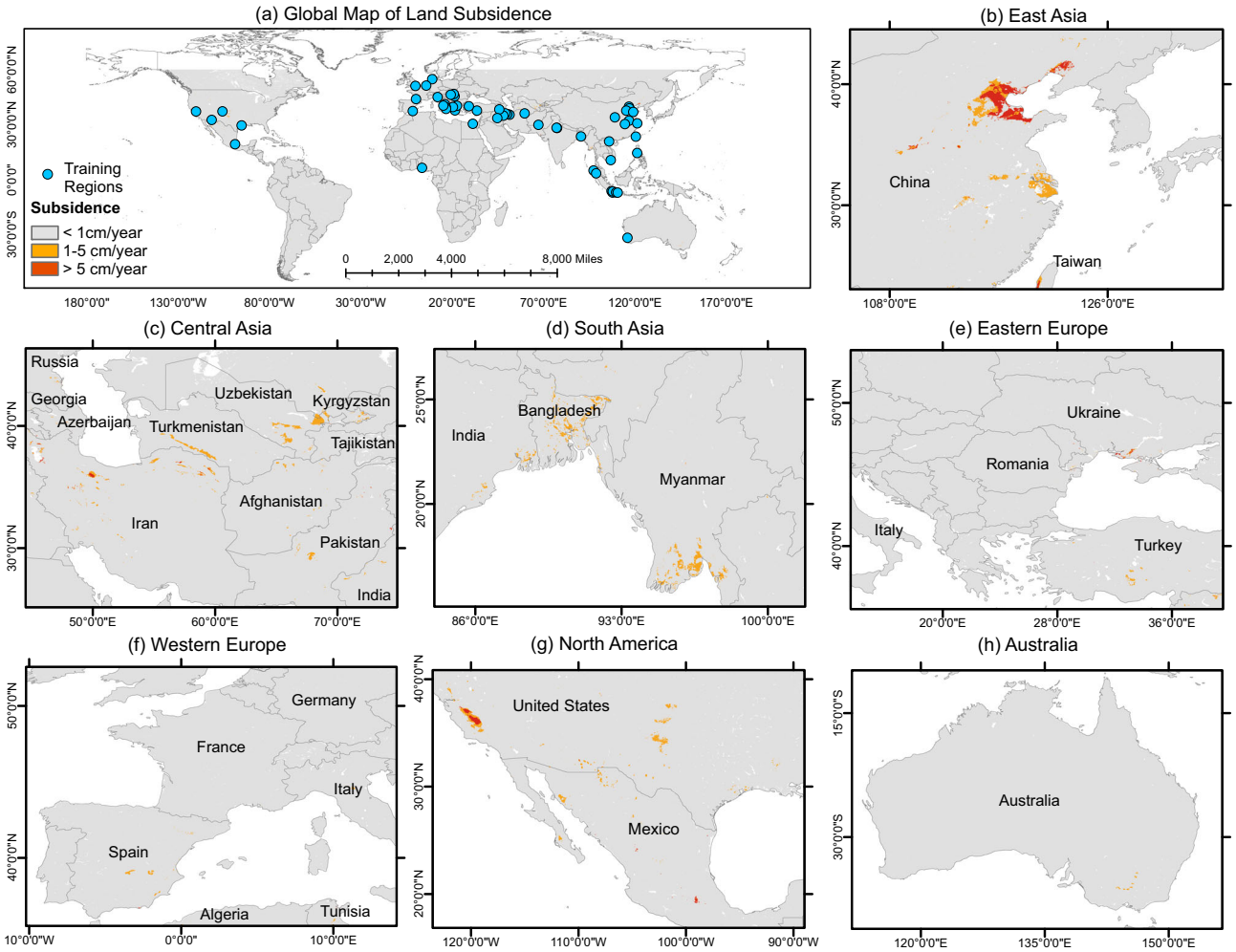

Fig. 1 | Groundwater withdrawal induced global land subsidence predicted by the random forest model. The model has been trained with Interferometric Synthetic Aperture Radar (InSAR)-derived subsidence data for 47 regions (blue dots) and a Global Navigation Satellite System (GNSS)-based coastal subsidence dataset to generate the **a** Global map of subsidence. 1–5 cm/year and >5 cm/year are considerable subsidence classes, and <1 cm/year is the nominal or no subsidence class. The model predicts significant subsidence across the globe that covers regions in **b** East Asia, **c** Central Asia, **d** South Asia, **e** Eastern Europe, **f** Western Europe, **g** North America, and **h** Australia. Source data are provided as a Source data file.

The model maps a considerable amount of subsidence in countries of East Asia: China, Taiwan, Vietnam, and the Philippines (Fig. 1 and Supplementary Fig. 1). In China, intensive irrigation activities have been mapped over the North China Plain aquifer by global irrigation mapping studies[18]. Major cities like Beijing, Shanghai, Wuhan, Xian, and Tianjin are in or nearby this region and are heavily dependent on groundwater to support agriculture and urban needs[8,19–22]. Our map shows a high subsidence signature in this whole region indicating significant groundwater storage decline. Countries in South, Central, and middle-East Asia, such as Bangladesh, Myanmar, India, Pakistan, Indonesia, Iran, and Turkey have areas of high subsidence signals as well. These predictions are in line with recent InSAR based groundwater studies for these regions[9,23–31]. Our model also predicts subsidence in irrigated and urban regions over Afghanistan, Turkmenistan, Uzbekistan, Azerbaijan, and Syria (Supplementary Fig. 1), where no previously published land subsidence studies due to groundwater withdrawal were available.

In Europe, recent studies have shown subsidence occurring in Spain, Italy, and England[32–34] with low magnitude of less or marginally higher than 1 cm/year. Our model categorizes the majority of subsidence in Europe as <1 cm/year. Vertical land movement data from GNSS[17] and European Ground Motion Service (EGMS) over Italy, Spain, France, Hungary, and Greece also show deformation lower than 1 cm/year, which is too low for our model to predict; however, subsidence of such magnitude can be significantly damaging in coastal areas, compounded with the impacts of sea level change. The map predicts subsidence between 1–5 cm/year, primarily due to groundwater irrigation, in Albacete and Ciudad Real province, and Alto Guadalentín basin in Spain. Comparison of model prediction in Albacete and Ciudad Real province with the EGMS data show that the model overestimates subsidence in these regions, possibly due to uncertainty in representing groundwater irrigation estimate in the model. Some 1–5 cm/year deformation signals have also been mapped in the dominantly agricultural region near Po Delta, Italy which were validated by the EGMS data. The model also predicts >1 cm/year subsidence in irrigated regions near the coast of Ukraine.

