Summer 2022

INTEGRATING REMOTE SENSING AND MODEL-BASED DATASETS IN A MACHINE LEARNING MODEL TO MAP GLOBAL SUBSIDENCE ASSOCIATED WITH GROUNDWATER WITHDRAWAL

Md Fahim Hasan
Missouri University of Science and Technology

Department:

INTEGRATING REMOTE SENSING AND MODEL-BASED DATASETS IN A MACHINE LEARNING MODEL TO MAP GLOBAL SUBSIDENCE ASSOCIATED WITH GROUNDWATER WITHDRAWAL

by

MD FAHIM HASAN

A THESIS

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN GEOLOGICAL ENGINEERING

2022

Approved by:

Ryan Smith, Advisor
Katherine Grote
Jeremy Maurer

ABSTRACT

Quantifying groundwater storage loss is becoming increasingly essential globally due limited availability of this major hydrologic component and its long recharge time. Groundwater overdraft gives rises to multiple adverse impacts including land subsidence and permanent groundwater storage loss. In absence of spatially dense monitoring network, publicly available in-situ data, and uniform monitoring strategies, it is challenging to assess the sustained losses from overexploitation of this resource. Remote sensing based techniques have the capacity to fill this gap to increase our groundwater monitoring capacities. Exploring the interrelation between groundwater pumping and land subsidence using remote sensing datasets can be a very effective technique to measure depletion of aquifers. In this study, we developed a machine learning model to explore this relationship with the help of gridded remotely sensed and model-based dataset, and Interferometric Synthetic Aperture Radar (InSAR) based land deformation data. InSAR generated land subsidence data from 36 different regions of the world were used to train a random forests model to map land subsidence globally at a high spatial resolution of ~2 km. The model predicted land subsidence magnitude in three classes: <1 cm/year, 1-5 cm/year and >5 cm/year. The model found realistic relationship between the driver variables, groundwater pumping and land subsidence with an overall score of 0.84 on the test set. Resulting maps from this model will be incredibly helpful in knowing the true spatial extents of subsidence in known subsiding areas and in locating unknown groundwater stressed regions where subsidence has not been documented before.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to the National Geospatial-Intelligence Agency for funding this project. Their support has been instrumental in this research and my master's degree at the Missouri University of Science and Technology.

I am deeply grateful to Dr. Ryan Smith for entrusting me with the responsibility of performing this research, for his guidance, and for keeping me motivated along the path. Dr. Smith has been an amazing mentor and will remain an inspiration throughout my career. I would also like to convey my gratitude to Dr. Kathrine Grote for her outstanding teachings on groundwater in her courses, which have helped me to build my foundation in hydrogeology.

There is no word to express my love and indebtedness to my wife, parents, and sister, and family members for their unwavering encouragement and belief in me. My parents have made outstanding sacrifices for my education and well-being throughout their life. Without their unconditional support, I could not have come so far in my life.

I would like to extend my sincere gratitude to my colleagues Dr. Sanaz Vajedian, Rahel Pommerenke, Sayantan Majumdar, and Dawit Asfaw for their contributions to this research. Your technical support, ideas, and suggestions have accelerated the progress of this work manyfold. I would also like to recognize the contributions of those who have blessed me with their advice and motivation along my path to graduate studies. Last but not the least, I am thankful to the Almighty for giving me the long-cherished opportunity to fulfill my higher study and career goals.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

Growing population, agricultural and energy demand have been laying stress on global water resources. Increasing temperature and changing precipitation patterns due to climate change have affected the existing balance between different components of the hydrologic cycle. The consequence is increased stress on groundwater resources, which currently contribute almost one-third to the global freshwater supply (Margat & Van der hun, 2013; Shiklomanov, 1993) and will likely contribute a greater share in the future due to increasing variability in surface water supplies. Regions where surface water availability is limited and groundwater is the dominant source of water (Butler et al., 2018) have the highest risk of groundwater depletion, risking food and water supply in these regions.

Excessive groundwater pumping is responsible for adverse impacts, such as permanent aquifer storage loss, land subsidence, arsenic contamination, saltwater intrusion, and infrastructure damage (Camp et al., 2014; Chaussard & Farr, 2019; Erban et al., 2013; Galloway & Burbey, 2011; Ojha et al., 2018). In spite of its importance, many regions of the world with intensive groundwater withdrawals and storage loss are poorly monitored, making it challenging to quantify groundwater storage loss. Fortunately, remote sensing datasets have the potential to fill this gap by providing global scale datasets that are proxies of drivers affecting groundwater storage change.

One of the most visible and harmful effects of groundwater depletion is land subsidence caused by compaction of aquifer materials following the loss of storage. Estimating the amount of subsidence can be a way to quantify storage loss in confined aquifer systems (Smith & Majumdar, 2020). In-situ measurement methods for quantifying

subsidence exist but are spatially far too sparse to be used in accurate subsidence estimation at large scale. Interferometric Synthetic Aperture Radar (InSAR) deformation data, processed from backscattering phase information obtained by Synthetic Aperture Radar (SAR) sensor, can be a very reliable source of subsidence data, providing data with ~1 cm accuracy and at a fine spatial resolution of ~100 m. Satellite-based subsidence measurements have been used to monitor groundwater storage loss in many aquifer systems (Chaussard et al., 2013, 2014; J. Chen et al., 2016; Faunt et al., 2016; Higgins et al., 2014). In spite of its success, processing, validating and time series analysis of InSAR data involves high computational effort. Therefore, InSAR-based groundwater as well as land subsidence studies have been limited to local or regional level.

Fan et al. (2013); Graaf et al. (2015) used physical groundwater models to simulate groundwater heads at shallow depth globally, using datasets from available government databases and outputs from global hydrological models, respectively. Wada et al. (2010) derived global groundwater depletion data from groundwater recharge and abstraction data. However, subsidence associated with groundwater withdrawal is difficult to measure using traditional groundwater models as they require variables that are readily available, especially at global scale. Remote sensing and global model-based datasets offer direct measurements of some of these variables or sometimes might be assembled to form proxies. Previously, Smith & Majumdar (2020) used machine learning approach for groundwater pumping associated subsidence mapping for the entire western United States combining remote sensing datasets and InSAR-derived subsidence data. Herrera-García et al. (2021) employed statistical technique to estimate global land subsidence susceptibility,

which provides a map of subsidence risk categorized by probability values. However, no existing study have quantified the magnitude of subsidence and associated groundwater storage loss globally.

In this study, we present a machine learning based method to map of pumping-induced land subsidence globally at a high spatial resolution (~2 km), using remote sensing and model-based hydrologic, land use, climatic and geologic datasets. Our method produces subsidence estimates in <1cm/year, 1-5 cm/year and >5 cm/year classes. To the authors' best knowledge, this is the first global scale subsidence study that maps subsidence magnitude. Additionally, the machine learning model provides a global subsidence probability map which is critical in identifying regions under future subsidence threat. As land subsidence is the primary mechanism for groundwater storage loss in unconsolidated confined aquifers, the produced global subsidence map can be interpreted as a first-order map of global confined aquifer groundwater storage loss. Results of this study might be helpful in identifying groundwater stressed areas where groundwater resources have not been critically studied before and can work as a baseline for regional studies over subsidence threatened areas.

2. BACKGROUND

2.1. MECHANISM OF LAND SUBSIDENCE

In this study, by land subsidence we refer to the downward vertical deformation of the ground surface, and we only consider subsidence caused by groundwater extraction. Groundwater withdrawal causes the pore pressure in the aquifer to decrease, resulting in an increase in effective stress and subsequent compaction in aquifer sediments (Fetter, 2001). Weak sediments, such as clay, compact more than stronger sediments, like sand and gravel. Unconsolidated sedimentary aquifers that have higher clay content are more susceptible to land subsidence whereas aquifers formed of consolidated rocks, such as carbonate rocks, sandstone or basalt, have very little compaction.

Subsidence can be inelastic (permanent), and elastic (temporary) based on the preconsolidation history of the sedimentary layer. The lowest hydraulic head experienced in a sedimentary layer is defined as the preconsolidation head, and corresponds to the highest effective stress experienced at that location. When the hydraulic head drops below this level, inelastic deformation occurs and subsidence is permanent. (Galloway & Burbey, 2011, Faunt et al., 2016; Smith et al., 2017; Sneed, 2001). Inelastic subsidence is of particular concern as layers undergoing such deformation do not rebound even after recharge. In contrast, sediments undergoing elastic subsidence can go back to their previous state with natural or managed groundwater recharge. At the depths and pressures of virtually all aquifer systems, sands do not experience inelastic deformation, and the inelastic deformation is essentially limited to clays.

The relationship between groundwater pumping and land subsidence can be expressed by the following equation (Fetter, 2001) -

$$\Delta b = \Delta h S_{sk} b_0 \quad (1)$$

where Δh is the change is hydraulic head (h) due to pumping, b_0 is the thickness of the aquifer experiencing the change in head, S_{sk} is the skeletal specific storage and Δb is the change in thickness of the aquifer, which results in surface deformation. For weaker sediments like clay, S_{sk} has higher values than stronger sediments. Therefore, subsidence (Δb) will be greater in a clay dominated aquifer than a sand-gravel aquifer if equal thickness (b_0) of both aquifers are experiencing similar change in head (Δh). Depending on whether subsidence is elastic or inelastic, the value of S_{sk} is typically on the order of 10^{-5} and $10^{-3}\ m^{-1}$, respectively (Smith & Knight, 2019; Sneed, 2001). Since the inelastic S_{sk} value is roughly two orders of magnitude higher than the elastic S_{sk} (Faunt, 2009) and only clays deform inelastically, clay has the dominant signal during inelastic subsidence (Smith & Knight, 2019) and permanent subsidence only happens in clay-rich layers, especially in areas with high groundwater demand but low recharge.

2.2. LINK BETWEEN SUBSIDENCE AND GROUNDWATER STORAGE LOSS

Land subsidence can be a measure of groundwater withdrawal and associated storage loss in aquifers. Aquifers can be broadly classified into confined and unconfined aquifers. Confined aquifers are geologic units under pressurized condition, and confined by aquitards (generally fine-grained sedimentary layers) that disconnect the aquifer from the ground surface (Fetter, 2001). Confining units can be leaky if they are thin or not

extensive and are often referred to as semiconfined aquifers. Groundwater withdrawal from confined or semiconfined aquifer causes it to compact (subside) while maintaining its saturated, pressurized state (Fetter, 2001; Freeze & Cherry, 1979; Jacob, 1940). Because of existence of clay and soil matrix controlling aquifer storage, these aquifers are at high risk of inelastic subsidence. Unconfined aquifer can also undergo inelastic subsidence though it might not be as significant as confined aquifer.

Water holding capacity in confined and semiconfined aquifers are controlled by Specific Storage (S_s) which is related to compressibility of soil matrix and elasticity of water (Faunt, 2009; Lohman, 1972). Following equation defines how S_s is connected to volume of water removed per unit area of aquifer (ΔS), change in hydraulic head (Δh) and confined aquifer thickness (b)-

$$S_s = \frac{\Delta S}{(\Delta h)b} \tag{2}$$

S_s is the combination of previously explained skeletal specific storage (S_{sk}) and specific storage for water (S_{sw}) following the equation:

$$S_s = S_{sk} + S_{sw} \tag{3}$$

S_{sk} and S_{sw} account for the compressibility of soil matrix and elasticity of water, respectively (Freeze & Cherry, 1979; Jacob, 1940; Kuang et al., 2020). Except some hard-rock aquifers where S_{sw} value dominates over S_{sk}, usually S_{sk} value is of 1 to 3 orders of magnitude higher than S_{sw} in aquifers with compressible sediments (Smith et al., 2017). For inelastic deformation, when S_{sk} is 2 to 3 orders of magnitude higher than S_{sw}. Under these conditions, S_{sk} controls specific storage and specific storage can be written equal to

skeletal specific storage ($S_s = S_{sk}$). Considering S_{sw} as minimum, equation 1 can be written as:

$$S_s = \frac{\Delta b}{(\Delta h)b} \tag{4}$$

Comparing equation 2 and 4, we get:

$$\Delta S = \Delta b \tag{5}$$

Equation 5 connects groundwater storage change to surface deformation (subsidence), considering S_{sw} is negligible. This explains how land subsidence can work as a proxy or direct estimate of groundwater storage loss. Inelastic subsidence in confined aquifer system is associated with significant time lag, therefore, change in aquifer storage ΔS is not reflected on the initial values of land subsidence. When aquifer is pumped, high permeability sediments like sand response quickly to groundwater pumping and resulting deformation is small compared to clay (Sneed, 2001), which is the lower bound of actual land subsidence. In the later stage, clay layers start to release water into the dewatered sand pores though the total water storage of the aquifer remains unchanged. Pore water pressure in clay drops increasing the effective stress followed by rearrangement of clay particles in dispersed formation and significant deformation. The release of water from clay continues until the expected total deformation Δb is reached balancing itself with the amount of water pumped ΔS. This delayed drainage time can range from several weeks to years (Chen et al., 2016; Smith et al., 2017, Smith & Majumdar, 2020).

2.3. WATER-BALANCING USING REMOTE SENSING AND MODEL-BASED DATASETS

Remote sensing is a spaceborne technique of collecting geographic data and extracting useful information from it, and usually refers to data collected by aircraft or satellite sensors. Such datasets can be optical or multispectral depending on the sensor. Sometimes datasets assimilated from multiple sensors or estimated from multispectral datasets using different methods are also referred to as remotely sensed datasets. In the last decade, remote sensing datasets have been increasingly used in wide-ranging fields of earth science, such as environmental monitoring, land use change detection, weather forecasting, natural hazard tracking, water resources monitoring, etc.

