MIDAS Bangladesh Application

Supported by SESYNC Pursuit - A forecast of the timing, locations, sequence and likeliest destinations of populations displaced by sea level rise and coastal extremes

Coordinating Author Andrew Bell

This document describes two applications of the MIDAS framework: 1) a calibration of MIDAS to district-level internal migration in Bangladesh, and 2) an application of the calibrated Bangladesh model to project migration flows up to 2100 under additional flooding driven by anticipated sea-level rise, and an estimation of these flood impacts on agricultural and non-agricultural income sources.

Section 1 provides additional details on application-specific submodels not treated in the MIDAS ODD+D document, and describes the calibration procedure and result. Section 2 introduces the additional datasets and submodels used for the forward projections to 2100, and describes the experimental design used to produce our set of simulations.

1 Calibration to Bangladesh inter-district migration

This application simulates the period from 2005 to 2015 as 44 quarterly annual timesteps with an additional (randomized) number of timesteps added to the beginning of the simulation as spin-up time. This document supplements the main MIDAS ODD documentation, describing data and submodels specific to this application and not general to MIDAS. Specifically, it summarizes our approach to estimating the following model inputs:

1. **Initial demographic distribution**
2. **Mortality and fertility**
3. **Additional modeled effects of demography**
4. **Social network structure**
5. **Utility layers**
6. **Moving costs**
7. **Remittance costs**

1.1 Initial demographic distribution

We model demography (age, gender, and location) at simulation start based on age- and gender-disaggregated district-level population data from the 2011 Bangladesh Population Census, made available via IPUMS International (Minnesota Population Center 2015). Specifically, for each agent initialized in a simulation, we assigned their location probabilistically based on census populations by district, then assigned gender probabilistically based on gender-disaggregated census populations within the assigned district, then finally assigned age probabilistically based on age-disaggregated census population within assigned gender and district. Age classification included the groups of 0 years, 1-4 years, 5-9 years and all 5-year groups up to 84, and a final group of 85-100 years.

1.2 Mortality and Fertility

Mortality data was taken from the Global Health Observatory data repository of the WHO (WHO 2015), with age- and gender-disaggregated mortality rates for age classes matching those of the Population Census data used for demographic distribution. Mortality rates are the same in all districts.

Fertility data was taken from the World Fertility Data 2012 dataset produced by the United Nations Department of Economic and Social Affairs (UNDESA 2013). Fertility rates are reported for females in age classes from 15-19 and all 5-year groups up to 49.

1.3 Additional modeled effects of demography

There is considerable flexibility in MIDAS to implement demography-specific effects on decision, structure, etc. In the current simulation, only one age-specific effect is implemented – a gradual increase in discount rate over time, such that older agents will discount the future more highly (and thus value possible returns from a change in portfolio less highly against the upfront costs). Agents' individual discount rates are randomly distributed with distribution parameters included as calibration parameters; however, the factor by which their individual discount rates change over time is not. It is presently a fixed data input (Figure 1), but there is room to explore this and other demographic effects on decisions and behavior in model extensions.