In North America, Fig. 1 shows considerable subsidence in California, Texas, Arizona, and central Mexico, which follows historic subsidence observed in these regions[5,7,35]. Subsidence in California and Arizona is due to excessive groundwater irrigation[35,36], while urban dependency on groundwater is responsible for deformation in Houston and Mexico City[7,37]. The model overestimates subsidence in some regions of the heavily pumped High Plains Aquifer[38] where a recent study has estimated <1 cm/year subsidence[39].

InSAR studies in Africa[40,41] have estimated less than 1 cm/year subsidence in the Nile river delta and in several coastal cities, such as Lagos, Banjul, Mombasa, and Mogadishu. The model's predictions show similar subsidence rates in these areas. The model predicts 1–5 cm/year subsidence in Morocco, Algeria, and Tunisia over irrigated lands (Supplementary Fig. 1). In Australia, subsidence <1 cm/year has been detected in Perth and irrigation dependent Murray-Darling basin, with few locations undergoing higher than 1 cm/year subsidence[42]. Our map detects 1–5 cm/year subsidence in the Murray-Darling basin region which might be happening in the alluvial aquifers consisting of clay and silt[43]. In the South American continent, >1 cm/year subsidence has been mapped over small, irrigated regions in Argentina and Peru. In Bolivia, >5 cm/year subsidence has been predicted over groundwater dependent cities[44,45] of Oruro and Cochabamba.

While the model generally shows good agreement with observed subsidence data, it overestimates subsidence in some regions, which happens due to uncertainty associated with the input variables. A full discussion of input variable uncertainty is provided in Supplementary Discussion 3 and summarized here. We developed a "confining layer" input variable to represent the presence of depositional lacustrine and marine confined aquifer conditions using a digital elevation model (DEM). Although this dataset has been validated successfully for the major aquifers in the United States (Supplementary Method 2), imprecise delineation of depositional settings in other regions can add some uncertainty to the model prediction. In addition, we developed the "normalized clay indicator" variable using datasets of percent clay content data at 200 cm[46] depth and average unconsolidated material thickness[47] (detail in the "Methods" section). While the normalized clay indicator is an effective proxy for the presence of thick clays at greater depths, it is limited by the shallow depth of clay content used and thus adds uncertainty to our model.

Drivers of land subsidence and groundwater storage loss

In confined (pressurized) aquifers, the aquifer matrix remains saturated even as the pressure head in the aquifer drops, and storage loss occurs. The two mechanisms for storage loss in confined aquifers are loss of pore space in the aquifer due to consolidation, and expansion of water[48]. In confined aquifers that are unconsolidated (which are commonly the most productive[49]), consolidation accounts for the vast majority of storage change[5]. Thus, subsidence in a confined or semi-confined aquifer system is a first-order estimate of total aquifer storage loss[1].

Land subsidence is driven by several factors, including aquifer skeletal specific storage (S_{sk}), thickness of compressible sediments (b), change in hydraulic head (Δh) due to pumping and recharge, and the consolidation history of the layer experiencing subsidence. More background on the mechanism of land subsidence has been described in Supplementary Note 1. For local and even for some regional areas, these datasets are available and can be coupled to estimate land subsidence. However, data scarcity, heterogeneity of collected data, and coarse resolution of available data make it difficult to compile all required datasets at a global scale. Global coverage of remotely sensed datasets can bridge this void by providing estimates of variables, such as precipitation, evapotranspiration (ET), soil moisture, and total water storage (TWS) data, in areas with heavy groundwater exploitation and can give an indication of hydrologic change. TWS anomaly data from the Gravity Recovery and Climate Experiment (GRACE) satellite along with other hydrologic variables of coarse scale have been used in multiple studies to model groundwater storage change at regional and global scales[50,51]. However, GRACE has an effective spatial resolution no better than about 100,000 km^2 at mid-latitudes[52]. An analysis of subsidence training data with GRACE TWS trend (over 2013–2017) revealed that ~36% subsidence training pixels fall in regions where TWS has a positive trend. TWS comprises both surface and groundwater fluxes, and a GRACE storage trend over a 100,000 km^2 region can have a positive trend even if groundwater storage is in decline within a 4 km^2 pixel within the larger GRACE region. Due to the coarse resolution and to avoid biases in model training, GRACE data was not incorporated in the model. Rather, InSAR processed land subsidence data can function as a proxy of groundwater storage change in aquifers, as discussed in the previous section. Our model assimilates multiple hydrologic and land use variables which can be related to groundwater storage change. Geologic variables, such as normalized clay indicator and existence of confining units, have also been used in the model as proxies of aquifer properties, and they represent the presence of fine sediments units in the subsurface which is a major driver of subsidence.

To assess if the input variables realistically illustrate the physical processes in the machine learning model, we analyzed the Partial Dependence Plots (PDP). Two-way PDP plots in Fig. 2 represent the contributions of input variables in model predictions of subsidence of the 1–5 cm/year class. Supplementary Fig. 4 contains individual PDP for some of the key variables of the model.

The input variables added to the model were direct measurements or proxies of principal drivers of land subsidence and groundwater withdrawal. The presence of clay and fine-grained confining

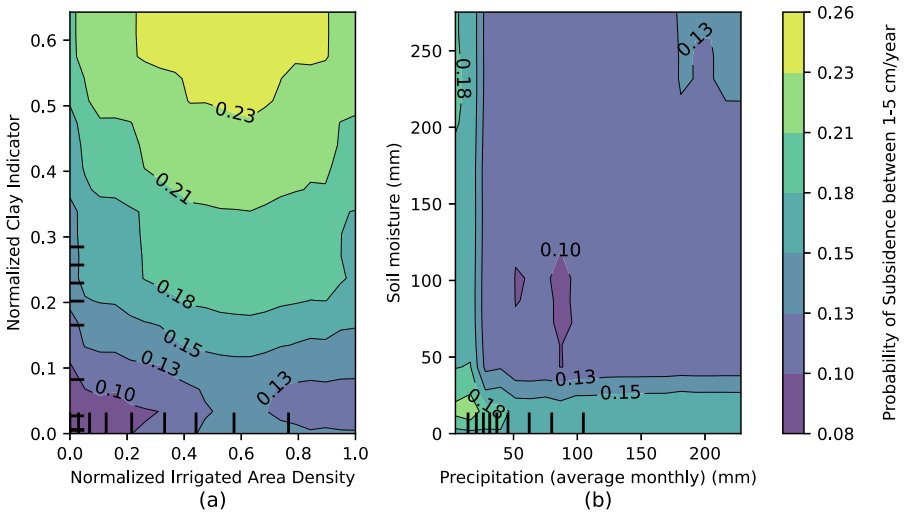

Fig. 2 | Two-way partial dependence plots of input variables. The plots represent contributions of input variable combinations **a** Normalized Irrigated Area Density and Normalized Clay Indicator and **b** Precipitation and Soil Moisture in predicting 1–5 cm/year subsidence. Warmer color indicates higher subsidence probability. The values in both axes have been plotted between 1st to 99th percentile as the model's response (in predicting subsidence) to the variables is more evident within this range. Normalized clay indicator and normalized irrigated area density are proxy variables that indicate the presence of fine sediments and groundwater irrigation density in the model, respectively, whereas soil moisture and precipitation, along with other hydrologic fluxes, represent water balance indicating areas with groundwater depletion.