Monitoring hydrologic fluxes using remote sensing techniques is rapidly rising due to their large spatial and temporal coverage. Available in-situ estimates are spatially sparse, have temporal gaps, and are highly inconsistent, restraining their capacity in analyzing hydrologic phenomena for larger spatiotemporal scales. Due to causes like spatially coarse, scant publicly available records, and unwillingness to share data, groundwater monitoring with in-situ records has been proven to be very challenging for large basins. Such difficulties have increased dependency on remote sensing datasets in tracking groundwater. Despite improvements in technology, capacity of airborne remote sensing sensors is limited to capture hydrologic and atmospheric fluxes up to a few centimeters of the subsurface (Babaeian et al., 2019). Therefore, there is no direct estimate of groundwater storage from airborne sensors. Some widely used alternatives are using groundwater models and water balancing with constituents of the hydrologic cycle. Groundwater model requires a large number of geologic and hydrologic variables as input. Though hydrologic

variables are relatively easier to compile, limited availability of geologic data makes it difficult to construct reliable groundwater models. Compared to such models, water balancing using remotely sensed datasets is a much more achievable approach and has been applied in different hydrologic applications including groundwater storage estimation (Gleeson et al., 2012; Mladenova et al., 2020; Munier et al., 2014). Such water balancing approaches are based on the principles of conservation of mass in the hydrologic cycle and can be expressed by the equation:

$$P = R + ET + \Delta S \tag{6}$$

where P is precipitation, R is streamflow, ET is evapotranspiration and ΔS is change is storage in soil. Some studies have also incorporated total water storage (TWS) data from Gravity Recovery and Climate Experiment (GRACE) satellite and GRACE-derived datasets in monitoring aquifer storage (Rateb et al., 2020; Yin et al., 2021). In this study, several remotely sensed and model-based hydrologic fluxes, such as evapotranspiration, precipitation, and soil moisture, were assimilated into a machine learning model to help the model develop an understanding of interrelations between these fluxes. Land subsidence data from InSAR, which is a proxy of groundwater storage change, was used in training the model. Soil moisture and land subsidence together represent ΔS in the model. We did not incorporate any streamflow data in the model, instead, a 'River Distance' dataset was used. Moreover, the other hydrologic datasets were not added in the same unit. Therefore, the water balance represented in the model can be seen as pseudo-water balance that helps the model interpret the interaction between the fluxes and how that drives land subsidence. Despite having access to remote sensing-based estimates of

hydrologic fluxes, model-based datasets were used in some cases to incorporate high resolution, gap-filled datasets.

2.4. INSAR IN AQUIFER MONITORING

InSAR has emerged as a powerful remote sensing technique for tracking land surface changes at very fine resolution and large spatial scale. It relies on data collected by SAR platform that emits electromagnetic signals at microwave wavelength and collects backscattering signals from earth's surface as amplitude and phase information. SAR signal can penetrate cloud cover and collect data day and night, being an active source, providing continuous data coverage irrespective of weather and daylight conditions. Higher wavelength SAR (L and P bands) can penetrate deeper into vegetation cover and subsurface, increasing its capacity to extract more information. Using the concept of synthetic aperture, where sensor's motion is used to simulate long synthetic antenna despite the short antenna onboard, SAR can collect high spatial resolution data of ground surface. Because of the side-looking characteristics of the sensor, data collected are in the line-of-sight (LOS) direction but can be distributed to vertical and horizontal components. The quality of SAR data depends on multiple factors like satellite viewing geometry, surface roughness, and dielectric permittivity. For example, high wavelength SAR (L-band) can penetrate deeper into soil with low $\in$ (low soil moisture) than soil with high $\in$ (high soil moisture). Radiometric and topographic corrections are applied to remove effects of erroneous topographic information (steep slope emitting bright signal) and distorted geometry from the data. InSAR uses interferometry to detect change in phase from pairs of images collected from successive satellite passes over the same location. At least two

images of the same geolocation are required to employ interferometric technique for change detection in topography. In addition to ground deformation, factors like viewing angle from satellite, water vapor in troposphere, change in orbital path, can contribute to the change in phase data. These effects must be removed during InSAR processing to account for the phase change contributed by topographic movement only. InSAR can accurately detect land disturbance from volcanic activities, earthquakes, landslides, and land subsidence. Despite having widespread advantages, the technique is prone to spatiotemporal decorrelation and vegetation, tropospheric, and ionospheric effects.

In hydrology, InSAR has been applied in aquifer storage quantification, analyzing aquifer physical properties, soil moisture estimation, etc. Groundwater withdrawal causes land surface to subside. Based on the property of geologic unit from where water has been extracted, the subsidence can be elastic or inelastic. Recent high temporal resolution data has enabled InSAR to monitor seasonal recharge and discharge dynamics in aquifers (Chaussard & Farr, 2019; Hoffmann et al., 2001). On the other hand, aquifers with inelastic subsidence undergo long-term subsidence which can also be detected. As mentioned in the previous section, no available remote sensing technique provides a measure of groundwater level. Exploring the relationship established in equation 5 can therefore function as an effective mechanism to indirectly estimate aquifer storage change. Land deformation is directly linked to the change in head in the aquifer. Thus, subsidence data from InSAR can serve as a proxy for groundwater storage change. Due to SAR's large spatial coverage and frequent temporal records, InSAR can be applied to monitor groundwater storage for large aquifers irrespective of regional and international boundaries with high accuracy.

2.5. DRIVERS OF LAND SUBSIDENCE AND POTENTIAL PROXIES

Land subsidence is contributed by several factors like aquifer skeletal specific storage (S_{sk}), thickness of compressible sediment (b), change in hydraulic head (Δh) due to pumping and consolidation record of the layer experiencing subsidence. For local and even for some regional levels, these datasets are available and can be coupled together to estimate land subsidence. However, data scarcity, unwillingness to share data publicly, heterogeneity of collected data and coarse resolution of available data make it increasingly difficult to compile all required datasets at global scale. Global coverage of remotely sensed datasets can bridge this void, such as balancing variables like precipitation, ET, soil moisture and TWS data in heavily groundwater exploited areas can give an estimate of hydrologic change. TWS data from GRACE satellite along with other hydrologic variables have been used in multiple studies to model groundwater storage change in regional scales (Rateb et al., 2020; Yin et al., 2021), but it has coarse spatial and temporal resolution (Leroux & Pellarin, 2016; Rodell, 2013). To avoid uncertainties arising from such coarse resolution, GRACE data was not incorporated in the model with other high resolution datasets. Rather InSAR processed land subsidence data can function as a proxy of groundwater storage change as discussed in previous section. The model assimilates multiple hydrologic, land use and geologic variables which can be related to groundwater storage change.

3. DATA

3.1. INPUT DATASETS

Input variables (predictors) of this model include remotely sensed and model-based global gridded datasets, within geographic boundary of -180° W, -60° S, 180° E, 90° N, that are proxies of principal hydrologic, geological, and anthropogenic processes that drive land subsidence. The variables were integrated in a supervised machine learning model named random forests. Random forests is a decision-tree based algorithm that can establish nonlinear relationship between predictors and response variable to generate final model prediction (Biau, 2012; Breiman, 2001, Smith & Majumdar, 2020).

High resolution (~4 km) precipitation, soil moisture, evapotranspiration (ET), Reference evapotranspiration (RET) and temperature (min and max) data were obtained from TerraClimate monthly climate and climatic water balance datasets (Abatzoglou et al., 2018). TerraClimate climate datasets are generated using climatologically aided interpolation (a downscaling technique) of WorldClim (Fick & Hijmans, 2017), CRU JRATs4.0 (Morice et al., 2012) and JRA-55 (Kobayashi et al., 2015) data while the water balance datasets come from a one-dimensional water balance model (Abatzoglou et al., 2018). Despite availability of finer spatial resolution remotely sensed ET data, such as MODIS global terrestrial evapotranspiration data (Mu et al., 2013), ET data from Terraclimate was selected because of its global coverage, including arid regions like Southern Africa, Western Asia, Central Australia where other datasets like MODIS has no coverage.

Table 3.1 in shows a list of the datasets used in the model along with their original spatial resolution and sources.

Table 3.1: List of datasets used in the model.

Variable	**Spatial resolution**	**Source**
Soil moisture	~ 4 km	Abatzoglou et al. (2018)
Precipitation	~ 4 km	Abatzoglou et al. (2018)
Evapotranspiration	~ 4 km	Abatzoglou et al. (2018)
Reference Evapotranspiration	~ 4 km	Abatzoglou et al. (2018)
Irrigated Area	~ 1 km	Meier et al.(2018)
Population Density	~ 1 km	CIESIN (2018)
Temperature	~ 4 km	Abatzoglou et al. (2018)
Percentage of Clay	250 m	Hengl (2018)
Sediment Thickness	~ 1 km	Pelletier et al.(2016)
Major River	Vector data	GRDC (2020)
EVI	250 m	Didan (2021)
NDWI	500 m	Vermote (2021)
Aridity Index	~ 1 km	Trabucco & Zomer (2019)
DEM	30 m	Farr et al. (2007)

Global irrigation area dataset at ~1 km resolution by Meier et al. (2018) was used in this study as one of the land use datasets. This dataset has been developed combining remote sensing technique and downscaled statistics-based irrigated area data (Siebert et al.,

2013), commonly known as Food and Agriculture Organization of the United Nations (FAO)'s irrigation dataset. Considering the strength and limitations of solely statistical or remote sensing approach (Siebert et al., 2013; Teluguntla et al., 2014; Thenkabail et al., 2009), this fusion technique of irrigated area classification provides the most reliable source of global irrigation data. Gridded population dataset of ~1 km resolution (CIESIN, 2018) was included in the model to capture subsidence in populated areas occurring from aquifer pumping.

As discussed in the previous section, existence of clay is the major geologic factor for inelastic subsidence. High resolution (250 m) % clay content data at different depths (0, 10, 30, 60, 100, 200 cm), generated from a machine learning model based on soil information (Hengl, 2018), was collected. In addition, we obtained ~1 km resolution global sediment thickness data (Pelletier et al., 2016). The highest value of sediment thickness in this dataset is 50 which denotes sedimentary deposit thickness of ≥ 50 meter.

Additional datasets included as predictors in the model were aridity index (Trabucco & Zomer, 2019), MODIS-derived Normalized Difference Water Index (NDWI) (Vermote, 2021) and Enhanced Vegetation Index (EVI) (Didan, 2021), SRTM-derived percent slope, and a vector dataset of major global rivers (GRDC, 2020).

3.2. LAND SUBSIDENCE DATA

Supervised machine learning algorithm requires training dataset to establish relationships between training data and input variables. Average vertical land subsidence rate was used as training data in our machine learning model. Subsidence data was collected for 36 regions of the world. InSAR can be a reliable tool for measuring

deformation in aquifers at high spatial resolution (Castellazzi & Schmid, 2021; Chaussard et al., 2013, 2014; Reeves et al., 2011). In this study, we collected deformation data using both primary and secondary methods. Primary method refers to downloading SAR sensor's data and processing it directly to extract inelastic deformation information over high groundwater stressed areas, such as Quetta valley in Pakistan, Qazvin region in Iran, North China plain and Hefei region in China, and San Luis Valley, Colorado in the United States. Processed subsidence data over California and Arizona in US were collected from California Natural Resources Agency and Arizona Department of Water Resources, respectively. Considering the computational effort required in processing InSAR data, we also collected InSAR data from secondary sources. These sources consist of groundwater studies that used InSAR information to determine aquifer vertical deformation. A list of these training data sources is provided in Table 3.2, where (1-7) were obtained from direct InSAR processing; (8-36) were georeferenced and classified from research articles with InSAR processed land deformation map and 37 is globally available GNSS-based coastal subsidence data.

Table 3.2: Sources of InSAR-based land subsidence data.

Primary Data Sources (InSAR Processing)			
No.	**Country**	**Region**	**Sources**
1	China	Hebei	processed
2	China	Hefei	processed

Table 3.2: Sources of InSAR-based land subsidence data (cont.)

3	Iran	Qazvin	processed
4	US	San Luis Valley, Colorado	processed
5	Pakistan	Quetta	processed
6	US	California	California Natural Resources Agency
7	US	Arizona	Arizona Department of Water Resources
Secondary Data Sources (Research Articles)			
8	Australia	Perth	(Castellazzi & Schmid, 2021)
9	Bangladesh	GBM Delta	(Higgins et al., 2014)
10	China	Beijing	(Chen et al., 2016)
11	China	Yellow River	(Higgins et al., 2013)
12	China	Shanghai	(Dong et al., 2014)
13	China	Wuhan	(Zhou et al., 2017)
14	China	Xian	(Qu et al., 2014)
15	China	Tianjin	(Luo et al., 2014)
16	Egypt	Nile Delta	(Gebremichael et al., 2018)
17	England	London	(Bonì et al., 2017)
18	Indonesia	Multiple Locations	(Chaussard et al., 2013)
19	Indonesia	Bandung	(Ge et al., 2014)
20	Iran	Marand Plain	(Andaryani et al., 2019)
21	Iran	Tehran	(Haghighi & Motagh, 2019)
22	Iran	Mashhad	(Khorrami et al., 2020)
23	Iraq	Tigris Euphrates Basin	(Rateb & Kuo, 2019)
24	Italy	Po Delta	(Corbau et al., 2019)
25	Italy	Venice	(Bock et al., 2012)
26	India	Delhi	(Garg et al., 2022)
27	Mexico	Mexico City	(Chaussard et al., 2014)
28	Nigeria	Lagos	(Cian et al., 2019)

Table 3.2: Sources of InSAR-based land subsidence data (cont.)

29	Philippines	Manila	(Zoysa et al., 2021)
30	Spain	Murcia	(Ezquerro et al., 2020)
31	Turkey	Karapinar	(Orhan et al., 2021)
32	Turkey	Bursa	(Aslan et al., 2019)
33	Taiwan	Yunlin	(Hsu et al., 2015)
34	US	Huston	(Miller & Shirzaei, 2019)
35	Vietnam	Ho Chi Minh	(Minh et al., 2015)
36	Vietnam	Hanoi	(Nguyen et al., 2022)
37	Coastal	-	(Shirzaei et al., 2021)

4. METHODOLOGY

4.1. PREPROCESSING INPUT DATASETS

Datasets in gridded format were downloaded from multiple sources, mostly from Google Earth Engine (GEE) platform. Timeline 2013-2019 was considered as the temporal window for aggregating (mean/median) time varying datasets before downloading from GEE. For integrating data in a machine learning model, all datasets need to be in a uniform format under the same coordinate system and similar number of pixels in the gridded format. We projected all the downloaded datasets in WGS 1984 geographic coordinate system. Depending on the original resolution, the datasets were down/upscaled to a resolution of 0.02 deg (~2 km) to achieve a uniform grid size.

We applied gaussian filter (kernel filtering window size 3σ pixels), available in Scipy library (Virtanen et al., 2020), on irrigated area and population datasets to add a smoothing effect and remove noise. Gaussian filter normalized the datasets within an interval range of 0 to 1, where value towards 1 represents higher density of respective land use class and vice versa. The vector dataset of major rivers was processed into a gridded format, representing distance from major global rivers, and incorporated in the model as well.