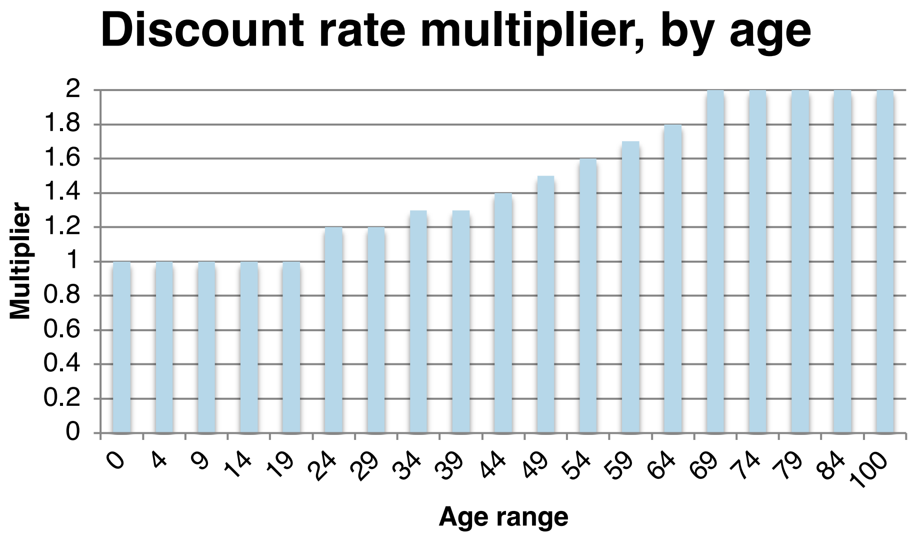

Figure 1: Discount rate multiplier

1.4 Social Network Structure

Social network connections in MIDAS are non-directed links between two agents *i* and *j,* described by a strength between 0 and 1 that influences the degree and likelihood of information and resource sharing in a simulation (see main ODD for more detail). Social network connections decay over time and are strengthened through interaction, via parameters set during calibration.

The network of connections among agents is constructed following a simple model. The number of network connections *n* that an agent *i* 'initiates' is drawn from the normal distribution $N(\mu_n, \sigma_n)$, where both μ_n and σ_n are calibration parameters. This is repeated for all *i*, generating a list of network connections for which one end of the link is defined. The ordering of this list is randomized, and then for each link, an agent *j* is selected probabilistically to be the other end of the link (note – links are not directed in MIDAS). The model used for identifying the likelihood of selection of an agent used in this application develops a weighted score of three factors: shared network connections, physical distance, and shared layers. The shared network connection score is the sum of the strengths of all links agent *i* has with other agents also connected to agent *j*, normalized by the highest raw connection score across all agents *j*. The distance score is calculated from the distance between agent *i* and agent *j*, raised to a polynomial (chosen by calibration) and normalized by all distances to agents *j*, such that shorter distances have nonlinearly higher scores. The shared layer score is estimated by the number of layers in the same place that agent *i* and agent *j* both occupy, again normalized by the highest number of shared layers between agent *i* and any agent *j*. The weights allocated to each of these three scores are also found by calibration.

1.5 Utility Layers

We derived all utility layers for this simulation from three waves (2005, 2010, 2015) of the Bangladesh Household Income and Expenditure Survey (HIES). Utility in timesteps representing the years 2005 to 2015 are developed as described below; spin-up time periods are a repeating cycle of the year 2005.

A complete specification for a utility layer includes i) the utility stream over time, ii) the time commitment of the layer as a fraction of total available time, iii) a classification of the number of openings in the layer as 'fixed' (meaning only a certain number of agents may access the layer – e.g., a limited number of teaching jobs in a district) or 'open' (meaning no restriction on access, though the value of the layer may decline with excess occupation – e.g., too many taxi drivers competing for fares drives down individual earnings) and iv) a cost for initially accessing the layer, which may allow an agent to access the same layer in other places (e.g., passing a certification valid anywhere in a state or country) or only in one place (e.g., buying an immovable asset like a house). We describe below our identification of the set of included income sources as utility layers, and the estimation of these three properties of the included utility layers.

1.5.1 Included layers

We identified all income layers recorded in the survey as:

1. Remittance income
2. Transfer income
3. Rest of other income (including rental income)
4. Annual wage income
5. Annual in-kind income
6. Annual salary income
7. Sugarcane income
8. Jute income
9. Wheat income
10. Boro rice income
11. Aman rice income
12. Aus rice income
13. Oilseed income
14. Pulse income
15. Maize income
16. Income from other crops
17. Total livestock sales
18. Total fish sales
19. Total tree sales

We summed all income sources at the household level, and classified households into four income quartiles (Q4 highest, Q1 lowest).

Next, we selected income layers for inclusion in our model – we chose to exclude income sources 1, 2, and 5 (remittance, transfer, and in-kind) as those sources would be treated through social network interactions in MIDAS, making these layers more appropriate as possible calibration data; additionally, we excluded income sources 17 through 19 (livestock, fish, and tree sales) as at least one of these streams (livestock) appeared to have a larger order of magnitude than other sources, suggesting that they may not be directly interpretable as income (and rather, as revenue).