units in confined or semi-confined aquifers are the major impetus of inelastic, high-magnitude subsidence. In addition, high density of irrigated agriculture in an area indicates higher probability of groundwater being used for irrigation in the presence of limited or no surface water resources. Land use with high groundwater irrigation is prone to subsidence if the water is being drawn from the fine-grained sediments in the subsurface. Figure 2a illustrates the model's ability to understand this relationship between normalized irrigated area density and normalized clay indicator (values ranging from 0 to 1, higher values indicate higher presence of irrigation density and clay), as the response shows higher probability of subsidence with high irrigated area density and normalized clay indicator.

The hydrologic variables represent water balance in the model indicating regions with groundwater depletion. Irrigation demands more groundwater in arid and semi-arid regions where precipitation is lower than evapotranspiration and surface water sources are scarce. In such regions, excessive groundwater pumping depletes groundwater storage which can lead to subsidence under favorable geologic conditions. Soil moisture is regarded as the most important variable by our model (Supplementary Fig. 8) because it is an effective indicator of the hydrologic conditions (sustained high temperature and low precipitation) that lead to subsidence if other principal drivers of subsidence are present (detailed discussion in Supplementary Discussion 2). Low soil moisture from a land surface model that does not account for irrigation, such as the one used in this study, indicates high irrigation water demand in croplands. Our model predicts higher subsidence probability in regions with low precipitation and soil moisture (Fig. 2b). The interaction between variables in predicting subsidence, as shown in the PDP plots, confirms that the response of land use, geologic variables, and hydrologic fluxes in the model are realistic. Supplementary Discussion 1 and 2 discuss the interaction between the input variables and their interpretation.

Country statistics of subsidence and groundwater storage loss

Results from the global subsidence model were used to generate country-level statistics of subsidence and groundwater storage loss to get a global perspective of groundwater stress. Figure 3a shows 10 countries with the highest percentage, by area, of land that is experiencing 1 cm/year of subsidence or greater. Our model has mapped significant subsidence over small island nations like Taiwan and the Philippines along with many other coastal regions, and over semi-arid and arid climates in Uzbekistan, Azerbaijan, Armenia, and Turkmenistan (Supplementary Fig. 1), where studies have reported high groundwater uses[27,53–56].

Subsidence is not limited to arid regions; it has been mapped in humid climates, including Bangladesh, India, Vietnam, and Indonesia (see Supplementary Fig. 10 for more rankings), indicating their dependency on groundwater despite having high precipitation supply. Additionally, the model results were used to estimate a permanent global groundwater storage loss value of ~17 km^3/year due to aquifer consolidation. China, the United States, and Iran account for the majority of this loss (Fig. 3b). Overall, a comparative assessment between the percentage area affected by subsidence and total aquifer storage loss is required to understand the comprehensive groundwater stress scenario of a particular country.

Land uses driving subsidence and implications in planning

Comparison of mapped subsidence with the Moderate Resolution Imaging Spectroradiometer (MODIS) land use product reveals that most of the predicted subsiding regions are on either croplands or urban areas (~60.5% are on croplands, ~12.5% are on urban and built-up lands), ~19% are on vegetated, uncultivated regions, and 8% on other land cover types. The MODIS land use product reports an overall accuracy of 75%[57]; therefore, there is a possibility of intermixing between the vegetation and cropland classes. Vegetated lands sometimes exists in the vicinity of croplands and might even be periodically used for agriculture, which may explain the high predicted subsidence rates in vegetated areas. The high probability of subsidence associated with increasing irrigation and population density (Supplementary Fig. 4) represents the model's ability to understand the relation between long-term subsidence and groundwater use in those regions. Moreover, subsidence predictions are high (~75%) in arid and semi-arid climates where climatic water deficit leads to higher groundwater dependency.

In regions where groundwater usage from aquifers, especially from confined and semi-confined layers, is significantly more than the volume of recharge over a long period, inelastic subsidence causes permanent aquifer storage loss due to consolidation[5]. This indicates

