A Data-Driven Analysis of Environmental Migration in Coastal Bangladesh

By

Kelsea Best

Thesis
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in fulfillment of the requirements
for the degree of

MASTER OF SCIENCE
in
Earth and Environmental Sciences

August 9, 2019

Nashville, TN

Approved:

Jonathan Gilligan, Ph.D.

John Ayers, Ph.D.

Acknowledgements

I would like to thank the Vanderbilt Department of Earth and Environmental Sciences for their continued support. I would especially like to thank my advisor, Dr. Jonathan Gilligan for his patience, support, and brilliant guidance. Thank you as well to Dr. Hiba Baroud and Dr. David Furbish for their wisdom and inspiration. I would also like to thank Dr. Amanda Carrico, Dr. Brooke Ackerley, Dr. Katharine Donato, and Dr. Bishawjit Mallick for allowing me to use their datasets and for helping me interpret my results. I am also extremely grateful to my fellow graduate students, especially Jess Raff and Chris Tasich for cheering me on always and making this process so much more fun and enjoyable.

I would like to thank our Bangladeshi colleagues for guiding us through their amazing country. It is truly an honor to be able to work in Bangladesh and to learn from the people there.

Finally, and most importantly, I would like to thank my family. Thank you to my mother, who has believed through every stage of my life, and whose wisdom I rely on daily. Thank you to Trey, Jen, Tony, and Caison, the best siblings anyone could ask for. Thank you to Shayan and Henry for your love.

Table of Contents

List of Tables

List of Figures

Introduction

Future climate change poses a wide variety of threats to human health and well-being ("IPCC - SR15"). This is especially true in low-lying coastal communities, where climate change is likely to affect a variety of natural phenomena including storms, sea level rise, coastal inundation, erosion, and precipitation (Nicholls et al. 2007). In addition, climatic changes and environmental stressors will impact livelihood opportunities in vulnerable coastal areas (Nicholls et al. 2008).

One possible response to climate change and other environmental stresses that has been studied is migration. Discussions of climate-induced refugees have traditionally been framed around a looming crisis of "climate refugees" (Myers et al. 2002). However, this narrative has been challenged as lacking complexity. Recent work has shown that, although climate change and environmental pressure can affect population mobility, those impacts may be nonlinear or even negative (Call et al. 2017; Paul 2005). Additionally, environmental factors are rarely the only causes of migration (Obokata et al. 2014). Rather, migration is complex, multi-causal phenomenon that is impacted by both "push" factors such as political instability, lack of economic opportunity, and lack of natural resources in the location of origin, as well as "pull" factors related to the destination location including availability of employment, resources, and social capital. Intervening factors such as transportation networks, social ties, and cultural norms can further complicate the decision to migrate (Amrith 2013; Black et al. 2011).

Similarly, migration is only one of many possible human responses to environmental change. As concern for community displacement increases, it is important to understand the factors that impact migration and what role migration might play in adaptation to environmental stress. Communities on the coastal delta plain of Bangladesh face a particular set of challenges as sediment transport, deposition, and erosion continuously create and destroy land and shape the areas where people live and work (Tessler et al. 2015). Additionally, both natural and human changes to the environment are causing shifts in livelihood choices (Ackerly et al. 2015; Tessler et al. 2015).

The complexity of human migration poses a challenge for researchers who aim to study the effects of environmental changes on population mobility (McLeman 2013). Questions remain about how to best model human migration to account for this complexity, as well as how to obtain appropriate and accurate data to test these models (Neumann and Hilderink 2015). Additionally, it is often difficult or impossible to isolate one driving factor of migration and to control for all other variables, especially in such a complex system. To address this, researchers who study migration will often use expert judgement or theory to select variables to assess. Though this can provide insights into how specific drivers might impact migration decisions, it does little to explain what the most important variables that drive decisions might be, especially when considering nonlinear dynamics and relationships between variables.

As researchers continue to collect large amounts of data with household surveys, challenges may arise in how best to analyze such datasets, especially where motivating theories are unclear or conflicting. To advance the study of environmental migration, especially as large datasets and surveys become more readily available, new methods will need to be employed (Neumann and Hilderink 2015). This work aims to address this need by applying machine learning, specifically random forests, to social survey data for the study of environmental migration in Bangladesh. Random forest is a machine learning approach that has been shown to perform well in environmental and ecological contexts (Cutler et al. 2007; Prasad et al. 2006). However, reviews of methodologies used in studying environmental migration did not mention machine learning techniques (Piguet 2010) and to our knowledge, our application of random forest methods to the topic of environmental migration is novel.

In this work, I present the application of random forests to determine the importance of each covariate in a large dataset for predicting migration outcomes (**Objective 1**). Though random forest models are able to identify drivers of migration, there exists a tradeoff between high predictive ability and low interpretability. To address this tradeoff, random forests and other complex machine learning algorithms may be especially useful in combination with more traditional, simpler methods. I conduct a survival analysis of household time to first migration using the important variables identified by the random forest algorithm, which provides deeper insight into how important variables impact mobility (**Objective 2**).

This mixed-methods approach of random forest models and survival analysis provides a data-driven approach to identifying and further investigating key variables that impact migration from social datasets.

To address this overall research goal, this thesis is divided into two distinct objectives:

Objective 1: Use random forests to determine the importance of covariates in a large dataset for predicting migration outcomes.

Research questions:

1. Are random forest models useful in identifying salient variables and nuances in two sets of ethno-survey data related to migration in Bangladesh?
2. Can random forest models improve predictive ability of migration, compared to other machine learning techniques?

Objective 2: Conduct a survival analysis of household time to first migration using the important variables identified by the random forest algorithm to provide deeper insight into how important variables impact mobility.

Research questions:

1. Is applying survival analysis to salient variables valuable in combination with random forest models to identify which factors contribute to an increase or decrease in household mobility?
2. Do livelihood variables positively impact the probability of a first migration trip, indicating that a certain level of wealth is necessary for migration?

Chapter 1

Literature Review

1.1 Environmental migration

Environmental migration is a rich topic of research because of concerns surrounding how future climate change and environmental stress might impact population mobility, as well as the potential implications for policy (Geddes et al. 2012). As mentioned, migration is a complex and multi-causal phenomenon that is driven by factors such as environmental stress and economic opportunity, both of which may be affected by future climate change (Perch-Nielson et al. 2008). Even where the environment drives migration, it can be compounded by social, economic, and political factors (Hunter et al. 2015; Walsham 2010).

A popular conceptual framework of environmental migration that highlights its complexity was proposed by Black et al (2011). The framework identifies economic, political, social, demographic, and environmental factors as the primary drivers that affect migration decisions. The unique contribution of the framework is that the effect of environmental drivers on a migration decision is dependent on the other factors and the context of the decision. In this way, environmental conditions can directly impact a migration decision, but also impact the decision indirectly through effects on the economic, social, political, and demographic drivers (Black et al. 2011).

In this work, Black et al. also highlight the existence of both "push" and "pull" factors that impact a migration decision. "Push" factors are those that may drive someone to leave their current location, such as political instability, lack of economic opportunity, and lack of natural resources in the location of origin. Alternatively, "pull" factors are positive conditions related to the destination location such as availability of employment, resources, and social capital, that may draw an individual to that destination. Intervening factors such as transportation networks, social ties, and cultural norms can further complicate the decision to migrate (Black et al. 2011). This approach of categorizing push, pull, and intervening factors has been adopted in various attempts to model environmental migration (Hassani-Mahmooei 2012; Henry et al. 2003).

One widely explored aspect of Black's framework is the impact of livelihood on the decision to migrate. In many areas, migration is a livelihood strategy to access work opportunities outside of the community (Gioli et al. 2014). Where environmental conditions negatively impact livelihood, migration may be a way to increase livelihood resilience (Tanner et al. 2015). This kind of migration may also be important for socioeconomic status of the community of origin, as migrants may send remittances back home (Gioli et al. 2014; Massey 1990). These remittances can be an important source of wealth and can also increase resilience for communities of origin (Warner et al. 2009; Redehegn et al. 2019). Additionally, migration may cause shifts in local market incentives and may work to reduce local income inequality (Shayegh 2017). However, in other cases, the influx of remittances into select households may increase inequality in a community (Black et al. 2005; Massey 1990).

In line with the importance of environmental impacts on livelihood, environmental effects on agriculture are thought to be an important driver of migration, especially in rural, agrarian communities (Cai et al. 2016; Dillon et al. 2011; Feng et al. 2010; Galvin 2009). In a study of climate variability and migration in the Philippines, Bohra-Mishra et al. find that temperature and typhoons have a larger impact on migration than precipitation, which they explain by the larger effects on rice yields (2017). The result was that temperature and typhoons caused a larger outmigration from regions that were more rural and agriculturally dependent (Bohra-Mishra et al. 2017). On a global level, Cai et al. found that temperature caused more international outmigration only in highly agriculture-dependent countries, reflecting the impact of temperature on agricultural productivity (2016).

The question of environmental migration can be further complicated when considering differences between temporary and permanent migration, as well as internal versus international. Some research has suggested that most cases of environmental migration are temporary and internal, but this also seems to be highly situational (Gray 2011; Bardsley and Hugo 2010). In another study, Bohra-Mishra et al. focus on permanent environmental migration, and find that climate variations have an impact on permanent migrations while natural disasters have minimal or no effect (2014). In some existing literature, researchers choose to focus specifically on internal or international migration (Willekens et al. 2016). For example, Donato and Massey specifically focus on how avoiding threats from climate change may increase illegal international migration from the world's poorest nations (Donato and Massey 2016). Obakata et al. provide a detailed review of research on international environmental migration (2014).

