

Predicting and Alleviating Road Flooding for Climate Mitigation

Amrita Gupta

School of Computational Science and
Engineering
Georgia Institute of Technology
agupta375@gatech.edu

Caleb Robinson

School of Computational Science and
Engineering
Georgia Institute of Technology
dcrobin@gatech.edu

Bistra Dilkina

School of Computational Science and
Engineering
Georgia Institute of Technology
bdilkina@cc.gatech.edu

ABSTRACT

Developing and maintaining resilient road networks is a key strategy for meeting several UN sustainable development goals. Currently about 75% of the road network in Africa is unpaved, making it especially susceptible to damage from precipitation (which accounts for 80% of their degradation). Global climate change will cause higher intensity weather events, which can take the form of stronger and more damaging floods. Infrastructure is a key aspect to economic development, and the World Bank has recently focused on “closing the infrastructure gap in Africa”. In this paper we propose a general framework for recommending road network improvements under different flooding scenarios in order to minimize the losses to network accessibility. This framework can consider different optimization objectives and budget constraints in order to give policy makers, investors, or relief organizations an idea of the trade-offs associated with these parameters. We demonstrate our framework in Senegal, by using flood data from Fathom.Global and mobility data from Orange S.A. We find that there are trade-offs between different budgets, and optimization objectives, and discuss their broader impact on sustainability objectives.

KEYWORDS

Computational sustainability, flood risk, Senegal

1 INTRODUCTION

Climate change is a global concern that is causing widespread disruptions to environmental, socioeconomic systems and human health. Recent studies have assessed the risks posed by extreme weather events, rising sea levels, and altered temperature and precipitation regimes to essential infrastructure systems. Damages to infrastructure are of particular concern for developing countries, where investment in energy, water, communication and transport infrastructure is a key strategy for meeting several UN sustainable development goals. For instance, estimated costs of road repair and maintenance across Africa under current climate change projections exceed \$150 billion, diverting significant funding from initiatives for expansion and development. Thus, climate change will exacerbate existing socioeconomic vulnerabilities and threaten the success of crucial development schemes unless steps are taken to proactively mitigate these costs.

Road networks are especially important to supporting socioeconomic development in the least developed countries, since they provide access to services like education and healthcare and enable overland trade flows that are integral to the growth of developing economies. Currently about 85% of the road network in Western Africa and 88% of that in south central Africa are comprised of

unpaved roads [3], making them highly susceptible to damage from precipitation which accounts for 80% of their degradation [2].

The first step toward developing climate-resilient road infrastructure is to analyze the exposure to climatic pressures. The probability of roads being inundated is a plausible measure of climate vulnerability given the prevalence of unpaved roads and the increasing incidence of flooding in Africa in recent decades. Similarly, the simulated maximum flood depth over roads, derived from hydrological models, can provide a worst case estimate of the difficulty of traversing a road. Secondly, the contribution of each road segment to regional connectivity must be assessed. Roads that are both critical thoroughfares and frequently flooded under historical or projected climatic conditions should be prioritized for weather-proofing and other upgrades. After flood risk and mobility impacts are determined, optimization techniques can be used to determine explicit plans for allocating road maintenance funds. Multiple sustainable development objectives can be explored within this framework, such as maximizing rural connectivity or minimizing the expected number of people isolated due to flooding. This approach has the potential to minimize the long-term cost of establishing a reliable road network while helping to buffer vulnerable populations from extreme weather events.

Our work aims to implement the framework described in the previous paragraph. Our objective is to find which roads in Senegal should be reinforced to prevent flooding, given a fixed budget for infrastructure investments, in order to best help the most people in Senegal. Our solution to this problem consists of three computational modules, pictured in Figure 1. Each module incorporates data from external sources as well as outputs from previous modules to perform its computation. The modules respectively calculate 1.) which roads in Senegal are likely to be flooded under different conditions; 2.) how different flooding scenarios will impact the mobility of persons in Senegal; and 3.) how to best fortify the road network against flooding to minimize the loss of mobility under different flooding scenarios. We use this framework in the context of Senegal by leveraging flood data from Fathom.Global (previous SSBN) and mobility data from Orange S.A., but note that our framework can be applied more broadly. Specifically, because the functionality of each module is self-contained, i.e. decoupled from the whole system, it would be straightforward to substitute modules that perform each computation differently, e.g. substituting module 1 for one that computes wildfire risk near each road segment.

The organization of this paper is as follows. In Section 2, specifically, 2.1, 2.2, and 2.3, we describe the operation of each of the three modules, we present computational results in Section 3, and finally discuss the results and broader application of our methods in Section 4.

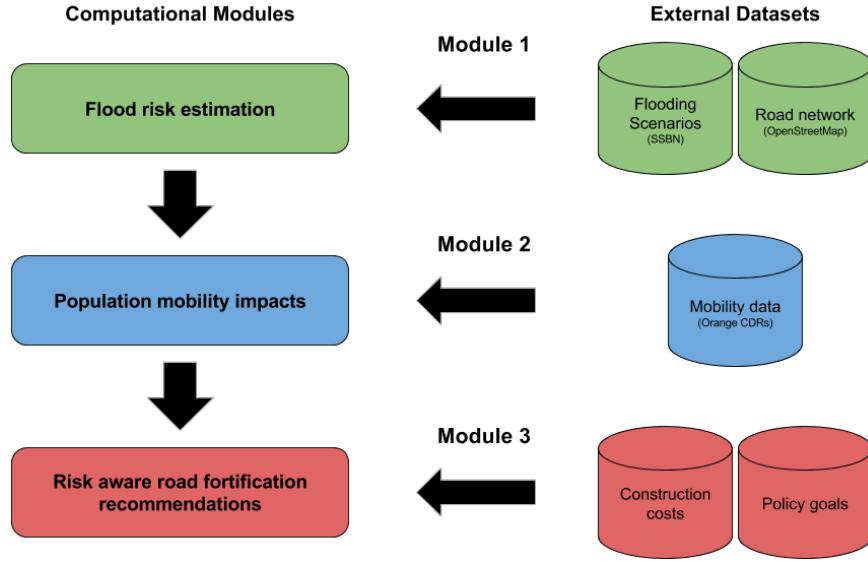


Figure 1: Our system is made up of three connected modules: a “flood risk computation” module, “population mobility impacts” module, and “risk aware road fortification recommendation” module. Each module relies on external data and outputs from the previous modules.

2 METHODS

We divide the problem of upgrading the road network in Senegal into three distinct components. The first task (Section 2.1) is to determine which parts of the Senegalese road network are at risk of inundation or damage from flooding events. It is possible to prioritize roads for repair simply on the basis of their exposure to flood risk, but this does not take into consideration the degree to which different roads are used by the population. Ideally, the upgraded roads should be those that are both at high risk of being flooded and have a high contribution to the accessibility afforded by the road network. Thus, the second component (Section 2.2) of our approach determines the travel demand between different regions of Senegal and how well the road network supports the mobility of the population under different flooding scenarios in terms of a number of accessibility measures or metrics. The final component (Section 2.3) of our approach optimizes the allocation of a fixed budget to upgrade a subset of roads such that the accessibility provided by the road network is maximized. Each of these components is handled in an independent module that can be modified to use different methods from the ones used in this work.

2.1 MODULE 1: Estimating Flood Risk

The purpose of this computational module is to determine which roads in Senegal are flooded under different scenarios. This module takes two GIS data sources as input—a road network and flood raster—and subsequently outputs both the *flooded* and *unflooded*, or traversable, parts of the road network. The data and methods used in this module are described below.

2.1.1 Road and Flooding Data. We obtained a shapefile of the highway system in Senegal from ArcGIS containing roads

classed as national and regional roads and highways [1]. We manually preprocessed and cleaned this road network data by connecting all roads that lead to the same settlement¹, adding in the ‘N4’ highway that passes through The Gambia, and joining several pairs of disconnected road endpoints that are likely data errors. We then represent this input road network as a weighted, undirected graph $G = (V, E)$ consisting of edges E representing road segments and vertices V representing the latitude-longitude coordinates of the endpoints of the road segments. Each edge has an associated weight denoting the length of the road segment. The graph representation of the Senegalese road network has 6,917 vertices and 7,175 edges all within a single connected component. We refer to this as the *business-as-usual* road network to distinguish it from networks when some of the edges are lost due to flooding.