https://doi.org/10.1038/s41467-023-41933-z

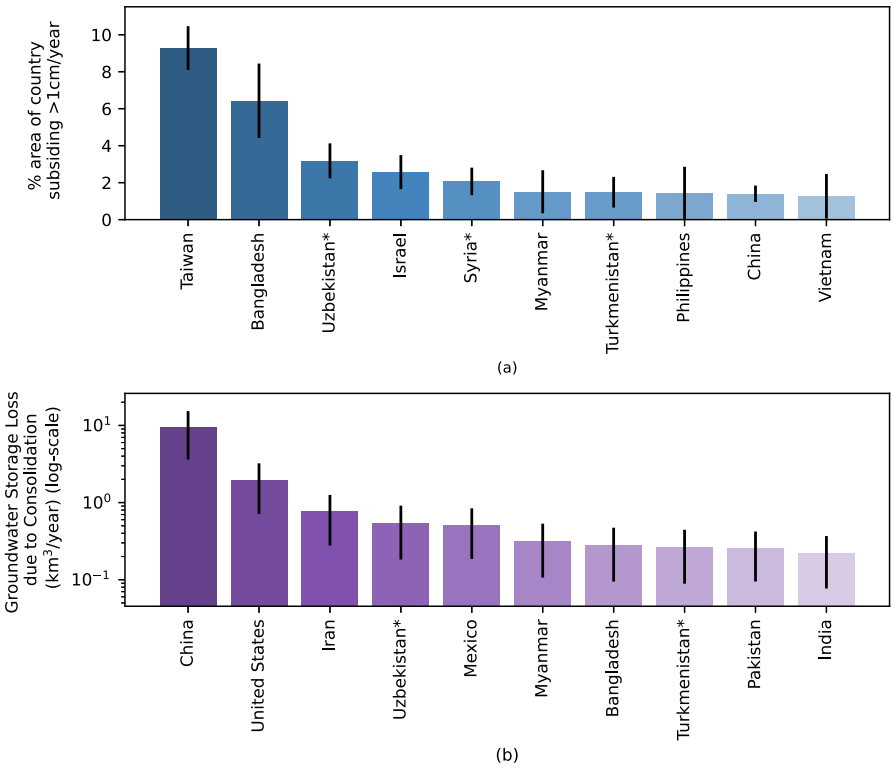

Fig. 3 | Rankings of country-level statistics of subsidence of magnitude >1 cm/year for the top 10 countries. a shows countries with the highest percentage of subsidence with respect to their country area. The error bars represent the standard deviation of the mean (standard error) of % area subsiding estimates. **b** shows countries with the highest groundwater storage loss, predicted by our model. China, the United States, and Iran account for the majority of permanent aquifer storage loss due to consolidation. The error bars represent the upper and lower bounds of groundwater storage loss estimates. In total, we have estimated an average ~17 km^3/year (a first-order estimate) of confined aquifer storage loss globally (lower and upper bounds ~11.5 km^3/year and ~22 km^3/year, respectively). Countries with asterisks are where no previously published land subsidence studies due to groundwater withdrawal were available. In both plots, Source data are provided as a Source data file.

that in the predicted subsiding locations, groundwater storage is permanently declining. Our model overestimates subsidence in some regions due to uncertainty in the input variables. Despite that, the generated map provides a first-order global estimate of subsidence due to groundwater overdraft. Comparison with documented subsidence locations and global groundwater studies (Supplementary Discussion 5 and Supplementary Fig. 13) shows that the machine learning model was able to reveal the true spatial extent of subsidence in many regions. These subsiding areas may continue to experience subsidence, and potentially experience an increase in subsidence magnitude and area, if water use practices are not modified. We also identified subsidence in irrigated and populated locations where groundwater related subsidence has not been studied and reported before. These regions are undergoing groundwater stress, and it is essential to develop effective, long term aquifer monitoring strategies to understand the true dynamics of groundwater resources in the affected regions. Additionally, regional studies incorporating InSAR data analysis should be undertaken for the mapped subsidence risk areas. Such efforts will help to formulate appropriate groundwater use, recharge, and long-term action plans for aquifer sustainability.

Methods

Processing input hydrologic and land use datasets

Input variables (predictors) of this model include remotely sensed and model-based global gridded datasets[58] that are proxies of principal hydrologic, geologic, and anthropogenic processes that drive land subsidence. Supplementary Table 1 shows a comprehensive list of all input datasets used in the model along with their original spatial resolution and sources. Depending on the original resolution, the datasets[58] were downscaled/upscaled to a resolution of 0.02 deg (~2 km) using the "nearest neighbor" algorithm to achieve a uniform grid size.

Direct, global estimates of groundwater resources are not available at the 2-km scale of our model from remote sensing sources. However, other water balance components such as precipitation, soil moisture and evapotranspiration are available, and correlate with withdrawals and recharge, important fluxes[59] whose relative magnitude is one control on subsidence[60]. We added these variables to account for these crucial water balance components.

A global irrigation area dataset Meier et al.[18] at ~1 km resolution was used in this study as one of the land use datasets. This dataset was developed by combining remote sensing datasets and downscaled statistics-based irrigated area data from an irrigation dataset[61] produced by the Food and Agriculture Organization of the United Nations (FAO). A gridded population dataset of ~1 km resolution[62] was included in the model to capture subsidence in populated areas occurring from aquifer pumping. A Gaussian filter was applied to both of the irrigated area and population datasets to add a smoothing effect that accounts for groundwater depletion in regions adjacent to aquifer pumping and to remove noise. The Gaussian filter normalized the datasets within an interval range of 0–1, where larger values represent higher density of respective land use class and vice versa. Supplementary Method 1 describes more about input dataset processing.

Processing input geologic datasets

Existence of fine sediments is a major geologic factor for inelastic subsidence[63]. Global geologic datasets indicating the presence of fine sediments is not readily available; therefore, existing global geologic

datasets have been modified to form proxy variables that indicate the presence and relative magnitude of fine sediments in the subsurface.

High-resolution (250 m) percent clay content data at 200 cm below ground surface, generated from a soil information based machine learning model[46], was multiplied with an average unconsolidated material thickness dataset (~1 km resolution)[47]. This product was normalized to form the "normalized clay indicator" dataset, a proxy dataset representing presence of clay in the subsurface.

Because the presence of a confining layer has a pronounced impact on the relationship between pumping and subsidence, a dataset indicating the likely presence or absence of a confining layer was produced as part of this study (Supplementary Fig. 11). This confining layer dataset was derived based on the depositional environment of basins and was produced using a globally available DEM from the Shuttle Radar Topography Mission (SRTM)[64]. The premise used in creating this dataset is that regions that are likely to have had lacustrine or oceanic depositional environments over the past several hundred thousand years are also likely to have extensive clay layers, which confine the aquifer and result in more subsidence from groundwater withdrawals. We validated our confining layer model using major aquifers in the United States as defined by the Groundwater Atlas of the United States[65], because this provides an extensive, geo-referenced map of aquifers with major confining layers. Supplementary Table 2 shows a summary of each major aquifer, along with the percent area of each aquifer that is estimated to have a confining layer present based on the methods outlined above. Detail discussion on developing this dataset along with its validation method have been presented in Supplementary Method 2.

Assembling land subsidence dataset for model training

Supervised machine learning algorithms require a training dataset to establish relationships between training data and input variables. Average vertical land subsidence rates (units in cm/year), in areas where significant groundwater pumping has been recorded historically, were used as training data in the machine learning model. Subsidence data were collected for 47 regions of the world from InSAR sources. In addition, a global coastal subsidence dataset was obtained from a GNSS-based study[17]. In this study, deformation data was gathered in three ways: processed by the authors; obtained as georeferenced, pre-processed data from public agencies of the United States and EGMS; and georeferenced from published studies.

We processed subsidence data in regions where no recent (2013 and later) subsidence data were available to the authors' knowledge in the published literature: Quetta Valley, Pakistan; Qazvin, Iran; and San Luis Valley, Colorado, United States. Our initial model revealed significant subsidence in the Hebei and Hefei regions of China; therefore, we processed InSAR data for these regions and included that in the training dataset. For processing InSAR, Small Baseline Subset (SBAS) InSAR time series analysis was conducted to estimate the average vertical subsidence rate in the Line of Sight (LOS) direction. LOS velocities were further decomposed into vertical and horizontal components using measurements from ascending and descending imaging geometries.