One reason that environmental migration is important to understand, is its potential benefit as a household or community adaptation strategy to environmental stress (Bardsley and Hugo 2010; Hunter et al. 2015; McLeman and Hunter 2010; McLeman and Smit 2006). As one example, Bardsley and Hugo claim that migration is an important aspect of a community's adaptation response to climate change or environmental degradation (2010). They claim that thresholds of tolerable environmental stress exist, after which migration will be a crucial adaptation strategy (2010). Though this may seem similar to the idea of displacement or forced relocation as a response to extreme environmental stress, this framing suggests that migration is actually a positive and proactive strategy to build resilience and counter declines in livelihood.

If environmental migration is a positive adaptation strategy, then a larger concern than mobility could be immobility, or "trapped" populations (Chen et al. 2017). While wealthier households or communities with access to more natural resources may better be able to incur costs associated with migration, poorer communities may not. This area of research suggests that lack of mobility is a major challenge for especially vulnerable communities to adapt to their environmental conditions, especially in the case of a loss of livelihood (Adger et al. 2015; Bennett et al. 2011; Hunter et al. 2014). This also points to the importance of disaster response activities in impacted areas that provide support and economic opportunities for those who are not able to move.

In contrast to the literature which considers migration as a possible adaptation strategy, there is also a body of literature that focuses on population displacement, forced relocation, or managed retreat as a result of challenging environmental or climatic conditions (Black et al. 2013; Sherbinin et al. 2011; Gray et al. 2014; Hino et al. 2017). This literature may use the term "climate refugees" to describe a possible future wave of environmentally displaced people (Biermann and Boas 2010; Piguet 2013). However, the term "climate refugees" is more often considered to be overly simplistic, as there is likely never a single driver of migration, even in the case of extreme environmental change (Black et al. 2013).

Black et al. discuss the implications of environmental change on population displacement, even while acknowledging the complexity of migration (2013). In this work, Black et al. perform a review of research assessing the impacts of weather-related extreme events on population mobility. In this work, Black et al. emphasize the importance of distinguishing between migration, displacement, and immobility in such research, all of which may be a result of an environmental event (2013). In this way, they further complicate and challenge the idea of "climate refugees". Taking it a step further, the authors claim that environmental migration in some cases may represent a failure of adaptation to environmental conditions, rather than a positive adaptation (Black et al. 2013).

In another study, Gray et al. use the example of Sumatra after a 2004 tsunami to study population displacement after a natural disaster (2014). The authors use survey data from respondents both before and after the tsunami, as well as satellite imagery of the area, to assess vulnerability to what the authors call "post-tsunami displacement" (Gray et al. 2014). This work suggests that other factors that influence migration normally, such as gender and age, may be less important in the case of disaster-induced displacement (Gray et al. 2014).

In general, there is little agreement between studies in terms of how different environmental impacts will affect migration. For instance, some scholars indicate that sea level rise could be a significant driver of migration (Hauer 2017), while others suggest that, even in vulnerable island nations such as the Maldives, sea level rise does not play a major role in influencing migration decisions (Speelman et al. 2017). In general, research has suggested that large variations in temperature have a positive impact on migration, but precipitation has a largely mixed effect (Henry et al. 2004; Marchiori et al. 2012; Mastorillo et al. 2016; Mueller et al. 2014; Thiede et al. 2016). Gray and Mueller found that drought in Ethiopia increases men's labor migration but decreases migration for women associated with marriage (2012b), while Findley found that drought increased migration for women and children in rural Mali (1994). Additionally, Hunter et al. found that drought increased migration only in communities in Mexico with strong migration histories, but decreased migration in communities without such networks of previous migration (2013).

One additional challenge in studying environmental migration is that the findings vary significantly by location (Gray and Wise 2016). Specific research has focused on climate variability in South America (Thiede et al. 2016), drought in Ethiopia (Gray and Mueller 2012b), land use in Ecuador, (Gray and Bilsborrow 2014), heat stress in Pakistan (Mueller, Gray et al. 2014), soil quality in Kenya (Gray 2011), tsunamis in Sumatra (Gray et al. 2014), weather anomalies across Africa (Marchiori et al. 2012), and Hurricane Katrina in New Orleans (Fussell et al. 2010) to name a few.

Beyond differences in location and type of environmental change, the literature has also explored implications of gender (Farris 2010; Gray and Mueller 2012a; Miletto et al. 2017), ethnicity (Massey et al. 2010), legal status (Donato et al. 2016), norms (De Jong 2000), social networks (Haug 2008; Hunter et al. 2013), satisfaction with local services (Dustmann and Okatenko 2014), and risk attitudes and perceptions (Hunter 2005; Jaeger et al. 2010; Koubi et al. 2016) on environmental migration. Again, the results of these analyses are highly mixed. As one example, a significant amount of work has suggested that the social networks can positively impact rates of environmental migration by decreasing uncertainty and lowering hurdles that inhibit migration (Haug 2008). Other work has found that social networks are not significant as determinants of migration behavior, while family norms were (De Jong 2000).

Just as the drivers, rates, and magnitudes of environmental migration are complex, the implications are similarly multifaceted. Some research emphasizes the importance of studying environmental migration in the context of the pressure that rural to urban migration may put on urbanization (Barbieri and Carr 2005; Barrios et al. 2006; Sherbinin et al. 2012; Qin 2010). The connection between environmental migration and urbanization could have important implications for urban planning, but the linkage is still highly dependent on location. For instance, Suckall et al. conclude that climate change would not contribute to increased urbanization in Malawi, but could actually contribute to the reverse (2015). Similarly, Cattaneo and Peri find that increases in temperature can drive migration to urban areas in middle-income countries, but reduce urban migration in poor countries due to limited resources (Cattaneo and Peri 2016).

Yet another body of work is interested in what is known as the "climate change, migration, and conflict nexus." (Burrows and Kinney 2016). This work focuses on the possible connection between climate-induced migration and violent conflict (Reuveny 2007). This theory suggests that an influx of migrants into urban areas due to climate variations may lead to increased competition for resources, and if government response is not adequate, violent conflict may arise (Abel et al. 2019). For instance, Abel et al. explores the impacts of drought on the climate change, migration, conflict nexus, but ultimately determine that the causal connection is highly dependent on region, time, and democratic indicators (2019).

Even within the community of researchers who study environmental migration, there is not a consensus surrounding language to describe possible impacts of environmental change on population mobility. Thornton et al. used an online survey to assess perceptions of environmental migration amongst more than 260 professionals around the world who work on the issue (2018). Their results show a wide range of conceptualizations of environmental migration, as well as a range of policy concerns. As one example, Thornton et al. found that the most preferred term to describe the scenario of interest was 'migration' (38%), followed by 'displacement' (20%), 'mobility' (19%), 'refugee' (7%), 'relocation' (3%), 'resettlement' (1%), and 'other' (11%). In many ways, this lack of consensus between experts on even how to refer to changes in population mobility due to environmental conditions highlights the complexity of the problem.

1.2 Environmental migration in Bangladesh

In Bangladesh, migration has been a way of life for centuries, as well as a common method of livelihood diversification and adaptation to stressful natural conditions (Alam et al. 2017; Amrith 2013; Black et al. 2005; Bryan et al. 2014; Lagakos et al. 2018; Martin et al. 2014). Bangladesh also faces a variety of challenging environmental conditions, which are discussed in more detail in Chapter 2. As a result, environmentally induced migration has also been widely studied in Bangladesh (Afsar et al. 2003; Ahsan et al. 2011; Call et al. 2017; Chen and Mueller 2018; Donato et al. 2016; Gray and Mueller 2012a; Islam 2017; Joarder and Miller 2013).

Much of this research focuses on extreme weather events representing rapid onset environmental change, such as cyclones (Gray and Mueller 2012a; Kartiki 2011; Lu et al. 2016; Mallick and Vogt 2012). For example, Gray and Mueller investigate the consequences of flooding and crop failure due to extreme weather events on long-term population mobility in rural Bangladesh (2012a). They use longitudinal survey data from 1,7000 households over a 15-year period, analyzed with multivariate event history models to estimate the effects of flooding and crop failure due to extreme weather events on long-distance migration. Their results indicate that flooding has a slight effect on mobility, while crop failures have a strong effect (Gray and Mueller 2012a).

In another study, Mallick and Vogt assess "disaster-induced population displacement" in the context of the 2009 cyclone Aila in Bangladesh (2014). They use a survey of 280 individuals from 12 villages in the southwestern part of Bangladesh in a period after Aila. Their results suggest that male household members migrated towards cities in order to access livelihood opportunities after the cessation of emergency aid (Mallick and Vogt 2014). This work highlights the gendered dynamics of migration, as well as the importance of livelihood. Similarly, Paul found that disaster-induced migration was prevented after a 2004 tornado in northern Bangladesh due to the effective distribution of emergency aid in effective regions (2005).

Other research considers slower onset environmental change such as salinity encroachment, temperature change, and precipitation (Call et al. 2017; Chen and Mueller 2018; Perch-Nielsen et al. 2008). Call et al. studied the impacts of temperature, precipitation, and flooding on temporary migration in a non-coastal area, Matlab, Bangladesh (2017). Their work showed that temporary migration declines immediately after a flood, but quickly recovers, while high temperatures consistently increase temporary migration, and precipitation has a strongly non-linear effect on migration rates (Call et al. 2017). This work supports other research that has indicated that environmental stress could actually decrease population mobility and limit the effectiveness of migration as an adaptation strategy, leading to potentially "trapped" populations by limiting the resources necessary for mobility (Adger 2015; Bennet and Beddington 2011). Call et al. further conclude that climate change is likely to disrupt existing migration patterns in Bangladesh rather than cause mass displacement in the case of non-coastal communities (2017).

Most recently, Chen and Mueller found that salinity encroachment into soil could be a powerful driver of migration within Bangladesh due to impacts on agriculture and associated loss of livelihood (Chen and Mueller 2018). Interestingly, Chen and Mueller found the opposite effect for international migration from Bangladesh, and their work shows that extreme soil salinity had a negative impact on international migration. Their work further reinforced the phenomenon of vulnerable trapped populations that are unable to move due to a lack of resources to do so.