1.5.2 Utility Stream

We use the following formula for utility or 'value' for layer i in a particular place j at time t:

$$V_{ijt} = V_{base,ijt} \cdot \frac{m \cdot n_{expected,ij}}{\left(max\left(0, n_{actual,ij} - m \cdot n_{expected,ij}\right) \cdot k + m \cdot n_{expected,ij}\right)} \quad (1)$$

where $V_{base,ijt}$ is a base rate for the utility value estimated from reported income, as described below; $n_{expected,ij}$ is the expected number of agents occupying the layer, obtained by scaling the number of observed occupants of a layer in the HIES by n_{agents} / n_{HIES}, the ratio of the number of agents in the simulation to number of people in the HIES sample; $n_{actual,ij}$ is the number of agents actually occupying the layer; and m and k are parameters that shape how utility declines as the number of actual occupants exceeds the number of expected occupants. All agents occupying layer i in a particular place j at time t derive the same value from the stream, but have agent-specific coefficients specifying their utility on a unit of value from a particular source.

To develop the values $V_{base,ijt}$ for the included layers, we used our initial classification of households into income quartiles to identify average earnings per layer by members of a particular quartile in a particular district (e.g., what do Q1 households earn on average in sugarcane income in Bogra?). Our dataset included values for each of the years 2005, 2010, and 2015 for each of 13 included income sources, each broken into four quartile layers, for each of 64 districts. We used Matlab's *interp* function to estimate the interior annual averages 2006-2009, 2011-2014.

At this point we have an array of 52 layers x 64 districts x 11 years. However, this application of MIDAS uses a quarterly annual time periods (i.e., four time periods per annual cycle, and 44 time periods in 11 years), so that annual income must be attributed to one or more of these steps. We assumed income sources 3, 4, 6, and 16 to be spread equally across all 4 quarterly periods (P), and consulted an agricultural calendar for Bangladesh (BBS 2017) for approximate harvest periods for crop income, such that the fraction of annual income earned in each quarter, for each layer, is as follows:

Table 1: Annual income spread across quarterly periods, by income source

	Rest of other income	Wage income	Salary income	Sugarcane income	Jute income	Wheat income	Boro rice income	Aman rice income	Aus rice income	Oilseed income	Pulse income	Maize income	Income from other crops
P1	0.25	0.25	0.25	1	0	1	0	0	0	0.25	1	0	0.25
P2	0.25	0.25	0.25	0	0	0	1	0	0	0.25	0	1	0.25
P3	0.25	0.25	0.25	0	1	0	0	0	1	0.25	0	0	0.25
P4	0.25	0.25	0.25	0	0	0	0	1	0	0.25	0	0	0.25

We spread each annual average across four periods according to the table above, giving us the complete array V_{base} of size 52x64x44.

Each simulation includes a spinup time of random length (8 to 20 periods in this application); these are treated as repeated cycles of the year 2005.

1.5.3 Time constraints

We estimated time constraints specific to layers from each income quartile, for each place, in a two step procedure.

First, we identified quarters in which income sources would demand an agent's time as follows: we assumed income sources 3, 4, 6, and 16 to take time across all four quarterly periods, and consulted an agricultural calendar for Bangladesh (BBS 2017) for approximate sowing and harvest periods, such that quarters in which a layer demands time from an agent, for each layer, is as follows:

Table 2: Quarterly periods requiring effort, by income source

	Rest of other income	Wage income	Salary income	Sugarcane income	Jute income	Wheat income	Boro rice income	Aman rice income	Aus rice income	Oilseed income	Pulse income	Maize income	Income from other crops
P1	1	1	1	1	1	0	1	0	0	1	1	1	1
P2	1	1	1	1	1	0	1	0	1	1	0	1	1
P3	1	1	1	1	0	0	0	1	1	1	0	0	1
P4	1	1	1	1	0	1	0	1	0	1	1	1	1

Second, we calibrated time constraints such that income diversity for agents in each income quartile in the simulation could match that observed in the HIES. Generally speaking, respondents in higher income quartiles in the HIES tended to have more diversified incomes (Figure XXX), with the poorest quartile having 0 or 1 sources on average, and the richest quartiles having 3 or 4.