In addition to the ArcGIS road network, we used Fathom-Global (previously known as SSBN-Global) flooding data [9, 10] for Senegal as input to this module. These data are comprised of rasters of the flood depth for floods of different severities. Specifically, each flooding raster has a size of 8000×6000 cells, where each cell represents an area of approximately $90m^2$ and the value of the raster cell is the maximum flood depth in meters within the $90m^2$ area during a flood of a specified severity. The severity of a flood is characterized by its return period λ , which is the estimated time interval between flooding events of a similar intensity. A flood with return period λ years has a $\frac{1}{\lambda}$ chance of occurring in any given year. Intuitively, a 500 year flood is more extreme and less likely to occur than a 200 year flood. The SSBN data includes fluvial (river based) and pluvial (rain based) flooding scenarios for 10 different flooding

¹In the raw ‘Senegal Roads’ dataset, many of the roads end at the outskirts of cities, whereas in reality they are connected through urban roads that are not included in the dataset.

return periods: 5, 10, 20, 50, 75, 100, 200, 250, 500, and 1000 years. Because the distinction between fluvial and pluvial flooding is not important for determining which populations could be affected by inundated roads, as a preprocessing step we sum the fluvial and pluvial flooding rasters for each return period. Formally we represent each flooding raster as a matrix, $F^\lambda \in \mathbb{R}_{\geq 0}^{8000 \times 6000}$, whose entries represent the combined maximum flood depth in meters at each cell in the case of a λ return period flooding scenario.

2.1.2 Classifying Roads as Flooded. Next, we describe the process used to designate whether or not a road is considered flooded during a specific flooding event. We introduce a parameter α which defines the flooding threshold level over which a cell is classified as flooded. That is, only raster cells with a flood depth $F_{ij}^\lambda \geq \alpha$ are considered flooded under a given return period flood scenario and flooding threshold. Flooding thresholds are commonly used to distinguish between minor, moderate and major flooding, and so by varying α from small to large threshold values we shift our focus from all flooded cells to only those with relatively severe flooding. The α parameter can also be regarded as a tolerance level for flood depth beyond which roads within a cell become impassable. Given that the Fathom-Global flooding data rasters contain the *maximum* flood depth within each cell and that we sum pluvial and fluvial flooding to obtain a single flood raster, the flooding data we use reflects the *worst-case* flooding level under a given return period scenario. However, the true flood depth observed within the cell may be lower than this worst-case level, e.g. if there are local features such as drainage or elevation that reduce the extent to which the roads are submerged. From this standpoint, the α parameter is the maximum flooding that can be expected to occur before mobility through the cell is affected. In this case considering low to high values of α equates to assessing impacts under conservative to optimistic estimates of flood tolerance.

Then, for a return period λ flood and a flooding threshold of α , we say that a road $(u, v) \in E$ is flooded if the road segment connecting u and v passes through any $\{i, j\}$ cells in F^λ for which $F_{i,j}^\lambda \geq \alpha$. This is based on the assumption that even if only a small stretch of a road segment is too flooded to traverse, the road segment is impassable. We obtain the subgraph of the BAU road network G that is considered *unflooded* under flood scenario λ and threshold α , $G_U^{\lambda, \alpha} = (V, E_U^{\lambda, \alpha})$, consisting of the original set of vertices V and any edges corresponding to road segments *not* affected by flooding $E_U^{\lambda, \alpha}$. In the second computational module (Section 2.2) we use this unflooded road network $G_U^{\lambda, \alpha}$ to calculate how different accessibility measures of the road network are impacted by a flooding scenario. Similarly, we can construct a *flooded* subgraph of G , $G_F^{\lambda, \alpha} = (V, E_F^{\lambda, \alpha})$, in which only edges representing roads that *are* flooded are included. For each edge in the flooded graph, we also compute the flooded distance or the length of the road that passes through flooded raster cells to estimate how much of the road would need to be upgraded in order to make the road segment traversable again. In the third computational module (Section 2.3) we consider the edges in the flooded road network $G_F^{\lambda, \alpha}$ as candidate roads to reinforce against flooding, in order to recover the accessibility provided by the road network. Note that the unflooded and flooded edge sets are mutually exclusive, i.e. $E_U \cap E_F = \emptyset$, and

$E_U \cup E_F = E$. An example of a road network with road segments classified as flooded or unflooded can be seen in Figure 2, which shows the predicted flood depth for a 100-year flood in Senegal and the road segments that would be considered impassable due to floodwater over 1m.

2.1.3 Effects of Flooding on Road Network. As the severity of flooding scenarios λ increases (with increasing return period), the number of unflooded road segments declines, the total length of flooded roads rises and the fragmentation of the road network into disconnected regions increases (Figure 3). As the flooding tolerance α increases, each of these effects becomes less pronounced.

2.2 MODULE 2: Quantifying Mobility over the Road Network

The second computational module focuses on quantifying the effects that flooding scenarios will have on the mobility of people living in Senegal. The module takes as input road networks obtained from the Flood Risk module (Section 2.1) and baseline travel demand data, derived from anonymized call detail records (CDRs) from the Orange cellular provider. Using these data, it calculates global road network accessibility metrics, such as average trip distance, that measure the ease with which individuals can reach desired destinations throughout Senegal. We compare the accessibility metrics of the business-as-usual road network to those from different flooding scenarios in order to quantify the disruption the floods will have on mobility in Senegal. This process is split up into 4 parts: 1.) dividing Senegal into zones based on Orange cell tower data; 2.) calculating the distances between each pair of zones; 3.) deriving travel demand between zones from the Orange data; and 4.) calculating the road network accessibility measures. Each of these parts is described in the following sections.

2.2.1 Cell Tower Zones. The CDR data from Orange consists of fine-grained trajectories for sets of anonymized users. These trajectories give a time indexed sequence of *cell towers* used by each user, thus the minimum resolution travel demand data we can derive will be between areas covered by these cell towers. Here the first step is defining the spatial area covered by each zone. The Orange data gives the approximate latitude-longitude locations of the 1666 towers that service Senegal. We use these points to divide Senegal into “cell tower zones” by constructing Voronoi regions from the locations of the cell towers. The zone for each cell tower is defined as the polygon that contains all points that are closer to that tower than to any of the other cell towers. This is a very common technique for deriving zones from point data, and has been used by several papers during the Orange Data for Development Challenge[8]. Figure 4 shows the location of each cell tower as a purple point, and the resulting zones.

2.2.2 Zone-to-zone Distance. The second step in calculating the global accessibility measures is calculating the distance between each pair of zones over the road network. Formally, we create a matrix D from a given road network $G^{\lambda, \alpha} = (V, E^{\lambda, \alpha})$, where an entry D_{ij} represents the average shortest path distance over paths with origin vertices in i and destination vertices in j . Recall from Section 2.1.1 that we represent the road network as a