Also noteworthy is Islam's work to understand the causes of migration decisions for Bangladeshi communities in vulnerable char lands, or riverine islands. This work studied the connections between livelihood vulnerabilities and climate change that could result in migration decisions (Islam 2017). The study used a mixed method approach of household interviews, focus group discussions, participant observation, and community mapping to collect a rich set of data on the char people. Islam found that natural and climate related threats, along with economic and social vulnerabilities had wide range of impacts on char peoples' migration decisions, though his work stops short of attempting to quantify those impacts on migration decisions through statistical learning (Islam 2017).

Davis et al. use a diffusion-based model combined with population, geographic, and climatic data to estimate fluxes of migrants between locations in Bangladesh driven by projected sea level rise in 2050 and 2100 (2018). Based on elevation and population data, they predicted that 0.9 million people in Bangladesh by 2050, and 2.1 million people by 2100 could be displaced by flooding (Davis et al. 2018). As part of this analysis, Davis et al. estimate strains on jobs, housing, and food supplies in destination locations within Bangladesh due to this displacement from coastal regions (2018).

Adams and Kay also studied the effects of sea level rise on migration in Bangladesh using survey data from1,500 households and a physical model of sea level projections (2019). Their analysis is informed by a behavioral framework that emphasizes that individuals may have unique migration thresholds and innate propensities to migrate that impact the ultimate migration decision in the event of exposure to environmental stress (Adams and Kay 2019). Rather than studying the household response, they focus on the village unit to identify responses specific to the community and to assess whether some villages had a higher propensity to move (Adams and Kay 2019). This work provides an example of quantitative analysis of environmental migration in Bangladesh that is highly theory driven.

This review identified one example of agent-based modeling being applied to studying environmental migration in Bangladesh (Hassani-Mahmooei 2012). Hassani-Mahmooei developed an agent-based model to simulate migration decisions between districts based on 10 heuristics or migration "rules", as well as "push", "pull", and "intervening" factors related to climate change scenarios, socioeconomic conditions, house ownership, and employment (2012). The model can impose climate shocks on agents, pushing them to decide to migrate and then select where to migrate. Combined with population growth and agent mortality, Hassani-Mahmooei uses the model to predict that between 3 and 10 million people in Bangladesh will migrate internally over 40 years (2012).

Even within the literature of environmental migration in Bangladesh, there is disagreement in terms of the potential of migration to be a positive adaptation strategy to environmental stress. Though temporary migration is common in Bangladeshi communities, some work has asserted that permanent migration due to environmental stress may be a last resort for households whose environment becomes inhospitable (Penning-Rowsell et al. 2013). Penning-Rowsell et al.'s study of five villages in Bangladesh impacted by natural hazards found that factors that "anchor" households to their homes are strong, and migrations to urban areas can come at a significant loss (2013).

1.3 Methods and future challenges

Accurate modeling and predictive ability of environmental migration are critical for informing future climate policy and adaptation strategies (Stern et al. 2006; Hugo et al. 1996; Biermann et al. 2010; Black et al. 2011; Ahsan et al. 2011). However, predictions of environmental migration, especially estimates of the magnitude of environmental migration, are controversial (Gemenne 2011). Especially because of the previously explained complexity of environmental migration and the lack of agreement in the literature, prediction may be highly uncertain. This uncertainty is further exacerbated by the uncertainty related to future climate and demographic scenarios (Hugo 2011). Apart from the complexity of the environmental-migration nexus, predicting human decision making in general is notoriously difficult (DeAngelis and Diaz 2019; Klabunde and Willekens 2016; Subrahmanian and Kumar 2017).

Additionally, current work uses a wide range of methods and models from strictly conceptual models (Perch-Nielsen et al. 2008; Renaud et al. 2011), to logistic regression (Koubi et al. 2016), multivariate regression (Hino et al. 2017), statistical analysis (Henry et al. 2003, 2004), and a few agent-based models (Cai and Oppenheimer 2013; Hassani-Mahmooei 2012; Kniveton et al. 2011; Silveira et al. 2006; Smith 2014; Thober et al. 2018). Some researchers choose to control for demographic factors, while others do not (Fussell et al. 2014). Several detailed reviews of existing methods and challenges call for the exploration of new methods that can improve prediction and better address nonlinearities in environmental migration (Neumann and Hilderink 2015; Obokata et al. 2014; Piquet 2010). As Obokata et al. indicate, existing quantitative methods of studying environmental migration often simplify complex variables and limit the number of variables studied (Obokata et al. 2014).

A major challenge to current efforts to model environmental migration is the "limited understanding of the environmental and non-environmental drivers of migration (including their interactions)" (Neumann and Hilderink 2015). As previously mentioned, this challenge means that researchers who study migration will often use expert judgement or theory to select which variables to assess. Though this can test theoretically motivated hypotheses and provide insights into how specific drivers might impact migration decisions, it does little to identify which variables might be the most important at driving decisions, especially when considering nonlinear interactions among variables.

Access to appropriate data poses another challenge to the study of environmental migration. Household surveys have been a common source of data for migration research (Bilsborrow and Henry 2012), which Neumann and Hilderink suggest is likely the most appropriate level for obtaining information about the causes of migration (Neumann and Hilderink 2015). Other work uses a combination of data sources. For example, Fussell et al. advocate for using a combination of population censuses, surveys, and multi-level modeling (Fussell et al. 2014). Call et al. used a combination of an existing demographic surveillance system and census data, as flood observatory data and biophysical databases (Call et al. 2017). In their Sumatra research, Gray et al. used a novel combination of population-based survey methods, satellite imagery, and multivariate statistical analysis to assess population mobility after the 2004 Indian Ocean tsunami (Gray et al. 2014). While modeling techniques continue to increase in sophistication, it is thought that obtaining reliable data to be used in those models will continue to be a challenge (McLeman 2013).

Recently, Lu et al. utilized mobile phone data from more than six million anonymous phone users in Bangladesh to track mobility across short time scales (Lu et al. 2016). Bell has also employed the use of mobile phone data to studying migration (Bell 2017; Bell et al. 2019). As big data, especially through mobile phones and other technologies, becomes increasingly available, it poses an opportunity for the study of human mobility and migration. Beyond providing new sources of high resolution and potentially high-quality data, technologies may add another dimension of complexity to the study of environmental migration by influencing the migration decision process (Boas 2017). Researchers will also need to consider how to incorporate big data into more traditional methods, including field-based research (Baos et al. 2019).

As researchers continue to collect large amounts of data, including with large household surveys, challenges may arise in how best to analyze such datasets, especially where motivating theories are unclear or conflicting. To advance the study of environmental migration, especially as large datasets and surveys become more readily available, new methods will need to be employed (Neumann and Hilderink 2015). This work aims to address this need by applying machine learning, specifically random forests, to social survey data for the study of environmental migration in Bangladesh. Random forest is a machine learning approach that has been shown to perform well in environmental and ecological contexts (Cutler et al. 2007; Prasad et al. 2006). However, reviews of methodologies used in studying environmental migration did not mention machine learning techniques (Piguet 2010). One previous article applied machine learning to satellite imagery for the study of environmental migration (Ahmed et al. 2018), but this application of random forest methods to social datasets in order to investigate the topic of environmental migration is novel.

In contrast, survival analysis is a common method for studying migration (Bailey 1993; Fussell et al. 2014). As an example, a review of the literature found a previous study that used proportional hazards models to study the effects of fertility on internal migration in Peru. The study used demographic and health data and showed that women with higher levels of education and fewer children had higher levels of mobility (White et al. 1995). This study also argued the importance of temporal data and survival analysis when studying human mobility (White et al. 1995). One additional study used a risk attitudes approach to analyze migration propensities for livelihood and labor in Germany (Jaeger et al. 2010). However, survival analysis to study human migration has not previously been informed by an analysis of variable importance using machine learning.

More generally, the study of migration is challenging because it is inherently interdisciplinary. Environmental migration represents what is called a “coupled human-natural system” or “social-ecological system” (Ostrom 2009). Not only do environmental conditions impact the human response to migrate or stay in a location, but further research has looked at the linkage in the other direction, where migration decisions have an effect on the environment (Gray and Bilsborrow 2014; Greiner and Sakdapolrak 2013). To address such questions related to coupled human-natural systems, both sophisticated understanding of the natural environment and social sciences are required (Fischer et al. 2015). As such, several researchers have called for the necessity of interdisciplinary teams to study environmental migration (Fussell et al. 2014; Hunter and O’Neill 2014). Other researchers have argued for the importance of an interdisciplinary, coupled human-natural systems approach to specifically studying the Ganges-Brahmaputra Delta region, where Bangladesh is located (Nicholls and Goodbred 2004). Computational modeling methods can serve as a bridge between interdisciplinary teams and a range of stakeholders aiming to understand environmental migration (Till et al. 2018).

Chapter 2

Study Area and Data

2.1 Study area

Bangladesh is located on the low-lying deltaic floodplain of the Ganges-Brahmaputra-Jamuna Delta, which includes the Ganges, Brahmaputra, Jamuna, Padma, and Meghna Rivers (Passalacqua et al. 2013). Bangladesh has a population of more than 164 million (World Bank n.d.) in an area of roughly 144,000 km^2, making it one of the most densely populated countries in the world (Black et al. 2008; Ahsan et al. 2011). Bangladesh is also a rapidly urbanizing country, and is estimated that more than 30% of Bangladesh's population lives in an urban setting (Ahsan et al. 2011). Bangladesh is also one of the poorest countries in the world, and its GDP is currently approximately 150 billion USD (World Bank n.d.). In the year 2000, urban poverty was estimated at 52.5%, and rural poverty was 47.1% (Ahsan et al. 2011).

