

Assessment of Inundation Dynamics in Nebraska Using AlphaEarth Foundations Embeddings

Aditya Kapoor

adityain2003@gmail.com

University of Nebraska–Lincoln

Research Article

Keywords: Google Earth Engine, Random forest classification, Water occurrence, Surface water monitoring, Lakes and reservoirs

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11254365/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Spatiotemporal assessment of surface water inundation is fundamental to water resource management, ecological monitoring, and flood risk characterization. Satellite remote sensing enables such assessments at the regional scale, without the logistical expense associated with field surveys. While assimilating data from multiple sources improves the performance of remote sensing-based classification, it introduces substantial preprocessing complexity and computational overhead. This study investigates the utility of AlphaEarth Foundations Embeddings (AFE) as an analysis-ready alternative for regional-scale inundation mapping. AFE is a precomputed, 10 m resolution, 64-dimensional annual dataset that reduces the need for multi-source data preprocessing and fusion. A random forest classifier was applied to AFE within Google Earth Engine to delineate annual inundation extents. The analysis covered 901 HydroLAKES surface water bodies in Nebraska, USA, from 2017 to 2025. This is one of the first large-scale inundation studies to use the complete AFE record. Classification performance was assessed visually against Sentinel-2 imagery and quantitatively against field-derived elevation-area relationships for five reservoirs (mean absolute difference 3.32%). At five example lakes, Dynamic World mapped between 1.0% less and 39.6% more inundated area than random forest. Mean inundated area showed no statistically significant trend. However, its year-to-year change followed annual rainfall ($r = 0.73$, or 0.91 when each water body counted equally). Inundation frequency analysis further revealed that smaller water bodies exhibit less persistent inundation than larger ones. These findings indicate that AFE can support annual-scale regional inundation monitoring with reduced preprocessing, with implications for water resource planning, wetland conservation, and flood hazard assessment.

Highlights

- AFE enables annual-scale inundation mapping across 901 HydroLAKES surface water bodies in Nebraska
- Random forest maps were compared with Dynamic World at five example lakes
- Year-to-year changes in inundated area follow annual rainfall
- Smaller water bodies exhibit less persistent inundation than larger counterparts

Introduction

The spatial extent of the surface water bodies can vary rapidly. Two types of factors primarily drive this variation: natural phenomena, such as precipitation and evapotranspiration (Hall et al., 2019), and anthropogenic activities, such as changes in land use and land cover and the construction of hydraulic structures, such as dams and barrages (Cai et al., 2024). The variation caused by these factors can have profound societal and ecological impacts (Pekel et al., 2016). A reduction in extent can reflect drought conditions, which may diminish regional water resources. Furthermore, the resulting loss of wetlands can lead to the loss of the ecological services they provide (Qi et al., 2009; Hall et al., 2019; Khare et al.,

2025; Tesfaye & Breuer, 2025). On the other hand, an expansion in spatial extent can lead to flooding (Huang et al., 2018), which may result in loss of life and assets.

These implications make it essential to monitor and assess the spatiotemporal dynamics of the surface water bodies. These assessments are performed using two distinct approaches. The first involves conducting field surveys and collecting in situ data. Although accurate, field surveys are cumbersome and logistically expensive (Munasinghe et al., 2018) and may not be feasible in inaccessible or remote areas. These limitations are overcome by the second approach, which utilizes remote sensing and geographic information system (GIS) techniques. Satellites revisit a region at regular intervals, providing temporally consistent observations of the Earth's surface that can be used to assess the spatiotemporal dynamics of surface water bodies.

Satellite remote sensing based assessments have been performed since the 1970s (Shastri et al., 2023) with satellite data sourced from Earth Resources Technology Satellite (Rango & Anderson, 1974; Deutsch & Ruggles, 1974; Reimold & Linthurst, 1975), Landsat (Qi et al., 2009; Wu & Liu, 2015; Tulbure et al., 2016; Pekel et al., 2016; Rad et al., 2021; Cai et al., 2024), the Moderate Resolution Imaging Spectroradiometer (MODIS; Di Vittorio & Georgakakos, 2018), Sentinel-1 (Torres et al., 2012; DeVries et al., 2020; Vanderhoof et al., 2023), Sentinel-2 (Drusch et al., 2012), WorldView-1 and WorldView-2 (Maglione et al., 2014). The classification of the inundated areas has been performed using the simple index-based threshold method (Gao, 1996; McFeeters, 1996; Xu, 2006; Munasinghe et al., 2018). Simple classification procedures use either optical datasets, such as Sentinel-2, or synthetic aperture radar data, such as Sentinel-1. In contrast, machine learning-based classification methods, such as random forests (Nguyen et al., 2018; Tikuye et al., 2023), support vector machines (Burgess, 1998; Singh et al., 2014; Aryal et al., 2023), and classification and regression trees (CART; Breiman et al., 2017; Pande et al., 2024; Tesfaye et al., 2024; Bayable et al., 2025), can combine multiple bands and water indices to map surface water with high accuracy (Tulbure et al., 2016). On the other hand, multi-source classification can incorporate multiple satellite products and auxiliary data, such as precipitation and topography (Birkett, 1995; Qi et al., 2009; Rossi & Vervoort, 2025). Such multi-source classifications can exhibit better performance, but often at the cost of extensive querying of input data and complex data fusion workflows. Furthermore, for large regional analysis, these workflows incur significant computational expenses (Xue et al., 2011; Gorelick et al., 2017; Tesfaye & Breuer, 2025) and may be infeasible on decentralized, single-node systems and are better processed using distributed computing systems (Wu et al., 2024).

While foundation models such as Prithvi have been applied to flood inundation mapping through fine-tuning on labeled datasets (Li et al., 2023; Kaushik et al., 2026), the use of precomputed foundation model embedding fields as analysis-ready inputs for long-term, regional-scale inundation monitoring of surface water bodies is still in its infancy (Ryan et al., 2026; Kapoor, 2026a). This work addresses this gap by proposing a framework based on the recently launched AlphaEarth Foundations Embeddings (AFE; The AlphaEarth Foundations team, 2025), to assess the inundation dynamics of water bodies in Nebraska - a region characterized by semi-arid climate variability and dependence on the declining

Ogallala Aquifer (Xu, 2023), making it particularly susceptible to long-term changes in surface water availability. These water bodies are part of the HydroLAKES dataset (Messenger et al., 2016), and the 901 such bodies that intersect Nebraska are depicted in Fig. 1. AFE is an annual, 10 m-resolution dataset that provides a ready-to-use, unified digital representation of petabytes of Earth observation data from multiple sources.

Inundation dynamics were characterized using inundation frequency (or water occurrence frequency), which is the percentage of years in which a pixel is classified as inundated. Inundation frequency (IF) reflects the persistence of mapped inundation across years and is an indicator of ecological functions and processes (Millennium Ecosystem Assessment, 2005; Pekel et al., 2016). Specifically, this study pursues the following objectives: i) to classify annual inundation extents across 901 HydroLAKES water bodies in Nebraska using a random forest (supervised) classifier applied to AFE; ii) to compare the classified inundation extents with the Dynamic World product at five example lakes and to evaluate the mapped inundated areas against estimates from field-derived elevation-area relationships at five reservoirs; iii) to compute inundation frequency across the study area to assess spatiotemporal variability; and iv) to estimate trends in inundation extent and to relate the annual extent and its year-to-year changes to rainfall over the 2017–2025 study period.

The workflow for generating the annual inundation layers and the resulting IF layer has been developed and employed in Google Earth Engine (GEE), an online, cloud-based, distributed-computing powered geospatial analysis platform that hosts this dataset. The details of this dataset and GEE are presented in the next section.

Materials and methods

The water bodies were taken from HydroLAKES (Messenger et al., 2016), which includes lakes and reservoirs with a surface area of at least 10 ha. The polygons were read in GEE from the community catalog asset projects/sat-io/open-datasets/HydroLakes/lake_poly_v10. All 901 polygons that intersect the 2016 U.S. Census Bureau TIGER boundary of Nebraska were selected. They comprise 877 lakes and 24 reservoirs, according to the HydroLAKES Lake_type attribute, so they are referred to here as water bodies. All 901 outlines come from the SRTM Water Body Data, derived from the Shuttle Radar Topography Mission of February 2000. Polygons that extend beyond the state line, chiefly Lewis and Clark Lake on the Missouri River, were measured only inside Nebraska. Water bodies smaller than 10 ha are not included, so wetlands are underrepresented.

The embeddings were taken from the Earth Engine collection GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL (Brown et al., 2025), and the classification was performed in January 2026. Each image holds the 64 embedding bands, A00 to A63, at a pixel size of 10 m. For each year from 2017 to 2025, the images intersecting Nebraska, 19 per year, were combined into one mosaic. The AlphaEarth Foundations Satellite Embedding dataset is produced by Google and Google DeepMind.

AlphaEarth Foundations Embeddings (AFE) and Google Earth Engine (GEE)

The embeddings have been derived from the AlphaEarth Foundations model, which is a member of the foundation model class (Blumenfeld, 2023). Foundation models leverage self-supervised learning techniques on massive amounts of unlabeled satellite data across a broad spectrum of applications (Xiao et al., 2025). The AlphaEarth Foundations model assimilated petabytes of remotely sensed and textual data from multiple sources, as presented in Fig. 2. It was trained on more than 3 billion observations from nine gridded data sources and one text source (Brown et al., 2025). In its developers' evaluations, AFE reduced error by 23.9% on average compared with the next-best featurization method (Brown et al., 2025). Additionally, it offers finer resolution than its counterparts, such as Prithvi-EO-2.0, which has a 30 m resolution (Szwarcman et al., 2026). Readers can refer to Brown et al. (2025) for additional details regarding the data sources, data preprocessing, training architecture and procedure, and validation of the AlphaEarth Foundations model.