Clay % data at 200 cm depth were multiplied with the sediment thickness dataset to create a new ‘Clay Thickness’ dataset. Additionally, a dataset indicating the likely presence or absence of a confining layer was produced as part of this study. This was produced based on the presence of basins surrounded by high elevation, and regions within

25 m elevation of sea level. These regions are likely to have higher and extensive deposition of fine-grained sediments.

4.2. LAND SUBSIDENCE TRAINING DATA

Figure 4.1 provides the workflow of our modeling steps. Average vertical deformation rate (unit in cm/year) of subsiding areas, where significant groundwater pumping has been recorded historically, were used as training data in the model. Subsidence data collection by directly processing InSAR data requires a significant time, being computationally expensive. Therefore, we relied on primary (directly processed/collected already processed data) as well as secondary InSAR data sources to accumulate land subsidence information for this study. Primary data include InSAR data over Quetta valley in Pakistan, Qazvin in Iran, North China Plain and Hefei region in China, California, and Arizona in US.

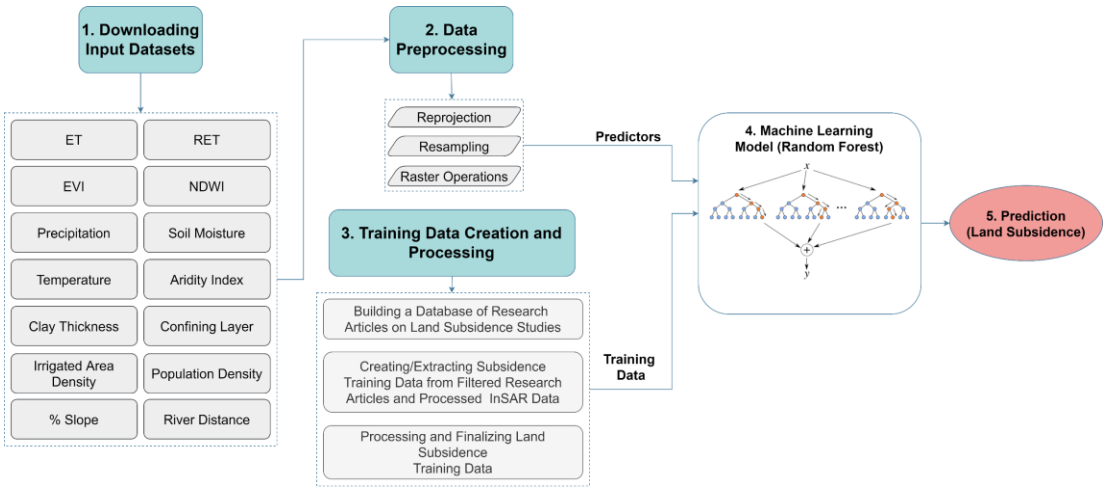

Figure 4.1: Workflow of data processing and machine learning model integration.

For collecting secondary data, we developed a database of almost 300 subsidence studies around the world. The database was filtered for studies that analyzes InSAR data to study subsidence that is associated with groundwater pumping only. As discussed in Section 2.2, inelastic subsidence can continue up to decades. So, we extended the timeline back to 2006 so that our model can capture long-term subsidence that is happening in current years. Inelastic subsidence happening in period 2013-2019 might be because of groundwater pumping around 2006 in pumped regions. We took images of average vertical deformation maps from filtered research articles. The maps were georeferenced using ArcGIS tools and ArcGIS built-in tree based algorithm was used to extract deformation information from the georeferenced maps. If deformation was reported in the Line of Sight (LOS) direction, we converted it to vertical values using satellite incidence angle. The extracted data were classified into three classes: <1 cm/year subsidence, 1-5 cm/year subsidence, >5cm/year subsidence. Subsidence data collected from research articles are referred as georeferenced subsidence data in this study, but they are based on InSAR processing. We also integrated vertical land deformation data, based on global navigation satellite system (GNSS), from Shirzaei et al. (2021) for coastal regions around the world and classified them into our subsidence categories. Out of nearly 3000 samples of this dataset, 98% falls into the <1 cm/year class. For this global study, <1 cm/year subsidence is considered as negligible to no subsidence class while the other classes represent medium to significant subsidence. But it should be noted that 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.

Finally, directly processed subsidence data (primary) and the georeferenced subsidence data (secondary) were merged and resampled to a spatial resolution of 0.02 deg (~ 2 km) to form training subsidence data. This training dataset was used to train the machine learning model so that the model can understand inter-relations between subsidence and its driving variables.

4.3. RANDOM FORESTS MODEL PREDICTION

Variables interplaying in land subsidence have complex non-linear relationships that can be explored using a machine learning model. Random forests model performs well with non-linear variables and creates randomness in data to minimize noise (Breiman, 2001). It does not require input variable normalization, being a tree based algorithm, so, datasets with values in varying units can be assimilated in such a model without concern. Based on the type of output, machine learning models can be designed as a classifier or a regressor. Our model predicts subsidence in three classes, hence can be categorized as a random forests classifier.

Multiple number of trees (*n_estimators*) are generated and trained by the random forests model. In a classifier type model, the individual results of these trees are ensembled by majority voting to produce the final model outcome. Each tree is trained with a subsample of the original training data and can be allowed to grow up to a certain depth represented by the parameter *max_depth*. The subsample data for each tree is generated by a technique known as bootstrap aggregating (bagging) which involves random subsampling from the original data without replacement and it helps in avoiding overfitting. One key feature of the random forests algorithm in random selection of

predictors while growing a decision tree. At each node (splitting point), the model selects a random *max_features* number of predictors to make further splits. Splitting continues until a split node holds minimum numbers of samples and generates leaf nodes containing a minimum number of samples, defined by the parameters *min_samples_split* and *min_samples_leaf,* respectively. Splits in random forests are made, using values of the randomly chosen predictors, in such a way that model variance is minimized. With bagging and random splitting techniques dealing with overfitting and variance reduction, random forests does not require a separate validation dataset from the training data to help with these issues (Breiman, 2001). *n_estimators*, *max_depth*, *max_features*, *min_samples_split* and *min_samples_leaf* are hyperparameters of the model and were optimized using K-fold cross validation to improve model accuracy and avoid overfitting.

Random forests model creation requires a primary training dataset (also referred as response variable), in this case land subsidence data collected by primary and secondary methods. Machine learning model learns the relationship between input variables using the response variable. To create the training dataset, only grid pixels having land subsidence information were filtered from the input datasets. A second filtering was performed on those pixels to select only the pixels that do not have no data value. The resulting filtered dataset is referred as the original training dataset of the model. This training dataset was randomly split into train and test sets, with 70% data on train set and 30% data on test set, for model calibration and validation purpose. 80% of the subsidence observations in our training dataset belongs to <1 cm/year class, while 13.5% in 1-5 cm/year and 6.5% in >5 cm/year class, creating imbalance in the dataset. A *'balanced'* class weight was assigned

to prevent the model from being biased towards the majority class (<1 cm/year here). *'Balanced'* assigns a lowest to highest class weight value to the most frequent to the least frequent class to deal with dataset imbalance.

5. RESULT ANALYSIS AND DISCUSSION

5.1. MODEL ACCURACY

Performance of the model was assessed with *F1-score*, being a machine learning classifier model, on testing set for individual classes and for all classes. *F1-score* for <1 cm/year, 1-5 cm/year and >5 cm/year are 0.95, 0.7 and 0.88, respectively. Considering all classes, the *macro F1-score* is 0.84. To avoid overfitting, we optimized the hyperparameters and compared model's performance with train and test set. Table 5.1 shows *F1-score* for discrete classes and for the entirety.

Table 5.1: Model performance score on test set.

	<1 cm/year	*1-5 cm/year*	*>5 cm/year*	*Macro F1-score*
Train Set	0.97	0.86	0.96	0.93
Test Set	0.95	0.70	0.88	0.84

F1-score for <1 cm/year and >5 cm/year are satisfactory. The difference between score values for train and test sets for these classes indicate that they have been optimally trained by the model. However, the *F1-score* for 1-5 cm/year class is not as high as the other two classes, and the difference between scores on train and test sets shows that there is some overfitting in this class. Machine learning models with imbalance datasets can often get biased towards the majority observation class (Johnson & Khoshgoftaar, 2019).

With only 20% observations in 1-5 cm/year and >5 cm/year classes in the original training dataset, majority observations are in <1 cm/year class, generating bias towards this class, which might have affected the metrices of 1-5 cm/year class.

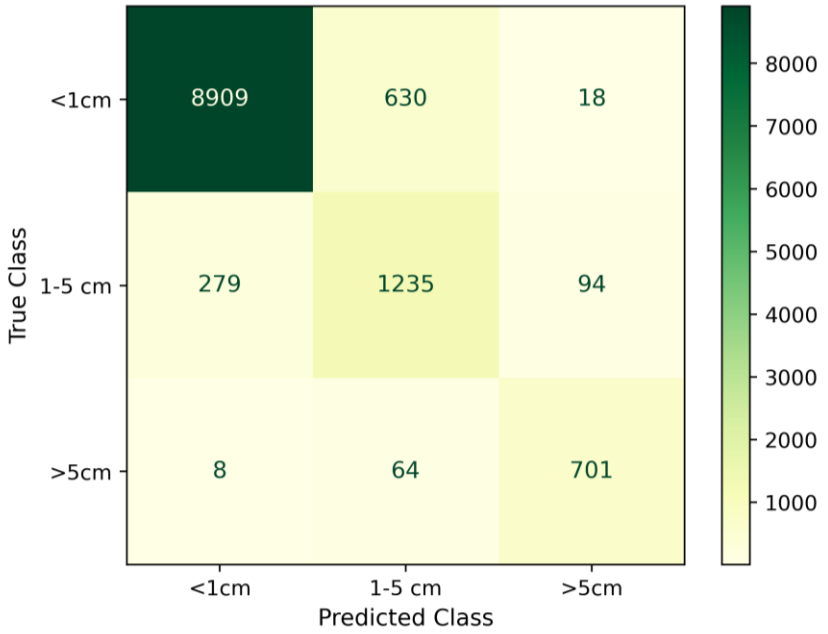

Figure 5.1: Confusion matrix of test set.

Confusion matrix (Figure 5.1) of the test set shows that the model misclassified approximately 23% of the 1-5 cm/year class, compared to 7% in <1 cm/year and 9% in >5 cm/year classes. We made random split of the original training dataset into train and test sets to ensure that both sets contain some pixels over all training regions. Therefore, even if some pixels are misclassified in one region, there are other pixels that are classified accurately to indicate ground deformation activity. Under these considerations, we accepted the slight overfitted metrices of the 1-5 cm/year class.

5.2. VARIABLE IMPORTANCE AND PARTIAL DEPENDENCE

To determine how our random forests model evaluated the input variables in predicting land subsidence, we assessed their importance in the model. Gini importance (mean decrease impurity), a method that is based on node purity (James et al., 2013), was chosen to measure the relative importance of variables. For each variable, mean decrease in impurity in all the trees due to that variable is estimated to find their contribution in the model.

Among the 15 variables used, the model considers soil moisture, temperature (min), clay thickness, confining layers, river distance, precipitation, and irrigated area density as the most important variables (Figure 5.2).

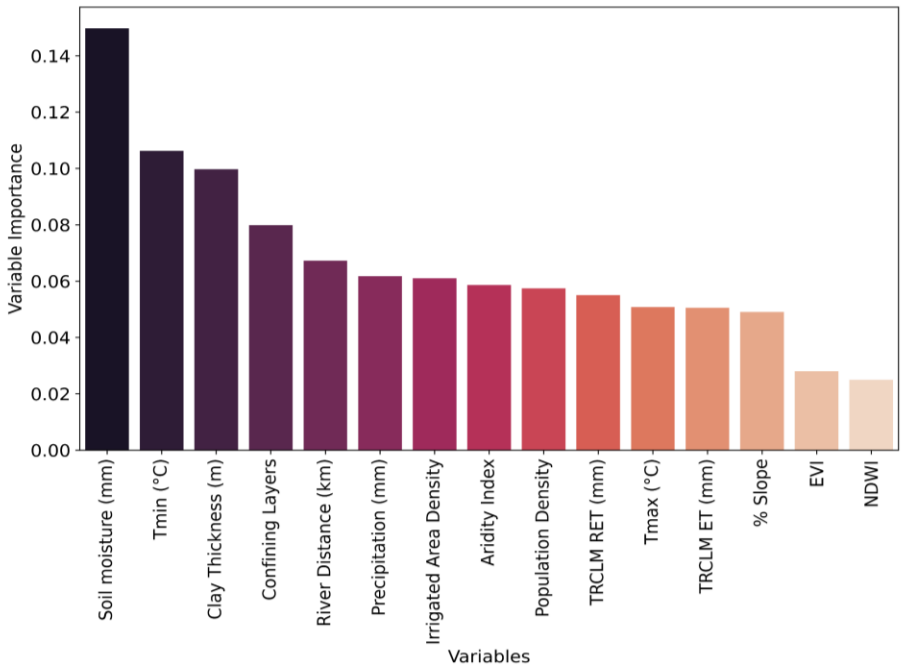

Figure 5.2: Variable importance plot for the model predictors.

Another essential evaluation technique of machine learning model is Partial Dependence Plots (PDP) that show if the input variables are able capture the known dynamics of physical processes in the model. Figure 5.3 presents the PDP plots of the key variables of our model for 1-5 cm/year subsidence class. The x-axis of the plots consists of variable values within 5-95th percentiles (except Confining Layers). The y-axis of the plots represents how subsidence prediction varies with a variable when all the other variables are also contributing to the model outcome. The input variables added to the model were direct measurements or proxies of principal drivers of land subsidence and groundwater withdrawal. While selecting each variable, we had a preliminary assumption on how they will affect subsidence prediction.