Bangladesh falls in a low elevation coastal zone, and as such faces environmental vulnerabilities (McGranahan, Balk, and Anderson 2007). Bangaldesh experiences regular flooding, extreme weather events, and sea level rise (Call et al. 2017; Dewan et al. 2007; Hallegatte 2013; Higgins et al. 2014; Islam and Sado 2000). In addition, communities on the coastal delta plain of Bangladesh face a particular set of challenges as sediment transport, deposition, and erosion continuously create and destroy land and shape the areas where people live and work, threatening to destroy homesteads and livelihoods (Ackerly et al. 2015; Auerbach et al. 2015; Bhuiyan et al. 2017). It is estimated that more than 50 million people live in the coastal areas of Bangladesh, where they are highly vulnerable to natural disasters and environmental shocks (Ahsan et al. 2011). More than one million people are estimated to lose their homesteads to river erosion every year (Black et al. 2008).

Bangladesh is commonly considered one of the most vulnerable countries to climate change in the world (Black et al. 2008; Walsham 2010). As in other delta regions, future climate change is expected to create additional stress and uncertaintly in Bangladeshi communities through its interactions with natural hazards such as cyclones, flooding, waterlogging, salinity encroachment, and land erosion, as well as with natural resources, such as accreting land and freshwater supplies (Ackerly et al. 2015; Auerbach et al. 2015; Benneyworth et al. 2016; Brammer 2014; Nicholls et al. 2007; Nicholls et al. 2008; Tessler et al. 2015, Xu et al 2009).

Environmental conditions in Bangladesh pose a severe challenge to rural communities, as approximately two-thirds of workers in rural areas and nearly half of all workers in Bangladesh depend on agriculture as their primary source of livelihood (World Bank 2016). Environmental changes, such as salinity encroachment, have already led to shifts in livelihood choices in some Bangladeshi communities from rice to shrimp aquaculture, which may, in turn, exacerbate negative impacts to the environment (Alauddin and Sharma 2013; Islam 2014; Paul and Vogl 2011). However, even the productivity of shrimp may be declining due to extreme salinity, changes in rainfall, high temperatures, and diseases associated with climate change (Rakib et al. 2019). Shrimp aquaculture itself may also be an unsustainable source of livelihood as it may exacerbate soil salinity levels (Azad, Jensen, and Lin 2009; Islam and Tabeta 2019).

In Bangladesh, migration has long been a way of life as a common method of livelihood diversification and adaptation to stressful natural conditions (Alam et al. 2017; Amrith 2013; Black et al. 2005; Martin et al. 2014). Rural to urban migration is the most prevalent form of migration in Bangladesh (Ackerley et al. 2015; Ackerley et al. 2017; Afsar 2003; Bryan et al. 2014; Lagakos et al. 2018), especially temporary migration to adapt to seasonal poverty (Khandker 2012). Remittance provided by household members who have migrated can increase livelihood stability in the midst of agricultural instability and seasonal poverty (Call et al. 2017). However, it is unclear how these existing mobility patterns will be impacted by climate change. **Figure 1** portrays some of the key coupled human-natural system dynamics present in Bangladesh that will be investigated as part of this research.

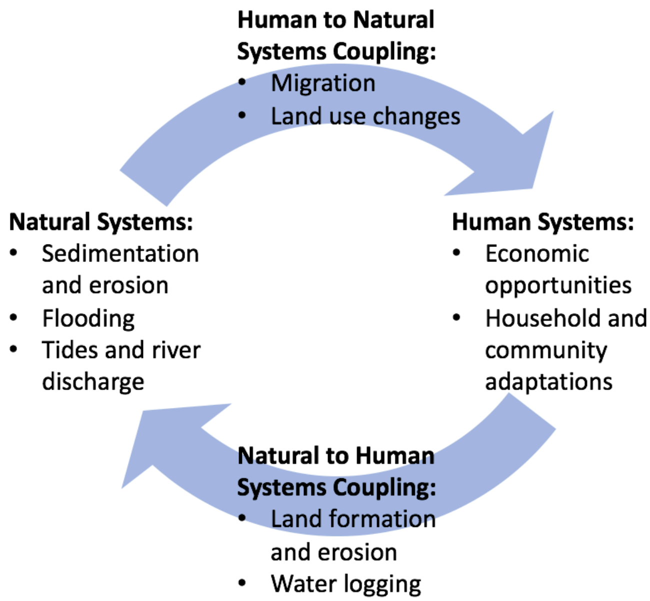


Figure 1: Coupled human-natural system components related to livelihood, migration and landscape change in coastal Bangladesh.

The study area for this research is located primarily in the southwest of rural Bangladesh. This area is positioned just north of the Sunderbans mangrove forest, the largest mangrove forest in the world (Benneyworth et al. 2016).

2.2 Data

Data for this work come from two distinct social surveys collected from households in southwestern Bangladesh. The first survey, Survey 1, comes from Integrated Social Environmental Engineering Bangladesh (ISEE-B), a multi-disciplinary collaborative project to study community resilience to environmental change in coastal Bangladesh (Ackerly et al. 2015). The data were collected in household interviews in 26 communities in the southwest region of Bangladesh from March through April 2014. The 26 communities are a purposive sample from a set of 75 communities identified based on the properties of their aquifer which gave them inadequate access to fresh safe water (Ackerly et al. 2015). The 26 study communities were identified for variability on three dimensions: nongovernment organization partner, geographic dispersion, and ground water quality. Each community was a neighborhood of approximately 100 households, generally sharing a common water source. After doing a geolocated photo census of each community, 20-

50% of the households were randomly selected for study. In total, 1,204 heads of household were interviewed about their household's demographics, sources of livelihood, sources of water, environmental stressors, and other factors. Additional questions measured their individual risk perception, sense of social cohesion, and political trust. The original dataset consists of 1,204 observations and 1,456 variables.

The second survey of households, Survey 2, was also collected in the southwest region of Bangladesh by the Bangladesh Environment and Migration Survey (BEMS). This survey contains migration, employment, and livelihood histories on more than 3,000 individuals affiliated with 1,695 households. The data represents 1,695 randomly sampled households in nine sites in Bangladesh, which were surveyed in 2014. The survey specifically asks for histories of migration within Bangladesh, to India, and to any other country (Donato et al. 2016). Here, I focus only on each household's reported migrations internal to Bangladesh. The original dataset consists of 1,695 observations of 1,997 distinct variables.

Because of their distinct purposes, Survey 1 and Survey 2 ask different questions and include data from different communities, so they present two unique opportunities to test the ability of random forest models to identify salient variables. **Figure 2** shows the geographic locations of households surveyed in Survey 1 and Survey 2.

The structure of Survey 1 is such that the outcome variables to be fit with models were Boolean variables indicating the respondent's answer to yes or no questions about migration: "Have you ever moved your household temporarily to another place within this village because of an environmental event?"; "Has anyone in your household ever moved for education?"; "Has anyone in your household ever moved for health care?"; "Has anyone in your household ever moved for commerce/ trading?"; and "Has anyone in your household ever moved to visit relatives?" These questions were used to assess migration for environmental reasons, for education, for health, for trade, and to visit relatives respectively. Thus, Survey 1 also allows us to assess random forests' ability to compare the salient variables associated with migration for different reasons.

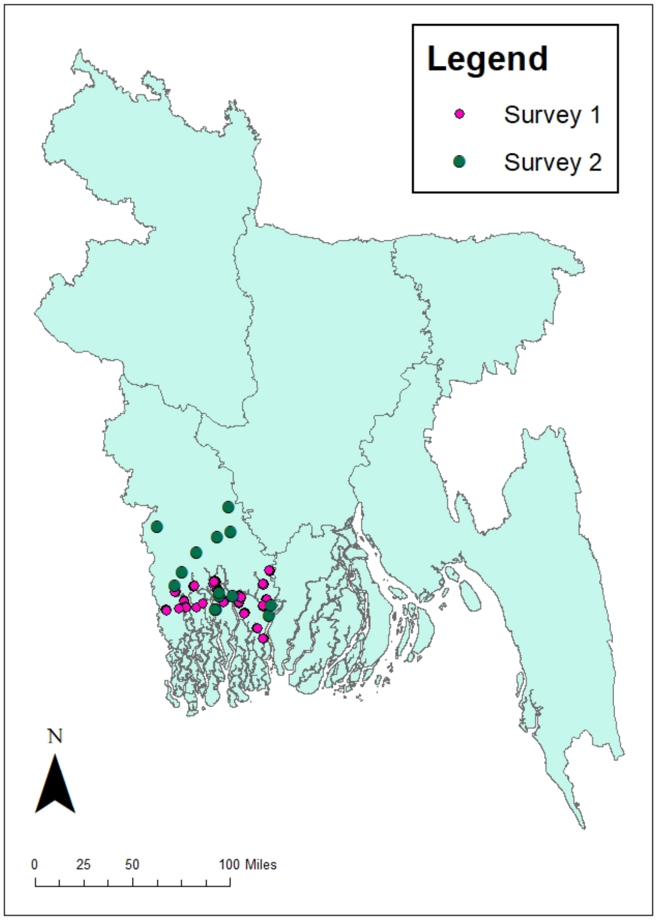


Figure 2: Map of Bangladesh with locations of households surveyed by Survey 1 and Survey 2.

Survey 2 asks respondents to recall the total number of migrations that any member of the household has made, without attributing underlying motivation. This provides the total number of migration trips per household, normalized by total person-years. Person-years were calculated for each member of the household, beginning at age 11, which is the age that many Bangladeshis begin migrating for livelihood opportunities, until 2014 when the survey was collected (Donato et al. 2016). Our analysis of Survey 2 takes as its dependent variable the annual probability of making a migration, which is represented as a continuous variable at the household level, and identifies salient variables that predict this probability.

Chapter 3

Methods

3.1 Methods for Objective 1

Objective 1: Use random forests to determine the importance of covariates in a large dataset for predicting migration outcomes.

