Flood Impact Analysis Methodology (Revised)

(This document was created in collaboration with Google's Gemini (2.5 Pro) generative AI.)

This document explains the methodology used to analyze the impact of flooding on road networks.
The approach is designed to support disaster planning by identifying critical road segments that could become isolated or disrupted during flood events.
The following sections provide a high-level summary followed by an intermediate-level walkthrough of the algorithm.

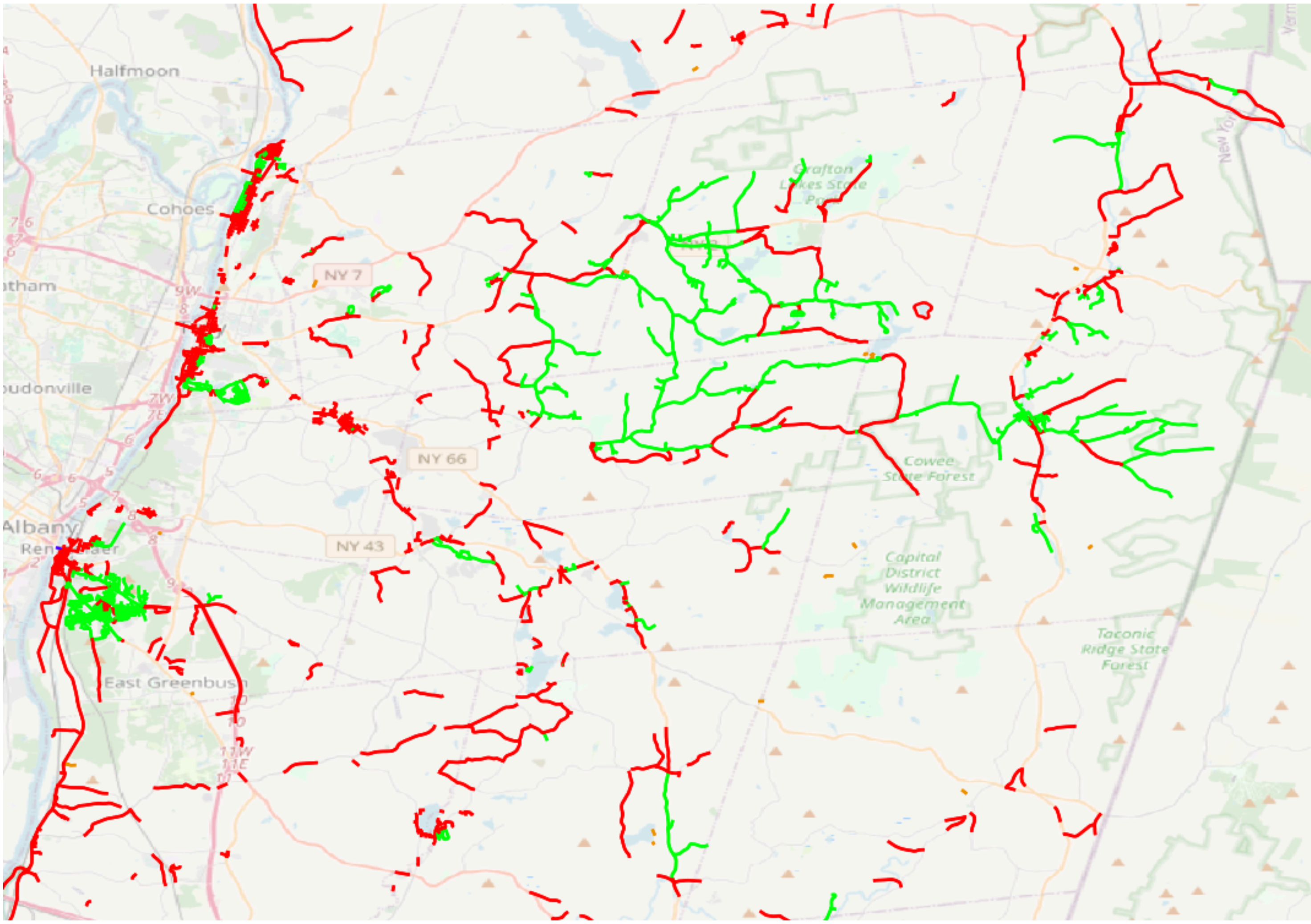
High-Level Summary

This methodology assesses the impact of flooding on road networks to inform disaster planning.
By identifying roads vulnerable to flooding and evaluating network connectivity, emergency planners can anticipate where roads and bridges are at risk of becoming nonfunctional, potentially isolating communities.
The analysis integrates flood risk data with road network topology (how roads connect) to determine which roads may become impassable and the network-level consequences of their loss.