The trained AlphaEarth Foundations model yields a global, annual embedding dataset at 10 m resolution, where each pixel is represented by a 64-dimensional vector encoding the underlying spectral, temporal, and spatial signatures of the multi-source input data (Brown et al., 2025). The embeddings are the compact, high-level representations of the parent, multi-source data (Kimutai, 2025). These embeddings offer multiple advantages. First, this embedding dataset provides analysis-ready data that does not require standard preprocessing steps such as atmospheric corrections, cloud masking, or speckle filtering (Ryan et al., 2026). Second, users can perform a similarity search by matching the 64-dimensional digital signatures of two locations (spatialthoughts, 2025). For example, a maize field will have a signature similar to that of another maize field but very different from that of an inundated or built-up surface. This property is instrumental for both unsupervised and supervised classification. Unsupervised classification (Cihlar et al., 2000) relies on clustering pixels based on the similarity of their underlying signatures, and the embeddings dataset is ideal for it. On the other hand, supervised classification (Maxwell et al., 2018) requires manual labeling of training data, and AFE was designed and evaluated for mapping from sparse labels (Brown et al., 2025). Furthermore, AFE enables change detection by comparing annual embedding datasets, facilitating hazard mitigation, environmental monitoring, and the assessment of crop yields, surface water dynamics, and land-use changes. However, the large-scale operation on this high-dimensional dataset requires distributed computing, as a single-threaded approach can be computationally inefficient.

Google Earth Engine (Hasan et al., 2024) is a cloud-based, distributed-computing powered remote sensing and GIS platform that hosts hundreds of datasets, including AFE. It also includes built-in supervised classification functions, making it an ideal platform for leveraging the aforementioned embedding capabilities. Furthermore, it allows the sharing of scripts, which makes collaboration easier, and the community forum provides crowdsourced support for users. It is freely available for non-commercial purposes, further lowering the barrier to large-scale geospatial analysis. GEE can be used in its native JavaScript-based code editor. This code editor includes a code input window, a pixel inspector to investigate the underlying bands and properties, and a layer visualizer. This code editor, with AFE for

2017 and pixel-level signatures, is illustrated in Fig. 3. The GEE script can be accessed at <https://code.earthengine.google.com/bfaf1067b39ea6b3ef8ad628b673921e>. GEE can also be used via its Python API (Amani et al., 2020), but this study used the native code editor.

Classification and IF analysis

The inundated pixels were obtained by classifying the annual embeddings in GEE using the built-in supervised classification functions. First, the embeddings dataset was filtered to retain tiles within the study area. Then, supervised classification was performed using random forests (RFs; Breiman, 2001) with 150 trees, as used by Yue et al. (2023). All 64 dimensions (bands) were used as input features for classification. Three distinct classes comprising i) inundation, ii) vegetation and iii) built-up area (constructed area) were considered. Mutually exclusive labeled data were prepared for training the classifiers. Then, an independent validation dataset was prepared for assessing the accuracy of the trained classifiers. True-color Sentinel-2 composites, high-resolution Google basemaps and the 10 m ESA WorldCover map (Zanaga et al., 2021) were used as references for this purpose. The training set comprised 30,000 data points, and the validation set comprised 7,500 points, with an equal number of data points per class. Separate training and validation polygons were delineated to keep the two sets independent. Training and validation polygons were separated by a minimum distance of 17 km for inundation, 38 km for vegetation, and 80 km for built-up area. The built-up validation points were derived from ESA WorldCover built-up areas in seven cities not used for training: Grand Island, Kearney, Hastings, North Platte, Scottsbluff, Norfolk and Columbus. The number of data points was selected with consideration for GEE's computational limits. The training was performed on the 2017 annual embedding data, and validation on the 2018 annual embedding data. The trained classifier was subsequently applied to all annual embedding datasets from 2017 to 2025. Its applicability to other years was evaluated by the 2018 validation and by the 2017–2025 comparison with elevation-area relationships at five reservoirs. The RF classifier achieved a training accuracy of 99.94%. Evaluation on the separate 2018 validation set of 7,500 points yielded an overall accuracy of 94.33% and a Kappa coefficient of 0.915. Both producer's and user's accuracies for the inundation class were 100% within this validation sample. In a Landsat-based random forest study, Yue et al. (2023) also reported near-perfect accuracy for permanent water in the centers of large lakes and reservoirs (average producer's accuracy of 99.8%), but 93.6% for seasonal water at the edges of water bodies. However, the high agreement for the inundation class may partly reflect the selection of lake-interior reference points and the use of three broad land-cover classes. These estimates apply to the selected reference polygons and do not establish classification accuracy at lake margins or across Nebraska as a whole. The other classification settings were three variables per split, a bag fraction of 0.1, at most 100 leaf nodes per tree, a minimum leaf population of 1 and a seed of 0 (`smileRandomForest` in GEE); none was tuned. Detailed confusion and error matrices are presented in Table S1 in the supplementary information.

The IF was calculated by classifying annual raster embeddings to identify submerged pixels, which were then aggregated over the study period. Each annual classified raster was converted to a binary raster, with inundated pixels set to 1 and non-inundated pixels to 0. A summing operation was performed on

these rasters, which gave the number of years in which a pixel was classified as inundated. Finally, this layer was divided by the number of years and multiplied by 100 to obtain the IF raster, in percent. This formulation is mathematically presented in the following equation.

$$IF_i = \frac{\sum_{Y=1}^{NY} IS_{iY}}{NY} \times 100 \quad (1)$$

Where IF_i is the IF (in percent) at the i th pixel over the study period, and NY is the number of years (nine, from 2017 to 2025). IS_{iY} is the inundation status for the i th pixel in the Y^{th} year. It is 1 for pixels classified as inundated and 0 for valid pixels classified as non-inundated. Pixels without a valid classification in all nine years were excluded from IF. An IF of 0% means the pixel was not classified as inundated in any of the nine annual layers, and an IF of 100% means it was classified as inundated in all nine. A value of 100% does not mean the pixel was under water all year.

This layer has also been used to perform zonal statistics. With nine annual layers, IF can take only ten values, from 0% to 100% in steps of one year (about 11.1%). The IF raster was reclassified into four IF categories: i) > 0 -<25%, ii) 25-<50%, iii) 50-<75% and iv) 75-100% (pixels classified as inundated in 1-2, 3-4, 5-6 and 7-9 of the nine years, respectively), and a zonal statistics operation was performed to find the share of each category in the area inundated in at least one year. Pixels never classified as inundated (IF of 0%) were left out. The 901 water bodies were classified by their HydroLAKES area (the Lake_area attribute) into three size groups to assess the effects of size on inundation: small ($< 1 \text{ km}^2$, 806 water bodies), medium (1 to $< 10 \text{ km}^2$, 88) and large ($\geq 10 \text{ km}^2$, 7). Inundated area was measured within a 250 m buffer around each HydroLAKES polygon, because the polygons can miss water when the water levels are high. The buffers were clipped at the state line, and areas were computed on a 10 m equal-area grid (EPSG:5070). The unbuffered polygons were also measured for comparison. A buffer can include nearby ponds or river reaches, and water shared by overlapping buffers is counted for each water body.

Validation and performance comparison with Dynamic World

Validation was performed in two ways. The first involved visually comparing the 2017 classified inundation extents of five example lakes with true- and false-color Sentinel-2 composites of the same year (Fig. 4). Since Sentinel-2 is among the source datasets used to generate AFE, this comparison is not fully independent and should be interpreted as a qualitative check rather than a rigorous validation. Furthermore, the classified inundation for five reservoirs was validated against the inundation obtained from their elevation-area relationships. These relationships take the water surface elevation as input and return the corresponding inundated area as output. In addition, the 2017 random forest map was compared with Dynamic World (GOOGLE/DYNAMICWORLD/V1 ; Brown et al., 2022), a Sentinel-2 based product that labels the land cover in each image. A pixel was mapped as inundated where the median of its 2017 Dynamic World labels was water (class 0). The two maps were compared at five example lakes, within the same 250 m buffers.

Trend analysis

A trend analysis was performed to assess variation in the mean inundation of water bodies and in average rainfall, which is one of the most important driving factors of inundation dynamics (Huang et al., 2022; Tesfaye & Breuer, 2025). The trend analysis used the mean inundated area per water body for each year, calculated by dividing the summed inundated area of the 901 water bodies by 901. The five largest water bodies accounted for 32% of the area summed across the nine years, highlighting the potential influence of large water bodies on this aggregate. This metric is not normalized by water-body size and represents only the selected inventory, rather than all surface water in Nebraska. Zonal statistics were applied to the annual inundation layers to first calculate the area inundated at each lake each year. Annual rainfall was the mean calendar-year rainfall over Nebraska from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS; Funk et al., 2015). Sen's slope technique (Sen, 1968; Gocic & Trajkovic, 2013) was applied to estimate the rate of change (overall trend) in inundation and rainfall. This is a non-parametric technique that is robust to outliers and thus preferred over the standard regression analysis. The Mann–Kendall test (Gocic & Trajkovic, 2013) was used to test whether each trend was statistically significant. A normalized area gave each water body equal weight: each water body's annual area was expressed as a percent departure from its 2017–2025 mean, and the departures were averaged. The year-to-year changes in the mean and normalized areas were compared with the rainfall of the same year, and the mean area with the rainfall of one and two years before, using the Pearson correlation coefficient (Machiwal & Jha, 2012).

Results and discussion

Comparison with Dynamic World

The classified inundated areas are sparsely distributed across Nebraska and are harder to discern at a large spatial and temporal scale in an article. Inundation maps for five lakes representing different sizes and locations are presented for 2017, the year used to train the random forest classifier. Three of these lakes (Lake McConaughy, Babcock Lake, and Mother Lake) contain water training polygons. These examples therefore illustrate the classification results and do not constitute an independent assessment of accuracy. The classified inundation extents and associated false-color composites for all 9 years are available via this web app: <https://adityain2003.users.earthengine.app/view/comparison>.