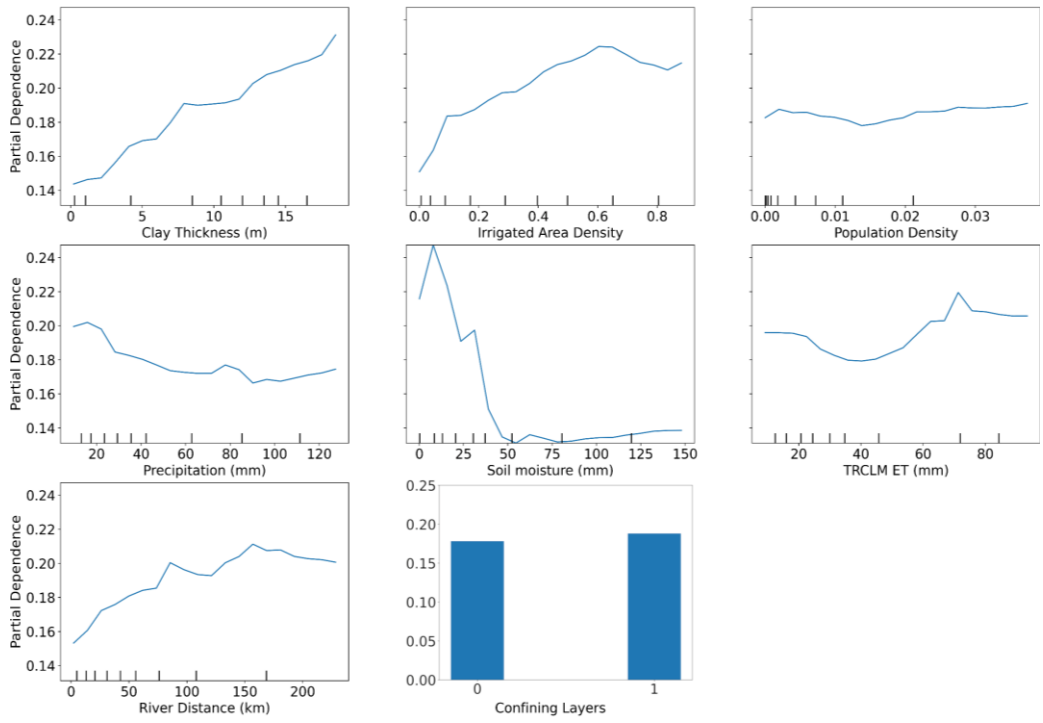

Figure 5.3: The partial dependence plot for the 1-5 cm/year class.

As discussed in Sections 2.1 and 2.2, presence of clay and fine-grained confining unit in confined or semi-confined aquifer are major impetus of inelastic subsidence. PDP plot of clay thickness demonstrates that subsidence probability increases with thickness of clay. Subsidence phenomena generally occurs over agricultural lands and developed areas where surface water availability is limited. Response from our irrigation, population density and river distance datasets justify this rational assumption. Irrigation demands more groundwater in arid and semi-arid regions where precipitation is lower than evapotranspiration and surface water sources are scarce, which increases subsidence susceptibility. Low soil moisture indicates high irrigation water demand in croplands. Soil moisture PDP plot shows increased subsidence probability at lower soil moisture values. Thus, the presented PDP plots confirm that response of hydrologic fluxes in the model are realistic. In contrast to other variables, variable confining layer only has 2 values (1 for presence of confining units, 0 for no confining units). Response of this variable in the model is more interpretable if PDP plots of this predictor for all classes (<1 cm/year, 1-5 cm/year, >5cm/year) are analyzed together (Appendix C). For <1 cm/year class, subsidence probability is high for no confining layer (value 0), signifying that- areas with no confining units tend to be less vulnerable to high subsidence. For the other classes, probability of subsidence increases with existence of confining units (value 1).

5.3. GLOBAL SUBSIDENCE MAP

Our model predicts subsidence in three classes: <1 cm/year, 1-5 cm/year, >5cm/year. As discussed before, we considered <1 cm/year class as the nominal or zero subsidence class. Figure 5.4 shows the global map of subsidence, zoomed in on regions

with high subsidence signatures, mapped by our random forests model. A larger map is showed in Appendix A. We want to clarify that our model only delineates subsidence related to aquifer system compaction from groundwater pumping, therefore, total subsidence estimates over some regions, which are undergoing subsidence from other sources, may not match.

The model maps considerable amount of subsidence in countries of East Asia: China, Taiwan, Vietnam, and Philippines. In China, high irrigation activities have been mapped over the North China Plain aquifer by global irrigation mapping attempts (Meier et al., 2018; Teluguntla et al., 2018). Major cities like Beijing, Shanghai, Wuhan, Xian, and Tianjin are in or nearby this region and heavily dependent on groundwater to support agriculture and urban needs (Chen et al., 2016; Dong et al., 2014; S. Higgins et al., 2013; Luo et al., 2014; Qu et al., 2014; Zhou et al., 2017). Our map shows high subsidence signature in this whole region indicating significant groundwater storage decline. Countries in South and middle-East Asia, such as, Bangladesh, India, Indonesia, Pakistan, Iran, Turkey, have high subsidence signals in particular regions as well. These predictions are in line with recent InSAR based groundwater studies for these regions (Andaryani et al., 2019; Aslan et al., 2019; Chaussard et al., 2013; Dang et al., 2014; Erban et al., 2014; Haghighi & Motagh, 2019; Higgins et al., 2014; Hung et al., 2010; Kakar et al., 2020; Khorrami et al., 2020; Minh et al., 2015; Nguyen et al., 2022; Orhan et al., 2021). Our model also predicts subsidence in irrigated regions over Afghanistan, Turkmenistan, Uzbekistan, Kyrgyzstan, Azerbaijan, and mixed land use areas over Syria, Jordan, and Israel, where we have not found previous land subsidence studies.

In Europe, recent studies have shown subsidence occurring in Spain, Italy and England (Bock et al., 2012; Bonì et al., 2017; Corbau et al., 2019; Polcari, 2019) with low magnitude of less or marginally higher than 1 cm per year. Vertical land movement data along the coasts of Italy, Spain, Portugal, and France from GNSS also shows deformation lower than 1 cm/year (Shirzaei et al., 2021), but subsidence of such magnitude can be significantly damaging considering the impacts of sea levels. Our model categorizes majority of subsidence in Europe as <1 cm/year. The map predicts large subsidence between 1-5 cm/year, primarily because of groundwater irrigation, in Albacete and Ciudad Real province, and Alto Guadalentín basin in Spain. Some 1-5 cm/year deformation signal has also been mapped in dominantly agricultural and industrial (Fabris et al., 2014) Po Delta region in Italy.

In North America, our map shows considerable subsidence in California, Huston and Arizona in US and in central Mexico, which follows historic subsidence observed in these regions (Borchers et al., 2014; Chaussard et al., 2014; Chaussard & Farr, 2019; Conway, 2015; Sharma et al., 2016). Subsidence in California and Arizona is due to excessive groundwater irrigation while urban dependency on groundwater is responsible for deformation in Huston and Mexico City. The model predicts subsidence in agricultural lands in Texas, within the boundary of the heavily pumped High Plains Aquifer (HPA) region (Butler et al., 2018), and in Washington state where Subsidence have not been reported before.

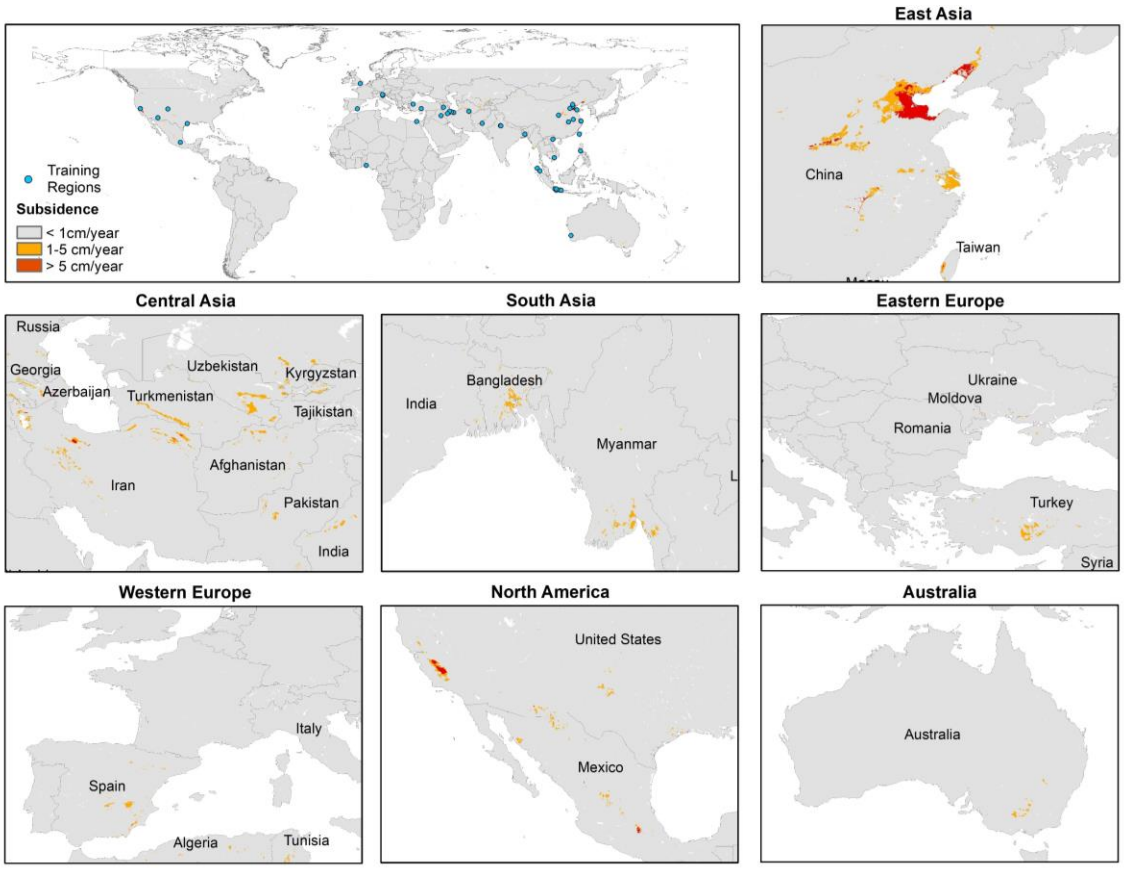

Figure 5.4: Global model prediction of subsidence.

InSAR studies in Africa (Cian et al., 2019; Gebremichael et al., 2018) have estimated less than 1 cm per year subsidence in Nile delta and in several coastal cities, such as Lagos, Banjul, Mombasa, Mogadishu. Prediction from the model shows similar subsidence rate in these areas. The model predicts 1-5 cm/year subsidence in Morocco, Algeria, and Tunisia over irrigated lands. 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 (Castellazzi & Schmid, 2021). Our map detects 1-5 cm cm/year subsidence in Murray-Darling region which might be happening in

the alluvial aquifers consisting of clay and silt (Lamontagne et al., 2014). No significant subsidence has been mapped in the South American continent except over some agricultural lands in Argentina.

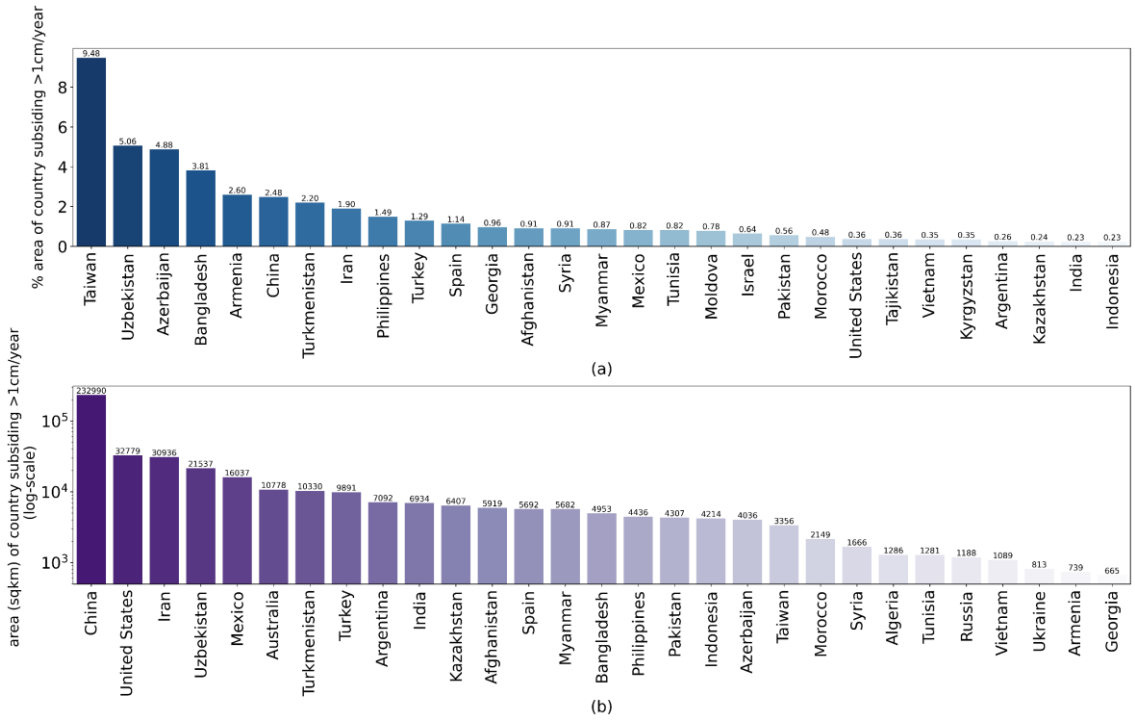

Figure 5.5: Country statistics of subsidence prediction of magnitude >1 cm/year.

Figure 5.5 shows (a) countries with highest percentage of subsidence with respect to their land area, (b) countries with highest area of subsidence predicted by our model. Percentage of subsiding area (>1 cm/year subsidence) (Figure 5.5(a)) is comparatively higher in Asian countries. Significant subsidence has been predicted in small island nations like Taiwan, Philippines. Subsidence has been mapped in humid regions as well, such as Bangladesh, Myanmar, China, Vietnam, Indonesia, indicating that high precipitation

supply is not sufficient to meet water demand in such populous regions. Globally, highest subsiding areas have been predicted in China followed by US and Iran (Figure 5.5(b)).

Subsidence prediction is dominating is Arid and Semi-Arid climates where climatic water deficit leads to higher groundwater dependency. Approximately, 84% of our >1 cm/year subsidence prediction lies on Arid and Semi-Arid regions. Comparison with MODIS land use product revealed that subsidence prediction is highest of approximately 60% on croplands. 10% of urban and built-up lands have subsidence of >1 cm/year. Significant subsidence, approximately 21%, has also been observed on vegetation land use. MODIS land use product has 75% overall accuracy (Friedl et al., 2010), hence there is possibility of intermixing between vegetation and cropland classes. Vegetation sometimes exists in vicinity of agricultural lands and might even periodically used for agriculture resulting in high subsidence on this class.