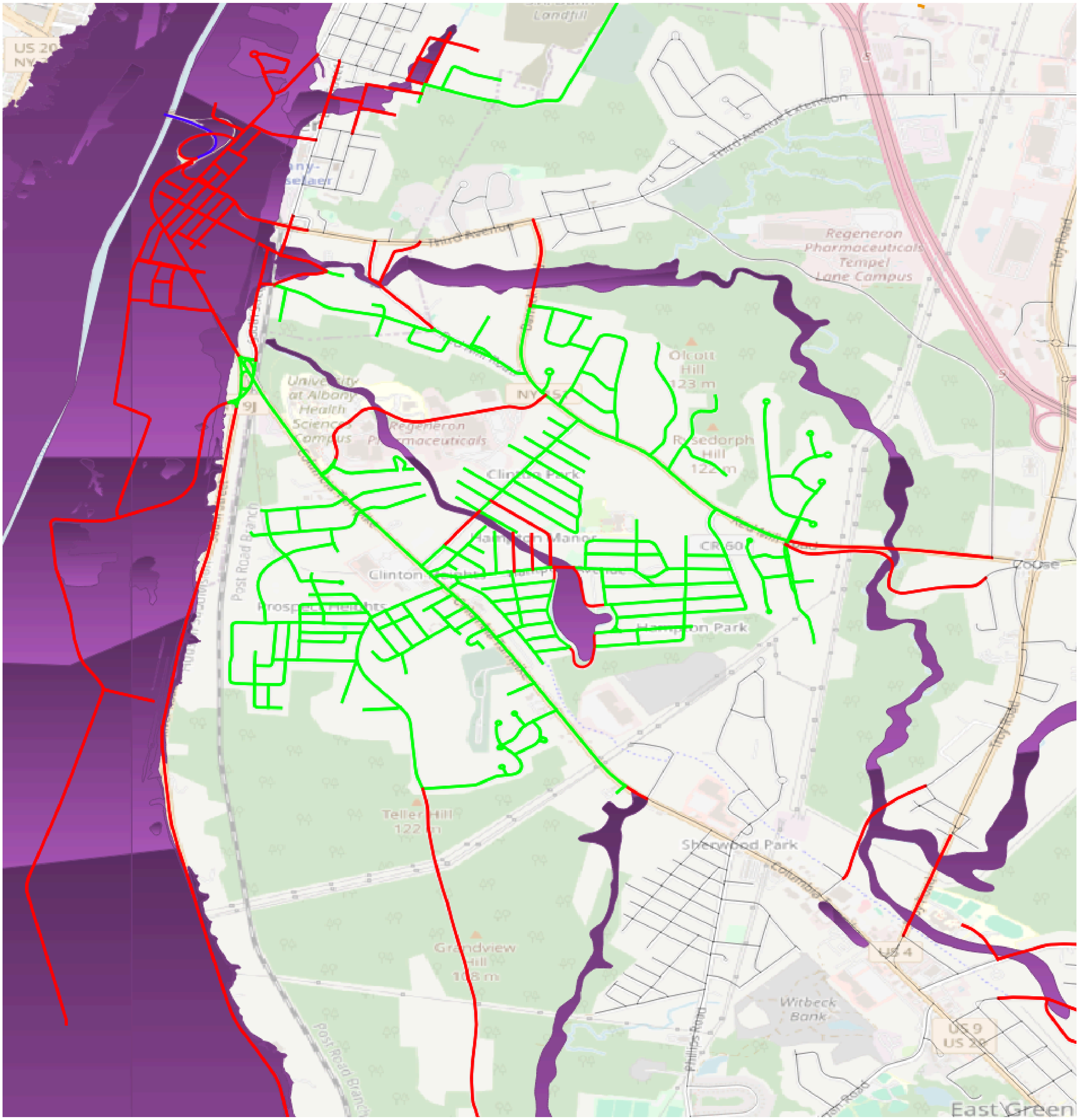
Example

Symbology

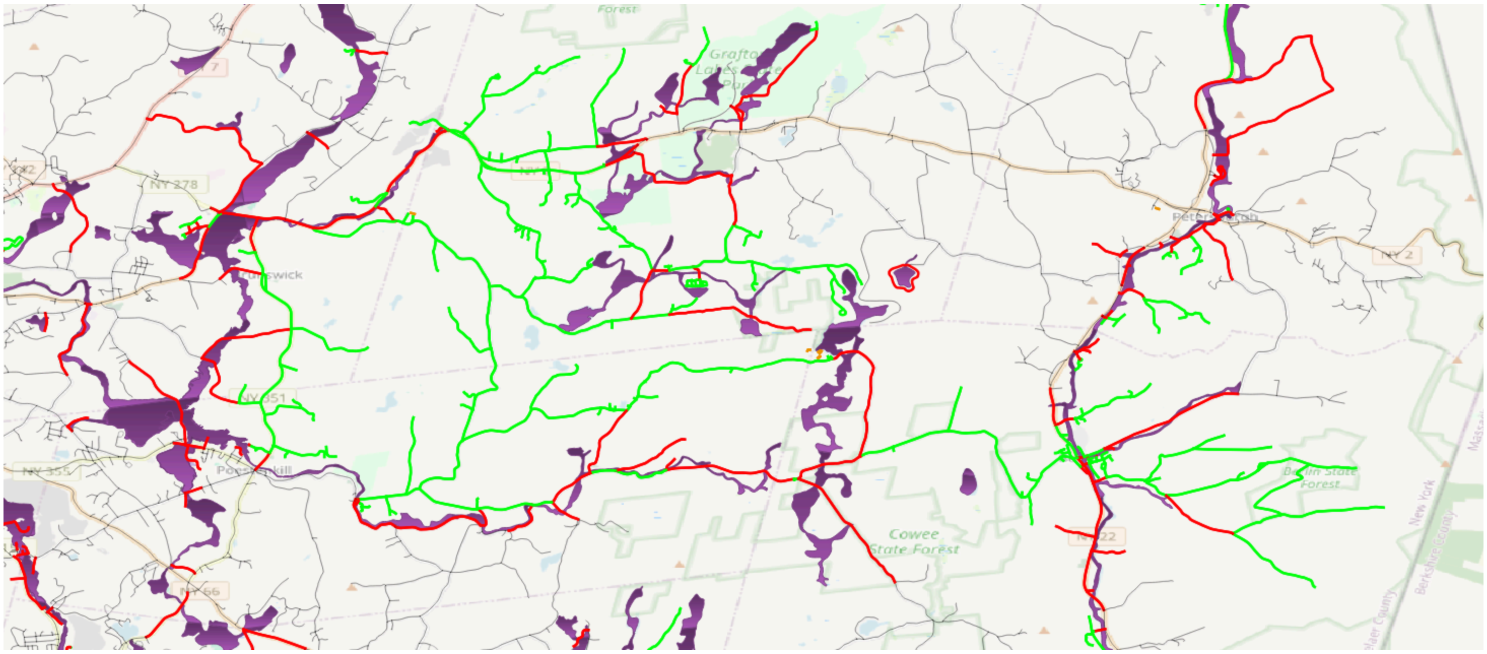
- **Red:** Roads predicted to be directly flooded.
- **Green:** Roads that are *not* flooded themselves but become cut off from the main road network due to flooding elsewhere (isolated).
- **Thin Black Lines:** Roads unaffected by flooding and still connected.
- **Purple Areas:** Flood hazard zones (e.g., 100-year or 500-year floodplains).



High-level view of the impact of 500-year flooding on Rensselaer County, New York.



The impact of 500-year flooding on the Clinton Heights neighborhood in Rensselaer County, New York. Notice how a large portion of the community becomes isolated because flooding along its perimeter cuts off access.



The impact of 500-year flooding on a large rural region in Rensselaer County, New York. Notice how a small number of flooded roads can isolate such a large area.

Methodology Overview

Step 1: Data Preparation

- Gather road network data (e.g., from OpenStreetMap) and floodplain maps (e.g., FEMA flood zones).
- Preprocess data to ensure accurate spatial alignment and prepare it for analysis.