The subset of lakes discussed here consists of i) Lake McConaughy (Lat: 41.2505, Long: -101.8158), ii) School Lake (Lat: 42.52707, Long: -100.61059), iii) Babcock Lake (Lat: 41.48498, Long: -97.36052), iv) Mother Lake (Lat: 42.14951, Long: -101.84197) and v) Pawnee Lake (Lat: 40.84741, Long: -96.87407). The RGB (column 1) and false color composite (column 2) derived from Sentinel-2 imagery are shown in Fig. 4, providing visual reference for comparison, as in previous studies (Li et al., 2023; Kaushik et al., 2026). Table 1 gives each difference as a percentage of the random forest area; the random forest serves as the comparator, not as a reference of known accuracy.

The classified inundations from the random forest were compared with those from the Dynamic World classification product for the same five water bodies. Columns 3 and 4 of Fig. 4 show the random forest-

and Dynamic World derived classified inundation regions, respectively. Dynamic World mapped more inundated area than the random forest at four of the five lakes (white pixels in column 5), but 1.0% less at Lake McConaughy. At the inflow end of the lake, shown in Fig. 4, Dynamic World mapped 20.5% less. This end slopes gently, so its shoreline shifts most as the water level changes, which likely explains the larger difference. The percentage differences for the five water bodies are provided in Table 1, with an average absolute difference of 15.31%. Especially for Babcock Lake (row 3), Dynamic World maps substantially more inundated area, classifying the centrally passing road (marked by a thin line) and the central region, which appears vegetated in the false-color composite, as inundated.

Table 1
Difference in 2017 inundated area within each lake's 250 m buffer: Dynamic World minus random forest, as a percentage of the random forest area

No.	Water body	% Difference Dynamic World
1	Lake McConaughy	-1.0%
2	School Lake	+ 19.7%
3	Babcock Lake	+ 39.6%
4	Mother Lake	+ 5.8%
5	Pawnee Lake	+ 10.4%

Validation with the field data

A more rigorous method of validating the classified inundations is to compare them against the inundated area estimated from the elevation-area relationship. The elevation-area relationship links the elevation of a water body's water level to its area and is obtained from rigorous field surveys. Such relationships, from surveys dating from the 1980s to 2003, were available for five reservoirs: Lake McConaughy (Kress et al., 2005), Harlan County Lake, Swanson Lake, Hugh Butler Lake and Enders Reservoir. Daily water-level elevations were obtained from the Central Nebraska Public Power & Irrigation District (CNPPID) for Lake McConaughy and from the U.S. Bureau of Reclamation for the other four reservoirs, and converted to annual average elevations for the years 2017 to 2025. Then the inundated area corresponding to these elevations was computed from the elevation-area relationship. The classified inundated area for these years was computed within the 250 m buffer by using the zonal statistics function. The inundated areas obtained from the elevation-area relationship and classification are presented in Fig. 5. Over the 45 reservoir-years, the classified area was on average 2.73% lower than the elevation-area value, with a mean absolute percentage difference of 3.32% and a root mean square error of 1.85 km², indicating close agreement. The mean difference ranged from + 0.02% at Swanson Lake to - 5.38% at Hugh Butler Lake, and R² from 0.794 at Swanson Lake to 0.995 at Lake McConaughy. An earlier analysis of the same five reservoirs within the HydroLAKES polygons found a mean difference of - 3.9% (Kapoor, 2026a). It may be noted that the elevation-area relationship-derived inundated areas are slightly higher than the classified inundation at four of the five reservoirs, especially in the high-water

year of 2019. Closer investigation revealed that the inundated area extends beyond the HydroLAKES boundary in such years, so measuring within the polygons alone underestimates the classified inundated area; within the polygons, the mean difference was -3.95% . Another possible reason for the mismatch is that the elevation-area relationships may be old and have changed over the last two decades due to sediment deposition (Kress et al., 2005; Gebregiorgis & Hossain, 2012). A third possible reason is the position of the water training points, which were placed in lake interiors; placing training points explicitly along the shorelines may better capture the characteristics of shallow shoreline water and improve its classification. With five reservoirs, the contributions of these three causes cannot be separated.

Computation of IF and zonal statistics

The computed IF raster layer for the study area is hosted as an interactive GEE web app at <https://adityain2003.users.earthengine.app/view/inundation-frequency-webapp>. It allows users to explore annual inundation status at any location. Figure 6a shows the IF for Lake McConaughy extracted from this web app. It may be noted that the core (central) region of the lake shows 100% IF, meaning that it was classified as inundated in all nine years, while the peripheral regions show lower IF on account of fluctuating inundated extent of the lake.

Figure 6b shows how the area classified as inundated in at least one year is distributed among the four IF categories. Areas were pooled across the water bodies within each size group. Seventeen small water bodies had no mapped inundation in any year, so they had no inundated area to divide among the categories and were therefore excluded. The 75–100% IF category, comprising pixels classified as inundated in at least seven of the nine years, accounts for the largest share in every size group. This category covers 73.4% of the pooled inundated area in small water bodies ($< 1 \text{ km}^2$), compared with 86.9% in medium ($1 \text{ to } < 10 \text{ km}^2$) and 85.6% in large ($\geq 10 \text{ km}^2$) water bodies. Small water bodies have a greater share of area classified as inundated in only a few years. For example, the lowest IF category ($> 0\text{--}25\%$), representing inundation in one or two years, accounts for 10.8% of this area in small water bodies, compared with 5.4% in medium and 2.8% in large water bodies. Each intermediate category (25–50% and 50–75% IF) accounts for less than 10% in every size group. These patterns indicate that inundation is less persistent in small water bodies than in medium and large ones, consistent with the higher inter-annual variability of small reservoirs reported by Donchyts et al. (2022).

Annual trend analysis

The temporal analysis of 901 water bodies in Nebraska from 2017 to 2025 shows no statistically significant trend in regional surface water extent or rainfall (Mann–Kendall test; Table S2 in the supplementary information). The application of the non-parametric Sen's slope estimator, chosen for its robustness against outliers, indicates a decline in both surface water extent and rainfall as presented in Fig. 7a, b.

Specifically, the Sen's slope analysis estimates a reduction in the annual mean inundation area of $7,073 \text{ m}^2/\text{year}$, or $2,895 \text{ m}^2/\text{year}$ without Lake McConaughy. This estimated reduction in surface water extent

coincides with an estimated decline in mean annual rainfall of 14.80 mm/year over the same period. The year-to-year change in inundation area followed the rainfall of the same year (Fig. 7c, d), with a correlation of 0.73 for the mean inundation area and 0.91 for the normalized area. This suggests that the regional hydrological footprint of the water bodies is sensitive to precipitation patterns. However, the relationship between these two variables is not always perfectly linear. In some years, the opposite trend is observed: the inundation area remains stable or even increases despite lower-than-average precipitation. This may reflect hydrological lag effects (Berghuijs et al., 2025), wherein surface water extent is influenced by cumulative storage from prior high-rainfall years. The mean inundation area correlated more closely with the rainfall of one and two years before ($r = 0.50$ and 0.72) than with that of the same year ($r = 0.13$), which is consistent with such lag effects.

The stronger response of the normalized area, together with the higher percentage of area in lower IF categories for smaller water bodies (Fig. 6b), suggests that smaller, more volatile water bodies respond more directly to wet and dry years than larger, more stable ones. By utilizing the AFE within GEE, this study quantified these year-to-year changes across 901 water bodies with minimal preprocessing.

Conclusions

This study demonstrates the utility of the AFE as an analysis-ready dataset for monitoring inundation dynamics across the HydroLAKES dataset, which consists of 901 water bodies in Nebraska. By leveraging the distributed computing power of Google Earth Engine, the research provides a scalable framework for assessing large-scale hydrological changes with less preprocessing and data fusion than typical multi-source remote sensing workflows.

Dynamic World mapped a larger inundated area than random forest at four of the five example lakes and a smaller area at Lake McConaughy. These differences indicate disagreement between the two maps; a common independent reference would be needed to establish their relative accuracy. Random forest also allows the target water class to be defined explicitly through labeled training samples. Validation against field-derived elevation-area relationships for five reservoirs showed a mean difference of -2.73% and a mean absolute percentage difference of 3.32% , providing quantitative accuracy estimates for reservoirs spanning 2 to 116 km^2 in area.

The estimated decline in surface water extent over the study period, with water bodies cataloged as losing an average of $7,073 \text{ m}^2$ annually, was not statistically significant. However, the year-to-year change in inundated area followed annual rainfall ($r = 0.73$, or 0.91 when each water body counted equally), highlighting the sensitivity of Nebraska's water bodies to wet and dry years. Furthermore, zonal statistics on IF underscored that smaller water bodies ($< 1 \text{ km}^2$) are less persistently inundated than larger, more stable water bodies.

This work shows how the AFE embeddings can be classified to monitor inundation across a regional inventory of water bodies at an annual scale. However, this study has several limitations. First, the temporal coverage of this analysis is constrained by the availability of AFE, which currently spans 2017

to 2025; as AFE is updated with additional years, the framework developed here can be directly extended to support longer-term trend assessments without modification. Second, visual comparison against Sentinel-2 composites is not fully independent, since Sentinel-2 is among the source datasets used to generate AFE; this is partially mitigated by the quantitative field validation using elevation-area relationships, which uses independent reference data, although the gauged reservoirs also contain water training polygons. Third, quantitative field validation was limited to five reservoirs due to the unavailability of elevation-area relationships for other water bodies; these reservoirs span 2 to 116 km², whereas 806 of the 901 water bodies are smaller than 1 km², so the accuracy of the framework for small water bodies remains untested. Fourth, the random forest classifier was trained on 2017 embedding data and applied across all years; temporal shifts in land cover could therefore introduce classification uncertainty, and apart from the 2018 validation and the comparison at the five gauged reservoirs, no years were held out to test this applicability.