Processed, georeferenced vertical subsidence data over California and Arizona in the USA were collected from the California Department of Water Resources and from the Arizona Department of Water Resources, respectively. Similarly, processed, georeferenced vertical subsidence data over 15 regions of Europe were collected from the EGMS. Considering the substantial computational effort required in processing InSAR data, and the challenges associated with interpreting the principal subsidence cause to be related to groundwater, data was also collected from secondary sources. These sources consist of groundwater studies that used InSAR or GNSS information to determine aquifer vertical deformation. Regions where secondary sources were used include China, Indonesia, Iran, Turkey, USA, and Vietnam. A comprehensive list of these training data sources is provided in the Supplementary Table 3.

Finally, the subsidence data collected from these three sources were classified into three classes: <1 cm/year subsidence, 1–5 cm/year subsidence, and >5 cm/year subsidence. Subsidence data collected from research articles are referred to as georeferenced subsidence data in this study; however, these data are primarily based on InSAR processing. The classified subsidence data from the three processing methods were merged to form a final training dataset and resampled to a spatial resolution of 0.02 deg (~2 km). Supplementary Fig. 12 provides a framework of our modeling steps. It should be noted that <1 cm/year subsidence is considered as negligible to no subsidence class while the other classes represent medium to significant subsidence for this global study. However, subsidence of <1 cm/year values can be significantly damaging for coast-side regions due to the impact of climate change and resulting sea level rise, but predicting this level of subsidence was out of the scope of this study. Supplementary Method 3 describes how the training subsidence dataset was formed from multiple sources.

Random forests model prediction

Variables interplaying in land subsidence have complex nonlinear relationships that can be explored using a machine learning model. In this study, random forests, a popular tree-based ensemble learning algorithm, was used to incorporate the input variables to predict land subsidence in three classes. Random forests algorithm performs with high efficiency without input variable scaling[66], so datasets with values in varying units can be assimilated in such a model without issues.

Random forests employ techniques like bootstrap aggregating (bagging) of training samples and random splitting of input features to reduce variance, thus minimizing model overfitting[67]. Ensemble results of multiple trees are summarized by majority voting (in a classification model) to produce the final model outcome. The model's key hyperparameters were optimized (detail in Supplementary Table 4), to improve model accuracy and avoid overfitting, using a random search 10-fold cross-validation approach. The optimized model has hyperparameter values of n_estimators = 300, max_depth = 14, max_features = 7, min_samples_split = 7, and min_samples_leaf = 1^{-5}, which resulted in a macro F1-score of 0.83 on the test set.

Random forests model creation requires a primary training dataset (also referred to as response variable), in this case, land subsidence data. Machine learning models learn the relationship between input variables using the response variable. To create the training dataset for our model, pixels containing a land subsidence classification (<1 cm/year, 1–5 cm/year, or >5 cm/year) were matched to the input variables at the co-located pixel. The resulting dataset is referred to as the original training dataset of the model. This original training dataset was randomly split into train and test sets, with 70% data on the train set and 30% data on the test set, for model calibration and validation purposes. Of the subsidence training samples, approximately 84.5% belong to the <1 cm/year class, 10.5% are in the 1–5 cm/year class, and 5% are in the >5 cm/year class, creating imbalance in the dataset. Machine learning models with imbalance datasets are often biased towards the majority observation class[68]. A "balanced" class weight was assigned to prevent the model from being biased towards the majority class (<1 cm/year here). This approach calculates class weight using an inverse relation to the number of observations in a class[69], and assigns the lowest class weight value to the most frequent class and the highest value to the least frequent class. The class weights are considered during node splitting and weighted majority voting of ensemble results to assign more penalty on misclassifying the least frequent classes (1–5 cm/year and >5 cm/year)[70], and thus help to compensate for the imbalance in the dataset.

The hyperparameter-tuned, weight adjusted random forest model was used to generate a global map of land subsidence and a

subsidence probability map (Supplementary Figs. 2 and 3). The generated land subsidence prediction was further refined with a land use filter to filter out subsidence predictions over areas where there are both low normalized irrigated area density (<0.06) and low normalized population density (<0.009). Values (unitless, ranging from 0 to 1) of normalized irrigated area density and population density where subsidence have been observed in training data are >0.1 and >0.005, respectively, but if the values are at the lower end for any one of them, the value of the other variable tends to be higher. For example, in Hefei, China, 1–5 cm/year subsidence have been observed in populated area with density ranging from 0.006–0.009 but normalized irrigated area density of >0.3. Therefore, the land use filter was only applied on areas where both of the land use variables have very low values. This land use filter removed ~7% of the 1–5 cm/year and >5 cm/year predicted subsidence pixels, mostly prediction noise generated by the model, resulting in the final global map of land subsidence (Fig. 1) induced by groundwater over-drafting. Finally, we used the subsidence map to estimate a permanent groundwater storage loss volume of ~17 km^3/year, assuming average subsidence values of 3 and 10 cm/year for the 1–5 cm/year, >5 cm/year subsidence classes, respectively. The lower bound of permanent groundwater storage loss volume was estimated to be ~11.5 km^3/year; assuming 2 and 7 cm/year subsidence for the 1–5 cm/year, >5 cm/year subsidence classes, respectively. The upper bound was estimated to be ~22 km^3/year; assuming 4 and 13 cm/year subsidence for the 1–5 cm/year, >5 cm/year subsidence classes, respectively.

Model performance

Evaluating model robustness by simply calculating the fraction of predictions that are correct can result in a biased view of model performance, particularly with an imbalanced dataset such as ours. For instance, if our model always predicts <1 cm/year of subsidence, it would be right 80% of the time, but would be ineffective. In such cases, the F1-score is an efficient metric to assess the model as it considers true positives, false negatives, and false positives in calculating accuracy. The F1-score is defined as the harmonic mean of precision and recall, where precision represents the number of true positives divided by the number of model-predicted positives, and recall represents the number of true positives divided by the sum of true positives and false negatives[69]. Performance of the model was assessed with the F1-score on the testing set for individual classes and for all classes. The F1-scores for <1 cm/year, 1–5 cm/year, and >5 cm/year were 0.96, 0.68, and 0.86, respectively. For all the classes combined, the macro F1-score (average of F1-score of individual classes) was 0.83. To avoid overfitting, the hyperparameters were optimized on the train set and the model's performance was evaluated on the test set. Supplementary Table 5 shows F1-score for discrete classes and for the entirety.