Step 2: Bridge Span Separation

- Before checking for intersections between roads and floodplains, the analysis distinguishes between regular road segments on the ground and elevated bridge spans.
- The assumption is that only road segments on the ground are directly susceptible to inundation. Bridges that completely span a floodplain are generally considered traversable unless their access points become flooded.

Step 3: Identifying Flooded Road Segments

- Overlay the ground-level road segments with floodplain data.
- Identify road segments that physically intersect flood zones.
- Classify these potentially flooded segments based on the severity of the flood zone they intersect (e.g., 100-year vs. 500-year risk).

Step 4: Evaluating Network Connectivity

- Represent the road network as a connected graph (like points connected by lines).
- Simulate the removal of road segments that become non-functional due to flooding:
  - Segments directly identified as flooded (Step 3).
  - Bridge segments whose entry/exit points connect only to flooded roads.
- Analyze the remaining network graph to identify which areas or road segments become disconnected or isolated from the main, functional part of the network.

Step 5: Risk Classification and Reporting

- Categorize all road segments based on how they are impacted:
  - Directly flooded (and at which flood level, e.g., 100-year).
  - Isolated (and by which flood level).
  - Unaffected.
- Generate maps and data outputs highlighting vulnerable areas and the reasons for vulnerability.
- Provide data to support planning for network resilience and emergency response.

This methodology provides critical insights for emergency planners, helping them anticipate and mitigate the impacts of flooding on transportation networks.
By identifying potential isolation risks, decision-makers can enhance emergency response strategies and infrastructure resilience.

Intermediate-Level Walkthrough of the Algorithm

The methodology is implemented through a sequence of computational steps that together assess and label road segments based on their flood risk and network connectivity.
Here's a breakdown of the process:

1. Data Initialization and Setup

- **Graph Representation:**
  - The road network is loaded and modeled using a `networkx.MultiDiGraph`.
    In this graph, intersections or road endpoints are "nodes," and the road segments connecting them are "edges."
- **Input Data:**
  - The process starts with the road network graph and a GeoDataFrame containing floodplain polygons.

The floodplain data must include a pre-defined column (`_flood_risk_level_`) indicating the severity of the flood risk for each polygon.

- **Bridge/Non-Bridge Separation:**
  - The road segments (edges) from the graph are converted into a GeoDataFrame.
    This is then processed to create two separate GeoDataFrames: one containing only bridge spans and another containing only non-bridge (ground-level) road segments.
    This separation is crucial because only non-bridge segments are checked for direct inundation by floodplains.
- **Initial Data Packaging:**
  - The initial graph, the three GeoDataFrames (all roads, bridge spans, non-bridge spans), and the input floodplains are bundled together into an `InitialImpactData` object.
    This object serves as the verified starting point for the analysis pipeline.

2. Integrity Hash Calculation and Verification

- **Purpose:**
  - In complex data pipelines like this, it's crucial to ensure that the *exact same* data structures (with consistent indexing and graph structure) are used between steps.
    Subtle changes or errors during processing could lead to incorrect results.
    The integrity hashing mechanism acts as a safeguard against such issues.
- **Mechanism:**
  - **Fingerprinting:** When key data objects (the graph, the road GeoDataFrames) are first created in the initialization step, a unique numerical "fingerprint" (a hash value) is calculated for each one.
    For DataFrames, this hash is based on their index; for the graph, it's based on its structure (nodes and edges).
  - **Tracking:** These initial hashes are stored in a dictionary (`integrity_hashes`) within the `InitialImpactData` object.
    Each hash is associated with a specific key (e.g., `IntegrityHashKey.ROADS`).
  - **Verification:** Before subsequent functions use a critical input object (like `roads_gdf`), they perform a check (if the `integrity_hashes` dictionary is provided).
    They recalculate the hash of the object they received and compare it to the expected hash value stored in the `integrity_hashes` dictionary.
  - **Error Prevention:** If the recalculated hash doesn't match the expected hash, it signifies that the data object has been unexpectedly altered.
    An error (`AssertionError`) is raised, stopping the workflow before incorrect data can lead to flawed results.
- **Memoization:**

    - To avoid repeatedly calculating hashes for the same object instance, the results of `calculate_integrity_hash` are cached using a `weakref.WeakKeyDictionary`.
      This cache uses a unique identifier (UUID) associated with each object as its key.