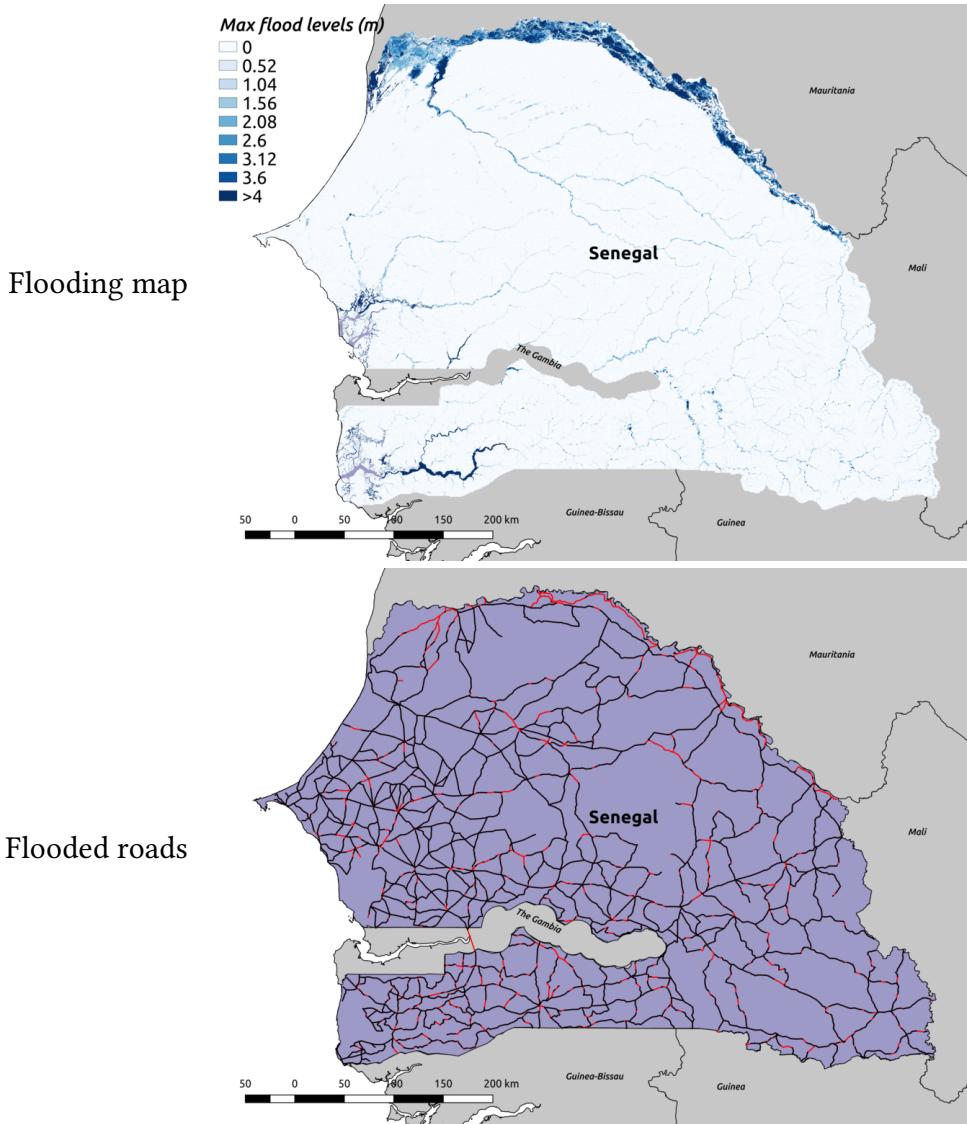


Figure 2: Flooding scenario for $\lambda = 100$ and $\alpha = 1.0$. (Top) map shows the maximum flood depth (in meters) over all of Senegal. (Bottom) map shows the road network, where the segments that are considered flooded under $\alpha = 1.0$ are colored red.

graph, where each vertex is indexed by its latitude-longitude location. Using this information, for each zone i , we calculate Z_i , the set of vertices from the road network that are within its boundaries. If a zone j has less than three vertices within its boundaries, i.e. $|Z_j| < 3$, we add the nearest vertices (as measured to the zone's centroid) until it has three associated vertices. Now, we calculate an entry D_{ij} as the average distance over all shortest paths that have an origin in Z_i and destination in Z_j . Note that for some given pairs of zones, especially in flooded road networks, there may be no paths available between some pairs of vertices from $Z_i \times Z_j$, i.e. it may be the case that there exists some $u \in Z_i$, and $v \in Z_j$ where u and v lie in different connected components of the given road network. We define D_{ij} as the average shortest path distance between

pairs of vertices for which there is a valid path. If there are no valid paths between i and j we set D_{ij} to an arbitrarily large number (because D_{ij} is only used in cases where there *are* feasible paths between i and j , the specific value does not matter). Separately, we calculate two matrices $C^{feasible}$ and $C^{infeasible}$, where an entry $C_{ij}^{feasible}$ represents the number of pairs of vertices in Z_i and Z_j between which a path exists, and similarly, an entry $C_{ij}^{infeasible}$ represents the number of pairs of vertices between which a path does not exist².

²Note that $C_{ij}^{feasible} + C_{ij}^{infeasible} = |Z_i \times Z_j|$.

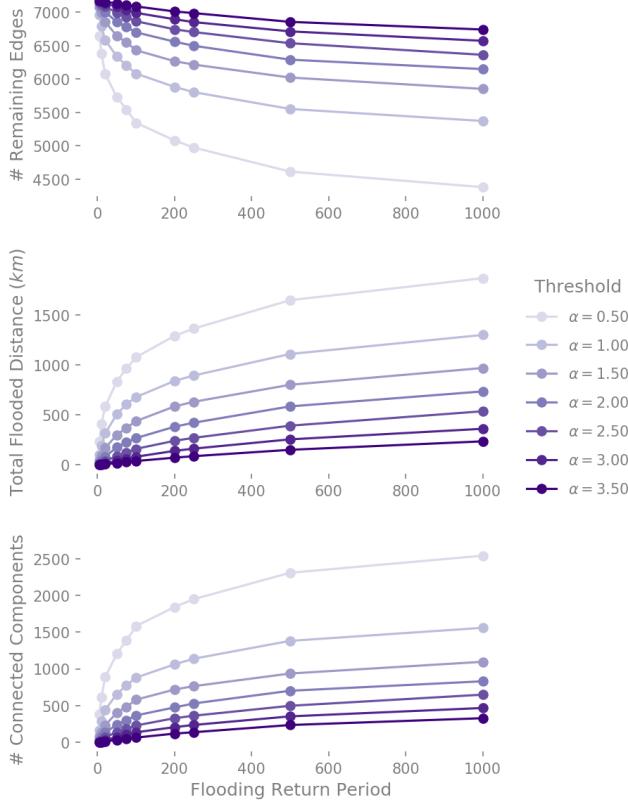


Figure 3: Effect of floods on number of traversable road segments, total distance of flooded roads and fragmentation of the road network, under varying α threshold levels.

2.2.3 Deriving Travel Demand. The final data component needed for calculating the accessibility metrics is an estimate of the actual travel demand between each pair of zones. Without this data, we can only calculate accessibility metrics related to the structure of the road network. That is, without the mobility data there is no way to quantify *how* the road network is used, and thus how flooding will directly disrupt the mobility patterns of people in Senegal.

To estimate inter-zone travel demand we use the fine-grained mobility dataset provided by Orange. This mobility dataset contains information from over 300,000 randomly sampled, anonymous customers for 25 biweekly periods in 2013³. The mobility data gives the spatial trajectory of each customer over a two week period as a sequence of cell tower ids, l_n , e.g. $\{l_1, l_2, \dots, l_N\}$ for a customer that moves through N zones. Each consecutive pair of cell tower locations (l_n, l_{n+1}) in a customer’s trajectory is considered as a trip from cell tower zone l_n to zone l_{n+1} . We construct an origin-destination (OD) trip matrix, $T \in \mathbb{Z}_{\geq 0}^{1666 \times 1666}$, where entry T_{ij} represents the total number of trips taken from cell tower zone i to zone j by all customers over all biweekly periods. More specifically, we iterate through the trajectory for each user and increment

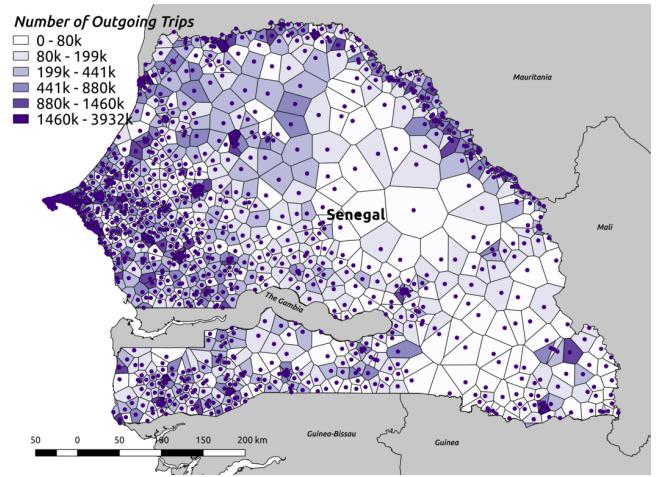


Figure 4: Map of Senegal that shows the given locations of the 1666 Orange data cell towers and their corresponding “cell tower zones”. Each zone is colored based on the number of outgoing trips from the derived trip data.