Future research should extend the temporal coverage as AFE data becomes available for additional years, implement in-field validation using drone-based imagery, expand the analysis to a seasonal scale to capture intra-annual inundation dynamics, apply the framework to other regions with contrasting hydrological regimes, and integrate in-situ data to quantify the societal and ecological impacts of changing surface water availability. Other classification techniques, such as unsupervised clustering (for example, K-means) and other supervised classifiers (for example, support vector machines and classification and regression trees), could also be compared against a common independent reference.

Declarations

Conflict of interest:

The author declares no conflict of interest.

Declarations of generative AI and AI-assisted technologies in the manuscript preparation process:

During the preparation of this work, the author used Grammarly, Claude (Anthropic), and Codex (OpenAI) to assist with correcting grammatical errors, refining language, and writing and reviewing the codebase. The author has reviewed the content and takes full responsibility for it.

Ethics approval:

All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Funding:

No funding was received for conducting this study.

Author Contribution

AK: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review and editing.

Acknowledgement

The author thanks the Central Nebraska Public Power & Irrigation District (CNPPID) for providing daily water-level elevation data for Lake McConaughy, which supported the field-based validation of the classification results.

Data Availability

The datasets generated in this study are publicly available on Zenodo at <https://doi.org/10.5281/zenodo.18741436> (Kapoor, 2026b). The repository includes annual inundation classification rasters from random forest (2017–2025), Dynamic World comparison data, and inundation frequency rasters, all at 10 m resolution covering 901 HydroLAKES water bodies in Nebraska. Interactive web applications presenting the annual classified inundation extents and the inundation frequency layer are available at <https://adityain2003.users.earthengine.app/view/comparison> and <https://adityain2003.users.earthengine.app/view/inundation-frequency-webapp>, respectively.

References

1. The AlphaEarth Foundations team. (2025, July 30). AlphaEarth Foundations helps map our planet in unprecedented detail. *Google DeepMind*. <https://deepmind.google/blog/alphaearth-foundations-helps-map-our-planet-in-unprecedented-detail/>
2. Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., Moghaddam, S. H. A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., & Brisco, B. (2020). Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5326–5350. <https://doi.org/10.1109/JSTARS.2020.3021052>
3. Aryal, J., Sitaula, C., & Frery, A. C. (2023). Land use and land cover (LULC) performance modeling using machine learning algorithms: A case study of the city of Melbourne, Australia. *Scientific Reports*, 13(1), Article 13510. <https://doi.org/10.1038/s41598-023-40564-0>

4. Bayable, G., Gebrie, G., Melese, T., & Melaku, A. (2025). Land use/cover classification using machine learning algorithms and their impacts on land surface temperature and soil moisture in the Alawuha Watershed, Ethiopia. *Environmental and Sustainability Indicators*, 27, Article 100797. <https://doi.org/10.1016/j.indic.2025.100797>
5. Berghuijs, W. R., Woods, R. A., Anderson, B. J., Hemshorn de Sánchez, A. L., & Hrachowitz, M. (2025). Annual memory in the terrestrial water cycle. *Hydrology and Earth System Sciences*, 29(5), 1319–1333. <https://doi.org/10.5194/hess-29-1319-2025>
6. Birkett, C. M. (1995). The contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes. *Journal of Geophysical Research: Oceans*, 100(C12), 25179–25204. <https://doi.org/10.1029/95JC02125>
7. Blumenfeld, J. (2023, August 3). *NASA and IBM openly release geospatial AI foundation model for NASA Earth observation data*. NASA Earthdata. <https://www.earthdata.nasa.gov/news/nasa-ibm-openly-release-geospatial-ai-foundation-model-nasa-earth-observation-data>
8. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
9. Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>
10. Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., & Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1), Article 251. <https://doi.org/10.1038/s41597-022-01307-4>
11. Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., & Kohli, P. (2025). *AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data* (Version 2). arXiv. <https://doi.org/10.48550/arXiv.2507.22291>
12. Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 121–167. <https://doi.org/10.1023/A:1009715923555>
13. Cai, Y., Shi, Q., & Liu, X. (2024). Spatiotemporal mapping of surface water using Landsat images and spectral mixture analysis on Google Earth Engine. *Journal of Remote Sensing*, 4, Article 0117. <https://doi.org/10.34133/remotesensing.0117>
14. Cihlar, J., Latifovic, R., & Beaubien, J. (2000). A comparison of clustering strategies for unsupervised classification. *Canadian Journal of Remote Sensing*, 26(5), 446–454. <https://doi.org/10.1080/07038992.2000.10855276>
15. Deutsch, M., & Ruggles, F. (1974). Optical data processing and projected applications of the ERTS-1 imagery covering the 1973 Mississippi River Valley floods. *Water Resources Bulletin*, 10(5), 1023–1039. <https://doi.org/10.1111/j.1752-1688.1974.tb00622.x>

16. DeVries, B., Huang, C., Armston, J., Huang, W., Jones, J. W., & Lang, M. W. (2020). Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine. *Remote Sensing of Environment*, 240, Article 111664. <https://doi.org/10.1016/j.rse.2020.111664>
17. Di Vittorio, C. A., & Georgakakos, A. P. (2018). Land cover classification and wetland inundation mapping using MODIS. *Remote Sensing of Environment*, 204, 1–17. <https://doi.org/10.1016/j.rse.2017.11.001>
18. Donchyts, G., Winsemius, H., Baart, F., Dahm, R., Schellekens, J., Gorelick, N., Iceland, C., & Schmeier, S. (2022). High-resolution surface water dynamics in Earth's small and medium-sized reservoirs. *Scientific Reports*, 12(1), Article 13776. <https://doi.org/10.1038/s41598-022-17074-6>
19. Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
20. Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, 2(1), Article 150066. <https://doi.org/10.1038/sdata.2015.66>
21. Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
22. Gebregiorgis, A. S., & Hossain, F. (2012). Hydrological risk assessment of old dams: Case study on Wilson Dam of Tennessee River Basin. *Journal of Hydrologic Engineering*, 17(1), 201–212. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000410](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000410)
23. Gocic, M., & Trajkovic, S. (2013). Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. *Global and Planetary Change*, 100, 172–182. <https://doi.org/10.1016/j.gloplacha.2012.10.014>
24. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
25. Hall, A., Thomas, R. F., & Wassens, S. (2019). Mapping the maximum inundation extent of lowland intermittent riverine wetland depressions using LiDAR. *Remote Sensing of Environment*, 233, Article 111376. <https://doi.org/10.1016/j.rse.2019.111376>
26. Hasan, R., Kapoor, A., Singh, R., & Yadav, B. K. (2024). A state-of-the-art review on the quantitative and qualitative assessment of water resources using Google Earth Engine. *Environmental Monitoring and Assessment*, 196(12), Article 1266. <https://doi.org/10.1007/s10661-024-13315-5>
27. Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, extracting, and monitoring surface water from space using optical sensors: A review. *Reviews of Geophysics*, 56(2), 333–360. <https://doi.org/10.1029/2018RG000598>

28. Huang, C., Smith, L. C., Kyzivat, E. D., Fayne, J. V., Ming, Y., & Spence, C. (2022). Tracking transient boreal wetland inundation with Sentinel-1 SAR: Peace-Athabasca Delta, Alberta and Yukon Flats, Alaska. *GIScience & Remote Sensing*, 59(1), 1767–1792.
<https://doi.org/10.1080/15481603.2022.2134620>
29. Kapoor, A. (2026a, December 1–4). *Evaluating AlphaEarth Foundations embeddings for annual inundation mapping: A case study of Nebraska* [Paper presentation]. IEEE India Geoscience and Remote Sensing Symposium (InGARSS 2026), Hyderabad, India.
30. Kapoor, A. (2026b). *A multi-year dataset of annual inundation dynamics in Nebraska (2017–2025) derived from AlphaEarth Foundations embeddings* [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.18741436>
31. Kaushik, S., Maurya, L., Tellman, B., & Zhang, Z. (2026). Assessing geo-foundational models for flood inundation mapping: Benchmarking models for Sentinel-1, Sentinel-2, and Planetscope. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 19, 5649–5665.
<https://doi.org/10.1109/JSTARS.2026.3656855>
32. Khare, A., Gupta, B. C., Rajib, A., Vanderhoof, M., & Wu, Q. (2025). Estimates of global surface water dynamics harnessing near real-time land cover observations and open science geospatial capabilities. *Environmental Research Letters*, 20(12), Article 124042. <https://doi.org/10.1088/1748-9326/ae137b>
33. Kimutai, L. (2025, July 31). AlphaEarth by Google DeepMind: What it means for GIS and remote sensing professionals. *Medium*. <https://medium.com/@kimutai.lawrence19/alphaearth-by-google-deepmind-what-it-means-for-gis-and-remote-sensing-professionals-0312d55797d2>
34. Kress, W. H., Seabee, S. K., Littin, G. R., Drain, M. A., & Kling, M. E. (2005). *Comparison of preconstruction and 2003 bathymetric and topographic surveys of Lake McConaughy, Nebraska* (Scientific Investigations Report 2005–5040). U.S. Geological Survey.
<https://doi.org/10.3133/sir20055040>
35. Li, W., Lee, H., Wang, S., Hsu, C.-Y., & Arundel, S. T. (2023). Assessment of a new GeoAI foundation model for flood inundation mapping. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery* (pp. 102–109). Association for Computing Machinery. <https://doi.org/10.1145/3615886.3627747>
36. Machiwal, D., & Jha, M. K. (2012). *Hydrologic time series analysis: Theory and practice*. Springer; Capital Publishing Company. <https://doi.org/10.1007/978-94-007-1861-6>
37. Maglione, P., Parente, C., & Vallario, A. (2014). Coastline extraction using high resolution WorldView-2 satellite imagery. *European Journal of Remote Sensing*, 47(1), 685–699.
<https://doi.org/10.5721/EuJRS20144739>
38. Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817.
<https://doi.org/10.1080/01431161.2018.1433343>

39. McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432.
<https://doi.org/10.1080/01431169608948714>
40. Messenger, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7(1), Article 13603. <https://doi.org/10.1038/ncomms13603>
41. Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. Island Press.
42. Munasinghe, D., Cohen, S., Huang, Y.-F., Tsang, Y.-P., Zhang, J., & Fang, Z. (2018). Intercomparison of satellite remote sensing-based flood inundation mapping techniques. *JAWRA Journal of the American Water Resources Association*, 54(4), 834–846. <https://doi.org/10.1111/1752-1688.12626>
43. Nguyen, H. T. T., Doan, T. M., & Radeloff, V. (2018). Applying random forest classification to map land use/land cover using Landsat 8 OLI. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3/W4, 363–367. <https://doi.org/10.5194/isprs-archives-XLII-3-W4-363-2018>
44. Pande, C. B., Diwate, P., Orimoloye, I. R., Sidek, L. M., Pratap Mishra, A., Moharir, K. N., Pal, S. C., Alshehri, F., & Tolche, A. D. (2024). Impact of land use/land cover changes on evapotranspiration and model accuracy using Google Earth Engine and classification and regression tree modeling. *Geomatics, Natural Hazards and Risk*, 15(1), Article 2290350.
<https://doi.org/10.1080/19475705.2023.2290350>
45. Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422.
<https://doi.org/10.1038/nature20584>
46. Qi, S., Brown, D. G., Tian, Q., Jiang, L., Zhao, T., & Bergen, K. M. (2009). Inundation extent and flood frequency mapping using LANDSAT imagery and digital elevation models. *GIScience & Remote Sensing*, 46(1), 101–127. <https://doi.org/10.2747/1548-1603.46.1.101>
47. Rad, A. M., Kreitler, J., & Sadegh, M. (2021). Augmented Normalized Difference Water Index for improved surface water monitoring. *Environmental Modelling & Software*, 140, Article 105030.
<https://doi.org/10.1016/j.envsoft.2021.105030>
48. Rango, A., & Anderson, A. T. (1974). Flood hazard studies in the Mississippi River Basin using remote sensing. *Water Resources Bulletin*, 10(5), 1060–1081. <https://doi.org/10.1111/j.1752-1688.1974.tb00625.x>
49. Reimold, R. J., & Linthurst, R. A. (1975). Use of remote sensing for mapping wetlands. *Transportation Engineering Journal of ASCE*, 101(2), 189–198. <https://doi.org/10.1061/TPEJAN.0000497>
50. Rossi, M. J., & Vervoort, R. W. (2025). Enhancing inundation mapping with geomorphological segmentation: Filling in gaps in spectral observations. *Science of the Total Environment*, 997, Article 180180. <https://doi.org/10.1016/j.scitotenv.2025.180180>

51. Ryan, S., Powell, M., Ling, J., & Wen, L. (2026). Streamlining wetland vegetation mapping with AlphaEarth embeddings: Comparable accuracy to traditional methods with cleaner maps and minimal preprocessing. *Remote Sensing*, 18(2), Article 293. <https://doi.org/10.3390/rs18020293>
52. Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
53. Shastry, A., Carter, E., Coltin, B., Sleeter, R., McMichael, S., & Eggleston, J. (2023). Mapping floods from remote sensing data and quantifying the effects of surface obstruction by clouds and vegetation. *Remote Sensing of Environment*, 291, Article 113556. <https://doi.org/10.1016/j.rse.2023.113556>
54. Singh, S. K., Srivastava, P. K., Gupta, M., Thakur, J. K., & Mukherjee, S. (2014). Appraisal of land use/land cover of mangrove forest ecosystem using support vector machine. *Environmental Earth Sciences*, 71(5), 2245–2255. <https://doi.org/10.1007/s12665-013-2628-0>
55. spatialthoughts. (2025, July 30). *Similarity search with Satellite Embedding dataset*. Google for Developers. <https://developers.google.com/earth-engine/tutorials/community/satellite-embedding-05-similarity-search>
56. Szwarcman, D., Roy, S., Fraccaro, P., Gíslason, Þ. E., Blumenstiel, B., Ghosal, R., de Oliveira, P. H., de Sousa Almeida, J. L., Sedona, R., Kang, Y., Chakraborty, S., Wang, S., Gomes, C., Kumar, A., Gaur, V., Truong, M., Godwin, D., Khallaghi, S., Lee, H.,... . Moreno, J. B. (2026). Prithvi-EO-2.0: A versatile multitemporal foundation model for Earth observation applications. *IEEE Transactions on Geoscience and Remote Sensing*, 64, Article 4400120. <https://doi.org/10.1109/TGRS.2025.3642610>
57. Tesfaye, M., & Breuer, L. (2025). Remote sensing with machine learning for multi-decadal surface water monitoring in Ethiopia. *Scientific Reports*, 15(1), Article 12444. <https://doi.org/10.1038/s41598-025-96955-y>
58. Tesfaye, W., Elias, E., Warkineh, B., Tekalign, M., & Abebe, G. (2024). Modeling of land use and land cover changes using Google Earth Engine and machine learning approach: Implications for landscape management. *Environmental Systems Research*, 13(1), Article 31. <https://doi.org/10.1186/s40068-024-00366-3>
59. Tikuye, B. G., Rusnak, M., Manjunatha, B. R., & Jose, J. (2023). Land use and land cover change detection using the random forest approach: The case of the Upper Blue Nile River Basin, Ethiopia. *Global Challenges*, 7(10), Article 2300155. <https://doi.org/10.1002/gch2.202300155>
60. Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A.,... . Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–24. <https://doi.org/10.1016/j.rse.2011.05.028>
61. Tulbure, M. G., Broich, M., Stehman, S. V., & Kommareddy, A. (2016). Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-

- arid region. *Remote Sensing of Environment*, 178, 142–157.
<https://doi.org/10.1016/j.rse.2016.02.034>
62. Vanderhoof, M. K., Alexander, L., Christensen, J., Solvik, K., Nieuwlandt, P., & Sagehorn, M. (2023). High-frequency time series comparison of Sentinel-1 and Sentinel-2 satellites for mapping open and vegetated water across the United States (2017–2021). *Remote Sensing of Environment*, 288, Article 113498. <https://doi.org/10.1016/j.rse.2023.113498>
 63. Wu, G., & Liu, Y. (2015). Capturing variations in inundation with satellite remote sensing in a morphologically complex, large lake. *Journal of Hydrology*, 523, 14–23.
<https://doi.org/10.1016/j.jhydrol.2015.01.048>
 64. Wu, J., Gan, W., Chao, H.-C., & Yu, P. S. (2024). Geospatial big data: Survey and challenges. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 17007–17020.
<https://doi.org/10.1109/JSTARS.2024.3438376>
 65. Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., & Yokoya, N. (2025). Foundation models for remote sensing and Earth observation: A survey. *IEEE Geoscience and Remote Sensing Magazine*, 13(4), 297–324. <https://doi.org/10.1109/MGRS.2025.3576766>
 66. Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025–3033.
<https://doi.org/10.1080/01431160600589179>
 67. Xu, Y. (2023, April 13). *Ogallala Aquifer, keeping most of state well-watered, continues to shrink in southwest Nebraska*. Flatwater Free Press. <https://flatwaterfreepress.org/ogallala-aquifer-keeping-most-of-state-well-watered-continues-to-shrink-in-southwest-nebraska/>
 68. Xue, Y., Chen, Z., Xu, H., Ai, J., Jiang, S., Li, Y., Wang, Y., Guang, J., Mei, L., Jiao, X., He, X., & Hou, T. (2011). A high throughput geocomputing system for remote sensing quantitative retrieval and a case study. *International Journal of Applied Earth Observation and Geoinformation*, 13(6), 902–911.
<https://doi.org/10.1016/j.jag.2011.06.006>
 69. Yue, L., Li, B., Zhu, S., Yuan, Q., & Shen, H. (2023). A fully automatic and high-accuracy surface water mapping framework on Google Earth Engine using Landsat time-series. *International Journal of Digital Earth*, 16(1), 210–233. <https://doi.org/10.1080/17538947.2023.2166606>
 70. Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E.,... Arino, O. (2021). *ESA WorldCover 10 m 2020 v100* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5571936>