3. Spatial Analysis: Road Segments vs. Floodplains

- **Spatial Join:**
    - A spatial join operation identifies intersections between the **non-bridge** road segments and the floodplain polygons.
      This step determines which ground-level roads lie within flood hazard areas.
      It's optimized using spatial indexing techniques (`pyogrio` engine, bounding box filtering) for better performance.
- **Flood Risk Assignment:**
    - For each intersection found, the corresponding non-bridge road segment is associated with the `_flood_risk_level_` from the specific floodplain polygon it intersects.
      This results in a new GeoDataFrame (`nonbridge_spans_floodplains_join_gdf`) linking non-bridge segments to flood risk levels.

4. Network Impact Assessment

- **Highest Risk Calculation:**
    - Since a single road segment might intersect multiple floodplains with different risk levels, the analysis determines the *highest* risk (lowest numerical value) associated with each non-bridge segment based on the results of the spatial join.
      This produces a pandas Series mapping each affected non-bridge segment to its single worst-case flood risk level.
- **Connected Components Analysis:**
    - The original road network graph is analyzed (treating roads as two-way for connectivity purposes) to identify distinct "components" or sections of the network that are connected internally but may or may not be connected to each other initially.
      This helps identify islands or naturally separate network parts even before flooding.
- **Non-Functional Edge Identification:**
    - The analysis iterates through the flood risk levels (from highest risk to lowest).
      In each iteration, it identifies road segments (edges in the graph) that become non-functional:
        - **FLOODED:** Non-bridge segments identified as intersecting a floodplain at the current risk level or higher.
        - **BRIDGE_FUNCTIONALLY_DISCONNECTED:** Bridge segments that become unusable because *all* paths onto or off them connect only to roads that are already marked as non-

functional (either flooded or previously disconnected bridges).
This check is iterative, as disconnecting one bridge might lead to another becoming disconnected.

- **Isolation Analysis:**
  - As segments are marked non-functional, the connectivity of the remaining network graph is re-evaluated.
  - Any segment (bridge or non-bridge) that is still functional but becomes part of a component that is newly separated from the main, largest functional component *at the current risk level* is marked as **isolated**.
- **Risk Level Propagation:**
  - Isolated road segments are assigned the flood risk level that *caused* their isolation (i.e., the risk level being processed when they became cut off).

5. Output Classification

- **Flood Frequency Classification:**
  - Based on the assigned `nonfunctional_risk_level` or `isolated_edge_risk_level` (which correspond to FEMA flood zone categories), segments are also classified by approximate flood frequency:
    - **100 YEAR:** Segments impacted by higher-risk flood zones (typically zones A, AE, V, VE).
    - **500 YEAR:** Segments impacted by lower-risk flood zones (typically zone X, 0.2 PCT).
- **Data Aggregation:**
  - All the calculated impact information is mapped back to the original `roads_gdf`.
    New columns are added to this GeoDataFrame.