5.4. SUBSIDENCE PROBABILITY

To assess potential subsidence risks, we generated a probability map (Figure 5.6) showing probability of subsidence happening in greater than 1 cm/year magnitude. Appendix B shows the global map of >1 cm/year subsidence probability. This map can also be interpreted as a map of future subsidence warnings if groundwater resources are exploited at current rate. High subsidence probabilities around the world are noticed in mostly agricultural and urban lands where groundwater dependency is significant. Probability values are considerably high in regions where subsidence is already happening. Areas showing probability within 40-60% are under future subsidence threat of higher magnitude, particularly if they are in the vicinity of currently highly subsiding areas.

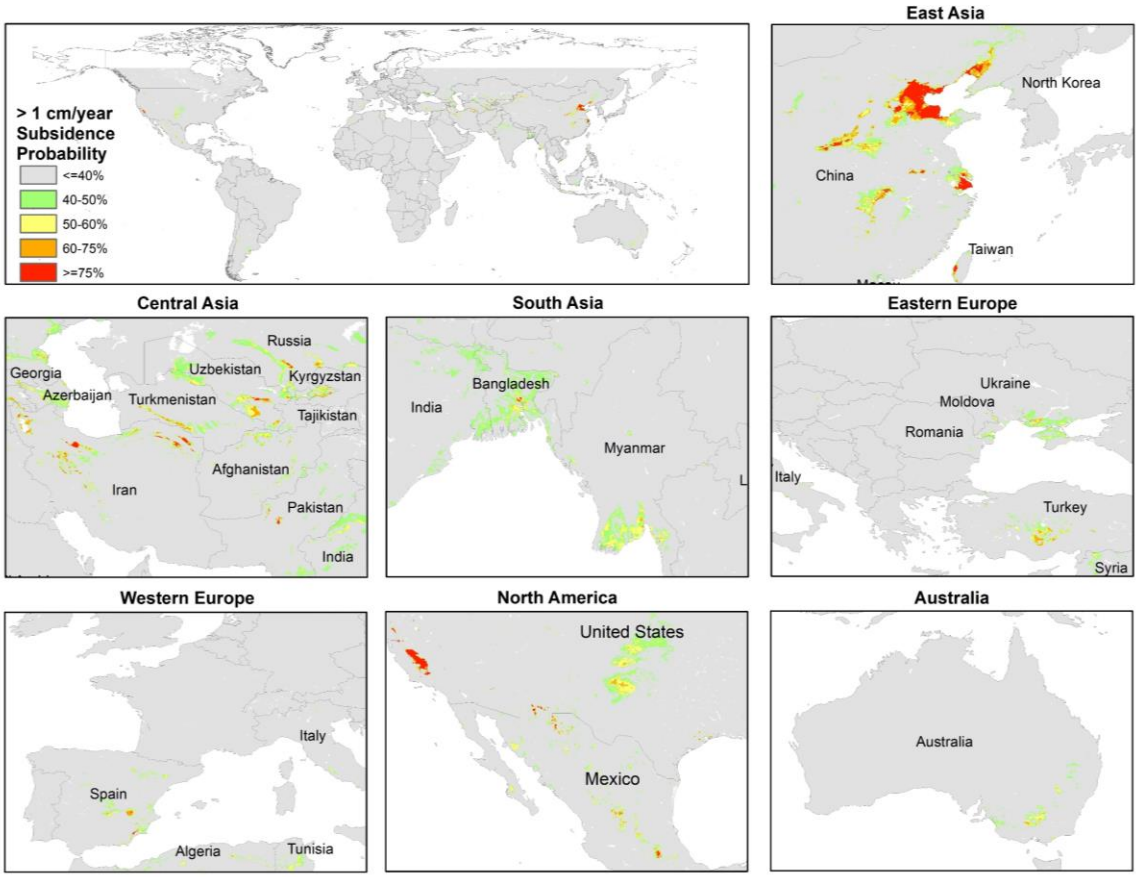

Figure 5.6: Probability of subsidence >1 cm/year.

In Asia, a vast region in North China Plain is significantly subsiding with more surrounding areas under deformation risk. Similar risk trend in observed in agricultural lands of Iran, Turkey, Israel, Jordan, and Syria in middle Eastern Asia and some countries in central Asia. 40-60% probability observed near the coasts of Bangladesh, Myanmar, Vietnam, Thailand, Philippines, Indonesia, Taiwan signify that increasing coast-side regions will be subsiding in near future if groundwater usage practice is not altered. In North America, high subsidence probability in California, Arizona, Texas in the US and in Mexico City in Mexico indicates already occurring deformation. Significant percentage of

lands in the HPA region shows probability (40 to 60%) of subsidence. Bulti-up areas in central Mexico also seem to be under high subsidence risk.

In South America, parts of Argentina show high probability of subsidence near coast. Land use classification products from MODIS and other cropland datasets indicate that this region in dominantly non-irrigated croplands (Friedl et al., 2010; Meier et al., 2018; Teluguntla et al., 2018), indicating less groundwater stress in the area. However, considering limitations of land use products, especially in identifying irrigated croplands, vertical deformation in this region is worth investigating using InSAR techniques. Large portion of Murray-Darling basin in South-western Australia shows probability of subsiding more than 1cm/year which supports results found in a previous study (Castellazzi & Schmid, 2021). In Europe, potential subsidence risk exists in Spain, Italy, and Ukraine. In Africa, parts of Egypt, Morocco, Libya, Algeria, and Tunisia shows subsidence risk.

5.5. LEAVE-ONE-AREA-OUT ACCURACY TEST

To test our model accuracy further, we designed the Leave-One-Area-Out (LOAO) accuracy test, inspired by a popular machine learning model evaluation technique called Leave-One-Out Cross-Validation. In previous sections, it was mentioned that we split the original training dataset randomly into train and test set for model fitting and evaluation purposes, respectively, as well as to ensure that the train set includes pixels from all regions to provide the model with better information. In the LOAO method, we did not make any random split, instead ran the model number of training regions times, leaving one training area completely out from training on each run and evaluating model performance on that

left out region (using it as test set). Thus, the model was iterated 36 times (did not consider coastal dataset during this test), each time without an area.

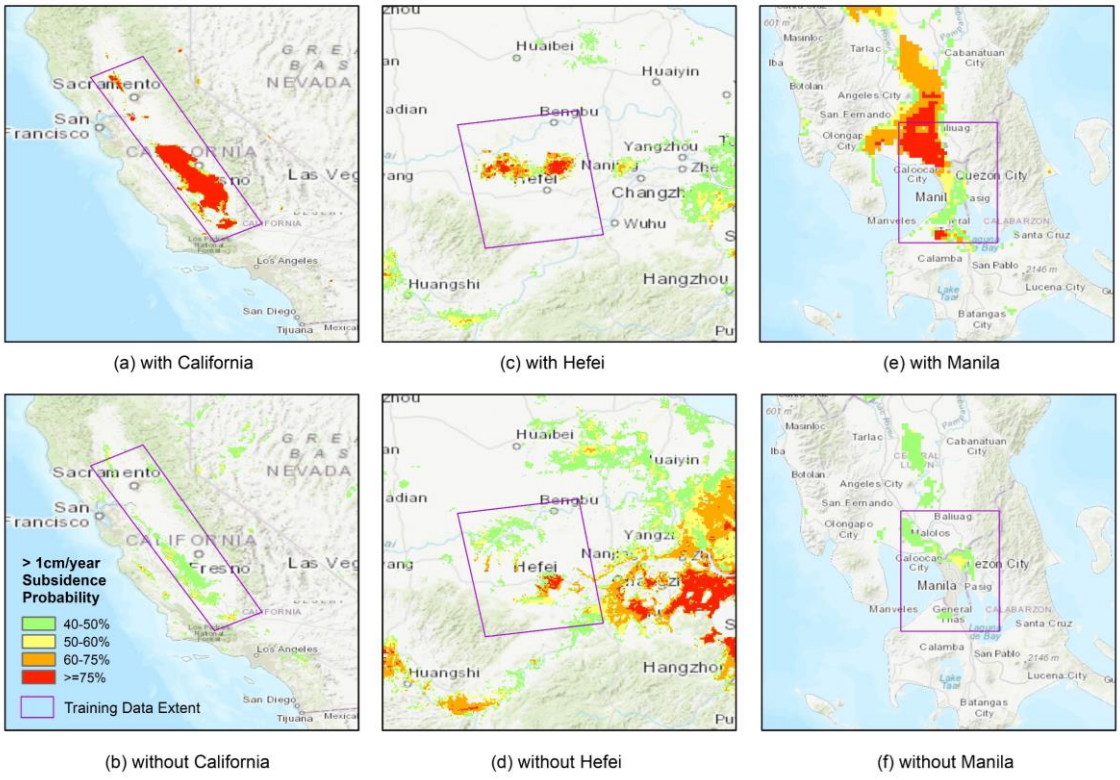

Figure 5.7: Subsidence probability map from Leave-One-Area-Out test compared to original subsidence probability.

Subsiding areas globally varies in terms of climate, hydrologic balance, and geologic formation. Removing one area from the training data might cost the model its predictive capacity for that region entirely. Therefore, the LOAO test can perform as a key evaluation technique for our model performance. We evaluated the model performance based on subsidence probability and checked if the model can detect some subsidence

signature over a training region without being trained with that training region data. We present the original model probability vs LOAO test probability for some regions of the world in Figure 5.7.

We evaluated the subsidence probability from the LOAO test outcomes for each training area. Probability of 40% of subsiding >1cm/year subsidence was chosen as the threshold value for categorizing the results using the polygon boundary extent of each training region (Figure 5.7(a, c, e) show respective polygon extents). We assigned 'satisfactory' accuracy if number of pixels with > 40% subsidence probability was higher than number of training subsidence pixels in that polygon extent. Result for areas where some subsidence probability of > 40% was observed was contemplated as 'acceptable' as the model was able to predict some high subsidence signature over that area without including the training data of that area. If these two criteria were not met, LOAO test performance was determined 'not satisfactory' for that region. Table 5.2 shows LOAO test results for 36 training regions of the model, out of which only 9 were categorized 'not satisfactory' according to the criteria described above. We had seven regions in the original training dataset where subsidence values all belonged to <1cm/year class. They were classified as 'satisfactory' if corresponding areas in respective models predicted <1cm/year subsidence majorly. For example, San Luis Valley, Colorado training data consisted of only <1cm/year class and the model in LOAO test without this data predicted only 10.5% pixels with higher subsidence probability, therefore, it was considered 'satisfactory'. 22 regions show 'satisfactory' and 4 show 'acceptable' accuracy. Result was 'not satisfactory

for 10 regions. The ‘bolded’ marked regions are where training data only consisted of <1 cm/year observations.

Table 5.2: Results of Leave-One-Area-Out accuracy test.

No.	*Country*	*Regions*	*Accuracy category*	*% Pixels with > 40% probability*
1	Philippines	Manila	satisfactory	8.17
2	China	Hefei	satisfactory	13.59
3	Spain	Murcia	satisfactory	16.44
4	Mexico	Mexico City	satisfactory	16.98
5	Turkey	Karapinar	satisfactory	21.08
6	Vietnam	Ho Chi Minh	satisfactory	22.40
7	Iran	Mashhad	satisfactory	33.91
8	Iran	Marand Plain	satisfactory	37.86
9	China	Beijing	satisfactory	42.38
10	Bangladesh	GBM Delta	satisfactory	46.56
11	China	Yellow River Delta	satisfactory	48.00
12	China	Xian	satisfactory	48.18
13	China	Tianjin	satisfactory	69.12
14	China	Hebei	satisfactory	74.32
15	Vietnam	Hanoi	satisfactory	79.41
16	Indonesia	Semarang	satisfactory	0.12
17	Iran	Qazvin	acceptable	0.36
18	United States	California	acceptable	3.55
19	Indonesia	Bandung	acceptable	3.86
20	Iran	Tehran	acceptable	4.76
21	Australia	Perth	satisfactory	0.00

Table 5.2: Results of Leave-One-Area-Out accuracy test (cont.)

No.	*Country*	*Regions*	*Accuracy category*	*% Pixels with > 40% probability*
22	England	London	satisfactory	0.00
23	**Italy**	**Venice**	satisfactory	0.00
24	**Nigeria**	**Lagos**	satisfactory	0.00
25	**Iraq**	**Tigris Euphrates Basin**	satisfactory	0.35
26	**US**	**San Luis Valley, Colorado**	satisfactory	10.49
27	**Egypt**	**Nile Delta**	not satisfactory	15.73
28	China	Shanghai	not satisfactory	0.00
29	China	Wuhan	not satisfactory	0.00
30	India	Delhi	not satisfactory	0.00
31	Italy	Po Delta	not satisfactory	0.00
32	Pakistan	Quetta	not satisfactory	0.00
33	Taiwan	Yunlin	not satisfactory	0.00
34	Turkey	Bursa	not satisfactory	0.00
35	US	Huston	not satisfactory	0.01
36	US	Arizona	not satisfactory	0.03

5.6. COMPARISON WITH GLOBAL STUDIES

There have been very few studies on global groundwater storage, especially for deep confined or semi-confined aquifers. Our model predicts groundwater storage loss in deep aquifers in terms of land subsidence. Therefore, we compared our model's result with global groundwater depletion study by Wada et al. (2010). Figure 5.8 shows the comparison between two studies in major groundwater stressed regions (1- both study results match; 2- our study predicts large groundwater storage loss while Wada et al. (2010)

does not; 3- Wada et al. (2010) estimates large groundwater decline while our study does not).

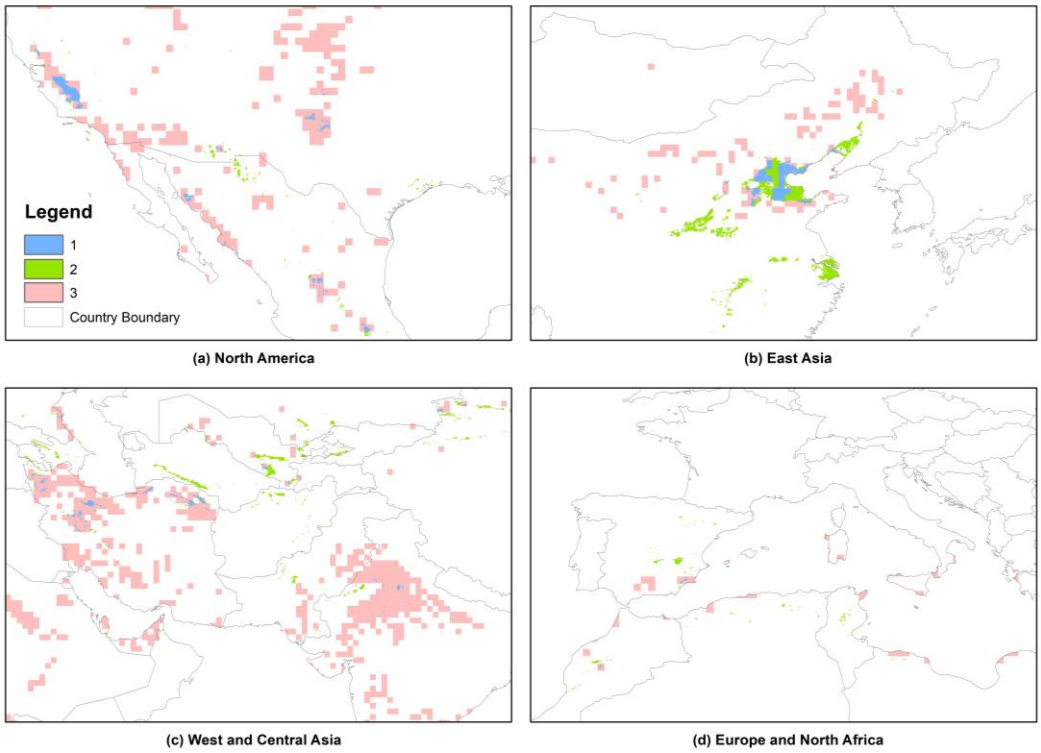

Figure 5.8: Comparison of model result with Wada et al. (2010).