3.1.1 Machine learning

While there are many different machine learning techniques, from simple linear regression to neural networks, machine learning, broadly, refers to a variety of methods that enable a computer or "machine" to automatically recognize patterns in data and use these patterns to build and refine a statistical model of the data without being explicitly programmed to do so and without theoretical or phenomenological preconceptions about the causal mechanisms that gave rise to the data. Machine learning methods are often categorized as *supervised* or u*nsupervised*. *Supervised* methods are used to predict one or more specified dependent variables. *Unsupervised* methods are used to identify patterns in the data (Jordan and Mitchell 2015). To give examples from common statistical methods, regression analyses are supervised methods and exploratory factor analyses are unsupervised methods. In order to guard against overfitting, machine learning models are trained using a subset of the complete data, known as the *training set*, while the remaining data, known as the *holdout* or *testing set*, is withheld and used for validating the model's performance after the model is fully trained.

Machine learning techniques often outperform standard regression analysis in predictive ability, especially when studying complex social problems (Hindman et al. 2015). Recently, there has been discussion of broadly incorporating machine learning into the social sciences, especially in the place of traditional regression analysis (Hindman et al. 2015; Mason et al. 2013). However, some machine learning algorithms can be very difficult to interpret due to their complexity and this complexity makes it difficult to assess how well a machine-learning model is likely to apply outside the specific context in which the data was gathered (Buolamwini and Gebru 2018). While a simple linear regression results in coefficients that can be easily interpreted, a more complex machine learning model may be "black box", making it difficult to draw insights from the model. As complexity of the model increases, interpretability may decrease, representing a tradeoff between model performance and interpretability (**Figure 3**). Where prediction is a priority, complex machine learning algorithms may perform very well, while they might not always be an appropriate tool for theory development. Furthermore, the greater predictive power that complex models often possess may arise from models reflecting details of the context in which the data set being analyzed was collected and the models may not transfer as well to other contexts as simpler models would. When the complexity of a model impedes interpretation, it can be difficult to draw on theory or other domain knowledge of the context to evaluate the applicability of a machine learning model to different contexts. Therefore, it is especially important for researchers to carefully consider the goals of their research when selecting a machine learning algorithm, as there is no one size fits all approach.

Machine learning should be incorporated into social scientists' toolkits for studying migration because of its ability to identify patterns in complex datasets. Various machine learning algorithms may have advantages over more traditional theory-driven regression methods, especially where theory is unclear.

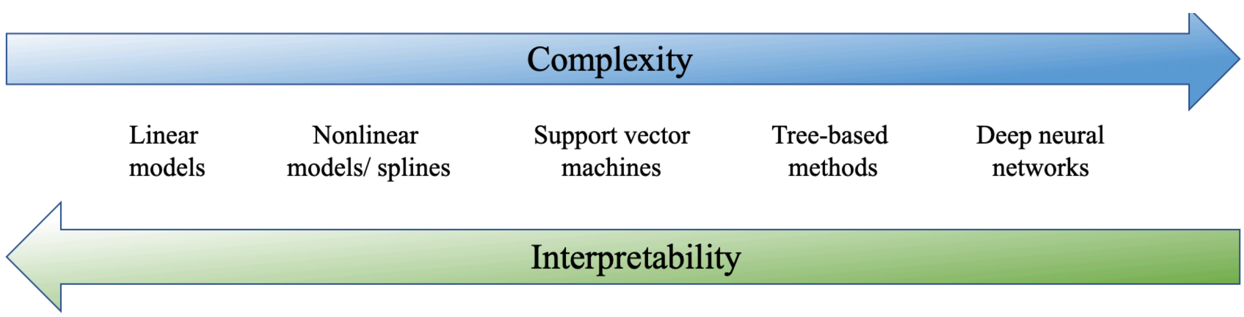


Figure 3: Schematic demonstrating the tradeoff between complexity and interpretability of common machine learning algorithms. For example, tree-based methods such as random forests are highly complex and sometimes challenging to interpret. Researchers should consider where a method falls on this continuum along with specific research goals when selecting an appropriate algorithm.

3.1.2 Model selection

The first step of the analysis was to compare different approaches to analyzing the survey data. I compared random forest models, multiple logistic regression, and support vector machines (SVMs) with a radial kernel. Random forest modeling, which is a tree-based method, is described in more detail in the next section. Multiple logistic regression is a generalized linear model that fits coefficients to predictors in order to fit the logit transformation of the probability of the event of interest, which is then converted to a dichotomous prediction of the outcome variable (Hosmer et al. 2013). SVMs are a machine learning method of classification which work by fitting a hyperplane in the parameter space to split data of different classes (Suykens and Vandewalle 1999).

All three models were fit to each of the five motivations of migration in Survey 1: environmental, education, health, trade, and to visit relatives, for a total of 15 models. Each model was trained on a random sample of 80% of the data set, and tested on the remaining 20% to assess predictive accuracy, as indicated in **Figure 4**. For random forests and SVMs, relevant model parameters were tuned by minimizing out-of-sample error. **Table 1** shows the prediction error for each model on the test data in percent error.

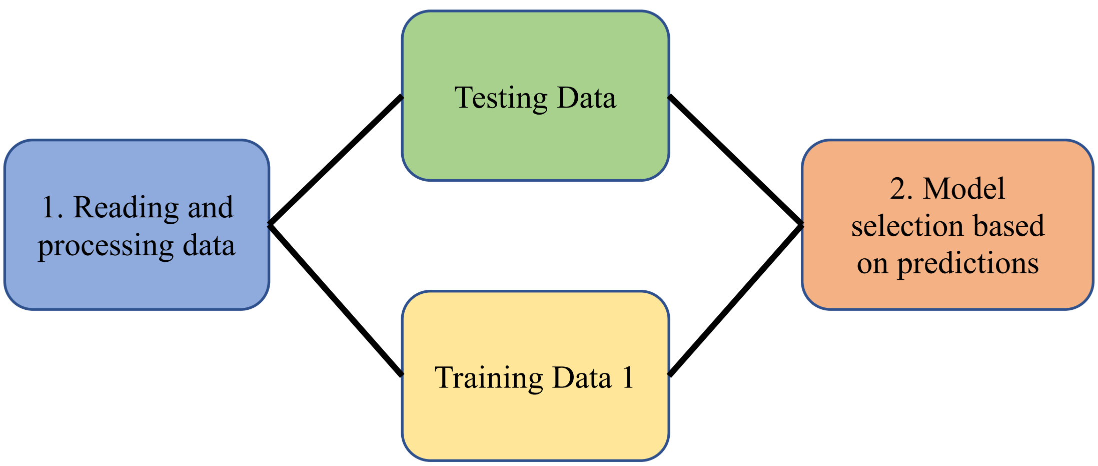


Figure 4: Schematic of method for model selection. Data is divided into testing and training data, and model predictive ability is assessed based on the model's ability to predict test data.

Table 1: Test data prediction error (percent error, calculated as average misclassification percentage) for logistic regression, support vector machine, and random forest models fit to each type of migration in Survey 1.

	Environmental	**Education**	**Health**	**Trade**	**Visit Relatives**
Logistic Regression	47.1	44.9	44.1	43.4	42.6
SVM	36.0	16.2	36.0	19.9	41.2
Random Forest	35.5	14.7	33.1	19.9	33.8

3.1.3 Imputing missing data in Survey 1

Before further analyzing Survey 1, data related to the household respondent was selected from the household roster, and summary variables related to household size, household education, and livelihood were developed. Questions that only applied to part of the sample were eliminated, keeping only variables that were relevant to the full data set. The remaining variables were then screened manually, and variables that were likely missing not at random, or for which there were known problems during data collection, were dropped.

The resulting subset of data consisted of 1184 observations of 730 variables. Within this subset of the original survey, approximately 1.5% of data across all variables and rows were missing. Even after dropping columns that were not relevant to all households from the subset of Survey 1, restricting the analysis to complete cases would have needlessly lost information in the partial cases. To address this, I imputed missing variables in partial cases using multiple imputation, which enables the assessment of the stochastic uncertainty associated with the imputation process (Liu and De 2015; van Buuren and Groothuis-Oudshoorn 2011).

Before imputing, the data was filtered to consider only variables with less than 12% missingness, which was a threshold that maintained 711 of the 730 variables. Imputations of missing data were then conducted using the *mice* (Multivariate Imputations by Chained Equations) package in R (van Buuren and Groothuis-Oudshoorn 2011). In order to accommodate both categorical and continuous data, a random forest imputation method was used to impute missing data 10 times. This resulted in 10 unique, complete datasets to be used in analysis. Imputations were only conducted on predictor variables, not outcome variables, which did not have significant degrees of missingness.

Survey 2 did not have significant missing data, and therefore imputations were not necessary before assessing variable importance.

3.1.4 Random forest models for variable importance

Random forest models are an ensemble method of decision trees, and represent a subset of machine learning known as tree-based methods. Tree-based methods, including random forests, can be used for classification of discrete outcome variables, or regression of continuous variables. They are especially powerful tools when there are strong nonlinearities or interactions between variables in the data.

Random forests models work by fitting many decision trees, where each tree uses a random subset of the predictor variables at each split in its decision tree. The final prediction is then calculated by averaging across the outputs of all of the individual decision trees (Hastie et al. 2009, Ch. 15). This allows random forest models to achieve high predictive accuracy without overfitting (James et al. 2013). One strength of random forest models, especially over other "black box" statistical models, is their ability to assess variable importance and account for complex, nonlinear interactions between variables. Random forest models are also able to use combinations of categorical, ordinal, and continuously-valued variables as inputs without requiring dummy variables or scaled data. This makes them especially appealing tools for analyzing large social surveys and studying complex challenges such as migration.