6. Output and Integration

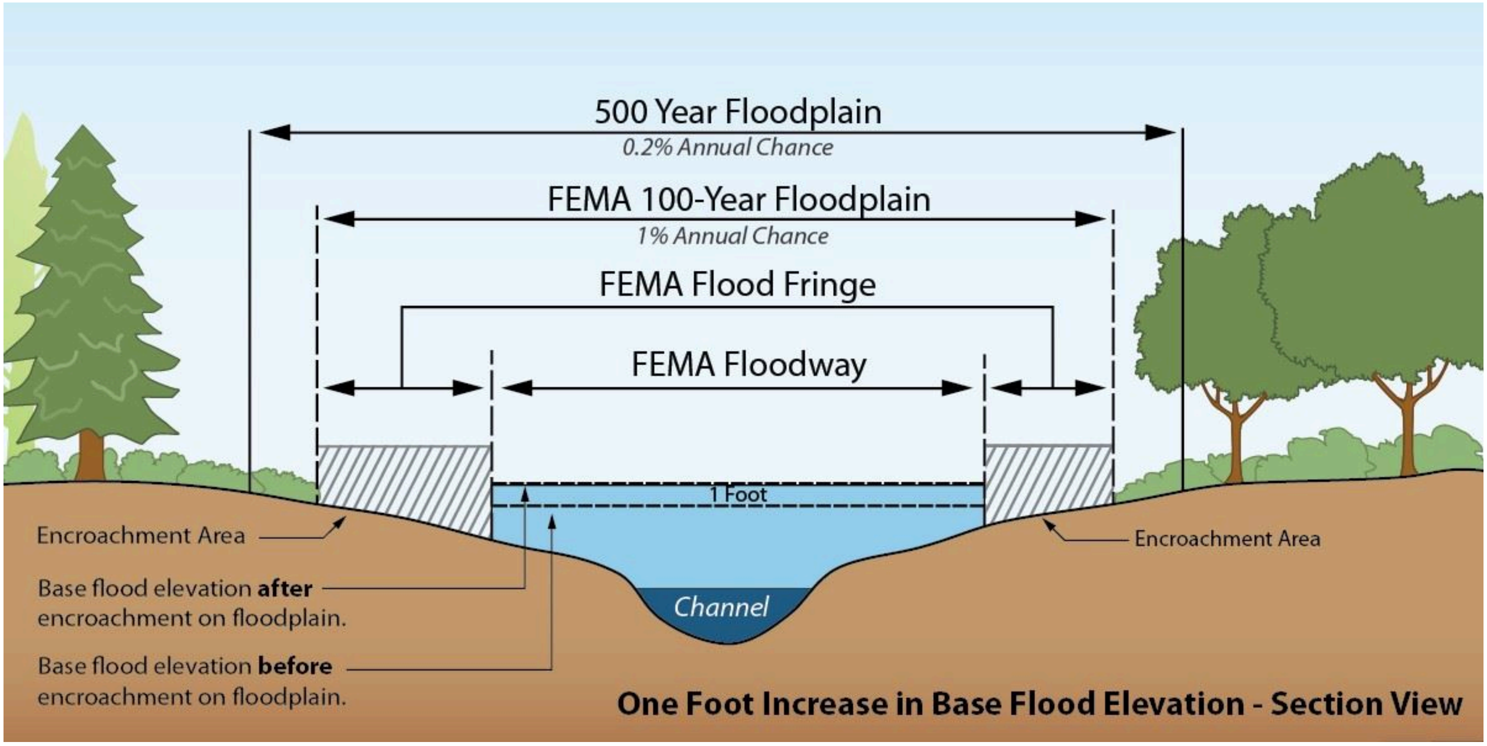
- **Comprehensive Output:**
  - The final primary output is a GeoDataFrame based on the original `roads_gdf`, but with added columns detailing the flood impact for each segment:
    - `nonfunctional_reason`: Why a segment might be unusable (`FLOODED`, `BRIDGE_FUNCTIONALLY_DISCONNECTED`, or `NOT_IMPACTED`).
    - `nonfunctional_risk_level`: The numeric risk level causing the segment to become non-functional.
    - `nonfunctional_frequency`: The flood frequency (100/500 YEAR) associated with becoming non-functional.
    - `isolated_edge_risk_level`: The numeric risk level at which a segment became isolated.

        - `isolated_edge_frequency` : The flood frequency (100/500 YEAR) associated with becoming isolated.
- **Saving:**
    - This final GeoDataFrame is saved to a GeoPackage file.
      The file is then set to read-only to prevent accidental modification.
- **Reporting and Visualization:**
    - This output data can be loaded into GIS software or other tools to generate reports, maps (like the examples shown earlier), and other visualizations to aid in disaster planning, infrastructure assessment, and emergency response routing.

Conclusion

This refactored floodplain network analysis methodology provides a structured and reliable approach to assessing flood impacts.
By separating bridge and non-bridge segments, analyzing network connectivity iteratively, and incorporating data integrity checks, the analysis produces detailed and trustworthy results.
The classified output helps planners understand not only which roads might flood but also which areas might become isolated, enabling more informed decisions to improve community resilience against flooding events.

Flood Zone Risk Hierarchy Explanation



From What is FEMA Floodway and How is it Different?