$T_{l_i, l_{i+1}}$ for each $i, i + 1$ pair of cell tower locations in the trajectory. Figure 4 shows each cell tower zone colored by the number of trips leaving that zone, i.e. a zone i is colored by $\sum_{j=1}^{1666} T_{ij}$. As expected, cell tower zones near major settlements are associated with higher numbers of outgoing trips. In particular, the cell tower zones along highways radiating out from Dakar, such as the N1 highway to Kaolack, the N2 highway to Thies up to St. Louis and along the northern border, and the N3 highway through Diourbel and Touba, have a large number of outgoing trips. In contrast, there is relatively low travel demand from within the Niokolo-Koba National Park and the Ferlo Nord Wildlife Reserve in eastern Senegal, reflecting the low population density in these regions.

2.2.4 Road Network Accessibility Measures. In review, we partition Senegal into a set of 1,666 zones based on the Orange cell tower locations, calculate the number of trips between each pair of zones, T , and, for a given road network, calculate the average distance between each pair of zones, D , number of infeasible paths between each pair of zones, $C^{infeasible}$, and number of feasible paths between each pair of zones, $C^{feasible}$. Note that for different road networks, i.e. the business-as-usual road network and flooded road networks from different scenarios, the D , $C^{infeasible}$ and $C^{feasible}$ matrices will all differ. We use these four matrices, and the structure of the road network itself to calculate four network accessibility measures as follows:

Number of connected components: A connected component in a network is a set of vertices such that there exists a path between every pair of vertices in the set. In the business-as-usual scenario, the road network of Senegal is made up of a single connected component, meaning that it is possible to reach any location in the network from any other location. However, under flooding scenarios the road network becomes fragmented, i.e. it is divided into many disjoint connected components. One of the accessibility measures

³Different samples of customers were taken for each biweekly period.

we consider in our analyses is a count of the number of connected components in a given road network, where the higher the number of connected components, the more fragmented is the road network and the worse the impact on mobility within Senegal.

Size of the largest connected component: Similar to the previous measure, this accessibility measure is simply the number of vertices that make up the largest connected component. The largest connected component in the road network will be the largest contiguous section of the road network where people can freely move around. Note that this measure and the previous measure do not consider the derived travel demand data and are purely structural metrics.

Percentage of infeasible trips: This measures how many of trips cannot be completed given a road network. Formally,

we let $I = \sum_{i=1}^{1666} \sum_{j=1}^{1666} \frac{T_{ij} C_{ij}^{\text{infeasible}}}{C_{ij}^{\text{feasible}} + C_{ij}^{\text{infeasible}}}$ be the number of infeasible trips. Here the number of infeasible trips between an origin i and destination j is counted as the total number of trips from i to j , multiplied the fraction of infeasible paths between the two zones. We assume that the number of impossible trips is directly proportional to the number of impossible paths between vertices in the origin zone, and vertices in the destination zone. Thus the percentage of infeasible trips becomes $I\% = I / \sum_{i=1}^{1666} \sum_{j=1}^{1666} T_{ij}$.

Average trip length: This measure also takes into account the trip matrix, and records the average trip distance over all possible trips. Formally, we form a matrix F , (similar to the scalar I value from the previous measure) where an en-

try, $F_{ij} = \frac{T_{ij} C_{ij}^{\text{feasible}}}{C_{ij}^{\text{feasible}} + C_{ij}^{\text{infeasible}}}$, represents the number of feasible trips between zone i and zone j . Now we let L be the global average trip distance given by $L = \frac{\sum_{i=1}^{1666} \sum_{j=1}^{1666} F_{ij} D_{ij}}{\sum_{i=1}^{1666} \sum_{j=1}^{1666} F_{ij}}$.

In other words, this is the total number of kilometers traveled over all trips, divided by the total number of trips, thus average trip length.

2.2.5 Accessibility Measure Results. In this section we show several results from computing the four accessibility measures for the business-as-usual road network, and the road networks induced by all of the flooding return periods described in Module 1, for several values of α . Figure 5 shows a graph for each accessibility measure that indicates how the metric changes under different flooding return periods. A notable result from these graphs is that the rate at which trips become infeasible due to flooding outpaces the rate at which trip lengths become longer due to flooding. The average trip distances generally *decrease* in more severe flooding scenarios, simply due to the fact that longer distance trips are not possible in those scenarios. Predictably, the number of connected components increases while the size of the largest connected component decreases as flooding becomes more severe.

Additionally, we show several maps in Figure 6 that show how the degradation of the road network will affect the mobility of people from specific cities. The first and second rows show the road network distance (D_{ij} values) in kilometers for two cities, Dakar and Tambacounda respectively, to all other zones, for different flood

scenarios. The first column shows the business-as-usual scenario with no flooding for both cities, while the second and third column show distances under $\lambda = 10, \alpha = 1.0$ and $\lambda = 100, \alpha = 1.0$ respectively. As the flooding scenarios become more severe larger sections of the network become unreachable, and sections of the network that remain accessible are generally harder to travel to. For example, travel from Dakar to the St. Louis region in the northern portion of Senegal is relatively short in the business-as-usual scenario, however, under the $\lambda = 10$ scenario, the shortest path is cut off and a longer route must be taken to make the same trip. Furthermore, under the $\lambda = 100$ scenario, the same St. Louis area is completely cut off from Dakar. The situation in the 100 year flooding return period is particularly bad for the city of Tambacounda, as can be seen in the bottom right map. Here, only the immediate area around the city is reachable on the road network.

2.3 MODULE 3: Optimizing Repairs over the Network

The final module provides a decision-support tool for designing road infrastructure improvement strategies that balance socioeconomic, environmental and economic considerations. It takes the graphs $G_F^{\lambda, \alpha}$ and $G_U^{\lambda, \alpha}$ representing the flooded and unflooded parts of the road network, the monetary costs of repairing or upgrading each road segment $c(e)$, and a budget B as inputs to determine the optimal allocation of funds that recovers the maximum amount of accessibility as quantified by a given accessibility measure. This is essentially a combinatorial optimization problem involving the selection of a subset of items (roads to fortify) from a superset (all affected roads). Many such problems fall into a class of problems known as “NP-complete”, for which there are no known polynomial-time algorithms for finding optimal solutions, and thus necessitate the use of heuristic solution strategies that are not guaranteed to yield optimal results. We begin by considering each of the accessibility measures discussed previously as optimization objectives, formulating each as a graph optimization problem and analyzing its computational complexity. Then we discuss algorithms for optimizing two of the accessibility measures, and present the solutions for recommended road upgrades in both cases.

2.3.1 Optimization Problem Formulations.

(1) Minimizing the Number of Connected Components

As flooding events render road segments impassable, either due to inundation or due to damage from exposure to frequent inundations, the overall road network becomes increasingly fragmented. This results in regions becoming isolated from one another, as trips between these regions are infeasible. One strategy for mitigating the loss in road network connectivity resulting from a flooding event is to re-establish connections between isolated regions, or to minimize the number of connected components. More formally, we upgrade flooded roads corresponding to a set E_{upgrade} of edges $e \in E_F^{\lambda, \alpha}$ such that $\sum_{e \in E_{\text{upgrade}}} c(e) \leq B$ and when they are added to the unflooded graph $G_U^{\lambda, \alpha}$ the number of connected components in the resulting graph $G = (V, E_U^{\lambda, \alpha} \cup E_{\text{upgrade}})$ is minimized.