A confusion matrix (Supplementary Fig. 13) of the test set shows that the model misclassified approximately 24% of the 1–5 cm/year class, compared to ~5.8% in the <1 cm/year and ~9.3% in the >5 cm/year classes. Despite a relatively lower F1-score, the accuracy of the 1–5 cm/year class is still quite high.

Leave-One-Area-Out accuracy test

Performance of a random forest model is greatly influenced by the quantity, distribution, and class balance of the training dataset. Class imbalance in the model can affect individual class accuracy and overall model performance[71]. Moreover, prediction of positive class in a region depends on the number and proximity of training samples in that vicinity[72]. The random forest model developed in this study suffers from all these challenges. These challenges were minimized by assigning class weights to reduce the effect of class imbalance and by expanding the subsidence training dataset with subsidence data collected from various sources and over many regions across the globe. Despite the effort, the quantity of training samples may not be sufficient considering the global extent of the model and they may not represent all combinations of groundwater stressed regions with variety of climate, hydrology, and geology. In such cases, our model might not be able to map subsidence in small regions whose characteristics were not represented in the training dataset. Therefore, to test the model accuracy and robustness further, a Leave-One-Area-Out (LOAO) accuracy test was designed, inspired by a popular machine learning model evaluation technique called Leave-One-Out cross-validation[73]. In the "Random forests model prediction" section, it was mentioned that the original training dataset was randomly split into train and test set for model fitting and validation purposes, respectively. The random split ensured that the train set included subsidence pixels from all regions to provide the model with varying information from respective regions. In the LOAO method, the dataset was not randomly split. Instead, the model iterated n times, where n is the number of training regions, leaving one training area completely out (performed as a test set) from training on each iteration and evaluating the model performance on the excluded region. Thus, the model ran 47 times (excluding the coastal datasets during this analysis), each time excluding an area. Subsiding areas globally vary in terms of climate, hydrologic balance, and geologic formation. Removing one area from the training data may decrease the model's prediction power over that region to some extent and sometimes entirely. Therefore, the model performance was evaluated based on subsidence probability, rather than original model subsidence prediction. The original model probability vs LOAO test probability for some regions of the world is presented in Supplementary Fig. 13. Supplementary Table 6 shows detail LOAO test results for the 47 training regions of the model, which were categorized based on criteria introduced in Supplementary Note 2. Out of the 47 regions only 10 were categorized "not satisfactory" according to the criteria. This means that the model cannot detect subsidence in these 10 regions without being explicitly trained with the deformation data of these regions. This may be due to the distinctive hydrologic, climatic, and geologic features of these regions that drive subsidence, which is not represented by any other area in the training dataset. However, the model performs satisfactorily for the rest of the 47 regions indicating robust model performance.

Data availability

The hydrological, geological, elevation, and remote sensing datasets have been cited throughout the paper, listed in Supplementary Table 1, and are publicly available. Sources of the secondary subsidence datasets have been listed in Supplementary Table 3 and are publicly available. The primary subsidence datasets (from InSAR) processed by the authors are available upon request from the corresponding author. The processed training subsidence data, processed input variables, training csv file, and reference files to run the modeling scripts, along with the global subsidence and subsidence probability prediction rasters by the model, are available at this HydroShare repository[58]—https://doi.org/10.4211/hs.dc7c5bfb3a86479b889d3b30ab0e4ef7. Data used for mapping purposes, such as global country-level shapefile and base map, are open-source datasets and have been used under appropriate license. Source data are provided with this paper.

Code availability

The project's modeling scripts[16] are available at the following GitHub repository: https://github.com/mdfahimhasan/Global-Subsidence-Groundwater.

References

1. Smith, R. & Majumdar, S. Groundwater storage loss associated with land subsidence in Western United States mapped using machine learning. *Water Resour. Res.* **56**, e2019WR026621 (2020).
2. Smith, R., Knight, R. & Fendorf, S. Overpumping leads to California groundwater arsenic threat. *Nat. Commun.* **9**, 1–6 (2018).