FEMA Floodplain Types and Risk Ranking

FEMA Flood Zone Designations				
Undetermined Risk	**Low Risk**	**Moderate Risk**	**High Risk**	**Coastal High Risk**
Increasing Risk →				
Zone D	Zones C and X (unshaded)	Zones B and X (shaded)	Zones A, AE, A1-30, AH, AO, A99	Zones V, VE, V1-30
	Non-Special Flood Hazard Area (NSFHA)		Special Flood Hazard Area (SFHA)	

From What are the Flood Zones in FEMA Maps - A-X

The Federal Emergency Management Agency (FEMA) designates various floodplains based on the likelihood and severity of flooding. These floodplains are categorized into different zones, each representing a distinct level of flood risk. In this analysis, we considered the following FEMA floodplain types:

- **100-Year Floodplain (1% Annual Chance Flood Hazard):** This area has a 1% chance of flooding in any given year. It's also known as the Special Flood Hazard Area (SFHA).
- **500-Year Floodplain (0.2% Annual Chance Flood Hazard):** This area has a 0.2% chance of flooding in any given year. It represents a moderate flood hazard.
- **Floodway:** The channel of a river or other watercourse and the adjacent land areas that must be reserved in order to discharge the base flood without cumulatively increasing the water surface elevation more than a designated height. This is the most dangerous area.
- **Riverine Floodplain:** Areas adjacent to rivers and streams that are subject to flooding.

Flood Zone Risk Hierarchy:

The risk ranking is based on the severity and frequency of flooding, considering factors such as:

- Coastal vs. riverine flooding: Coastal high-hazard areas (V and VE zones) are ranked highest due to the destructive force of wave action and storm surge.
- Floodways: Floodways (AE with FLOODWAY or FW sub-types) are ranked high due to the high velocity of floodwaters.
- Base Flood Elevation (BFE): Zones with BFEs (AE, VE) are ranked higher than those without (A). VE zones are ranked higher than V zones.
- Shallow flooding: AO and AH zones indicate shallow flooding and are ranked accordingly.
- 0.2% annual chance flood hazard (X zone): These areas are ranked lower than 1% annual chance flood hazard areas (A, AE, VE, etc.).
- Levee protection: Areas with reduced flood risk due to levees (X zone) are ranked lower.
- Open water: Open water is ranked the lowest.

The numeric risk levels (`nonfunctional_risk_level` , `isolated_edge_risk_level`) used in this analysis are derived from standard FEMA flood zone designations (`fld_zone` and `zone_subty` in the source data).
A lower rank value indicates a higher relative flood risk.
The ranking logic considers factors like coastal wave action, floodways, base flood elevations (BFEs), and shallow flooding potential:

1. **VE (Coastal High Hazard with BFE) + Riverine Floodway:** Highest risk due to combined wave action and riverine flooding.
2. **VE (Coastal High Hazard with BFE) + Coastal Floodplain:** High risk due to wave action.

3. **VE (General):** High risk due to wave action.
4. **V (Coastal High Hazard, no BFE):** High risk, but less specific than VE.
5. **AE (1% Annual Chance with BFE) + Riverine Floodway in Coastal Zone:** High riverine risk in coastal area.
6. **AE (1% Annual Chance with BFE) + Riverine Floodplain in Coastal Area:** High riverine risk in coastal area.
7. **AE (1% Annual Chance with BFE) + Floodway (FW):** High risk due to high-velocity riverine floodwaters.
8. **AE (1% Annual Chance with BFE) + Coastal Floodplain:** Standard 1% risk in coastal area.
9. **AO (Shallow Flooding, depth specified) + Coastal Floodplain:** Shallow flooding risk in coastal area.
10. **AO (Shallow Flooding, depth specified):** General shallow flooding.
11. **AE (General 1% Annual Chance with BFE):** Standard 1% flood risk.
12. **AH (Shallow Ponding, depth specified):** Shallow ponding risk.
13. **A (1% Annual Chance, no BFE):** Approximate 1% flood risk.
14. **X (0.2% Annual Chance):** Moderate risk (500-year flood).
15. **X (Shallow Flooding < 1 foot):** Minimal shallow flooding risk.
16. **X (Levee Protection):** Area with reduced risk due to levee, but residual risk remains.
17. **X (General):** Minimal flood hazard area.
18. **Open Water:** Lowest rank.

(Note: This hierarchy is based on interpreting FEMA guidelines. Refer to official FEMA documentation for definitive zone descriptions. See Citations and Resources section for links.)