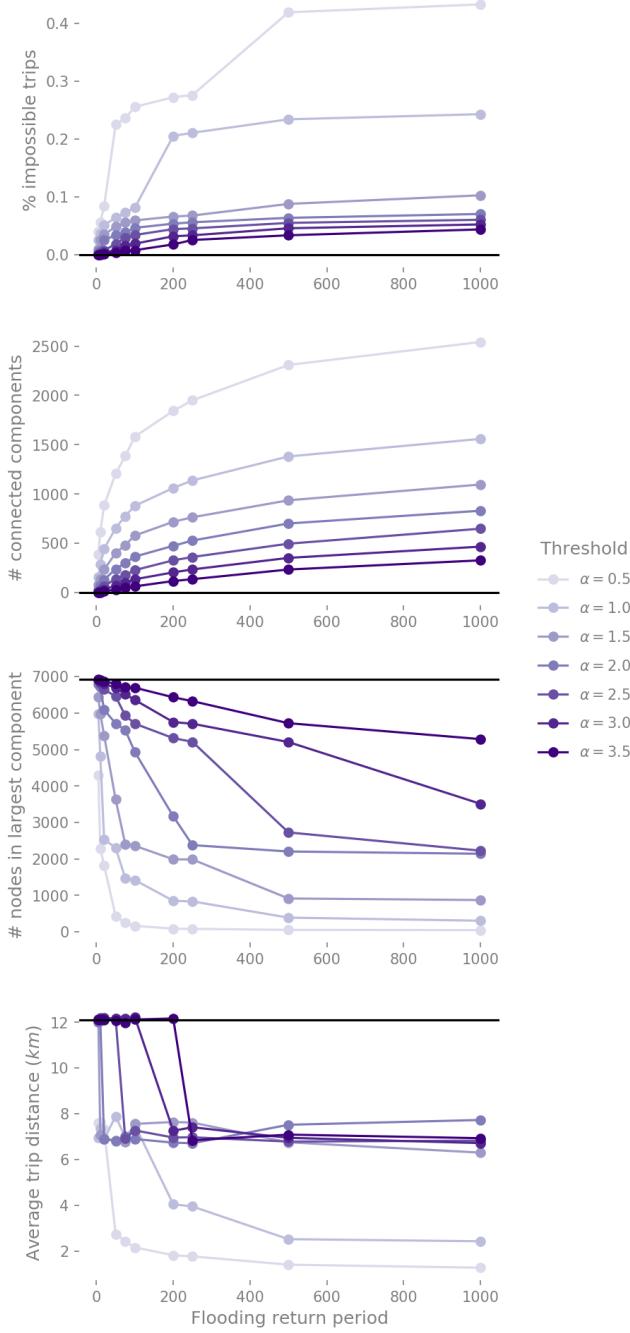


Figure 5: Road network accessibility metrics for different values of α over all return periods. The business-as-usual value is shown as a black horizontal line in each graph. From top to bottom the accessibility metrics shown are: number of infeasible trips, number of connected components, number of vertices in the largest connected component, and average trip length.

First, we specify that only edges $e \in E_F^{\lambda, \alpha}$ whose endpoints are in *different* connected components are candidates for upgrading, since otherwise if the edge connects two vertices in the same connected component then upgrading the corresponding road segment does not reduce the fragmentation in the road network. Furthermore, we only consider the minimum cost edge connecting a pair of connected components, since any other edge between the same pair of connected components offers the same benefit but at higher cost. More formally, given two connected components p and q with corresponding vertex sets V_p and V_q , the set of edges $E_{pq} = \{e_{ij} \in E_F^{\lambda, \alpha} | e_i \in V_p, e_j \in V_q\}$, we only consider $\arg \min_{e_{ij} \in E_{pq}} c(e_{ij})$ as a candidate for connecting p and q .

Lastly, note that we should not add edges that create cycles between connected components. For example suppose we upgrade a road between connected components p and q , and another road between q and r . We should no longer consider any roads between p and r as candidates for repair, since they are already connected via q .

Suppose the graph $G_U^{\lambda, \alpha}$ consists of K disjoint connected components V_k such that $V = \cup_{k=1}^K V_k$ and $V_i \cap V_j = \emptyset$ for $i \neq j$. We construct a second graph $G' = (V', E')$ in which there is a single vertex $v_k \in V'$ for each connected component in $G_U^{\lambda, \alpha}$, and there is at most one edge between any pair of vertices $v_p, v_q \in V'$ whose weight is equal to $\min_{e_{ij} \in E_{pq}} c(e_{ij})$. Our goal is to find a tree or collection of trees

(a forest) in G' such that the total weight of the tree edges does not exceed the given budget B and that spans the maximum number of vertices in V' . This problem can be solved optimally using a greedy algorithm (see Algorithm 1) based on the widely known Kruskal's algorithm for finding the minimum spanning tree of a graph. The edges in E' are considered in ascending order of their repair cost, and each edge is added as long as it does not introduce a cycle and there is enough budget left.

Algorithm 1 Minimizing the Number of Connected Components

```

1: procedure MINNUMCC( $G', B$ )
2:   Candidates  $\leftarrow$  sorted( $E'$ , ascending)
3:   Upgrade  $\leftarrow \emptyset$ 
4:    $i \leftarrow 0$ 
5:   while  $B \geq 0$  and  $i < |E'|$  do
6:     if Candidates[ $i$ ] does not form cycle then
7:       Upgrade  $\leftarrow$  Upgrade  $\cup$  Candidates[ $i$ ]
8:        $B \leftarrow B - \text{Candidates}[i][\text{'weight'}]$ 
9:     end if
10:     $i \leftarrow i + 1$ 
11:   end while
12:   return Upgrade
13: end procedure

```

(2) Maximizing Size of a Specific Connected Component
Rather than minimize the number of disconnected parts of

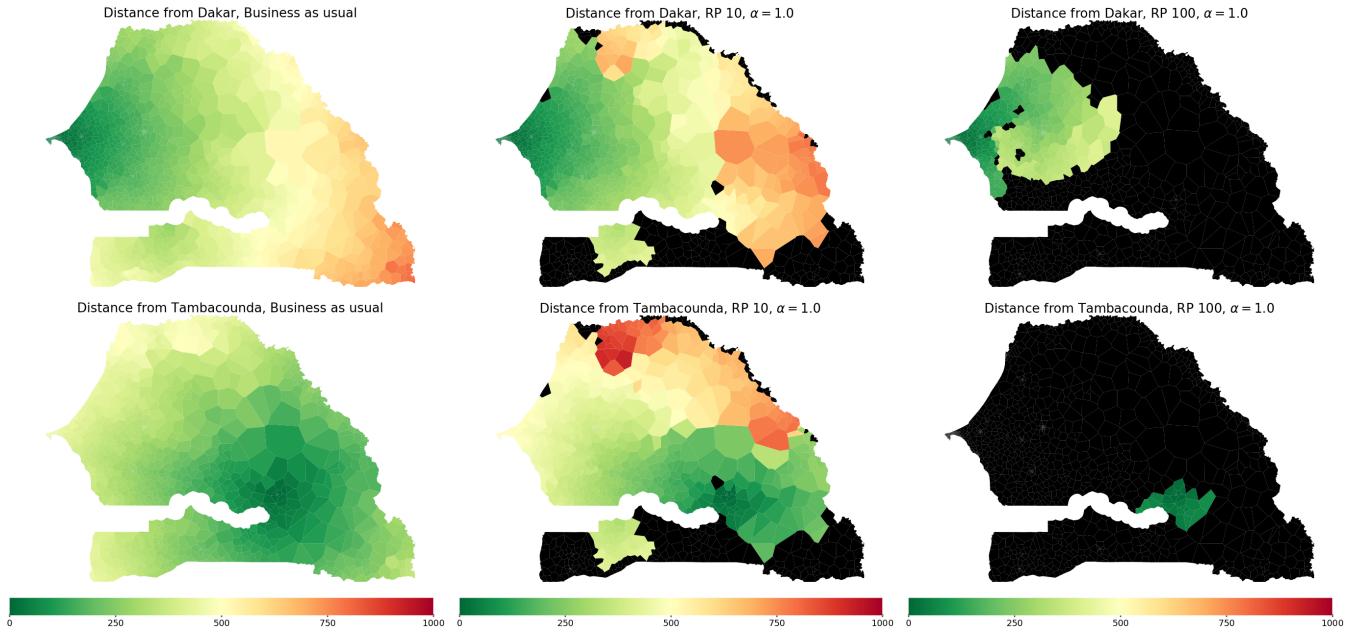


Figure 6: Maps showing the average road network distance from two cities in Senegal (top row: Dakar, bottom row: Tambacounda) under no flooding (business-as-usual), a 10-year flood scenario and a 100-year flood scenario, with inaccessible regions colored black.