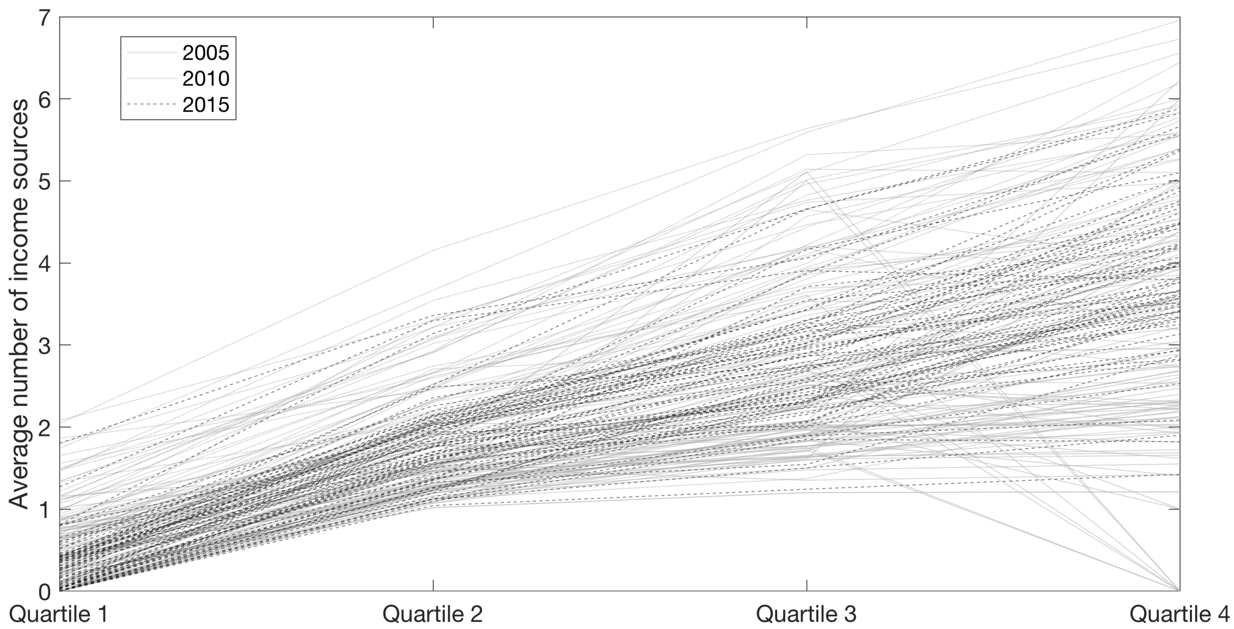

Figure 2: Average number of income sources, by income quartile, in the HIES. Each line represents one district in one wave (2005, 2010, or 2015) of the HIES.

We used the data displayed in Figure 2 to estimate appropriate time constraints for each district, averaging across survey waves to generate one set of average income source counts for each quartile *i*. We specified that for each income source, the *i-1th* income quartile layer was a **prerequisite** (see main ODD for explanation of prerequisites) for the i^{th} layer (e.g., to be a Q3 boro rice earner, an agent must also be occupying Q1 and Q2 for boro rice), and estimated time constraints such that the cumulative time constraint of layers 1 through *i* would allow an agent to occupy the expected number of layers at income quartile *i,* with a small amount of random noise added to the calculation. Consider the following example:

Table 3: Example time constraint calculation along quartiles from same income source

Quartile	Additional time constraint	Cumulative time constraint	Maximum possible layers within time constraint (< 1)
Q1	0.95	0.95	1
Q2	-0.475	0.475	2
Q3	-0.15833	0.31667	3
Q4	-0.0792	0.23747	4

In this example, the first-quartile layer for the income source has a time constraint of 0.95 (95% of an agent's time), such that it is the only layer the agent can occupy. The second-quartile layer of the same source has a marginal time constraint of -0.475, such that the combined time constraint is 0.475, allowing up to 2 layers of similar time constraint to be occupied. In the agricultural case, this might represent an agent purchasing more property but also mechanizing and capturing economies of scale (in both time and income, perhaps).