Figures

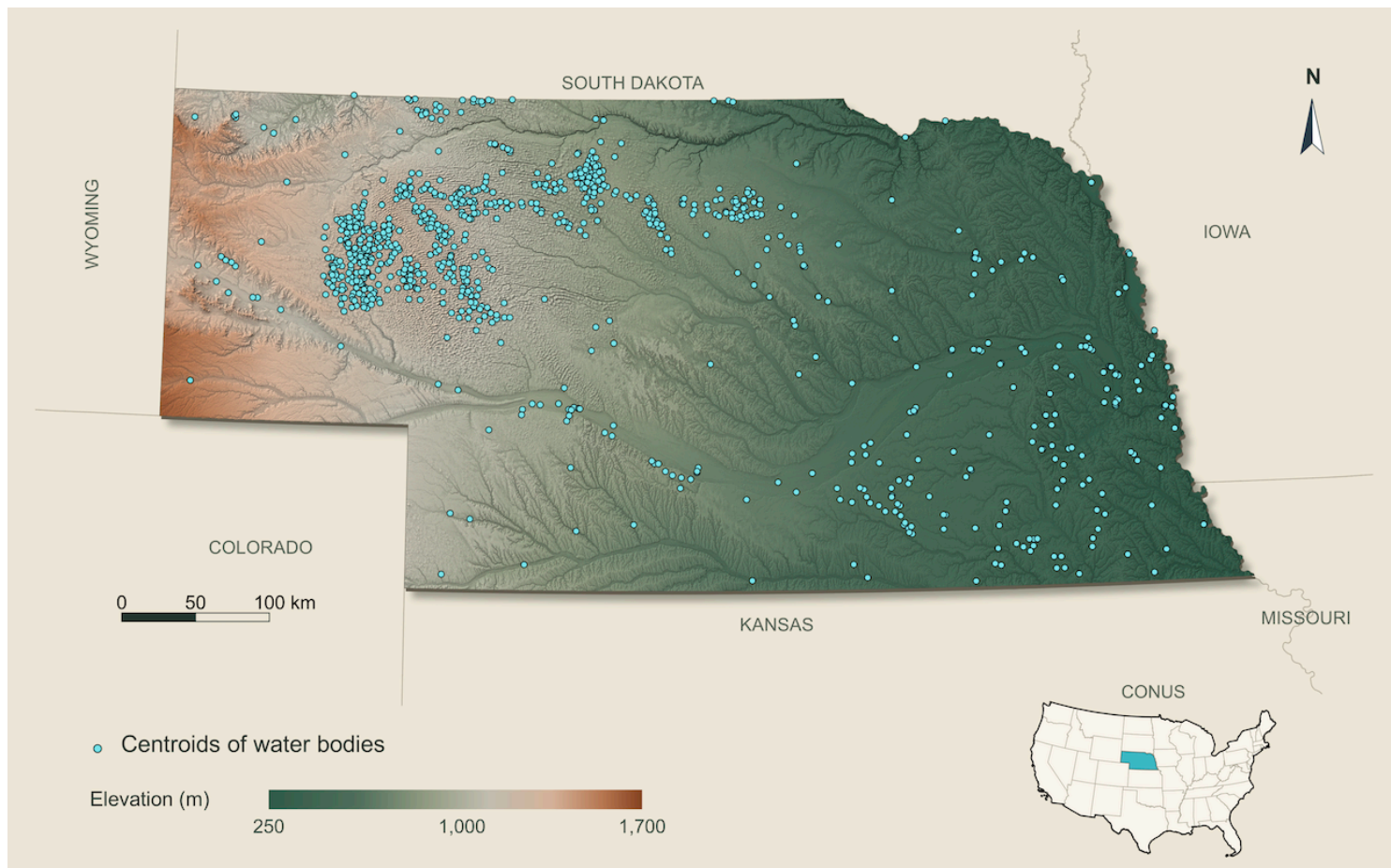


Figure 1

Geographical setting of Nebraska and locations of the 901 HydroLAKES water bodies

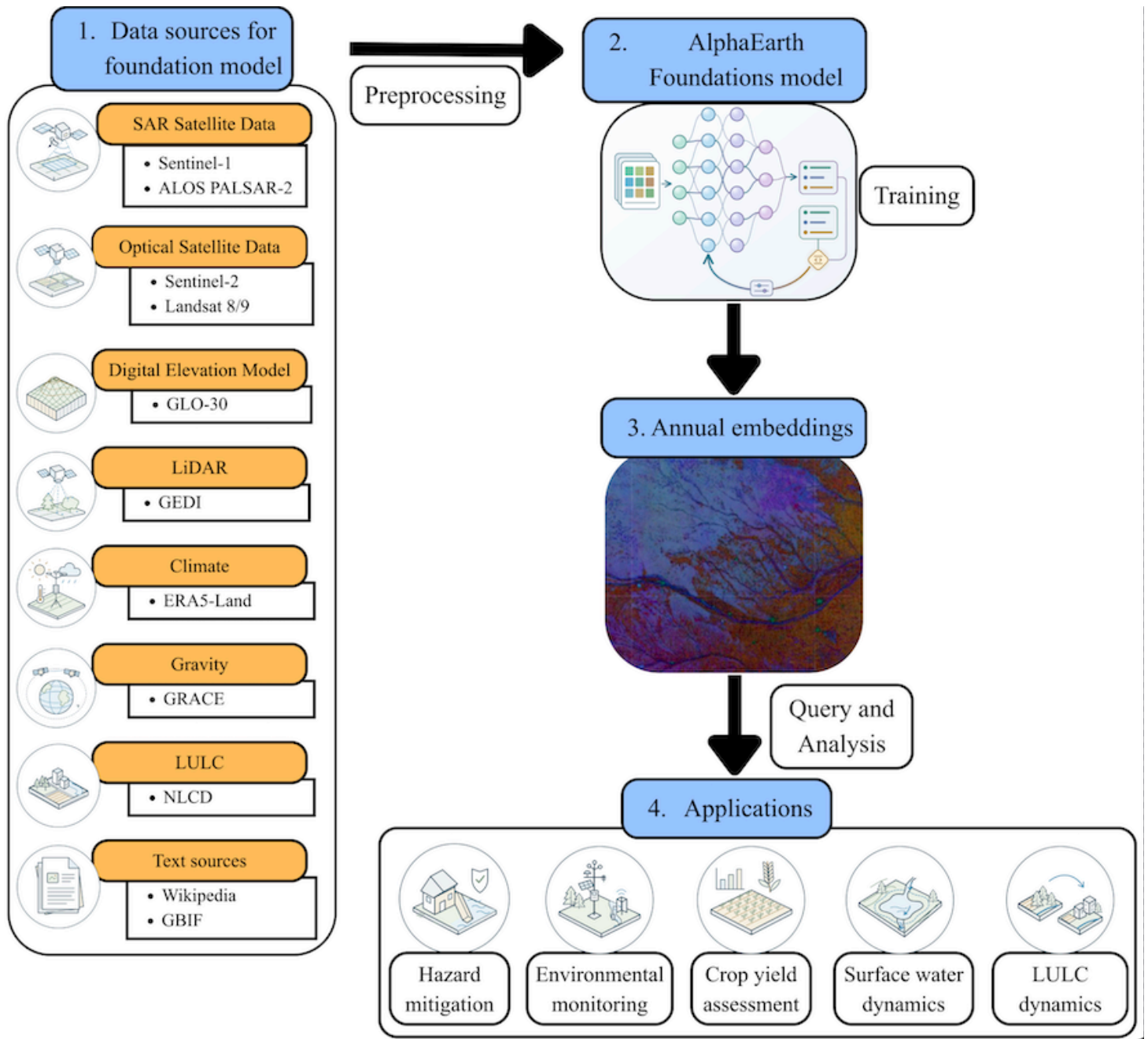


Figure 2

Steps involved in the development of the AFE data and its applications

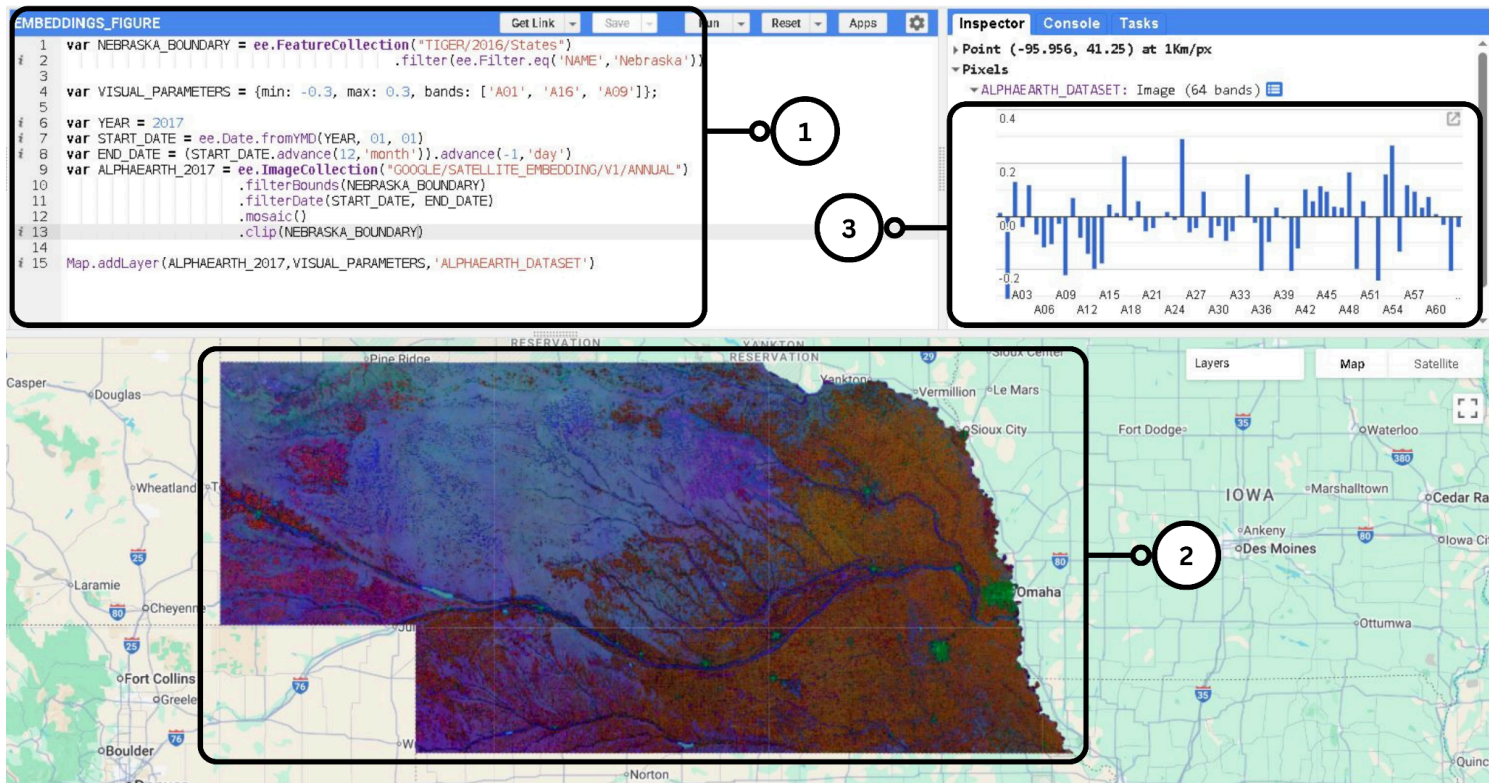


Figure 3

Different elements of the Google Earth Engine worksheet. The code editor window is labeled 1, the embeddings dataset for the year 2017 for the state of Nebraska is labeled 2, and label 3 shows the 64-dimensional signature values obtained by inspecting a pixel from the embeddings dataset.