In North China (Figure 5.8(b)), similar groundwater storage loss pattern (>1 cm/year) is observed in both studies. In North America (Figure 5.8(a)), both studies predict aquifer loss in central valley (California) and central Mexico. Wada et al. (2010) estimates large depletion in HPA region in US but our model only shows subsidence in parts of Texas. In central Asia, our results show aquifer loss in Uzbekistan, Afghanistan, Turkmenistan, Kazakhstan, and Azerbaijan in contrast to Wada et al. (2010). Both models predict groundwater depletion over Iran while Wada et al. (2010)'s estimate covers larger

spatial extent than our study. Large groundwater depletion in India is estimated in Wada et al. (2010) as well where our model predicts no significant subsidence.

Regardless the approach, both studies have been able to identify groundwater decline in key water-stressed regions. Number of reasons might have affected the inconsistency in results by the studies in some areas. Wada et al. (2010) assimilated global hydrological model generated groundwater recharge and publicly available groundwater use data, available from multiple sources, to map annual global groundwater depletion (for year 2000) in sub-humid and arid climates (Wada et al., 2012). Groundwater abstraction records from scattered sources may not be coherent and might fail to represent complete groundwater usage scenario being spatially inconsistent. However, we believe that the study was able to recognize regional pattern of groundwater depletion trend of that timeline. The reported spatial resolution of the depletion data is 0.5 deg (~55km) which is much coarser than our model resolution. This might have an impact on higher spatial extent reported by the study compared to our results in some places. Also, the depletion study did not segment their estimate into confined and unconfined portions rather reported the total depletion, while in our study storage loss in only confined or semi-confined aquifers have been modeled, which might have caused the difference in values in the studies. Future studies can be focused on regions where both studies do not converge. Comparison of our study result in North China Plain with a regional study (Gong et al., 2018) also found similar groundwater storage loss trend in that area by both studies.

For this research, no in-situ data was incorporated in the model rather we depended on open source remotely sensed and model-based datasets, which made the input datasets

spatially and temporally consistent, and easily collectable. High quality was maintained in accumulating and processing the training InSAR data. To ensure that our model recognizes groundwater storage loss associated inelastic deformation, we only considered studies that reported long-term vertical deformation from InSAR. Adding directly processed InSAR data with georeferenced InSAR data made the training data more robust and helped our model recognize subsidence pattern in vulnerable groundwater stressed regions, such as California in US, regions in Iran, North China Plain in China.

5.7. STUDY UNCERTAINTIES

Machine learning model performance is related to quality and quantity its training data. Training data used in our model has reasonable number of observations though class imbalance restricts the model's ability to predict minority classes (1-5 cm/year and >5 cm/year) to some extent. Significant number of our training data were extracted from research articles. This process involved georeferencing and accurately segmenting deformation rate into three classes. Both processes are based on human perception and prone to some level of uncertainties. We maintained a georeferencing residual value below 0.004 in most cases for ensuring uniformity.

InSAR processing is the primary source of land deformation data in the research articles from which we extracted training data. We also processed InSAR data over some regions directly to form the model's training dataset. Data processed from InSAR have uncertainties from atmospheric disturbance (Higgins et al., 2014; Reeves et al., 2014), irregularity in surface scattering (Erban et al., 2013) and tropospheric hindrance (Fattahi & Amelung, 2015), which were accumulated in the training data we formed.

Gridded datasets used in this study as input variables have inherent uncertainties from their sources. Remotely sensing based estimates are bound by their sensor's limitation. Sometimes post-processing, filtering, and statistical methods are applied on remote sensing datasets to refine the data or to come up with a new dataset from multiple remote sensing estimates, such as land use classification. Such techniques transfer the parent dataset's uncertainties to the new dataset formed. TerraClimate climatic estimates inherit uncertainties from their forming datasets. Water balance components of TerraClimate comes from a 1-D land surface model and limitations of the model assumptions can add uncertainties to the resulting estimates (Abatzoglou et al., 2018).

6. CONCLUSIONS

In this study, we developed a machine learning model that maps groundwater withdrawal associated inelastic land subsidence and subsidence probability at a high spatial resolution (~2 km) based on open source remotely sensed and model-based datasets. To the author's knowledge, this is the first attempt to map subsidence magnitude globally integrating machine learning and open source datasets. We trained our model with InSAR-derived land deformation data for 36 regions of the world to capture climatic and water use characteristics of representative regions. The model shows realistic response to land use patterns and hydrologic fluxes in partial dependence plots. Most of the predicted subsiding regions, approximately 70%, are on croplands and urban areas, which indicates high groundwater dependency of the predicted regions. High probability of subsidence with increase in irrigation and population density represents the model's ability to contemplate the relation between long-term subsidence and groundwater use in those land use. As inelastic subsidence causes permanent loss of aquifer storage, the predicted regions are where groundwater storage is permanently declining. Comparison with documented subsidence locations shows that the machine learning model was able to reveal the true spatial extent of subsidence in some regions. These subsiding areas might keep experiencing future subsidence, with increasing risk to spatially extend subsidence extent as shown in our subsidence probability map, if water use practice is not modified. We also identified subsidence in locations where groundwater related subsidence has never been studied and reported before. These regions are undergoing critical groundwater stress and it is essential to partake effective long term aquifer monitoring 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 plan for aquifer sustainability.

APPENDIX A.

GLOBAL MAP OF LAND SUBSIDENCE

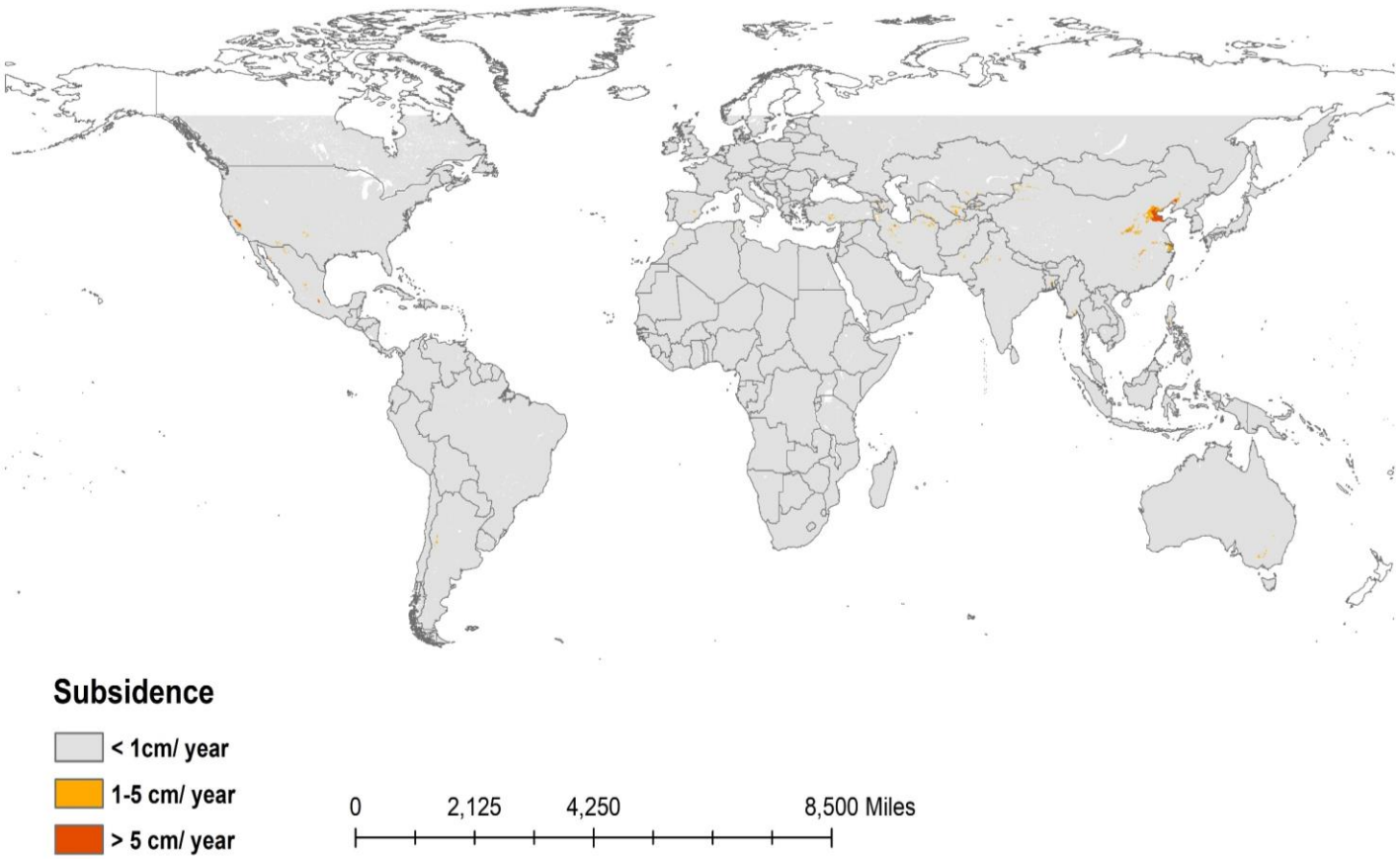

Figure A1: Global Subsidence Map

APPENDIX B.

GLOBAL MAP OF PROBABILITY OF >1 CM/YEAR SUBSIDENCE

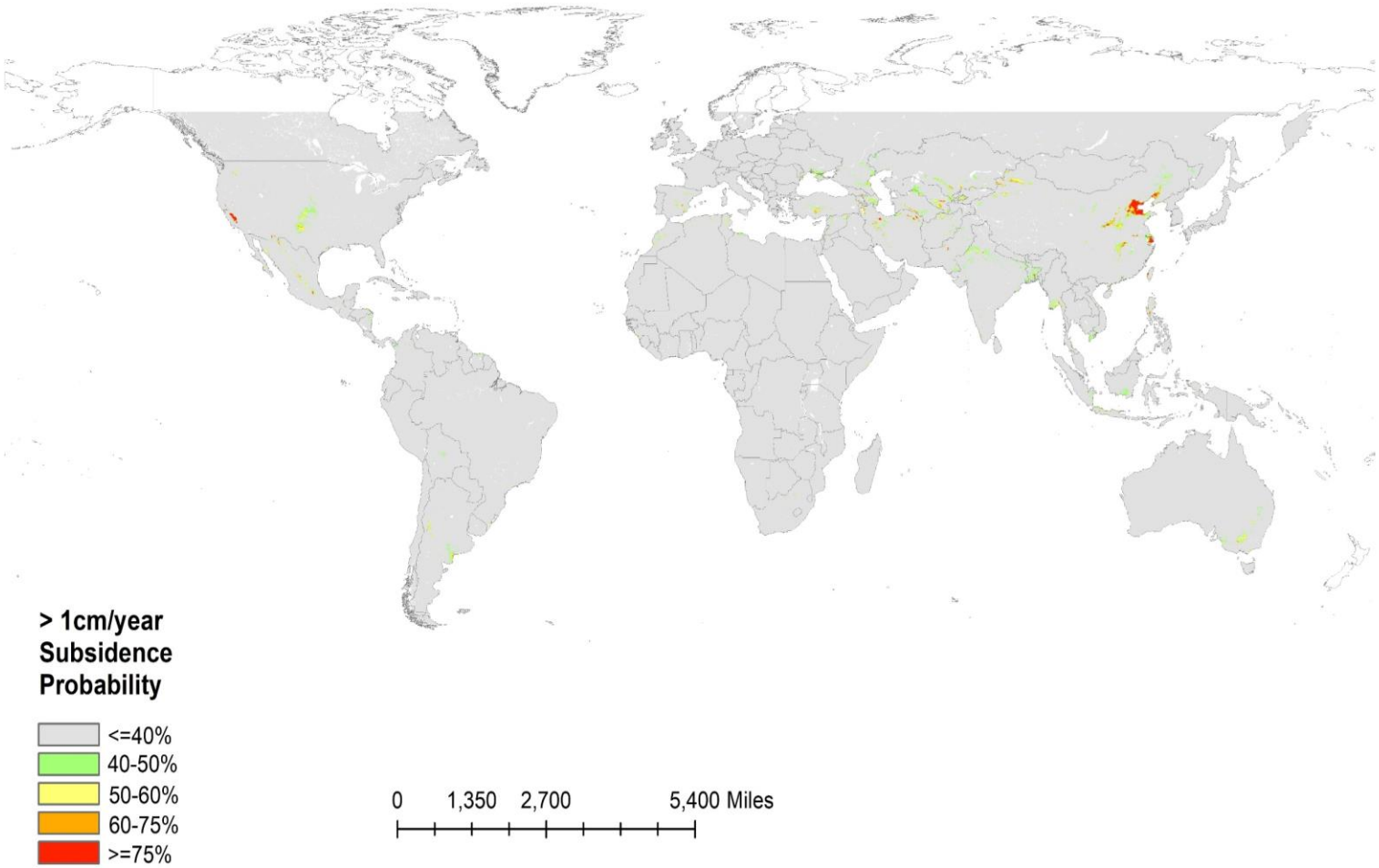

Figure B1: Global map of probability of >1 cm/year subsidence.

APPENDIX C.