Data Dictionary Explanation

This section provides an overview of columns added to the output GeoDataFrame by the analysis. These columns capture and explain why certain road segments might become unusable or inaccessible due to flood-related impacts.

Output Columns

1. `nonfunctional_reason`

- **Description:**
  - Indicates the reason why a particular road edge might become unusable during a flood event.
  - `NOT_IMPACTED`: The road edge is not directly flooded and remains connected to the main network.

    - **`FLOODED`:** The road edge is predicted to be inundated by floodwaters because it intersects a flood zone and is not a bridge span.
    - **`BRIDGE_FUNCTIONALLY_DISCONNECTED`:** The road edge is part of a bridge, but all access points to/from the bridge are predicted to be flooded, rendering the bridge unusable.
- **Data Type:** String
- **Possible Values:** `NOT_IMPACTED`, `FLOODED`, `BRIDGE_FUNCTIONALLY_DISCONNECTED`

2. `nonfunctional_risk_level`

- **Description:**
    - A numeric value (1-18, based on the hierarchy above) representing the flood risk level associated with the reason the edge became non-functional.
      Lower numbers indicate a higher risk.
      This is only populated if `nonfunctional_reason` is not `NOT_IMPACTED`.
- **Data Type:** Integer (Nullable)
- **Possible Values:** Risk level values (e.g., 1-18) or Null.

3. `nonfunctional_frequency`

- **Description:**
    - A simplified classification of the flood frequency associated with the edge becoming non-functional, derived from the `nonfunctional_risk_level`.
    - **`100 YEAR`:** Typically corresponds to higher-risk zones (risk levels 1-13). Represents areas with a 1% or greater annual chance of flooding.
    - **`500 YEAR`:** Typically corresponds to moderate or lower-risk zones (risk levels 14-17).
- **Data Type:** String (Nullable)
- **Possible Values:** `100 YEAR`, `500 YEAR`, or Null.

4. `isolated_edge_risk_level`

- **Description:**
    - A numeric value (1-18) representing the flood risk level that *caused* this edge (which is not directly flooded itself) to become disconnected from the main functional road network.
      For example, a road might be isolated because all connecting roads around it are flooded by a 100-year event (corresponding to risk levels 1-13).
- **Data Type:** Integer (Nullable)
- **Possible Values:** Risk level values (e.g., 1-18) or Null.

5. `isolated_edge_frequency`

- **Description:**
  - The simplified flood frequency (100/500 YEAR) associated with the event that *caused* this edge to become isolated, derived from `isolated_edge_risk_level`.
- **Data Type:** String (Nullable)
- **Possible Values:** `100 YEAR`, `500 YEAR`, or Null.

Summary

The addition of these columns provides a systematic classification of flood impacts on the road network. By capturing not only direct inundation (`FLOODED`) but also indirect impacts like bridge disconnection (`BRIDGE_FUNCTIONALLY_DISCONNECTED`) and network isolation (indicated by populated `isolated_edge_` columns), this data supports comprehensive analysis for flood resilience and emergency planning.

Citations and Resources

- FEMA Flood Zones: https://www.fema.gov/flood-zones
- FEMA Flood Map Service Center: https://msc.fema.gov/portal/home
- FEMA Guidelines and Standards for Flood Risk Analysis and Mapping: https://www.fema.gov/guidelines-and-standards-flood-risk-analysis-and-mapping
- FEMA P-259, Engineering Principles and Practices for Retrofitting Flood-Prone Residential Structures: https://www.fema.gov/sites/default/files/2020-08/fema259_complete_rev.pdf
- NOAA Coastal Flood Hazard: https://oceanservice.noaa.gov/hazards/natural-hazards/
- Association of State Floodplain Managers (ASFPM): https://www.floods.org/
- FEMA How to Read a Flood Map: https://www.fema.gov/sites/default/files/documents/how-to-read-flood-insurance-rate-map-tutorial.pdf
- FloodSmart What are Flood Zones and Maps?: https://www.floodsmart.gov/flood-zones-and-maps
- FEMA FIRM_Database_Technical_Reference_Feb_2019: https://www.fema.gov/sites/default/files/2020-02/FIRM_Database_Technical_Reference_Feb_2019.pdf