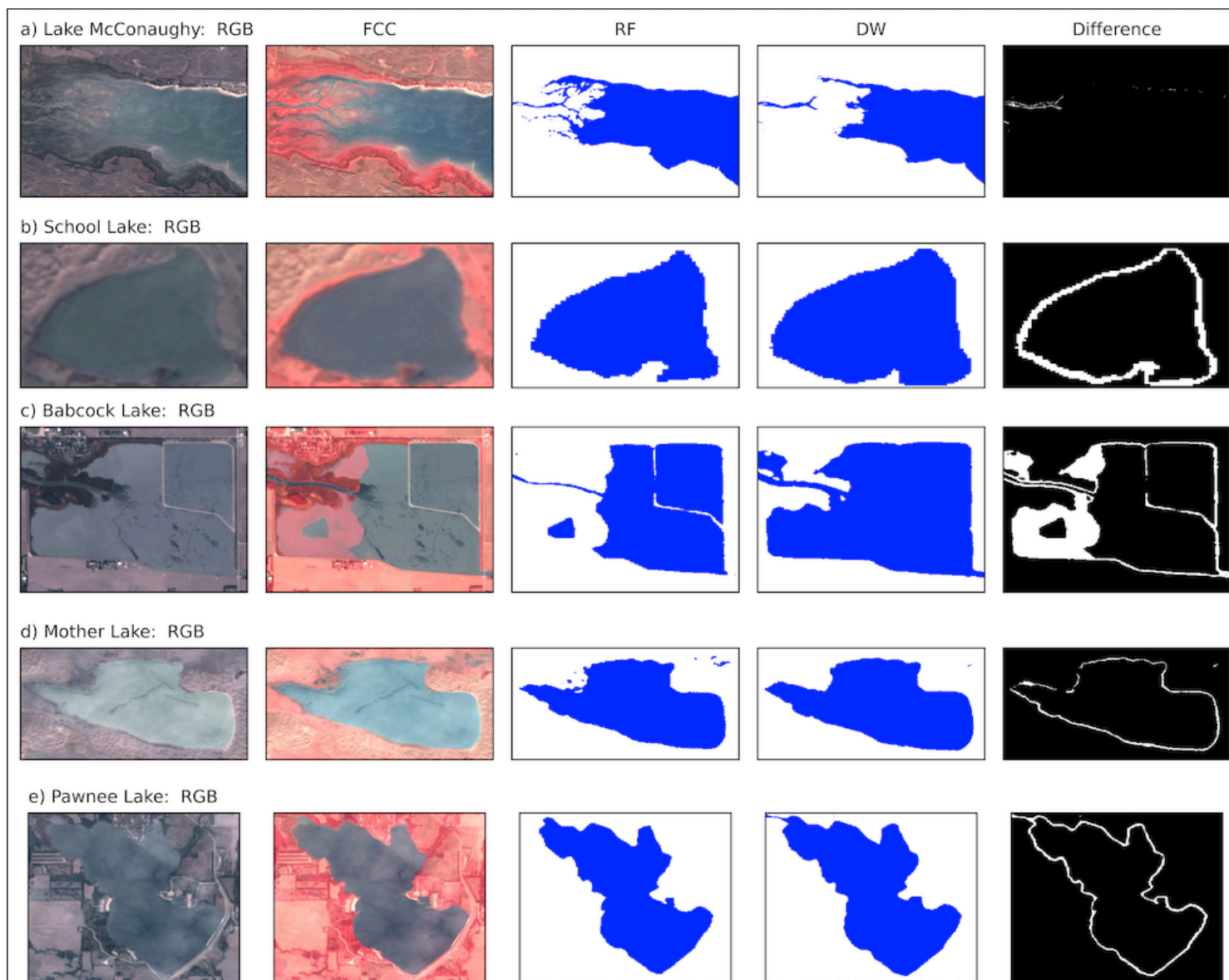


Figure 4

Comparison of classified inundation resulting from random forest (column 3) and Dynamic World (column 4) with RGB data (Column 1) and false color composite (column 2). Pixels classified as inundated by Dynamic World but not by random forest are shown in white in column 5. All maps are for 2017. The first row shows only the inflow end of Lake McConaughy.

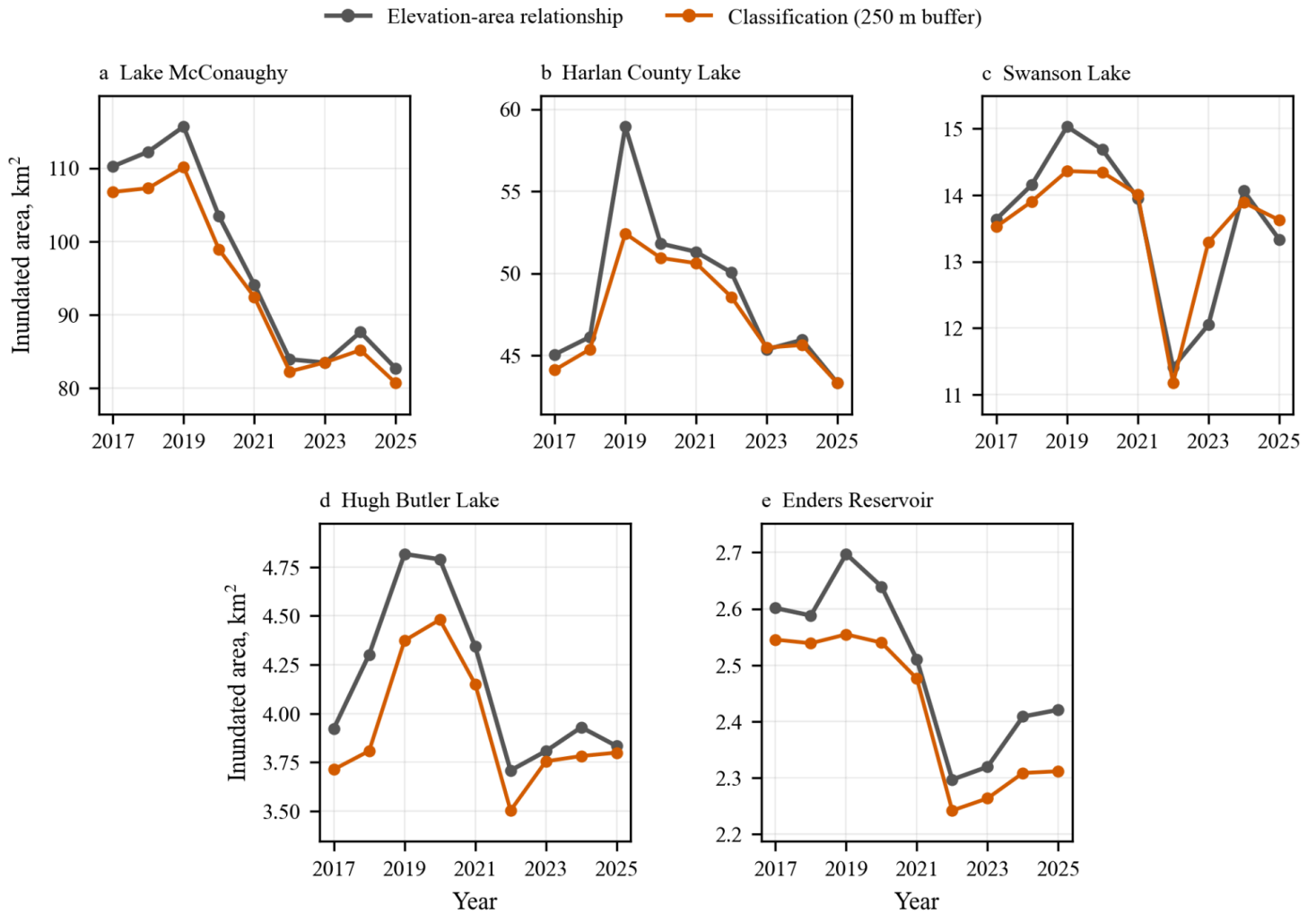
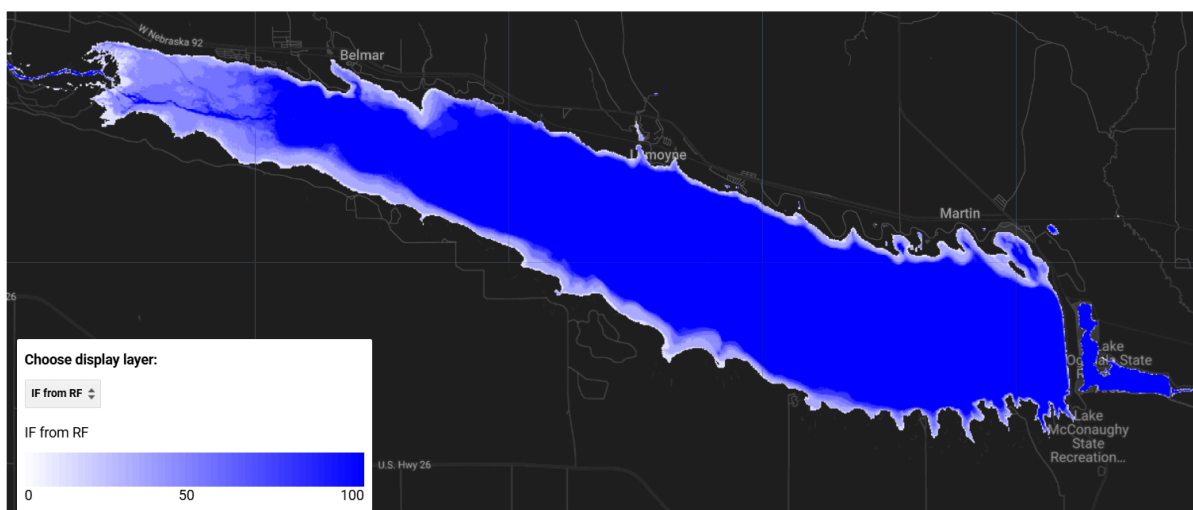