PARTIAL DEPENDENCE PLOTS OF VARIABLE ‘CONFINING LAYERS’ FOR ALL MODEL CLASSES

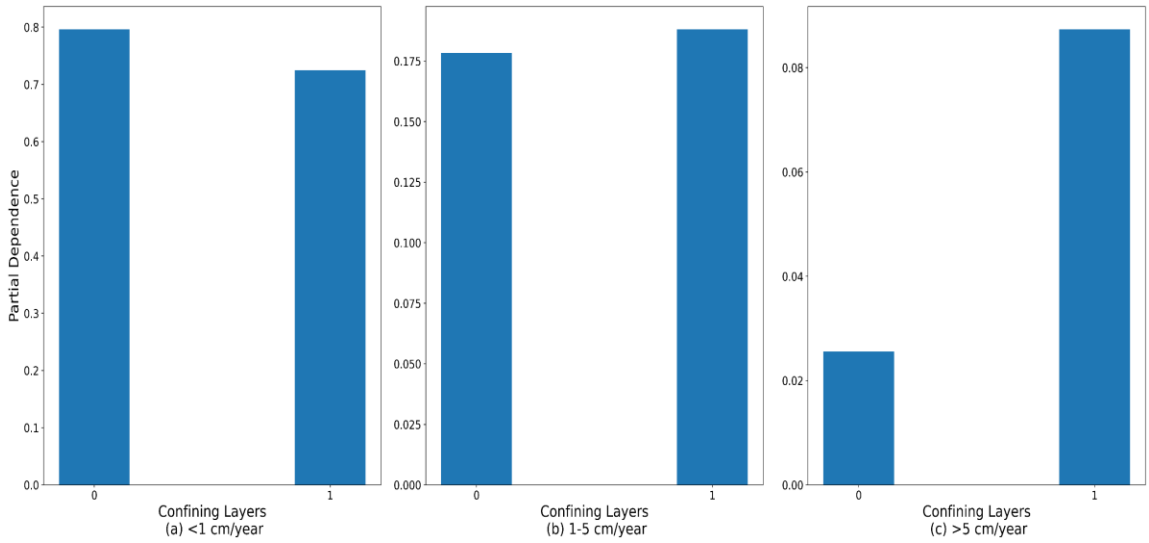

Figure C1: Partial dependence plots of variable ‘confining layers’ for all model classes.

APPENDIX D.

FUNDING AND RELEASE INFORMATION

This research has been funded by the National Geospatial-Intelligence Agency (NGA), United States and has been approved by the agency for public release, 22-555.

REFERENCES

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Scientific Data*, *5*. https://doi.org/10.1038/sdata.2017.191

Andaryani, S., Nourani, V., Trolle, D., Dehgani, M., & Asl, A. M. (2019). Assessment of land use and climate change effects on land subsidence using a hydrological model and radar technique. *Journal of Hydrology*, *578*(August), 124070. https://doi.org/10.1016/j.jhydrol.2019.124070

Aslan, G., Cakir, Z., Lasserre, C., & Renard, F. (2019). Investigating subsidence in the Bursa Plain, Turkey, using ascending and descending sentinel-1 satellite data. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010085

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. *Reviews of Geophysics*, *57*(2), 530–616. https://doi.org/10.1029/2018RG000618

Biau, G. (2012). Analysis of a Random Forests Model. In *Journal of Machine Learning Research* (Vol. 13, pp. 1063–1095).

Bock, Y., Wdowinski, S., Ferretti, A., Novali, F., & Fumagalli, A. (2012). Recent Subsidence of the Venice Lagoon from Continuos GPS and Interferometric Synthetic Aperture Radar. *Geochemistry, Geophysics, Geosystems*, *13*.

Bonì, R., Meisina, C., Cigna, F., Herrera, G., Notti, D., Bricker, S., McCormack, H., Tomás, R., Béjar-Pizarro, M., Mulas, J., & Ezquerro, P. (2017). Exploitation of satellite A-DInSAR time series for detection, characterization and modelling of land subsidence. *Geosciences (Switzerland)*, *7*(2). https://doi.org/10.3390/geosciences7020025

Borchers, J. W., Grabert, V. K., Carpenter, M., Dalgish, B., & Cannon, D. (2014). Land Subsidence from Groundwater Use in California. *California Water Foundation*, *4*(April), 1–115.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

Butler, J. J., Whittemore, D. O., Wilson, B. B., & Bohling, G. C. (2018). Sustainability of aquifers supporting irrigated agriculture: a case study of the High Plains aquifer in Kansas. *Water International*, *43*(6), 815–828. https://doi.org/10.1080/02508060.2018.1515566

Camp, M. Van, Mtoni, Y., Mjemah, I. C., Bakundukize, C., & Walraevens, K. (2014). Investigating seawater intrusion due to groundwater pumping with schematic model simulations: The example of the Dar es Salaam coastal aquifer in Tanzania. *Journal of African Earth Sciences*, *96*, 71–78. https://doi.org/10.1016/J.JAFREARSCI.2014.02.012

Castellazzi, P., & Schmid, W. (2021). Interpreting C-band InSAR ground deformation data for large-scale groundwater management in Australia. *Journal of Hydrology: Regional Studies*, *34*(November 2020), 100774. https://doi.org/10.1016/j.ejrh.2021.100774

Chaussard, E., Amelung, F., Abidin, H., & Hong, S. H. (2013). Sinking cities in Indonesia: ALOS PALSAR detects rapid subsidence due to groundwater and gas extraction. *Remote Sensing of Environment*, *128*, 150–161. https://doi.org/10.1016/j.rse.2012.10.015

Chaussard, E., & Farr, T. G. (2019). A New Method for Isolating Elastic From Inelastic Deformation in Aquifer Systems: Application to the San Joaquin Valley, CA. *Geophysical Research Letters*, *46*(19), 10800–10809. https://doi.org/10.1029/2019GL084418

Chaussard, E., Wdowinski, S., Cabral-Cano, E., & Amelung, F. (2014). Land subsidence in central Mexico detected by ALOS InSAR time-series. *Remote Sensing of Environment*, *140*, 94–106. https://doi.org/10.1016/j.rse.2013.08.038

Chen, J., Knight, R., Zebker, H. A., & Schreüder, W. A. (2016). Confined aquifer head measurements and storage properties in the San Luis Valley, Colorado, from spaceborne InSAR observations. *Water Resources Research*, *52*, 3623–3636. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Chen, M., Tomás, R., Li, Z., Motagh, M., Li, T., Hu, L., Gong, H., Li, X., Yu, J., & Gong, X. (2016). Imaging Land Subsidence Induced by Groundwater Extraction in Beijing (China) Using Satellite Radar Interferometry. *Remote Sensing*, *8*(6), 468. https://doi.org/10.3390/rs8060468

Cian, F., Blasco, J. M. D., & Carrera, L. (2019). Sentinel-1 for monitoring land subsidence of coastal cities in Africa using PSInSAR: A methodology based on the integration of SNAP and staMPS. *Geosciences (Switzerland)*, *9*(3). https://doi.org/10.3390/geosciences9030124

CIESIN. (2018). *GPWv411: UN-Adjusted Population Density (Gridded Population of the World Version 4.11)*. https://doi.org/10.7927/H4F47M65

Conway, B. D. (2015). Land subsidence and earth fissures in south-central and southern Arizona, USA. *Hydrogeology Journal*, *24*(3), 649–655. https://doi.org/10.1007/s10040-015-1329-z

Corbau, C., Simeoni, U., Zoccarato, C., Mantovani, G., & Teatini, P. (2019). Coupling land use evolution and subsidence in the Po Delta, Italy: Revising the past occurrence and prospecting the future management challenges. *Science of the Total Environment*, *654*, 1196–1208. https://doi.org/10.1016/j.scitotenv.2018.11.104

Dang, V. K., Doubre, C., Weber, C., Gourmelen, N., & Masson, F. (2014). Recent land subsidence caused by the rapid urban development in the Hanoi region (Vietnam) using ALOS InSAR data. *Natural Hazards and Earth System Sciences*, *14*(3), 657–674. https://doi.org/10.5194/nhess-14-657-2014

Didan, K. (2021). *MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061 EVI.* NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.061

Dong, S., Samsonov, S., Yin, H., Ye, S., & Cao, Y. (2014). Time-series analysis of subsidence associated with rapid urbanization in Shanghai, China measured with SBAS InSAR method. *Environmental Earth Sciences*, *72*(3), 677–691. https://doi.org/10.1007/s12665-013-2990-y

Erban, L. E., Gorelick, S. M., & Zebker, H. A. (2014). Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environmental Research Letters*, *9*(8). https://doi.org/10.1088/1748-9326/9/8/084010

Erban, L. E., Gorelick, S. M., Zebker, H. A., & Fendorf, S. (2013). Release of arsenic to deep groundwater in the Mekong Delta, Vietnam, linked to pumping-induced land subsidence. *Proceedings of the National Academy of Sciences*. https://www.pnas.org/content/110/34/13751

Ezquerro, P., Tomás, R., Béjar-Pizarro, M., Fernández-Merodo, J. A., Guardiola-Albert, C., Staller, A., Sánchez-Sobrino, J. A., & Herrera, G. (2020). Improving multi-technique monitoring using Sentinel-1 and Cosmo-SkyMed data and upgrading groundwater model capabilities. *Science of the Total Environment*, *703*, 134757. https://doi.org/10.1016/j.scitotenv.2019.134757

Fabris, M., Achilli, V., & Menin, A. (2014). Estimation of Subsidence in Po Delta Area (Northern Italy) by Integration of GPS Data, High-Precision Leveling and Archival Orthometric Elevations. *International Journal of Geosciences*, *05*(06), 571–585. https://doi.org/10.4236/ijg.2014.56052

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science*, *339*(6122), 940–943. https://doi.org/10.1126/science.1229881

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2), RG2004. https://doi.org/10.1029/2005RG000183

Fattahi, H., & Amelung, F. (2015). InSAR bias and uncertainty due to the systematic and stochastic tropospheric delay. *Journal of Geophysical Research: Solid Earth*, *120*(12), 8758–8773. https://doi.org/10.1002/2015JB012419

Faunt, C. C. (2009). *Groundwater availability of the Central Valley Aquifer, California*. U.S. Geological Survey.

Faunt, C. C., Sneed, M., Traum, J., & Brandt, J. T. (2016). Water availability and land subsidence in the Central Valley, California, USA. *Hydrogeology Journal*, *24*(3), 675–684. https://doi.org/10.1007/s10040-015-1339-x

Fetter, C. W. (2001). Applied Hydrogeology Fourth Edition. In *Applied Hydrogeology*.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Freeze, R. A., & Cherry, J. A. (1979). *GROUNDWATER*.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, *114*(1), 168–182. https://doi.org/10.1016/j.rse.2009.08.016

Galloway, D. L., & Burbey, T. J. (2011). Review: Regional land subsidence accompanying groundwater extraction. *Hydrogeology Journal*, *19*(8), 1459–1486. https://doi.org/10.1007/s10040-011-0775-5

Garg, S., Motagh, M., Indu, J., & Karanam, V. (2022). Tracking hidden crisis in India's capital from space: implications of unsustainable groundwater use. *Scientific Reports*, *12*(1), 1–17. https://doi.org/10.1038/s41598-021-04193-9

Ge, L., Ng, A. H. M., Li, X., Abidin, H. Z., & Gumilar, I. (2014). Land subsidence characteristics of Bandung Basin as revealed by ENVISAT ASAR and ALOS PALSAR interferometry. *Remote Sensing of Environment*, *154*, 46–60. https://doi.org/10.1016/j.rse.2014.08.004

Gebremichael, E., Sultan, M., Becker, R., El Bastawesy, M., Cherif, O., & Emil, M. (2018). Assessing Land Deformation and Sea Encroachment in the Nile Delta: A Radar Interferometric and Inundation Modeling Approach. *Journal of Geophysical Research: Solid Earth*, *123*(4), 3208–3224. https://doi.org/10.1002/2017JB015084

Gleeson, T., Wada, Y., Bierkens, M. F. P., & van Beek, L. P. H. (2012). Water balance of global aquifers revealed by groundwater footprint. *Nature*, *488*(7410), 197–200. https://doi.org/10.1038/nature11295

Gong, H., Pan, Y., Zheng, L., Li, X., Zhu, L., Zhang, C., Huang, Z., Li, Z., Wang, H., & Zhou, C. (2018). Long-term groundwater storage changes and land subsidence development in the North China Plain (1971–2015). *Hydrogeology Journal*, *26*(5), 1417–1427. https://doi.org/10.1007/s10040-018-1768-4

Graaf, I. E. M. De, Sutanudjaja, E. H., Beek, L. P. H. Van, & Bierkens, M. F. P. (2015). A high-resolution global-scale groundwater model. *Hydrology and Earth System Sciences*, *19*(2), 823–837. https://doi.org/10.5194/hess-19-823-2015

GRDC. (2020). *Major River Basins of the World / Global Runoff Data Centre, GRDC. 2nd, rev. ext. ed. Koblenz, Germany: Federal Institute of Hydrology (BfG)*. https://www.bafg.de/GRDC/EN/02_srvcs/22_gslrs/221_MRB/riverbasins_node.html

Gridded Population of the World, Version 4 (GPWv4.11): Population Density Adjusted to Match 2015 Revision of UN WPP Country Totals, R. 11. (2018). *Center for International Earth Science Information Network - CIESIN - Columbia University*. https://doi.org/https://doi.org/10.7927/H4F47M65

Haghighi, M. H., & Motagh, M. (2019). Ground surface response to continuous compaction of aquifer system in Tehran, Iran: Results from a long-term multi-sensor InSAR analysis. *Remote Sensing of Environment*, *221*, 534–550. https://doi.org/10.1016/j.rse.2018.11.003

Hengl, T. (2018). *Clay content in % (kg / kg) at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution*. https://doi.org/https://doi.org/10.5281/zenodo.1476854

Herrera-García, G., Ezquerro, P., Tomás, R., Béjar-Pizarro, M., López-Vinielles, J., Rossi, M., Mateos, R. M., Carreón-Freyre, D., Lambert, J., Teatini, P., Cabral-Cano, E., Erkens, G., Galloway, D., Hung, W.-C., Kakar, N., Sneed, M., Tosi, L., Wang, H., & Ye, S. (2021). Mapping the global threat of land subsidence. *Science*, *371*(6524), 34–36. https://doi.org/10.1126/science.abb8549

Higgins, S., Overeem, I., Steckler, M. S., Syvitski, J. P. M., Seeber, L., & Akhter, S. H. (2014). InSAR measurements of compaction and subsidence in the Ganges-Brahmaputra Delta, Bangladesh. *Journal of Geophysical Research F: Earth Surface*, *119*(8), 1768–1781. https://doi.org/10.1002/2014JF003117

Higgins, S., Overeem, I., Tanaka, A., & Syvitski, J. P. M. (2013). Land subsidence at aquaculture facilities in the Yellow River delta, China. *Geophysical Research Letters*, *40*(15), 3898–3902. https://doi.org/10.1002/grl.50758

Hoffmann, J., Zebker, H. A., Galloway, D. L., & Amelung, F. (2001). Seasonal subsidence and rebound in Las Vegas Valley, Nevada, observed by synthetic aperture radar interferometry. *Water Resources Research*, *37*(6), 1551–1566. https://doi.org/10.1029/2000WR900404

Hsu, W. C., Chang, H. C., Chang, K. T., Lin, E. K., Liu, J. K., & Liou, Y. A. (2015). Observing land subsidence and revealing the factors that influence it using a multi-sensor approach in Yunlin County, Taiwan. *Remote Sensing*, *7*(6), 8202–8223. https://doi.org/10.3390/rs70608202

Hung, W.-C., Hwang, C., Chang, C.-P., Yen, J.-Y., Liu, C.-H., & Yang, W.-H. (2010). Monitoring severe aquifer-system compaction and land subsidence in Taiwan using multiple sensors: Yunlin, the southern Choushui River Alluvial Fan. *Environmental Earth Sciences*, *59*(7), 1535–1548. https://doi.org/10.1007/s12665-009-0139-9

Jacob, C. E. (1940). ON THE FLOW OF WATER IN AN ELASTIC ARTESIAN AQUIFER. In *U. S. Geol. Surv. W.-S. Paper* (Vol. 779).