The *randomForest* package in R was used to fit random forest models to the training data, which consisted of a random subset representing 80% of the total data (Cutler 2018). For Survey 1, a binary variable representing whether or not a respondent migrated for a given reason was used as the outcome variable. For Survey 1, 10 random forest classification models were fit (one for each imputed dataset) for each of the five types of migration (environmental, education, health, trade, and to visit relatives). For Survey 2, 10 random forest regression models, each with a different subset of the data as a training set, were fitted to the continuous outcome variable of total internal migration trips per household normalized by person-years. For each model, the parameter for the number of variables randomly sampled at each split was tuned by minimizing the out-of-sample error. A schematic of these steps of the methodology are shown in **Figure 5**.

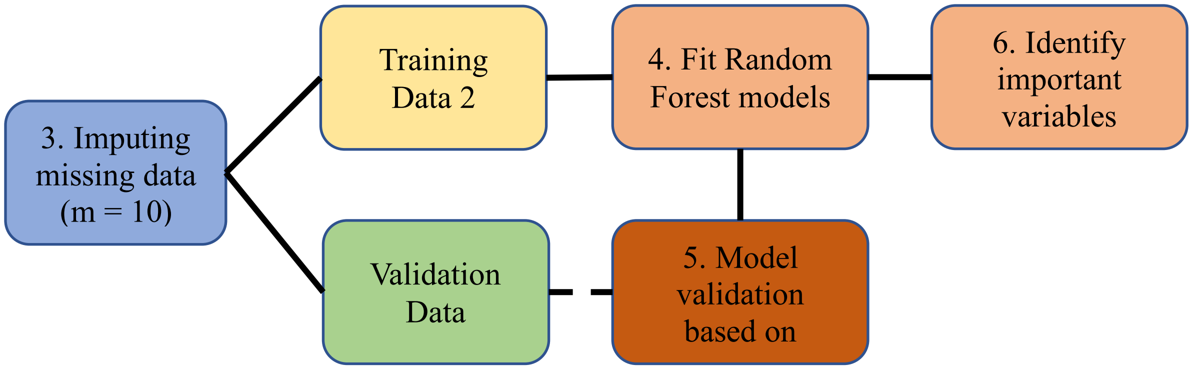


Figure 5: Schematic of method after model selection, which includes imputing missing data and fitting and validating random forest models. Random forest models are then used to assess variable importance.

For each of the five types of migration in Survey 1, variable importance was ranked by average across the 10 imputed datasets. Again, imputations were only conducted on predictor variables, not outcome variables. Variable importance for the model of Survey 2 was also ranked and averaged across the 10 complete models. Tuned and fitted models were then validated using the testing data, which consisted of the remaining 20% of the data that was not used for training.

For Survey 1, variable importance is given by mean decrease in Gini index, which is an index that measures a specific variables ability to correctly classify an outcome. The Gini index is a weighted measure of how much a split in the decision trees by a specific variable can decrease variance in the outcome, and therefore measures how much of the final model's predictive ability can be attributed to a specific variable. The Gini index is specific to a given analysis, so it does not allow comparisons between different analyses. Within an analysis, the Gini index is useful as a metric for comparing the importance of different variables included within a given analysis (Hastie et al. 2009). For regression random forest models, unlike classification random forest models, importance is calculated by node impurity, though it is a similar calculation of how much a split in the decision trees by a specific variable can decrease variance in the outcome (Hastie et al. 2009). Therefore, the analysis of Survey 2 uses node impurity to assess and compare variable importance.

3.2 Methods for Objective 2

Objective 2: Conduct a survival analysis of household time to first migration using the important variables identified by the random forest algorithm to provide deeper insight into how important variables impact mobility.

3.2.1 Survival analysis

This work uses Kaplan-Meiers estimators and Cox proportional hazards models to study overall survival probabilities and fit coefficients to a survival model with relevant variables. This analysis contributes to further understanding of what environmental, social, and demographic variables impact an individual's probability of having migrated at least once over time. Such insight will be critical for assessing which communities and households are most at risk of migration.

Survival analysis is a technique used to study the occurrence of a discrete event where the time until the event matters (Harrell 2015). The response variable in survival models is time until the event, usually referred to as failure time, survival time, or event time. Survival analysis has been commonly used in the medical field to describe times to a disease event (Bull and Spiegelhalter 1997; Crowley and Hu 1977; Prentice et al. 1981), failure or recovery times in engineering systems (Ansell and Phillips 1997; Barker and Baroud 2014), and binary events in social sciences, including the time of a woman's first child (Teachman 1985). It is also a common tool used in demography to study migration (Bailey et al. 1993; Fussell et al. 2014). Survival analysis also allows for some responses to be incomplete, meaning that the event of interest has not occurred within the observed time. Such data is called "censored", while data where the event of interest did occur within the study time is called "uncensored."

Survival analysis can be used to define a survival function, S(t), where *t* is time. The survival function quantifies the probability that a subject will survive past a time t (i.e. an event will has not occurred). The survival function can be written as

$$S(t) = Prob[T > t] = 1 - F(t) \tag{1}$$

where T is the response variable, or the time of the event of interest, and $F(t)$ is the cumulative distribution function for T.

Survival analysis also yields a hazard function $h(t)$, which describes the risk of an event occurring at time t. The hazard function can be found by

$$h(t) = f(t)/S(t) \tag{2}$$

where $f(t)$ is the probability density function of T.

Survival analysis may incorporate the effects of covariates to understand how variables impact the risk of an event occurring. Such estimates can be calculated using proportional hazards models (PHM) (Barker and Baroud 2014). The general equation for PHM is

$$h(t, X) = h_o(t)exp(\beta^T X) \tag{3}$$

where $h_o(t)$ is the baseline hazard function, X is a vector of covariates, and β is a vector of regression coefficients describing the effect of covariates on the baseline hazard. PHM can generally be derived parametrically, where $h_o(t)$ is estimated from a probability density function fit to the data, or semi-parametrically where the form of $h_o(t)$ is not restricted (Harrell 2015).

3.2.2 Distribution fitting

From this ethnosurvey data, time in person-years for each head of household to that head's first internal migration was calculated. This generated a discrete-time person-year file that followed the male head of household from age 11, the age when many Bangladeshi males begin engaging in paid work, to his first migrant trip or the year of the survey. Each male head of household received a 1 if they did complete a trip and a 0 if they did not complete a trip. In this way, the individual migration data was divided into censored and uncensored data for a survival model, as some heads have not completed their first migration by the time of the data collection. Only 17.3% of the data was uncensored, while the remaining 82.7% was censored.

A histogram of the uncensored data by years until the first trip is shown in **Figure 6**. Gamma, Weibull, normal, lognormal, and exponential distributions were fit to this histogram. The Weibull distribution was the only distribution that fit the data, with a χ^2 value of 16.4 and a p-value of 0.127. **Figure 7** shows a graphical summary of the Weibull fit to the uncensored data with a shape parameter of 1.59 and a scale parameter of 22.75.

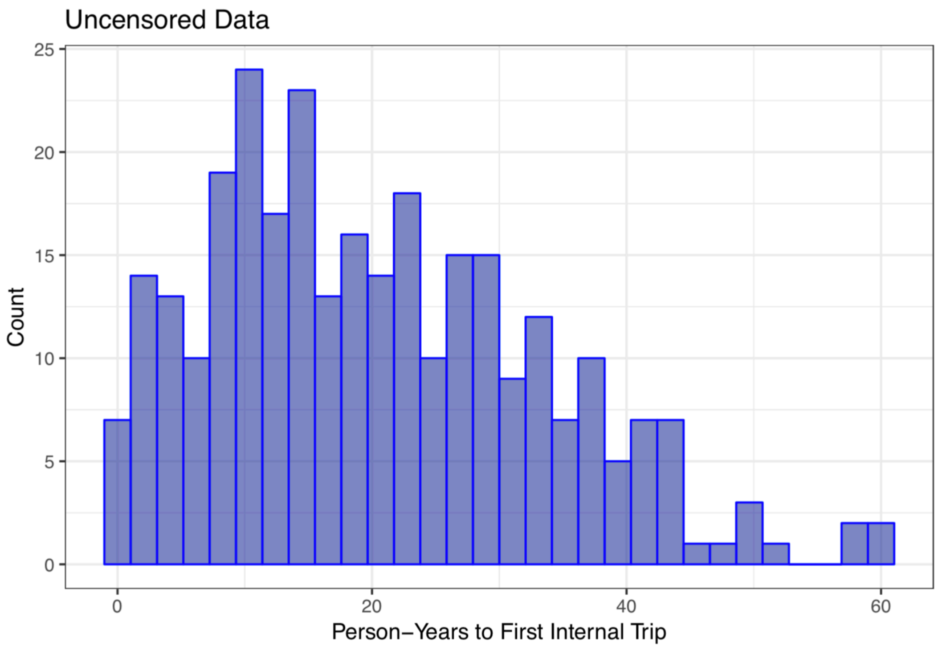


Figure 6: Histogram of uncensored data for time (in person-years) to first migration trip.

