The covariates in this dataset are derived from authoritative national datasets and remote sensing products, aggregated to hydrologically and ecologically meaningful spatial units. Climate variables are based on PRISM 1991–2020 climatological normals, while vegetation indices such as NDVI come from MODIS 2016 imagery. Land cover metrics are sourced from the National Land Cover Database (NLCD) and MODIS land classification. Topographic variables were derived from the 10-meter National Elevation Dataset (NED), and soil properties from SSURGO and STATSGO2 soil surveys. Watershed characteristics utilize stream and catchment features from NHDPlus. Spatial aggregation was performed using EPA Level II and III Ecoregions and NHDPlus catchments to ensure ecological and hydrologic relevance for regional flood frequency analysis.

**Design Structure and Spatial Stratification.**  
The covariate framework is hierarchically organized across five spatial scales, ranging from **station-level** (scale 0) to **local**, **subregional**, **regional**, and **macroregional** levels (scales 1–4). A total of **63 covariates** span four major domains: **Climate**, **Land Cover**, **Topography**, and **Watershed Metrics**. The design balances spatial resolution with thematic depth, placing heavier emphasis on regional and local metrics—particularly within the Climate and Topography domains—to capture scale-dependent drivers of flood skew. Macroregional variables provide broader context (e.g., climate zones, land cover fractions), while fine-scale local metrics (e.g., slope, flow accumulation) support mechanistic modeling of runoff generation and hydrograph shape. This structure enables multiscale modeling that integrates climatic gradients, landscape patterns, and geomorphic controls on hydrologic response.

**Station-Level Geospatial and Watershed Metadata.**  
These point-based covariates describe intrinsic site characteristics used as fixed inputs in hydrologic modeling. Longitude and Latitude capture spatial position and are commonly used to model climatic or physiographic gradients across the study domain. Station Altitude indicates elevation above sea level, influencing temperature regime, snow persistence, and runoff timing. Watershed Area defines the total upstream contributing area and is a key spatial predictor in flood frequency analysis and hydrologic scaling.

**Macroscale Environmental Covariates.**  
These covariates provide broad contextual information at the macroregional scale, capturing dominant climate regimes, land cover patterns, and terrain characteristics relevant for regional hydrologic analysis. Climate Zone and PHZM Zone (Dominant) summarize prevailing climatic and cold-hardiness conditions, while PHZM Zone Count reflects transitional climate diversity across zones. Land cover indicators—Cropland, Forest, Grassland, and Urban Fractions—characterize human and vegetative land use, informing runoff generation, evapotranspiration, and infiltration processes. Terrain attributes including Mean Slope, Median Slope, and Altitude Zone describe elevational and slope regimes that influence snow persistence, runoff velocity, and hydrologic timing.

**Regional-Scale Environmental Covariates (Level II Ecoregions).**  
These covariates summarize climate, landscape, and hydrologic characteristics across Level II ecoregions, offering regionally coherent inputs for spatial modeling. Annual Temp and Annual Precip represent baseline thermal and moisture regimes, while warm-season precipitation fractions (Pct May–Aug) capture convective rainfall seasonality. Vegetation dynamics are described by NDVI Amplitude, IQR, Peak NDVI, and Growing Season Length, which together reflect ecosystem productivity and climatic responsiveness. Terrain is characterized by Mean Slope Distribution, Slope Skewness, and measures of Slope Variability, providing insight into hydrologic energy and flow pathways. Surface soil texture (Clay, Silt, and Sand Fractions) informs infiltration and runoff potential. Finally, stream complexity is represented by Median and Max Stream Order, quantifying typical and maximum network scales across watersheds.