the road network, it may be preferable to maximize the number of regions that can be accessed from a particular location, such as the capital or a major city or a base of operations for the emergency mobilization of disaster response assets. In this case the goal is to maximize the number of vertices that can be accessed from a particular root vertex, or alternatively to maximize the size of a given connected component. In formal terms, we upgrade flooded roads corresponding to a set $E_{upgrade}$ of edges $e \in E_F^{\lambda, \alpha}$ such that $\sum_{e \in E_{upgrade}} c(e) \leq B$ and when they are added to the unflooded graph $G_U^{\lambda, \alpha}$ the size of a specified connected component V_p in the resulting graph $G = (V, E_U^{\lambda, \alpha} \cup E_{upgrade})$ is maximized. As with the problem of minimizing the number of connected components, we only consider upgrading edges that are flooded under the scenario in question, and only consider the minimum cost edge between pairs of connected components as repair candidates.

This problem can be formulated as an instance of the Budget Prize-Collecting Steiner Tree Problem [5] on the transformed graph G' , where the prize $p(v)$ associated with each vertex in V' is equal to the size of the connected component it represents. That is, we seek to find a subtree that maximizes the vertex prize subject to a budget constraint on the edge costs⁴. This problem is computationally challenging (NP-hard) due to both the requirement that the resulting subtree be connected and due to the goal of maximizing its value. One way

to solve this problem is to encode it as mathematical program that can be solved using mathematical optimization algorithms. The choice of whether or not to select each candidate edge is represented as a binary decision variable, e.g. $x_e = 1$ if edge e is selected for fortification and $x_e = 0$ otherwise. Then, the budget constraint and the objective can be written as linear functions of these decision variables. However, the connectivity constraints are less straightforward to encode, and a number of different formulations for encoding them have been proposed (see e.g. [4]). Alternatively, there have been some approximation algorithms proposed for solving the Budget Prize-Collecting Steiner Tree and the related Budget-Constrained Steiner Connected Subgraph with Node Profits and Node Costs problems [6, 7]. These algorithms have a more favorable runtime than exact solution approaches, but they cannot guarantee that they will find an optimal solution to the problem of selecting edges to upgrade. Out of the two structural accessibility measures, we focus on minimizing the number of connected components.

(3) Minimizing Number of Infeasible Trips

Minimizing the number of infeasible trips (or equivalently, maximizing the number of feasible trips) uses both the zone-by-zone distance matrix, D , and the zone-by-zone trip demand matrix, T . Similar to the other problem formulations, we aim to choose the subset of edges within the given budget that maximally reduces the number of infeasible trips in a network.

We begin the analysis of this problem by considering the effects of adding a single edge, $(u, v) \in E_{upgrade}$, back into

⁴This can also be formulated as a Budget-Constrained Steiner Connected Subgraph Problem with Node Profits and Node Costs, see Dilkina and Gomes [4].

$E_U^{\lambda, \alpha}$, on the accessibility measures for $G_U^{\lambda, \alpha}$. The addition of this edge could possibly change the shortest paths in the network, thus necessitating the recomputation of D , $C^{feasible}$, and $C^{infeasible}$. Indeed, since we do not know which shortest paths the newly added edge (u, v) might be part of, the shortest paths between all pairs of vertices in V must be recomputed. If we let P_{ij} represent the set of shortest paths that go from zone i to zone j in the original graph $G_U^{\lambda, \alpha}$, and let P'_{ij} represent the same for $G = (V, E_U^{\lambda, \alpha} \cup \{(u, v)\})$, then we only need to recompute entries in D , $C^{feasible}$, and $C^{infeasible}$ for pairs of zones i and j for which P_{ij} and P'_{ij} differ. The running time of the all pairs shortest path computation dominates this update, taking $O(|V|^2 \log |V|)$ with Dijkstra's algorithm.

As our goal is to minimize the number of infeasible trips, we will need to exactly evaluate the effect of adding different sets of edges back into $E_U^{\lambda, \alpha}$. Since there are combinatorially many possible sets of edges to upgrade, it is impractical to attempt a brute-force search for the optimal set of edges. A common strategy for solving this type of problem is to iteratively build a solution, which in our case is a set of edges to fortify that are within our budget constraint, by greedily choosing the single “best” candidate, adding it to the solution, and repeating until the budget is exhausted. As we have shown in the paragraph above, each evaluation of the quality of a solution to our problem is computationally expensive, therefore, even this naive approach is too costly to run in a reasonable amount of time. As an example, in the $\lambda = 100, \alpha = 1.0$ case there are ≈ 1000 edges that are considered flooded. Here we must evaluate the quality of a solution that considers adding the first edge, then the second edge, etc. This first iteration will take 1000 calls to the evaluation method, and choose a single edge to add to the solution. This leaves 999 edges, each of which must be evaluated again, etc. Assuming that each edge will cost 1 unit of budget to fix and that our evaluation computation takes 5 seconds to run, given a budget of 100, this iterative greedy algorithm will take at least 52 days to complete.

Considering this, we implement a further simplification of the greedy algorithm described above. This algorithm performs a single pass over the candidate edges, $E_{upgrade}$, and calculates the benefit of adding each to $G_U^{\lambda, \alpha}$ in turn. With this information we sort the edges by order of their cost benefit ratio, then simply pick the edges with the largest cost benefit ratio until the budget is exhausted.

(4) Minimizing Average Trip Distance

One outcome of flooding is that the original shortest path between a pair of locations in the business-as-usual may be unavailable and is replaced by a longer unflooded path. This effect causes an increase in the average trip distance within the road network, meaning that routes within the road network are less efficient. In order to counteract this effect, we consider upgrading the set of roads that maximally reduces the average trip distance in a network, subject to the same budget constraint on the cost of repairs as before.

Note that this problem suffers from the same computational difficulties as that of minimizing the number of infeasible trips. Each time an edge is considered as a candidate for upgrading, the number of feasible and infeasible trips and the average shortest path distances between cell tower zones must be recomputed, since both of these affect the overall average trip distance. It is possible to implement a similar greedy approach to tackle this problem. However, with this objective there is the added factor that adding an edge can actually increase the average trip distance by making previously infeasible *longer* trips possible. For this reason, out of the two travel demand-aware accessibility measures, we focus on minimizing the number of infeasible trips.

3 RESULTS

In Section 2 we have proposed a series of computational modules that can be used to plan road network improvements that improve different mobility related objectives under different budgets and flooding scenarios. Here, we discuss both the results from this process for a particular flooding scenario, and, more broadly, considerations to take when using optimization tools to achieve policy objectives.

The results we show are for the 100 year flooding return period with $\alpha = 1.0$. In Figure 7 we show the four accessibility metrics computed on the flooded road network without any improvements, and with the improvements suggested by our two optimization algorithms for four different budgets: {100, 200, 300, 400}. In these graphs, the business-as-usual accessibility metric is shown as a black horizontal line. The two optimization algorithms are minimizing different objectives, namely the number of connected components in the network and the percent of impossible trips over the network. As discussed in the previous section, the solutions obtained using the algorithm that minimizes the number of connected components are optimal, i.e they reduce the number of connected components by the largest amount possible under a given budget. The solutions obtained for minimizing the percent of impossible trips are *not* optimal, but constructed naively through a simple greedy heuristic. This discrepancy is reflected in the top graph of Figure 7, where, with a budget of 300 and 400, the algorithm that minimizes the number of connected components is able to find solutions that have smaller numbers of infeasible trips than the algorithm that actually minimizes the number of infeasible trips.