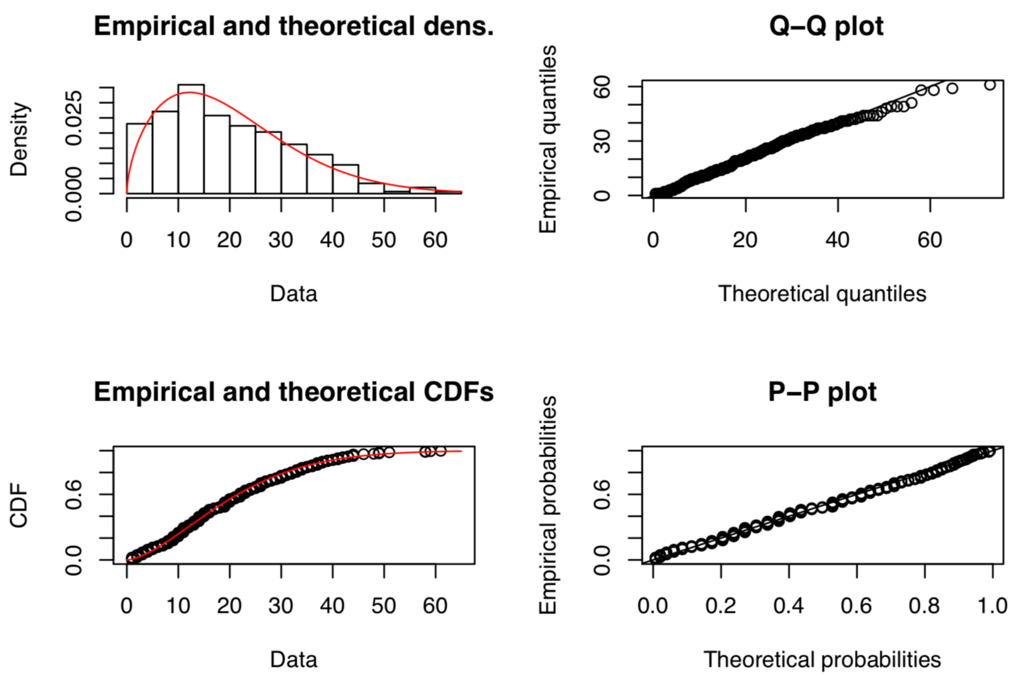


Figure 7: Summary of Weibull distribution fit to uncensored data.

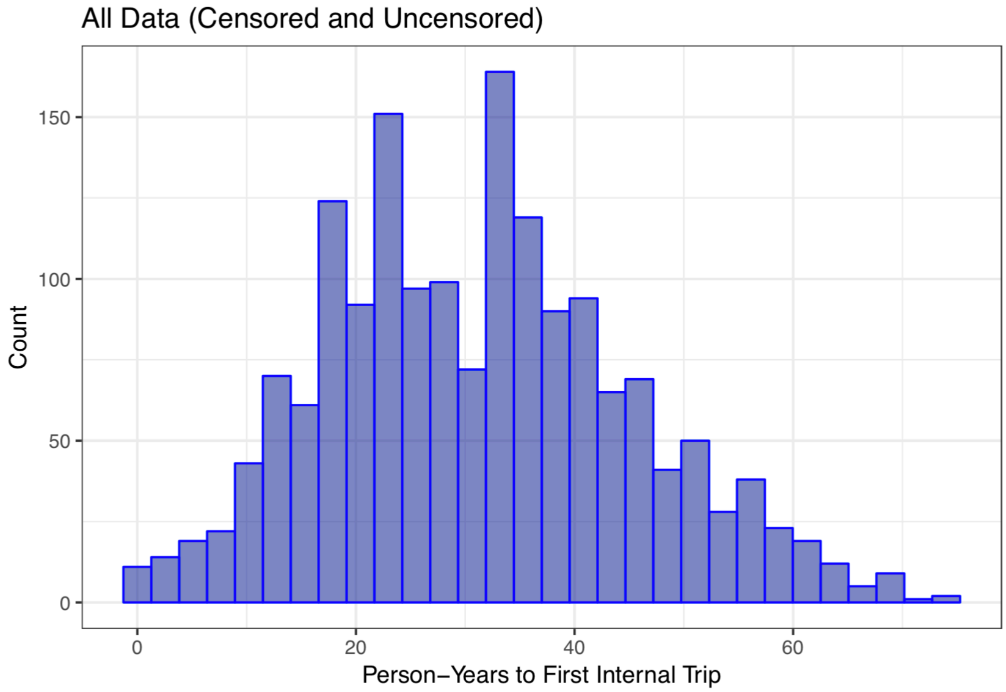


Figure 8: Histogram of uncensored and censored data for time (in person-years) to first migration trip. This data is not easily fit by a distribution.

This process of distribution fitting was repeated for the complete data, with censored and uncensored observations. A histogram of the complete data is shown in **Figure 8**. Again, gamma, Weibull, normal, lognormal, and exponential distributions were tested. However, none of these distributions were able to pass a goodness of fit test. Because of this, a semi-parametric or nonparametric approach was deemed to be the most appropriate for estimating the hazard function.

3.2.3 Kaplan-Meier estimator and Cox proportional hazards model

This work uses a Kaplan-Meier estimator (nonparametric) and Cox proportional hazards model (semi-parametric) to estimate the survival and hazards function to describe the risk of internal migration.

The Kaplan-Meier estimator is a nonparametric estimation that does not make any assumptions about the survival distribution form. The Kaplan-Meier estimator develops an estimate of the survival function without incorporating the effects of covariates, but simply estimates directly from the time-event data. A full and thorough explanation of Kaplan-Meier estimates can be found in Harrell 2015.

Because the complete data could not be fit by a standard probability density function, the effects of covariates were analyzed using a semi-parametric PHM. One very common semi-parametric PHM is the Cox proportional hazards model (Ansell and Phillips 1997; Harrell 2015). The Cox model assumes that covariates impact the base hazard function $h(t)$, but does not assume that $h(t)$ is constant with respect to time (ex: exponential, Weibull, or any other form). Rather, $h(t)$ may vary with time in a complex way that cannot be represented by a specific parametric functional form. However, the regression portion of the model is parametric and assumes that covariates are linearly related to the log of the hazard. This approach is ideal when data is not easily fit to a distribution and when the form of the true hazard function is complex. It is also a useful approach when the key question of concern is how covariates impact the hazard, rather than the shape of the hazard itself (Harrell 2015).

Chapter 4

Results

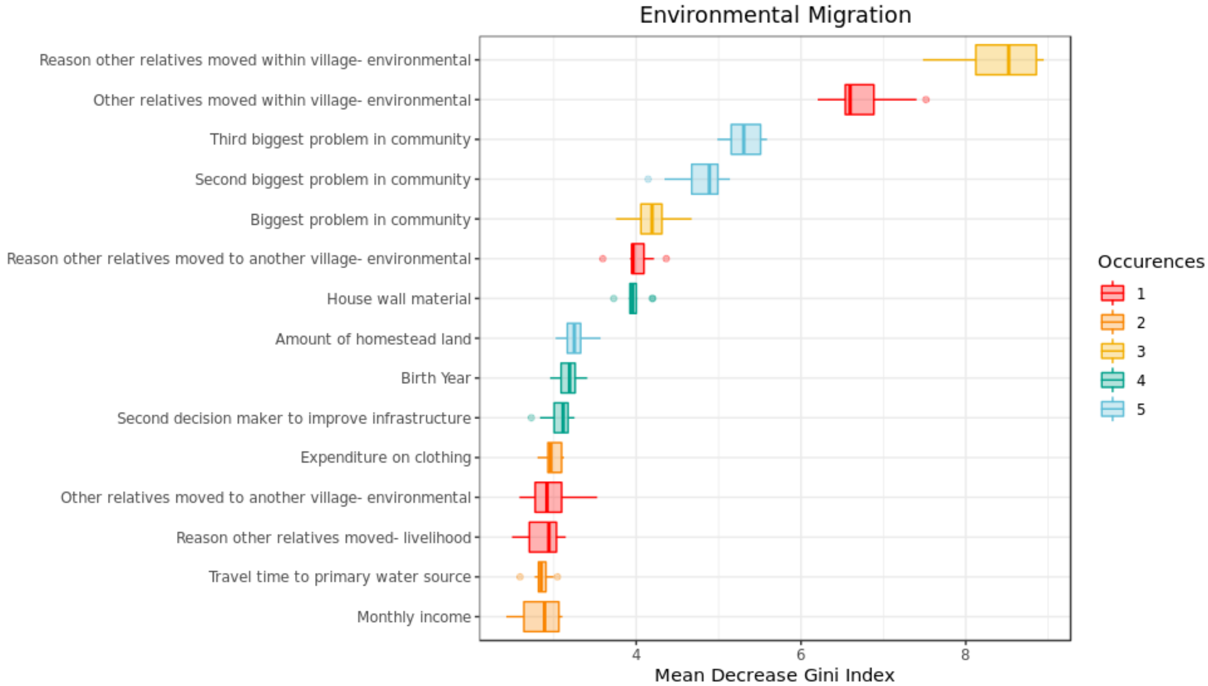
4.1 Results of Objective 1

4.1.1 Variable importance

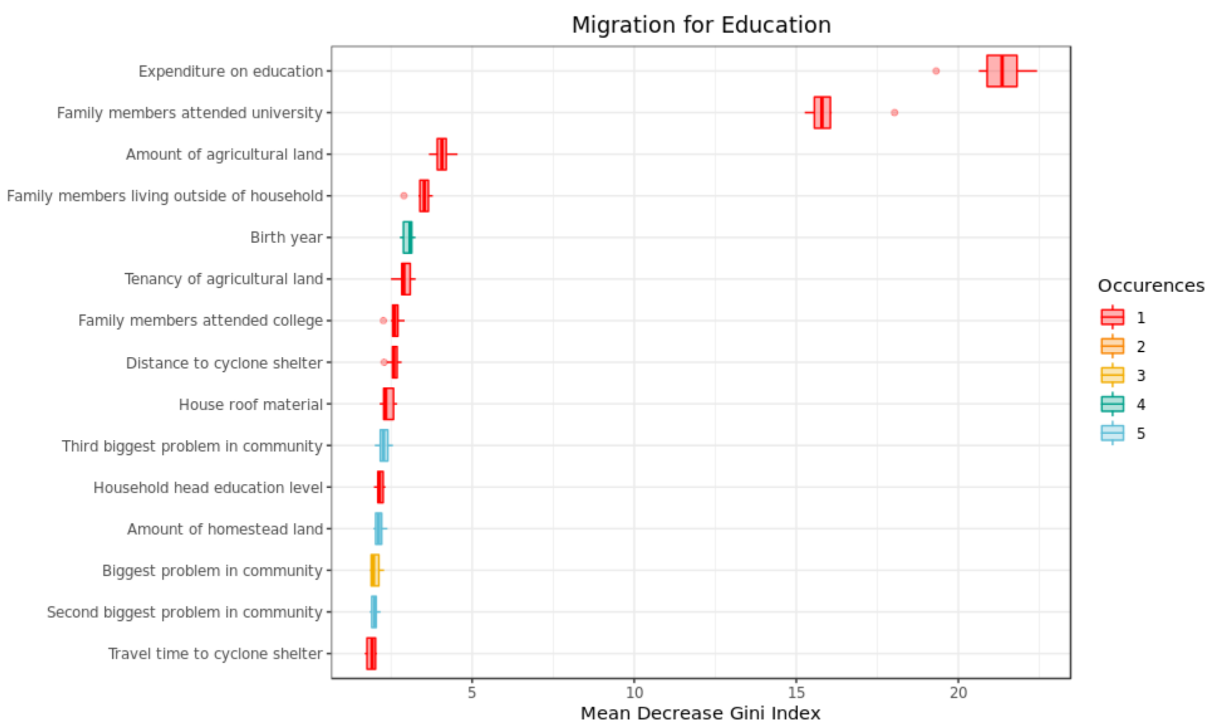
Figure 9 shows the top 15 important variables for random forest models predicting environmental migration, migration for education, migration for health, migration for trade, and migration to visit relatives. 15 variables are displayed because it was consistently found that after this cutoff there was minimal variability in variable importance. However, full results from this analysis provide a ranked list of the importance of every survey variable.