James, G., DanielaWitten, Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning wth application in R*.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1), 27. https://doi.org/10.1186/s40537-019-0192-5

Kakar, N., Kakar, D. M., & Barrech, S. (2020). Land subsidence caused by groundwater exploitation in Quetta and surrounding region, Pakistan. *Proceedings of the International Association of Hydrological Sciences*, *382*, 595–607. https://doi.org/10.5194/piahs-382-595-2020

Khorrami, M., Abrishami, S., Maghsoudi, Y., Alizadeh, B., & Perissin, D. (2020). Extreme subsidence in a populated city (Mashhad) detected by PSInSAR considering groundwater withdrawal and geotechnical properties. *Scientific Reports*, *10*(1), 1–16. https://doi.org/10.1038/s41598-020-67989-1

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., & Kiyotoshi, T. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kuang, X., Jiao, J. J., Zheng, C., Cherry, J. A., & Li, H. (2020). A review of specific storage in aquifers. *Journal of Hydrology*, *581*. https://doi.org/10.1016/j.jhydrol.2019.124383

Lamontagne, S., Taylor, A. R., Cook, P. G., Crosbie, R. S., Brownbill, R., Williams, R. M., & Brunner, P. (2014). Field assessment of surface water-groundwater connectivity in a semi-arid river basin (Murray-Darling, Australia). *Hydrological Processes*, *28*(4), 1561–1572. https://doi.org/10.1002/hyp.9691

Leroux, D., & Pellarin, T. (2016). Remote Sensing Data Assimilation: Applications to Catchment Hydrology. *Land Surface Remote Sensing in Continental Hydrology*, 363–399. https://doi.org/10.1016/B978-1-78548-104-8.50011-5

Lohman, S. W. (1972). *Ground-Water Hydraulics*.

Luo, Q., Perissin, D., Zhang, Y., & Jia, Y. (2014). L- and X-band multi-temporal InSAR analysis of tianjin subsidence. *Remote Sensing*, *6*(9), 7933–7951. https://doi.org/10.3390/rs6097933

Margat, J., & Van der hun, J. (2013). *Groundwater Around the World*.

Meier, J., Zabel, F., & Mauser, W. (2018). A global approach to estimate irrigated areas - A comparison between different data and statistics. *Hydrology and Earth System Sciences*, *22*(2), 1119–1133. https://doi.org/10.5194/hess-22-1119-2018

Miller, M. M., & Shirzaei, M. (2019). Land subsidence in Houston correlated with flooding from Hurricane Harvey. *Remote Sensing of Environment*, *225*(June 2018), 368–378. https://doi.org/10.1016/j.rse.2019.03.022

Minh, D. H. T., Van Trung, L., & Le Toan, T. (2015). Mapping ground subsidence phenomena in Ho Chi Minh City through the radar interferometry technique using ALOS PALSAR data. *Remote Sensing*, *7*(7), 8543–8562. https://doi.org/10.3390/rs70708543

Mladenova, I. E., Bolten, J. D., Crow, W., Sazib, N., & Reynolds, C. (2020). Agricultural Drought Monitoring via the Assimilation of SMAP Soil Moisture Retrievals Into a Global Soil Water Balance Model. *Frontiers in Big Data*, *3*(April), 1–16. https://doi.org/10.3389/fdata.2020.00010

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, *117*(8). https://doi.org/10.1029/2011JD017187

Mu, Q., Zhao, M., & Running, S. W. (2013). *MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3) Collection 5. NASA Headquarters MOD16A2/A3) Collection 5. NASA Headquarters*. https://scholarworks.umt.edu/ntsg_pubs/268

Munier, S., Aires, F., Schlaffer, S., Prigent, C., Papa, F., Maisongrande, P., & Pan, M. (2014). Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi basin and closure correction model. *Journal of Geophysical Research*, *119*(21), 12,100-12,116. https://doi.org/10.1002/2014JD021953

Nguyen, M., Lin, Y. N., Tran, Q. C., Ni, C. F., Chan, Y. C., Tseng, K. H., & Chang, C. P. (2022). Assessment of long-term ground subsidence and groundwater depletion in Hanoi, Vietnam. *Engineering Geology*, *299*, 106555. https://doi.org/10.1016/j.enggeo.2022.106555

Ojha, C., Shirzaei, M., Werth, S., Argus, D. F., & Farr, T. G. (2018). Sustained Groundwater Loss in California's Central Valley Exacerbated by Intense Drought Periods. *Water Resources Research*, *54*(7), 4449–4460. https://doi.org/10.1029/2017WR022250

Orhan, O., Oliver-Cabrera, T., Wdowinski, S., Yalvac, S., & Yakar, M. (2021). Land subsidence and its relations with sinkhole activity in karapınar region, turkey: A multi-sensor insar time series study. *Sensors (Switzerland)*, *21*(3), 1–17. https://doi.org/10.3390/s21030774

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., Williams, Z. C., Brunke, M. A., & Gochis, D. (2016). *Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1304

Polcari, M. (2019). Anthropogenic subsidence along railway and road infrastructures in Northern Italy highlighted by Cosmo-SkyMed satellite data. *Journal of Applied Remote Sensing*, *13*(02), 1. https://doi.org/10.1117/1.jrs.13.024515

Qu, F., Zhang, Q., Lu, Z., Zhao, C., Yang, C., & Zhang, J. (2014). Land subsidence and ground fissures in Xi'an, China 2005-2012 revealed by multi-band InSAR time-series analysis. *Remote Sensing of Environment*, *155*(December), 366–376. https://doi.org/10.1016/j.rse.2014.09.008

Rateb, A., & Kuo, C. Y. (2019). Quantifying vertical deformation in the Tigris-Euphrates basin due to the groundwater abstraction: Insights from GRACE and Sentinel-1 satellites. *Water (Switzerland)*, *11*(8). https://doi.org/10.3390/w11081658

Rateb, A., Scanlon, B. R., Pool, D. R., Sun, A., Zhang, Z., Chen, J., Clark, B., Faunt, C. C., Haugh, C. J., Hill, M., Hobza, C., McGuire, V. L., Reitz, M., Schmied, H. M., Sutanudjaja, E. H., Swenson, S., Wiese, D., Xia, Y., & Zell, W. (2020). Comparison of Groundwater Storage Changes From GRACE Satellites With Monitoring and Modeling of Major U.S. Aquifers. *Water Resources Research*, *56*(12). https://doi.org/10.1029/2020WR027556

Reeves, J. A., Knight, R., & Zebker, H. A. (2014). An analysis of the uncertainty in InSAR deformation measurements for groundwater applications in agricultural areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(7), 2992–3001. https://doi.org/10.1109/JSTARS.2014.2322775

Reeves, J. A., Knight, R., Zebker, H. A., Schreüder, W. A., Shanker Agram, P., & Lauknes, T. R. (2011). High quality InSAR data linked to seasonal change in hydraulic head for an agricultural area in the San Luis Valley, Colorado. *Water Resources Research*, *47*(12). https://doi.org/10.1029/2010WR010312

Rodell, M. (2013). Application of Satellite Gravimetry for Water Resource Vulnerability Assessment. *Climate Vulnerability: Understanding and Addressing Threats to Essential Resources*, *5*, 151–159. https://doi.org/10.1016/B978-0-12-384703-4.00521-9

Sharma, P., Jones, C. E., Dudas, J., Bawden, G. W., & Deverel, S. (2016). Monitoring of subsidence with UAVSAR on Sherman Island in California's Sacramento-San Joaquin Delta. *Remote Sensing of Environment*, *181*(August), 218–236. https://doi.org/10.1016/j.rse.2016.04.012

Shiklomanov, I. (1993). World fresh water resources. In P. H. Gleick (Ed.), *Water in Crisis: A Guide to the World's Fresh Water Resources*.

Shirzaei, M., Freymueller, J., Törnqvist, T. E., Galloway, D. L., Dura, T., & Minderhoud, P. S. J. (2021). Measuring, modelling and projecting coastal land subsidence. *Nature Reviews Earth & Environment*, *2*(1), 40–58. https://doi.org/10.1038/s43017-020-00115-x

Siebert, S., Henrich, V., Frenken, K., & Burke, J. (2013). Update of the digital global map of irrigation areas to version 5. *Rheinische Friedrich-Wilhelms-Universituy, Bonn, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy, August 2014*, 171. https://doi.org/10.13140/2.1.2660.6728

Smith, R, & Knight, R. (2019). Modeling Land Subsidence Using InSAR and Airborne Electromagnetic Data. *Water Resources Research*, *55*(4), 2801–2819. https://doi.org/10.1029/2018WR024185

Smith, R, & Majumdar, S. (2020). Groundwater Storage Loss Associated With Land Subsidence in Western United States Mapped Using Machine Learning. *Water Resources Research*, *56*(7). https://doi.org/10.1029/2019WR026621

Smith, Ryan, Knight, R., Chen, J., Reeves, J. A., Zebker, H. A., Farr, T., & Liu, Z. (2017). Estimating the permanent loss of groundwater storage in the southern San Joaquin Valley, California. *Water Resources Research*, *53*(3), 2133–2148. https://doi.org/10.1002/2016WR019861

Sneed, B. M. (2001). Hydraulic and Mechanical Properties Affecting Ground-Water Flow and Aquifer- System Compaction , San Joaquin Valley , California. In *U. S. Geological Survey* (Issue January 2001).

Teluguntla, P., Thenkabail, P. S., Xiong, J., Gumma, M. K., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Tilton, J., Sankey, T. R., Massey, R., Phalke, A., & Yadav, K. (2014). *NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) 1 km datasets*.

Teluguntla, P., Thenkabail, P. S., Xiong, J., Gumma, M. K., Giri, C., Milesi, C., Tilton, J., & Massey, R. (2018). Global Food Security Support Analysis Data at Nominal 1 km (GFSAD1km) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities. In *Remote Sensing Handbook - Three Volume Set* (Vol. 2, pp. 865–894). CRC Press. https://doi.org/10.1201/b19355-67

Thenkabail, P. S., Biradar, C. M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O. R. P., Turral, H., Cai, X., Vithanage, J., Schull, M. A., & Dutta, R. (2009). Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *International Journal of Remote Sensing*, *30*(14), 3679–3733. https://doi.org/10.1080/01431160802698919

Trabucco, A., & Zomer, R. (2019). *Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2*. https://doi.org/10.6084/m9.figshare.7504448.v3

Vermote, E. (2021). *MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V061*. NASA EOSDIS Land Processes DAAC. https://doi.org/https://doi.org/10.5067/MODIS/MOD09A1.061

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wada, Y., Van Beek, L. P. H., & Bierkens, M. F. P. (2012). Nonsustainable groundwater sustaining irrigation: A global assessment. *Water Resources Research*, *48*(1). https://doi.org/10.1029/2011WR010562

Wada, Y., Van Beek, L. P. H., Van Kempen, C. M., Reckman, J. W. T. M., Vasak, S., & Bierkens, M. F. P. (2010). Global depletion of groundwater resources. *Geophysical Research Letters*, *37*(20), 1–5. https://doi.org/10.1029/2010GL044571

Yin, W., Fan, Z., Tangdamrongsub, N., Hu, L., & Zhang, M. (2021). 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. *Journal of Hydrology*, *602*, 126735. https://doi.org/10.1016/J.JHYDROL.2021.126735

Zhou, L., Guo, J., Hu, J., Li, J., Xu, Y., Pan, Y., & Shi, M. (2017). Wuhan surface subsidence analysis in 2015-2016 based on sentinel-1A data by SBAS-InSAR. *Remote Sensing*, *9*(10). https://doi.org/10.3390/rs9100982

Zoysa, R. S. De, Schöne, T., Herbeck, J., Illigner, J., Haghighi, M., Simarmata, H., Porio, E., Rovere, A., & Hornidge, A. K. (2021). The "wickedness" of governing land subsidence: Policy perspectives from urban southeast Asia. *PLoS ONE*, *16*(6 June), 1–25. https://doi.org/10.1371/journal.pone.0250208

VITA

Md Fahim Hasan grew up in Bangladesh, a beautiful country in South Asia. In 2017, he received his Bachelor of Science in Civil Engineering with a major in Environmental Engineering from the Bangladesh University of Engineering and Technology. He always wanted to pursue a career in water resources and therefore, served as a professional water resource engineer in Bangladesh before joining the Missouri University of Science and Technology in 2021. For his Master's degree, he used remote sensing and machine learning techniques to study groundwater resources. He was a student member of the American Geophysical Union (AGU) and the American Society of Civil Engineers (ASCE). In 2021, Fahim was awarded Mike Alizadeh Geotechnology Scholarship by the ASCE St. Louis section for his early career involvements and research in environmental/geotechnical engineering. In July 2022, he received his Master’s degree in Geological Engineering from the Missouri University of Science and Technology. Following the degree, he accepted a Ph.D. position in the Department of Civil and Environmental Engineering at Colorado State University.