1.5.4 Fixed and open layers

We specified all agricultural layers as 'open,' meaning that there were not a defined number of slots, but that the earned income from each layer may decline (depending on the values *m* and *k* in Eqn. 1) as the number of agents occupying the layer exceeded those expected (i.e., proportional to the HIES). We specified the layers 'rest of other income' (which we believe to largely be property income), wage income, and salary income as 'fixed,' with a hard number of slots that could not be exceeded:

Table 4: Fixed or open layers, by income source

	Rest of other income	Wage income	Salary income	Sugarcane income	Jute income	Wheat income	Boro rice income	Aman rice income	Aus rice income	Oilsed income	Pulse income	Maize income	Income from other crops
Fixed?	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No

For layers that are fixed, agents have a belief about the likelihood of actually gaining a slot, which allows them to estimate the expected utility from that layer. This belief is updated through their social networks (i.e., I heard from my cousin that company X in place Y was hiring; see main ODD for mechanism); agents may choose to migrate based on an expected utility, but in the end not gain access to the layer.

1.5.5 Access Costs

Access costs are coded in MIDAS as a list of discrete cost items that may be associated with one or more layers; in turn, layers may be associated with one or more access costs. This allows representation, for instance, of the cost of multiple certifications in order to be able to practice law in a particular place (and the additional costs at the margin to transfer those credentials to a new place). It also allows representation of immobile assets such as properties, by specifying unique and separate access costs to access a 'property' layer in different places, for example.

In the current application, we include only income streams, but make the following assumption – that the costs of accessing sources 4 and 6 (wage and salary income) allow agents to access those sources in all parts of Bangladesh; i.e., they are transferrable qualifications. All other sources represent earnings from assets (properties in the case of source 3; farm land in all other cases), which we treat as immovable. **The present application does not include the possibility of selling assets to facilitate a move, though this is a possible area for extension if model findings indicate that it is a relevant constraint.**

Rather than try to identify from an additional data source the likely costs of purchasing land, buildings, or accessing wage and salary levels – which could be unavailable or inconsistent – we applied a simple model to generate plausible access costs for each layer, based on an underlying assumption of a reasonable expected rate of return on investment. The fraction of an investment that must be recovered each year of a project in order to break even, known as the uniform series capital cost recovery factor, is defined as:

$$\frac{A}{P} = \frac{i \cdot (1+i)^n}{(1+i)^n - 1} \quad (2)$$

Where A is the annual return in each of n project years, P is the upfront cost, and i is the discount rate. We leverage this model to estimate the reasonable investment costs to earn a return on investment r as:

$$P = A \cdot (1+r) \cdot \frac{(1+i)^n - 1}{i \cdot (1+i)^n} \quad (3)$$

The value of P is used as the cost of accessing a layer with average annual return A. Values for r, i, and n are estimated as calibration parameters, with the same values applied to all layers in the simulation.

1.6 Moving costs

Moving costs were estimated by calibration, as specified in Section 1.8.

1.7 Remittance costs

Remittance costs were estimated by calibration, as specified in Section 1.8.

1.8 Calibration procedure

We applied an Approximate Bayesian Calibration (ABC) approach over 65 model parameters of varying scope (Table 5), calibrating against the relative inter-district migration rate Q, with flows weighted by the populations of both source and destination districts, as:

$$SSE_Q = \frac{\sum_i \sum_j w_i w_j \left(Q_{ij,modeled} - Q_{ij,observed}\right)^2}{\sum_i \sum_j w_i w_j}$$

where i and j are the source and destination districts, and w is population. We conducted 4 rounds of Monte Carlo simulation across the 65 parameters, with parameter values drawn from a uniform distribution between some minimum and maximum value. At the end of each round, we selected the highest 1% of calibration results, and used the minimum and maximum values of each calibration parameter observed in this set to define the bounds of the uniform distribution for the following round. In total, we conducted 5,373 simulations in Round 1; 2,969 simulations in Round 2; 3,547 simulations in Round 3; and 3,565 simulations in Round 4 (variation across rounds due to priority and speed in the high-performance cluster). The final 1% of calibrations have weighted Pearson Correlations (for relative migration rates) ranging from 0.1752 to 0.2561.