**Subregional Environmental Covariates (Level III Ecoregions).**  
These variables capture finer-grained variation within Level III ecoregions, emphasizing seasonal climate behavior, land use heterogeneity, and hydrologic responsiveness. Seasonal precipitation metrics (Fall, Winter, Spring, Summer Precip, Seasonal StDev, IQR) reflect intra-annual moisture variability, informing recharge and flood timing. Land cover indicators include MODIS Land Cover % and Diversity Index, which quantify dominant vegetation classes and fragmentation, relevant for infiltration and evapotranspiration processes. Soil infiltration dynamics are further described by Soil Permeability and Runoff Class, while terrain–moisture interactions are represented using the Topographic Wetness Index (TWI) via Mean, Modal, and Class values. Finally, Flow Accumulation indicates upslope contributing area, offering insight into drainage concentration and runoff potential.

**Local Hydrologic and Terrain Covariates (Catchment Scale).**  
These high-resolution variables are derived at the catchment scale (~1–5 sq-km), capturing detailed physiographic and hydrologic characteristics that influence site-level flood response. Freeze–Thaw Days, Precipitation Intensity, and Wet Day Frequency reflect fine-scale climatic variability and storm behavior. NLCD Land Cover % and Diversity Index characterize dominant land use and fragmentation, affecting infiltration, surface runoff, and evapotranspiration. Detailed terrain metrics—including Elevation Range, Aspect (Cos/Sin), Curvature (Planform, Profile, IQR), and Relief Ratio—describe slope orientation, concavity, and drainage steepness. Watershed geometry metrics like Elongation Ratio and Circularity quantify basin shape and potential hydrograph response. Finally, drainage network structure is detailed via Stream Density, Flow Length, and Stream Slope, all of which govern runoff velocity, timing, and erosive energy.

**Methods Notes**

Polygons smaller than 1,000 sq-km or containing fewer than 30 gages were merged with the most ecologically similar adjacent Level III units, prioritizing units with matching Level II classification. Where multiple candidates existed, similarity was evaluated using Euclidean distance in a multivariate space defined by land cover composition, terrain metrics (mean/SD of slope), and vegetation seasonality (NDVI amplitude and peak).

The year **2016** is commonly used in environmental modeling because it offers a reliable balance of data quality, completeness, and representativeness. It falls near the midpoint of standard climatological normals (e.g., 1991–2020), making it ideal for capturing typical conditions without the edge effects of more recent or early years. Datasets from 2016—such as MODIS NDVI, land cover, and PRISM climate products—are widely available, preprocessed, and free of major anomalies or interruptions. Unlike years marked by extreme events (e.g., droughts or pandemics), 2016 provides a stable reference year that supports reproducibility and consistency across modeling efforts. As a result, it is frequently adopted by agencies like NASA, USGS, and NOAA as a benchmark year in geospatial and ecological analysis.

**Attribution of Prairie Macrozone Delineation:**  
The delineation of Mixed-Grass and Shortgrass Prairie macrozones in this study is based on the aggregation of EPA Level III and IV Ecoregions (U.S. EPA, 2013), informed by ecological, climatic, and physiographic criteria. Ecoregion boundaries were refined using expert knowledge of prairie vegetation transitions, precipitation gradients, soil characteristics, and land use patterns. Where necessary, Level IV subdivisions were used to distinguish ecological transitions within broader Level III units. The classification emphasizes hydrologically relevant distinctions in vegetation structure, infiltration dynamics, and climate variability, consistent with literature on prairie ecosystem function and landscape hydrology. This approach allows for spatial generalization while preserving ecologically meaningful variability across the Great Plains region.

### Tallgrass Prairie Macrozone Delineation

The **Tallgrass Prairie macrozone** was delineated to represent the mesic end of the prairie continuum, where high annual precipitation, fertile soils, and dense herbaceous vegetation support distinct ecohydrological regimes. This macrozone was derived by aggregating relevant Level II and Level III EPA ecoregions, focusing on the **Temperate Prairie** and **Texas–Louisiana Coastal Plain** regions. Further refinement was made using Level IV ecoregions to ensure consistency with climatic, physiographic, and vegetative characteristics indicative of tallgrass prairie systems.

The delineation prioritized areas with:

* Mean annual precipitation generally exceeding 850 mm,
* Dominant land cover of tallgrass vegetation (e.g., Andropogon gerardii, Panicum virgatum, Sorghastrum nutans),
* Deep soils with high water retention,
* Hydrologic behavior marked by high infiltration capacity and extended baseflow support.