3. Van Camp, M., Mtoni, Y., Mjemah, I. C., Bakundukize, C. & Walraevens, K. Investigating seawater intrusion due to groundwater pumping with schematic model simulations: the example of the Dar es Salaam coastal aquifer in Tanzania. *J. Afr. Earth Sci.* **96**, 71–78 (2014).
4. Galloway, D. L. & Burbey, T. J. Review: Regional land subsidence accompanying groundwater extraction. *Hydrogeol. J.* **19**, 1459–1486 (2011).
5. Smith, R. et al. Estimating the permanent loss of groundwater storage in the southern San Joaquin Valley, California. *Water Resour. Res.* **53**, 2133–2148 (2017).
6. Wright, T. J., Parsons, B. E. & Lu, Z. Toward mapping surface deformation in three dimensions using InSAR. *Geophys. Res. Lett.* **31**, L01607 (2004).
7. Chaussard, E., Wdowinski, S., Cabral-Cano, E. & Amelung, F. Land subsidence in central Mexico detected by ALOS InSAR time-series. *Remote Sens. Environ.* **140**, 94–106 (2014).
8. Chen, M. et al. Imaging land subsidence induced by groundwater extraction in Beijing (China) using satellite radar interferometry. *Remote Sens.* **8**, 468 (2016).
9. Higgins, S. et al. InSAR measurements of compaction and subsidence in the Ganges-Brahmaputra Delta, Bangladesh. *J. Geophys. Res. F Earth Surf.* **119**, 1768–1781 (2014).
10. Reeves, J. A., Knight, R. & Zebker, H. A. An analysis of the uncertainty in InSAR deformation measurements for groundwater applications in agricultural areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 2992–3001 (2014).
11. Fattahi, H. & Amelung, F. InSAR bias and uncertainty due to the systematic and stochastic tropospheric delay. *J. Geophys. Res. Solid Earth* **120**, 8758–8773 (2015).
12. Reinecke, R. et al. Challenges in developing a global gradient-based groundwater model (G3M v1.0) for the integration into a global hydrological model. *Geosci. Model Dev.* **12**, 2401–2418 (2019).
13. Wada, Y. et al. Global depletion of groundwater resources. *Geophys. Res. Lett.* **37**, 1–5 (2010).
14. Naghibi, S. A., Hashemi, H. & Khodaei, B. An integrated InSAR-machine learning approach for ground deformation rate modeling in arid areas. *J. Hydrol.* **608**, 127627 (2022).
15. Herrera-García, G. et al. Mapping the global threat of land subsidence. *Science* **371**, 34–36 (2021).
16. Hasan, M. F., Smith, R., Vajedian, S., Pommerenke, R. & Majumdar, S. Global land subsidence mapping reveals widespread loss of aquifer storage capacity. GitHub https://doi.org/10.5281/zenodo.8280482 (2023).
17. Shirzaei, M. et al. Measuring, modelling and projecting coastal land subsidence. *Nat. Rev. Earth Environ.* **2**, 40–58 (2021).
18. Meier, J., Zabel, F. & Mauser, W. A global approach to estimate irrigated areas - a comparison between different data and statistics. *Hydrol. Earth Syst. Sci.* **22**, 1119–1133 (2018).
19. Dong, S., Samsonov, S., Yin, H., Ye, S. & Cao, Y. Time-series analysis of subsidence associated with rapid urbanization in Shanghai, China measured with SBAS InSAR method. *Environ. Earth Sci.* **72**, 677–691 (2014).
20. Higgins, S., Overeem, I., Tanaka, A. & Syvitski, J. P. M. Land subsidence at aquaculture facilities in the Yellow River delta, China. *Geophys. Res. Lett.* **40**, 3898–3902 (2013).
21. Qu, F. et al. Land subsidence and ground fissures in Xi'an, China 2005–2012 revealed by multi-band InSAR time-series analysis. *Remote Sens. Environ.* **155**, 366–376 (2014).
22. Zhou, L. et al. Wuhan surface subsidence analysis in 2015–2016 based on sentinel-1A data by SBAS-InSAR. *Remote Sens.* **9**, 982 (2017).
23. Chaussard, E., Amelung, F., Abidin, H. & Hong, S. H. Sinking cities in Indonesia: ALOS PALSAR detects rapid subsidence due to groundwater and gas extraction. *Remote Sens. Environ.* **128**, 150–161 (2013).
24. Dang, V. K., Doubre, C., Weber, C., Gourmelen, N. & Masson, F. Recent land subsidence caused by the rapid urban development in the Hanoi region (Vietnam) using ALOS InSAR data. *Nat. Hazards Earth Syst. Sci.* **14**, 657–674 (2014).
25. Erban, L. E., Gorelick, S. M. & Zebker, H. A. Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environ. Res. Lett.* **9**, 084010 (2014).
26. Haghighi, M. H. & Motagh, M. Ground surface response to continuous compaction of aquifer system in Tehran, Iran: results from a long-term multi-sensor InSAR analysis. *Remote Sens. Environ.* **221**, 534–550 (2019).
27. Hung, W. C. et al. Monitoring severe aquifer-system compaction and land subsidence in Taiwan using multiple sensors: Yunlin, the southern Choushui river Alluvial fan. *Environ. Earth Sci.* **59**, 1535–1548 (2010).
28. Khorrami, M., Abrishami, S., Maghsoudi, Y., Alizadeh, B. & Perissin, D. Extreme subsidence in a populated city (Mashhad) detected by PSInSAR considering groundwater withdrawal and geotechnical properties. *Sci. Rep.* **10**, 1–16 (2020).
29. Minh, D. H. T., Van Trung, L. & Le Toan, T. Mapping ground subsidence phenomena in Ho Chi Minh City through the radar interferometry technique using ALOS PALSAR data. *Remote Sens.* **7**, 8543–8562 (2015).
30. Orhan, O., Oliver-Cabrera, T., Wdowinski, S., Yalvac, S. & Yakar, M. Land subsidence and its relations with sinkhole activity in karapınar region, turkey: a multi-sensor insar time series study. *Sensors* **21**, 1–17 (2021).
31. Kakar, N., Kakar, D. M. & Barrech, S. Land subsidence caused by groundwater exploitation in Quetta and surrounding region, Pakistan. *Proc. Int. Assoc. Hydrol. Sci.* **382**, 595–607 (2020).
32. Corbau, C., Simeoni, U., Zoccarato, C., Mantovani, G. & Teatini, P. Coupling land use evolution and subsidence in the Po Delta, Italy: revising the past occurrence and prospecting the future management challenges. *Sci. Total Environ.* **654**, 1196–1208 (2019).
33. Bock, Y., Wdowinski, S., Ferretti, A., Novali, F. & Fumagalli, A. Recent subsidence of the Venice lagoon from continuos GPS and interferometric synthetic aperture radar. *Geochem. Geophys. Geosyst.* https://doi.org/10.1029/2011GC003976 (2012).
34. Bonì, R. et al. Exploitation of satellite A-DInSAR time series for detection, characterization and modelling of land subsidence. *Geosciences* **7**, 25 (2017).
35. Conway, B. D. Land subsidence and earth fissures in south-central and southern Arizona, USA. *Hydrogeol. J.* **24**, 649–655 (2015).
36. Faunt, C. C. *Groundwater Availability of the Central Valley Aquifer, California*. (USGS, 2009).
37. Yu, J., Wang, G., Kearns, T. J. & Yang, L. Is there deep-seated subsidence in the Houston-Galveston area? *Int. J. Geophys.* **2014**, 942834 (2014).
38. Butler, J. J., Whittemore, D. O., Wilson, B. B. & Bohling, G. C. Sustainability of aquifers supporting irrigated agriculture: a case study of the High Plains aquifer in Kansas. *Water Int.* **43**, 815–828 (2018).
39. Overacker, J., Hammond, W. C., Blewitt, G. & Kreemer, C. Vertical land motion of the high plains aquifer region of the United States: effect of aquifer confinement style, climate variability, and anthropogenic activity. *Water Resour. Res.* **58**, e2021WR031635 (2022).
40. Cian, F., Blasco, J. M. D. & Carrera, L. Sentinel-1 for monitoring land subsidence of coastal cities in Africa using PSInSAR: a methodology based on the integration of SNAP and staMPS. *Geosciences* **9**, 124 (2019).