By comparison, in a separate exercise in which income values for the same set of utility layers were replaced with random numbers (i.e., a null model), the same weighted Pearson Correlation had a maximum of 0.0326.

Table 5 - Calibration Parameters

Scope	Parameter	Short Description
model	spinupTime	Time steps run before 'time 0' in simulation, for agents to develop expectations about utility layers
	numAgents	Number of agents in simulation
	utility_k	Utility model parameter from Section 1.5.2
	utility_m	Utility model parameter from Section 1.5.2
	utility_noise	Random noise in utility model from Section 1.5.2
	utility_iReturn	Access cost model parameter from Section 1.5.5

	utility_iDiscount	Access cost model parameter from Section 1.5.5
	utility_iYears	Access cost model parameter from Section 1.5.5
	creditMultiplier	Scalar multiple of paid access costs that gives proxy for access to credit, in Submodel 5
	remitRate	Fraction of remitted resources taken as fee
map	movingCostPerMile	Costs per mile to move between two places
	minDistForCost	Shortest distance at which costs accrue
	maxDistForCost	Greatest distance to which costs rise
network	connectionsMean	Mean and standard deviation in number of network connections per agent
	connectionsSD	
	weightLocation	Weight of sharing same location in likelihood of new link forming
	weightNetworkLink	Weight of having shared network connections in likelihood of new link forming
	weightSameLayer	Weight of occupying the same utility layer in likelihood of new link forming
	decayPerStep	Amount by which a network link strength decays in each timestep
	interactBump	Amount by which a network link strength increases due to social interaction
	shareBump	Amount by which a network link strength increases due to sharing of resources
	incomeShareFractionMean	Mean and standard deviation in fraction of new income shared across an agent's social network each time step
	incomeShareFractionSD	
agent	shareCostThresholdMean	Mean and standard deviation in the fraction of overall remittance that is lost to fees, above which agents choose not to share
	shareCostThresholdSD	
	interactMean	Mean and standard deviation in likelihood that agent participates in interactions in a time step
	interactSD	
	meetNewMean	Mean and standard deviation in probability that agent will meet a new agent during a time step
	meetNewSD	
	probAddFitElementMean	Mean and standard deviation in likelihood that algorithm to generate new utility layer portfolios for consideration will try to add an additional layer to a portfolio
	probAddFitElementSD	
	randomLearnMean	Mean and standard deviation in likelihood that agent will randomly learn new information about utility
	randomLearnSD	

		layers outside of a social interaction during a time step
	randomLearnCountMean	Mean and standard deviation in the number of pieces of information that an agent will randomly learn, if the agent learns outside of a social interaction
	randomLearnCountSD	
	chooseMean	Mean and standard deviation in the likelihood that an agent will reconsider their portfolio during a time step
	chooseSD	
	knowledgeShareFracMean	Mean and standard deviation in the fraction of knowledge about utility layers that an agent will share during an interaction
	knowledgeShareFracSD	
	bestLocationMean	Mean and standard deviation in the number of high-utility locations an agent will keep in memory after considering them
	bestLocationSD	
	bestPortfolioMean	Mean and standard deviation in the number of high-utility portfolios an agent will keep in memory for each location it considers
	bestPortfolioSD	
	randomLocationMean	Mean and standard deviation in the number of locations an agent will randomly select for consideration when evaluating new portfolios
	randomLocationSD	
	randomPortfolioMean	Mean and standard deviation in the number of portfolios an agent will generate randomly in each location when evaluating new portfolios
	randomPortfolioSD	
	numPeriodsEvaluateMean	Mean and standard deviation in the number of time periods an agent uses to consider future utility stream when comparing portfolios
	numPeriodsEvaluateSD	
	numPeriodsMemoryMean	Mean and standard deviation in the number of past time periods agent keeps in memory
	numPeriodsMemorySD	
	discountRateMean	Mean and standard deviation in agent-specific discount rate
	discountRateSD	
	rValueMean	Mean and standard deviation in agent-specific constant relative risk aversion coefficient
	rValueSD	
	bListMean	Mean and standard deviation in agent-specific coefficients on utility in different forms (income, etc.)
	bListSD	
	prospectLossMean	Mean and standard deviation in agent-specific disutility of loss relative to utility of gain
	prospectLossSD	
	informedExpectedProbJoinLayerMean	Mean and standard deviation in agent’s expectation to be able to join a layer, given that they have information about job openings
	informedExpectedProbJoinLayerSD	
	uninformedMaxExpectedProbJoinLayerMean	Mean and standard deviation in agent’s expectation to be able to join
	uninformedMaxExpectedProbJoinLayerSD	