Included regions:

* **Flint Hills (Level III: 28)** – A preserved tallgrass core with steep, rocky terrain and shallow limestone soils. This region retains native vegetation due to its resistance to plowing and remains under frequent fire management.
* **Eastern Iowa and Minnesota Drift Plains (24a)** – Glaciated terrain with productive soils and tallgrass cover in the wetter northern subregions.
* **Des Moines Lobe (24e)** – Included selectively where precipitation and historical land cover meet tallgrass criteria.
* **Rolling Sand Plains (25a, eastern portions)** – Incorporated where soil depth and precipitation gradients support tallgrass structure along the macrozone’s eastern fringe.

This delineation captures the unique hydrologic function of tallgrass systems, characterized by strong vegetation–soil feedbacks, enhanced infiltration, and moderated runoff. It forms one of three prairie macrozones used for stratification in regional skew estimation modeling.

### Mixed-Grass Prairie Macrozone Delineation

The **Mixed-Grass Prairie macrozone** was delineated to represent transitional prairie systems with moderate precipitation, variable soil texture, and a blend of tall and short grass species. These areas are characterized by intermediate hydroclimatic conditions that produce a dynamic balance between infiltration and runoff processes, with high sensitivity to interannual variability, land use, and disturbance regimes.

This macrozone was constructed by aggregating portions of several Level III ecoregions—particularly:

* **Central Irregular Plains (24)** – Includes wetter subregions such as the Eastern Iowa and Minnesota Drift Plains (24a), Central Kansas Plains (24b), and Northern Rolling Plains (24c), where precipitation (500–800 mm annually) and soil conditions support tallgrass–midgrass mixtures.
* **Western High Plains (25)** – Eastern margins, including Rolling Sand Plains (25b) and Northern Rolling High Plains (25d), where deeper soils and slightly higher precipitation support species like little bluestem (Schizachyrium scoparium) and sideoats grama (Bouteloua curtipendula).
* **Southwestern Tablelands (26)** – Upland areas such as the Canadian/Cimarron High Plains Breaks (26a) and Pecos-Capulin Uplands (26d), which exhibit a blend of short- and midgrass species due to localized elevation and topographic complexity.
* **Central Great Plains (27)** – Core mixed-grass zone with interspersed tall- and shortgrass species responding dynamically to precipitation variability and disturbance.
* **Flint Hills (28)** – Included conceptually for its ecological adjacency and transitional role, though primarily managed as a tallgrass remnant.

This aggregation captures regions with moderate infiltration capacity, spatial heterogeneity in vegetation structure, and variable flood response regimes, making it an ecologically and hydrologically coherent unit for regional modeling.

### Shortgrass Prairie Macrozone Delineation

The **Shortgrass Prairie macrozone** represents the driest portion of the prairie continuum, characterized by low annual precipitation, sparse vegetative cover, and limited infiltration capacity. Hydrologically, these areas are dominated by rapid runoff generation and higher flood skew, driven by reduced canopy cover, shallow soils, and arid-adapted plant communities.

This macrozone was delineated by aggregating Level III and IV ecoregions where shortgrass vegetation and arid hydroclimatic conditions dominate:

* **Western High Plains (25)** – Especially the Flat to Rolling Plains (25e) and Shallow Loess Plains (25f), where dominant species include blue grama (Bouteloua gracilis) and buffalograss (Buchloe dactyloides).
* **Southwestern Tablelands (26)** – Flat mesas and dissected plains with shrub–shortgrass communities, low precipitation, and highly erosive soils.
* **Central Great Plains (27)** – Western segments exhibiting lower precipitation and short-statured grasses.
* **Western Loess Hills (24d)** – Selectively included due to more xeric conditions and vegetation consistent with shortgrass prairie.

The delineation emphasizes regions with minimal canopy structure, reduced evapotranspiration buffering, and strong surface runoff response, aligning with known flood frequency behavior in semiarid grassland systems.