The third graph in Figure 7 shows the non-linear trade-offs between the optimization budget and solution quality. In particular, there is a large jump in solution quality, in terms of the number of nodes in the largest connected component, when increasing the budget from 100 to 200 (in the solution found by minimizing the number of connected components). Further budget increases do not afford the same large jumps in solution quality, i.e. they exhibit diminishing returns.

We observe another trade-off between solution qualities and optimization objectives in the third and fourth graphs. Here, the solutions found by the two algorithms are fundamentally different. As we discussed in the previous paragraph, the algorithm that minimizes the number of connected components quickly creates a large connected component (the jump between budgets of 100

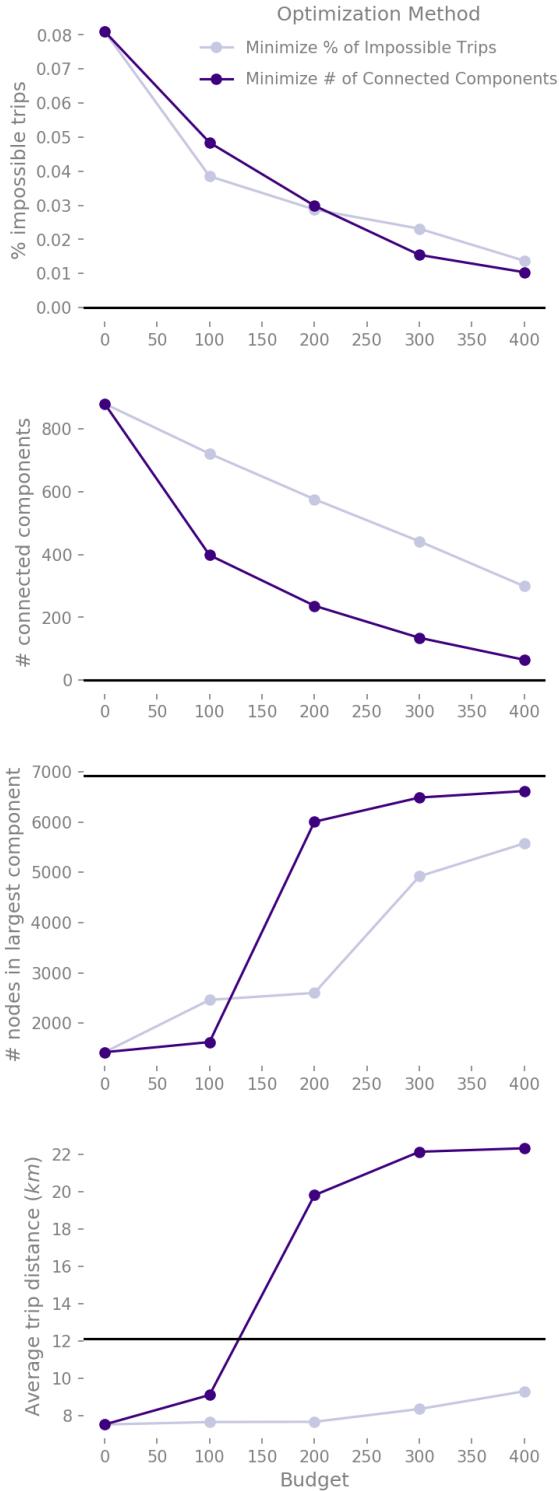


Figure 7: Results of optimization methods.

and 200), without any consideration to numbers of feasible trips, or average trip distances. This naivety is reflected by the corresponding large increase in the average trip distance. This behavior is *not* present in solutions found through the other algorithm. The algorithm that minimizes the percent of infeasible trips finds solutions with slightly fewer infeasible trips but with smaller average trip distances.

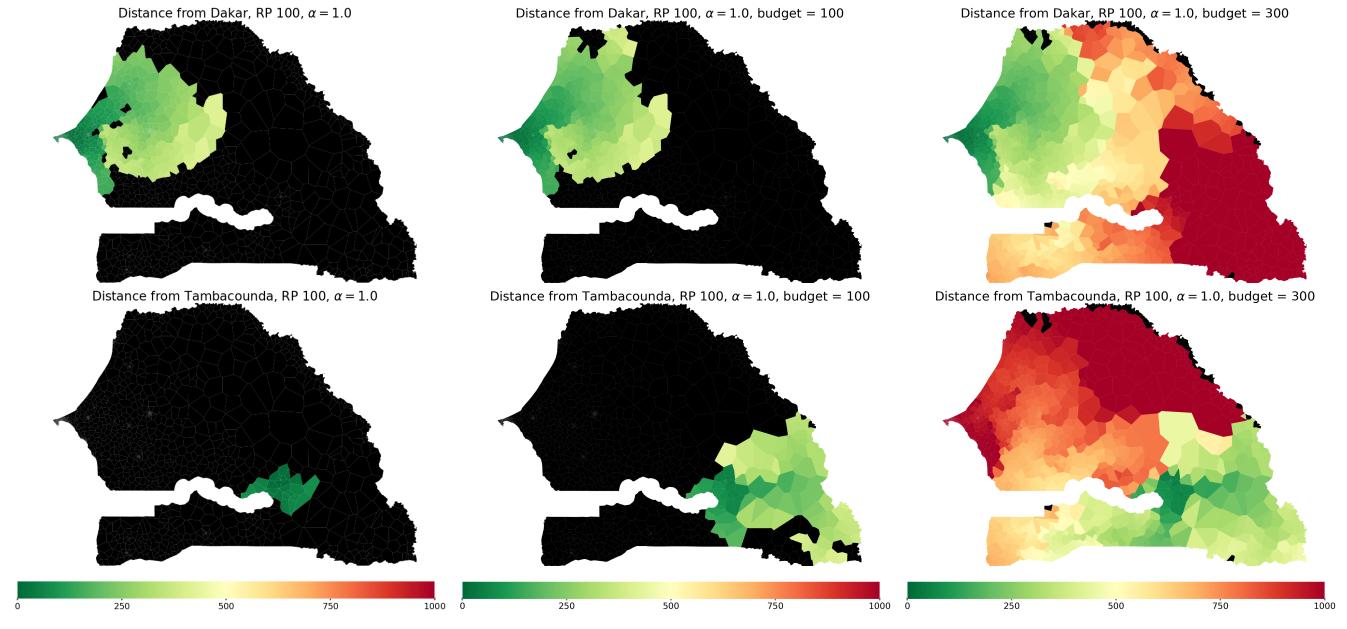
In Figure 8 we show similar maps to those from Figure 6, but with the optimized solutions found by both optimization algorithms for budgets of 100 and 300. The first column of Figure 8 shows the distance from each city (Dakar and Tambacounda) under the road networks given in the $\lambda = 100$, $\alpha = 1.0$ flooding scenario. The second and third columns show how the improvements, suggested with a budget of 100 and 300 respectively, translate into larger areas becoming accessible from each city, and shorter trip distances. The maps in the top two rows are created with the algorithm that minimizes the number of connected components, while the maps in the bottom two rows are created with the algorithm that minimizes the percent of infeasible trips. By comparing the two maps in the last column from the first two rows, to the same maps from the last two rows, we can see the difference in behaviours of the optimization strategies. There are few places in the first set of results that remain unreachable from Dakar or Tambacounda, because the algorithm minimizes the number of connected components. This is not the case in the second set of results, as the algorithm that minimizes the percent of infeasible trips will prioritize areas with high volumes of trips. This difference must be carefully understood for any optimization algorithm used in planning purposes.

4 DISCUSSION

Our results show that the accessibility provided by the Senegalese highway system will be significantly impacted even under flooding scenarios with return periods in the range of 20 to 50 years. Thus, the success of transportation development schemes in Senegal may be threatened by potentially near-future hazards and it is ever more urgent that present efforts to build climate-resilient infrastructure take these risks into account. Our framework integrates flood hazard, societal impact and financial consideration to provide recommendations to policy-makers for regions in which to prioritize investments.