		a layer, given that they do not have information about job openings
	expectationDecayMean	Mean and standard deviation in the rate at which expectation about job openings decays per time step
	expectationDecaySD	

2 Projections of inter-district migration to 2100 under anticipated sea-level rise

We conducted a set of experiments using our calibrated model application to estimate the effects of flooding (via impacts to income) on inter-district migration out to 2100 under three of the representative concentration pathways (RCPs; 2.6, 4.5, and 8.5).

We used exceedance maps of annual peak flooding for Bangladesh at decadal intervals out to 2100 (likelihoods of annual flooding exceeding depths of 0.1m to 5m in increments of 0.1m) to instantiate plausible annual peak floods for Bangladesh. For each year in a simulation we took one random uniform draw between 0 and 1, and used this number to estimate the peak flood depth in each district. Draws were independent between years, so that inter-district flooding is highly correlated within a year but any autocorrelation over time occurs only due to correlation in likelihoods.

We predicted the future economy for Bangladesh out to 2100 by randomly drawing one of our three periods of available income data (2005, 2010, and 2015) with replacement at five-year intervals from 2020 to 2100, interpolating for the years in-between. We used a damage function for flood depth on agricultural and non-agricultural income derived from Quisumbing and Mueller (2011), as follows:

	% drop in income per foot of flood shock, same year	% drop in income per foot of flood shock, following year
Agricultural income	45	2.5
Non-agricultural income	0	13.8

Because this damage function is defined based on the relative concept of ‘flood shock,’ and not the absolute concept of flood depth, we included a parameter for the ‘normal flood depth’ (as a fraction of the statistical expected flood depth) in order to calculate a flood shock (as total flood minus normal flood depth). The value of this parameter is unknown, so that our simulation experiments included a sensitivity analysis on this parameter – a random value between 0 and 1.5 (e.g., up to 150% of statistical expected flood depth) was drawn for every simulation run.

We conducted a total of 871 simulation runs across RCP 2.6 (255 runs), RCP 4.5 (294 runs), and RCP 8.5 (322 runs). Relative importance of parameters in shaping migration outcomes was estimated using Matlab’s treebagger algorithm using forests of 1000 randomly generated regression trees; parameter importance is estimated by the relative increase in prediction error when that parameter is excluded from the generation of trees, compared to the error rate when it is included in the generation of trees.

Literature Cited

BBS. 2017. “Yearbook of Agricultural Statistics-2016.” Dhaka, Bangladesh.

Minnesota Population Center. 2015. “Integrated Public Use Microdata Series, International: Version 6.4 [Dataset].” Minneapolis, MN: University of Minnesota. https://doi.org/10.18128/D020.V6.4.

Quisumbing, Agnes, and Valerie Mueller. 2011. “How Resilient Are Labour Markets to Natural Disasters? The Case of the 1998 Bangladesh Flood.” *Journal of Development Studies* 47 (12): 1954–71. http://www.tandfonline.com/doi/abs/10.1080/00220388.2011.579113.

UNDESA. 2013. “World Fertility Data 2012.” 2013.

WHO. 2015. “Global Health Observatory Data Repository.” 2015.