41. Gebremichael, E. et al. Assessing land deformation and sea encroachment in the Nile Delta: a radar interferometric and inundation modeling approach. *J. Geophys. Res. Solid Earth* **123**, 3208–3224 (2018).
42. Castellazzi, P. & Schmid, W. Interpreting C-band InSAR ground deformation data for large-scale groundwater management in Australia. *J. Hydrol. Reg. Stud.* **34**, 100774 (2021).
43. Lamontagne, S. et al. Field assessment of surface water-groundwater connectivity in a semi-arid river basin (Murray-Darling, Australia). *Hydrol. Process.* **28**, 1561–1572 (2014).
44. Van Den Bergh, K., Du Laing, G., Montoya, J. C., De Deckere, E. & Tack, F. M. G. Arsenic in drinking water wells on the Bolivian high plain: field monitoring and effect of salinity on removal efficiency of iron-oxides-containing filters. *J. Environ. Sci. Health Part A* **45**, 1741–1749 (2010).
45. Gonzales Amaya, A., Ortiz, J., Durán, A. & Villazon, M. Hydrogeophysical methods and hydrogeological models: basis for groundwater sustainable management in Valle Alto (Bolivia). *Sustain. Water Resour. Manag.* **5**, 1179–1188 (2019).
46. Hengl, T. Clay content in % (kg/kg) at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution. zenodo https://doi.org/10.5281/zenodo.1476854 (2018).
47. Pelletier, J. D. et al. A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *J. Adv. Model. Earth Syst.* **8**, 41–65 (2016).
48. Freeze, R. A. & Cherry, J. A. *Groundwater* (Prentice-Hall, 1979).
49. Margat, J. & van der Gun, J. *Groundwater Around the World* (CRC Press, 2013).
50. Rateb, A. et al. Comparison of groundwater storage changes from GRACE satellites with monitoring and modeling of major U.S. aquifers. *Water Resour. Res.* **56**, e2020WR027556(2020).
51. Yin, W., Fan, Z., Tangdamrongsub, N., Hu, L. & Zhang, M. Comparison of physical and data-driven models to forecast groundwater level changes with the inclusion of GRACE – a case study over the state of Victoria, Australia. *J. Hydrol.* **602**, 126735 (2021).
52. Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C. & Landerer, F. W. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res. Solid Earth* **120**, 2648–2671 (2015).
53. Valder, J. F., Carter, J. M., Medler, C. J., Thompson, R. F. & Anderson, M. T. Hydrogeologic framework and groundwater conditions of the Ararat Basin in Armenia. Scientific investigations report. https://doi.org/10.3133/sir20175163 (2018).
54. Liu, Y. et al. Sustainable use of groundwater resources in the transboundary aquifers of the five central Asian countries: challenges and perspectives. *Water* **12**, 2101 (2020).
55. Aliyev, F. S. & Askerov, F. S. in *Urban Groundwater Management and Sustainability* (eds Tellam, J. H., Rivett, M. O., Israfilov, R. G. & Herringshaw, L. G.) 59–77 (Springer Netherlands, 2006).
56. De Zoysa, R. S. et al. The 'wickedness' of governing land subsidence: policy perspectives from urban southeast Asia. *PLoS ONE* **16**, 1–25 (2021).
57. Friedl, M. A. et al. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **114**, 168–182 (2010).
58. Hasan, M. F., Smith, R., Vajedian, S., Pommerenke, R. & Majumdar, S. Global land subsidence mapping reveals widespread loss of aquifer storage capacity datasets. *HydroShare* https://doi.org/10.4211/hs.dc7c5bfb3a86479b889d3b30ab0e4ef7 (2023).
59. Gleeson, T., Wada, Y., Bierkens, M. F. P. & van Beek, L. P. H. Water balance of global aquifers revealed by groundwater footprint. *Nature* **488**, 197–200 (2012).
60. Tiwari, V. M., Srinivas, N. & Singh, B. Hydrological changes and vertical crustal deformation in south India: inference from GRACE, GPS and absolute gravity data. *Phys. Earth Planet. Inter.* **231**, 74–80 (2014).
61. Siebert, S., Henrich, V., Frenken, K. & Burke, J. *Update of the Digital Global Map of Irrigation Areas to Version 5* (Rheinische Friedrich-Wilhelms-Universituy, Food Agricultural Organganization, United Nations, 2013).
62. CIESIN. GPWv411: UN-adjusted population density (gridded population of the world version 4.11). https://doi.org/10.7927/H4F47M65 (2018).
63. Smith, R. & Knight, R. Modeling land subsidence using InSAR and airborne electromagnetic data. *Water Resour. Res.* **55**, 2801–2819 (2019).
64. NASA Shuttle Radar Topography Mission (SRTM). Shuttle radar topography mission (SRTM) global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF (2013).
65. Hydrologic Atlas 730. Ground Water Atlas of the United States. http://pubs.er.usgs.gov/publication/ha730 (2000).
66. Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D. & Siddique, Z. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies* **9**, 52 (2021).
67. Breiman, L. Random forests. *Mach. Learn.* **45**, 1–33 (2001).
68. Johnson, J. M. & Khoshgoftaar, T. M. Survey on deep learning with class imbalance. *J. Big Data* **6**, 27 (2019).
69. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
70. Chen, C., Liaw, A. & Breiman, L. *Using Random Forest to Learn Imbalanced Data* (University of California, Berkeley, 2004).
71. Mellor, A., Boukir, S., Haywood, A. & Jones, S. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS J. Photogramm. Remote Sens.* **105**, 155–168 (2015).
72. Collins, L., McCarthy, G., Mellor, A., Newell, G. & Smith, L. Training data requirements for fire severity mapping using Landsat imagery and random forest. *Remote Sens. Environ.* **245**, 111839 (2020).
73. Magnusson, M., Andersen, M., Jonasson, J. & Vehtari, A. Bayesian leave-one-out cross-validation for large data. In *Proc. 36th International Conference on Machine Learning* (eds Chaudhuri, K. & Salakhutdinov, R.) 4244–4253 (PMLR, 2019).

Acknowledgements

This research is funded by an academic grant from the National Geospatial-Intelligence Agency (Award No. # HM0476-21-1-0001, Project Title: Global Land Subsidence Mapping Reveals Widespread Groundwater Storage Loss and Supplemental. Approved for public release, 22–787). The authors would like to express gratitude to the National Geospatial-Intelligence Agency for funding this project. We acknowledge the providers of all open-source datasets, research articles, and Google Earth Engine platform for the contribution with the datasets and for providing public access. We thank Jiawei Li for providing an InSAR dataset and Dr. Yoshihide Wada for sharing his research outcome which was used in evaluating the model result.

Author contributions

M.F.H. and R.S. developed the methodology of the study. R.S. conceptualized the study, managed funding, and supervised the project. M.F.H. collected and processed datasets, coded and developed the model, and performed result analysis. M.F.H. wrote the draft manuscript and R.S., R.P., S.V., and S.M. reviewed it. S.V. processed InSAR data. R.P. helped in initial database formulation and data collection. S.M. helped with modeling/coding guidelines and suggested methodology.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-41933-z.

Correspondence and requests for materials should be addressed to Md Fahim Hasan.

Peer review information *Nature Communications* thanks Matthew Rodell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.