In order to derive a cost-effective road improvement solution, it is important to explicitly incorporate financial constraints into the optimization process. The reasoning for this is clear: any sub-optimal solution can be replaced with an optimal one that achieves greater benefit with no additional cost; or alternatively, the same amount of benefit could be achieved with significant savings. Funding agencies and decision makers may be interested in a range of metrics for quantifying the resilience of transportation infrastructure to flooding. Our framework allowed us to compare optimizing road network improvements according to several different accessibility measures. However, we found that these accessibility objectives differed significantly in terms of the ease with which optimal solutions could be obtained. Minimizing the number of connected components in the road network was the easiest accessibility measure to optimize, compared to maximizing the size of the largest connected component, minimizing the number of infeasible trips

Optimization results for minimizing the number of connected components



Optimization results for minimizing the percent of infeasible trips

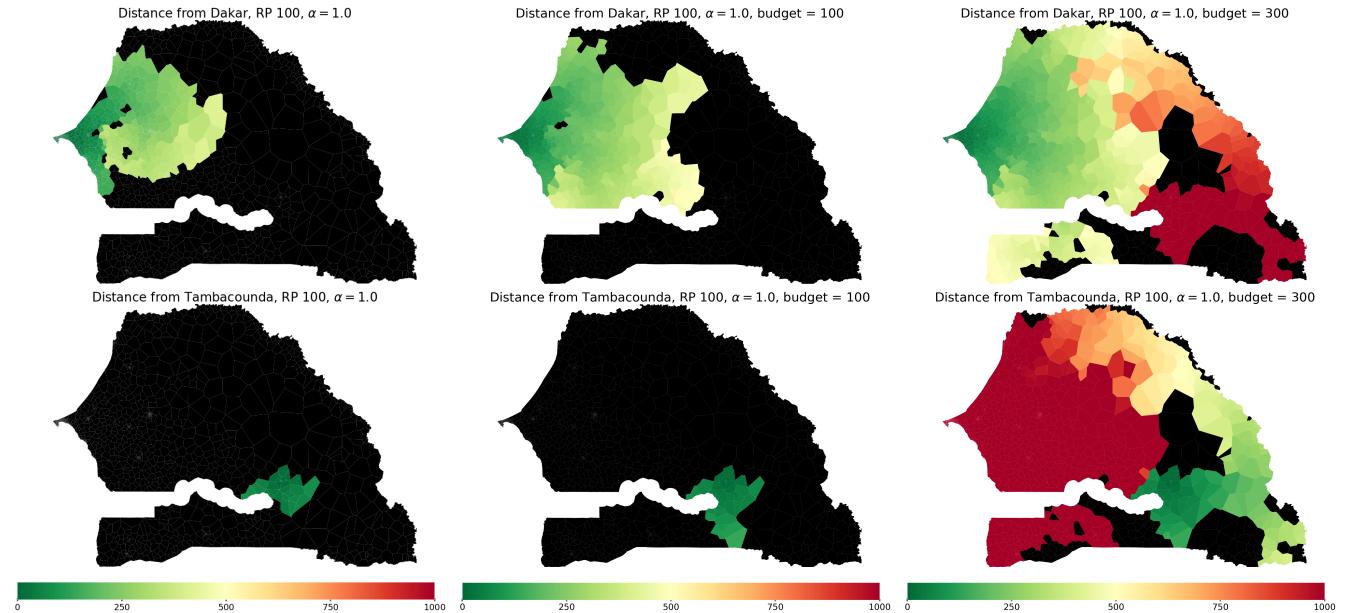


Figure 8: Optimization results for minimizing the number of connected components (top two rows) and for minimizing the percent of infeasible trips (bottom two rows). These results show how the road network becomes more accessible given larger budgets.

and minimizing the average trip distance. In contrast, optimizing with the other three accessibility measures resulted in computationally hard problems, and finding the optimal solution to these problems is not guaranteed. In fact, the reduction in infeasible trips obtained by minimizing the number of connected components was comparable to that obtained by the (sub-optimal) solutions obtained by directly minimizing the number of infeasible trips.

Aside from ease of computability, the different accessibility measures prioritize regions of Senegal differently. This may be of interest to policy makers and government agencies concerned with trade-offs between strategies that benefit the population as a whole versus those that attempt to equitably distribute benefits to more vulnerable regions. For instance, we found that minimizing the number of connected components resulted in greater accessibility to rural regions in the east of Senegal from Tambacounda, whereas minimizing infeasible trips led to a greater focus on accessibility around Dakar and the relatively densely-populated west. This result highlights a need for careful consideration into the intended outcomes of national-scale road network improvements and how these outcomes are reflected in the choices of measures of regional accessibility or vulnerability.

Finally, the framework we present here is flexible enough to be extended and adapted in a number of ways to build climate change resilience into development planning. One obvious refinement of our current approach is to replace the flood risk module with a predictive module that can take weather data as inputs to predict flood extents and depths, which can then be used in combination with climate change models to evaluate flood-related mobility impacts under different climate change scenarios. Vulnerability to other climate change-related hazards, such as droughts or wildfires, can also be incorporated to develop a more complete picture of climate change exposure and risks faced in different regions. The travel demand module could also be adapted to focus on roads and zones within cities, or the accessibility measures could be modified to weigh rural regions more heavily or prioritize routes needed for the distribution of disaster relief resources, for example. Finally, the optimization module can be extended to balance multiple policy objectives, to determine an optimal schedule for road network improvements in a long-term setting, or to find solutions that are robust to uncertainties in flooding scenario, travel demand or repair costs.

ACKNOWLEDGMENTS

Thanks to Orange/Sonatel for access to the mobility data. Thanks to Fathom.Global for access to the flooding data.

REFERENCES

- [1] 2013. Senegal Roads - ArcGIS. (2013). <https://www.arcgis.com/home/item.html?id=7a4486a4903b42c3ba3ce3a750c3bcca> Accessed: 2017-06-22.
- [2] Paul Chinowsky and Channing Arndt. 2012. Climate change and roads: A dynamic stressor-response model. *Review of Development Economics* 16, 3 (2012), 448–462.
- [3] Paul Chinowsky, Amy Schweikert, Niko Strzepek, Kyle Manahan, Kenneth Strzepek, and C Adam Schlosser. 2013. Climate change adaptation advantage for African road infrastructure. *Climatic change* 117, 1-2 (2013), 345–361.
- [4] Bistra N Dilkina and Carla P Gomes. 2010. Solving Connected Subgraph Problems in Wildlife Conservation.. In *CPAIOR*, Vol. 6140. Springer, 102–116.
- [5] David S Johnson, Maria Minkoff, and Steven Phillips. 2000. The prize collecting steiner tree problem: theory and practice. In *SODA*, Vol. 1. 4.
- [6] Asaf Levin. 2004. A better approximation algorithm for the budget prize collecting tree problem. *Operations Research Letters* 32, 4 (2004), 316–319.
- [7] Anna Moss and Yuval Rabani. 2007. Approximation algorithms for constrained node weighted steiner tree problems. *SIAM J. Comput.* 37, 2 (2007), 460–481.
- [8] Orange Data for Development Challenge. 2013. Book of abstract : scientific paper. http://www.d4d.orange.com/fr/content/download/43453/406503/version/1/file/D4DChallengeSenegal_Book_of_Abstracts_Scientific_Papers.pdf. (2013). Accessed: 2017-06-24.
- [9] Christopher C Sampson, Andrew M Smith, Paul D Bates, Jeffrey C Neal, Lorenzo Alfieri, and Jim E Freer. 2015. A high-resolution global flood hazard model. *Water resources research* 51, 9 (2015), 7358–7381.
- [10] SSBN. 2017. SSBN-Global Senegal. <http://www.ssbn.co.uk/ssbnglobal>. (2017). Accessed: 2017-06-25.