Figure 5

Comparison of the annual inundated areas obtained from the elevation-area relationship and random forest based supervised classification for the five gauged reservoirs, with the classification measured within the 250 m buffer.

a



b

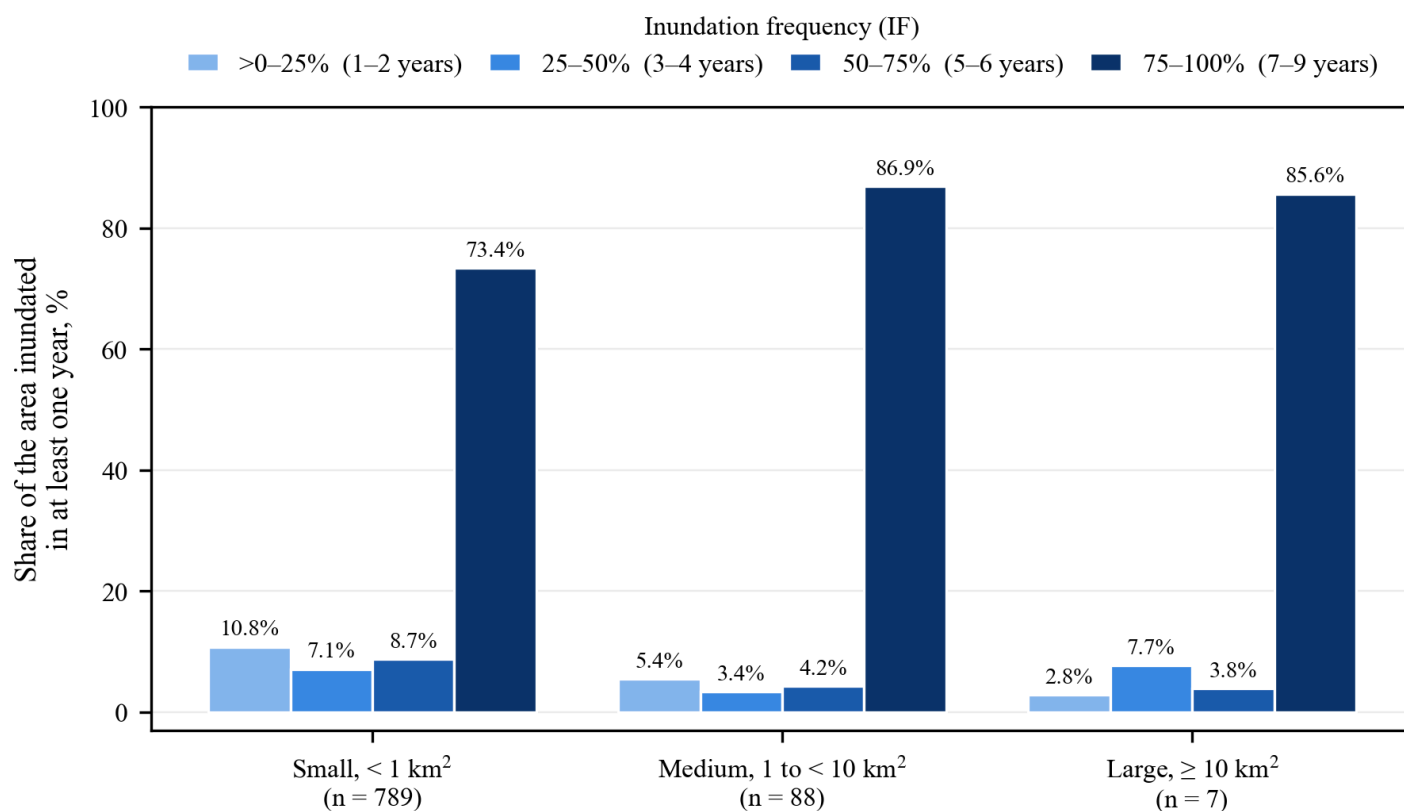


Figure 6

Figure 6a. Random forest derived IF distribution of Lake McConaughy. This figure has been extracted from the GEE-based web app (<https://adityain2003.users.earthengine.app/view/inundation-frequency-webapp>).

Figure 6b. Share of each IF category in the area inundated in at least one year, by size group, within the 250 m buffers. Area is pooled across the water bodies in each group, and n is the number of water

bodies shown.

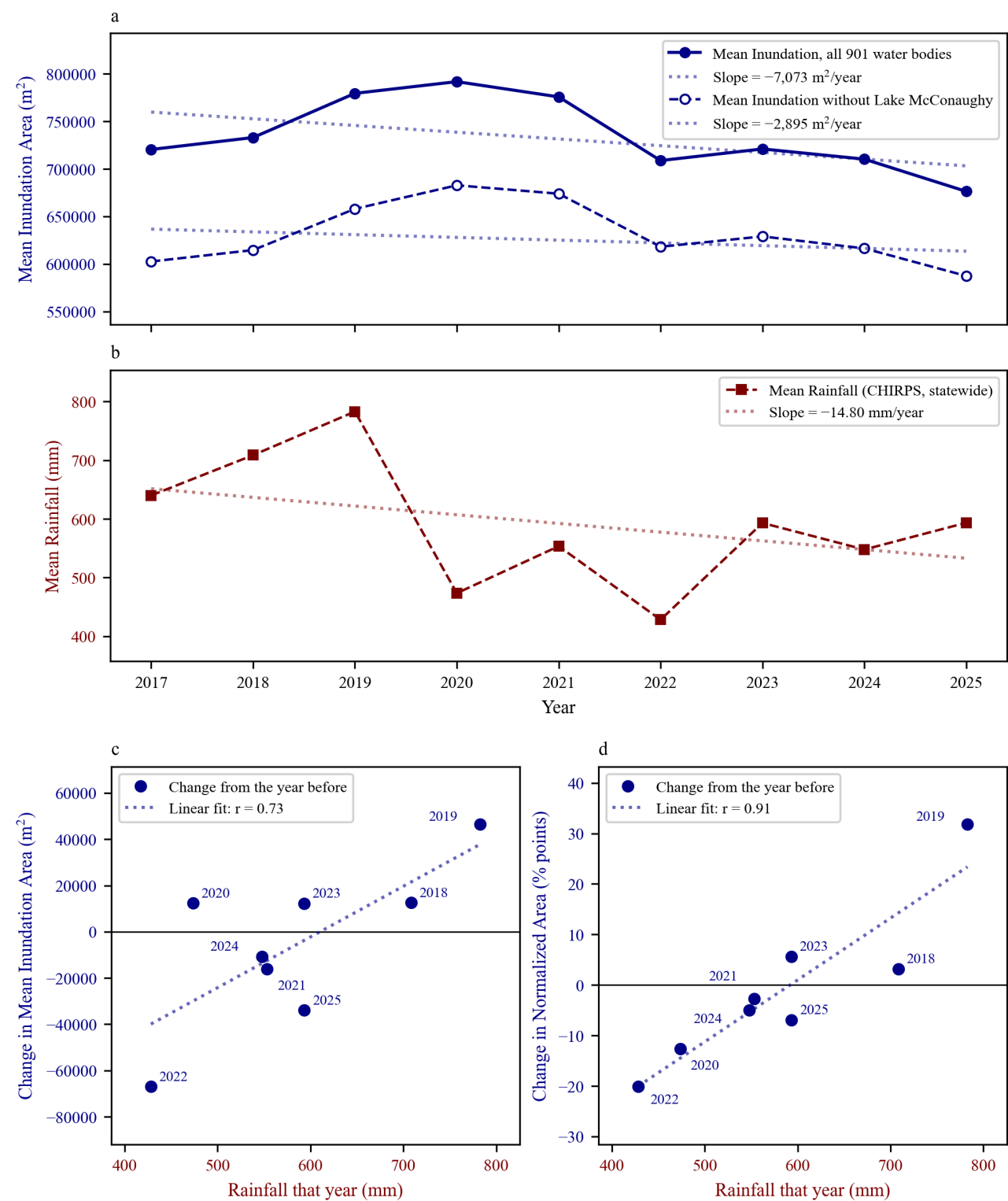


Figure 7

Comparison of annual mean inundation area (m^2) of 901 water bodies, measured within the 250 m buffers, and mean rainfall (mm) across Nebraska (2017-2025). Trend magnitudes (a, b) were calculated using the non-parametric Sen's slope estimator, with estimated rates of change of $-7,073 m^2/year$ for

inundation area (-2,895 m²/year without Lake McConaughy) and -14.80 mm/year for rainfall. The change from the previous year is plotted against that year's rainfall for the mean inundation area (c) and for the normalized area, in which each water body counts equally (d). The legends give the Pearson correlation coefficient (r).

Supplementary Files

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

- [ESM1.pdf](#)