In these figures, variable importance decreases from top to bottom. Colors in the figures are used to show similarities and differences across the five types of migration studied and to highlight the uniqueness of the variable. Colors represent occurrences, or how many times a specific variable was in the top 15 most important variables for another model. Therefore, an occurrence of one (red) means that a variable was important only for the models of that type of migration. Conversely, an occurrence of five (blue) means that the variable was important across all models of all five types of migration. **Table 2** contains full definitions of each variable.

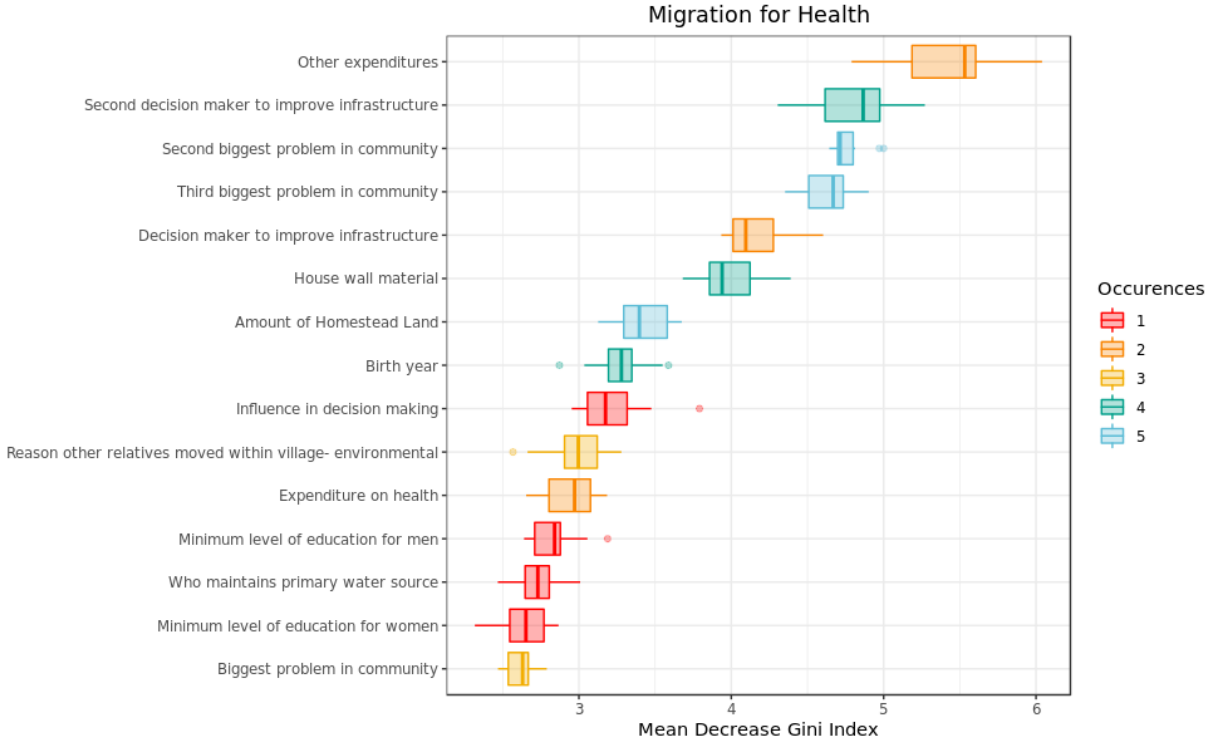
a.)


Environmental Migration
Reason other relatives moved within village- environmental
Other relatives moved within village- environmental
Third biggest problem in community
Second biggest problem in community
Biggest problem in community
Reason other relatives moved to another village- environmental
House wall material
Amount of homestead land
Birth Year
Second decision maker to improve infrastructure
Expenditure on clothing
Other relatives moved to another village- environmental
Reason other relatives moved- livelihood
Travel time to primary water source
Monthly income
4
6
8
Mean Decrease Gini Index
Occurences
1
2
3
4
5

b.)

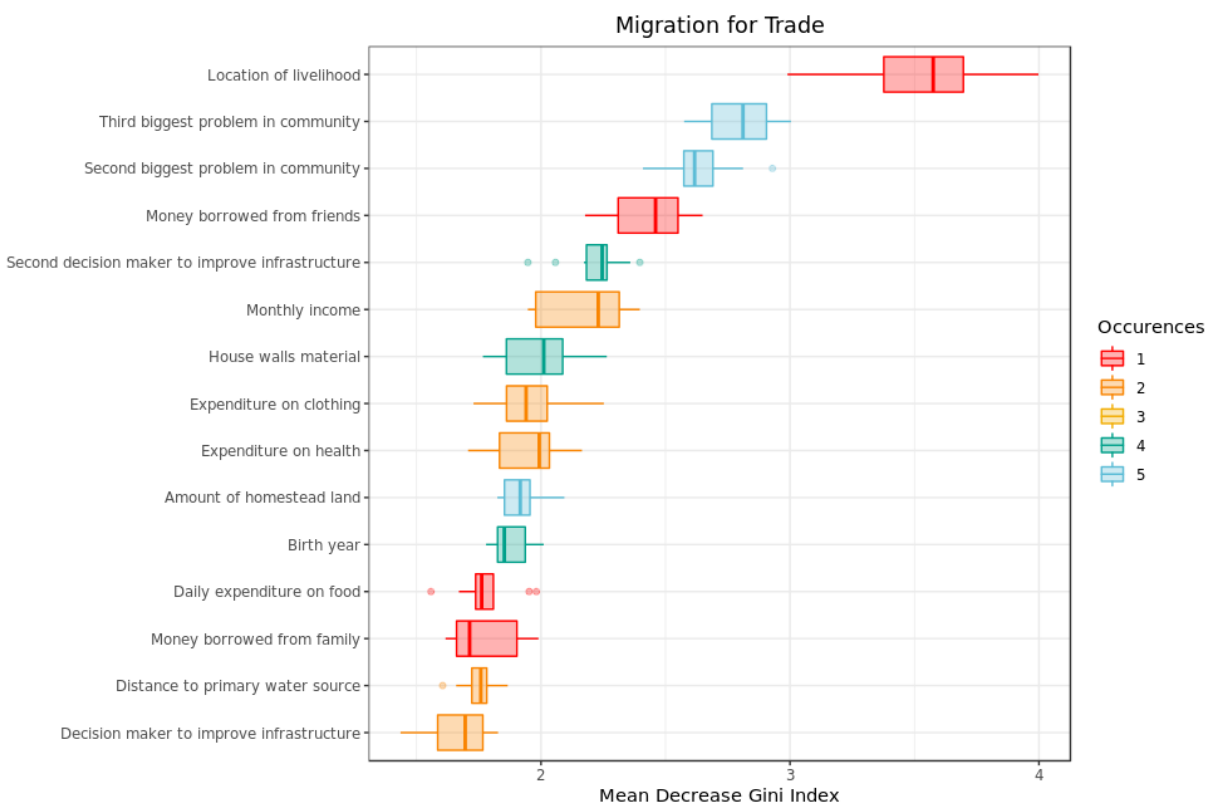

Migration for Education
Expenditure on education
Family members attended university
Amount of agricultural land
Family members living outside of household
Birth year
Tenancy of agricultural land
Family members attended college
Distance to cyclone shelter
House roof material
Third biggest problem in community
Household head education level
Amount of homestead land
Biggest problem in community
Second biggest problem in community
Travel time to cyclone shelter
5
10
15
20
Mean Decrease Gini Index
Occurences
1
2
3
4
5

c.)



Migration for Health
Other expenditures
Second decision maker to improve infrastructure
Second biggest problem in community
Third biggest problem in community
Decision maker to improve infrastructure
House wall material
Amount of Homestead Land
Birth year
Influence in decision making
Reason other relatives moved within village- environmental
Expenditure on health
Minimum level of education for men
Who maintains primary water source
Minimum level of education for women
Biggest problem in community
3
4
5
6
Mean Decrease Gini Index
Occurences
1
2
3
4
5

d.)



Migration for Trade
Location of livelihood
Third biggest problem in community
Second biggest problem in community
Money borrowed from friends
Second decision maker to improve infrastructure
Monthly income
House walls material
Expenditure on clothing
Expenditure on health
Amount of homestead land
Birth year
Daily expenditure on food
Money borrowed from family
Distance to primary water source
Decision maker to improve infrastructure
2
3
4
Mean Decrease Gini Index
Occurences
1
2
3
4
5

e.)

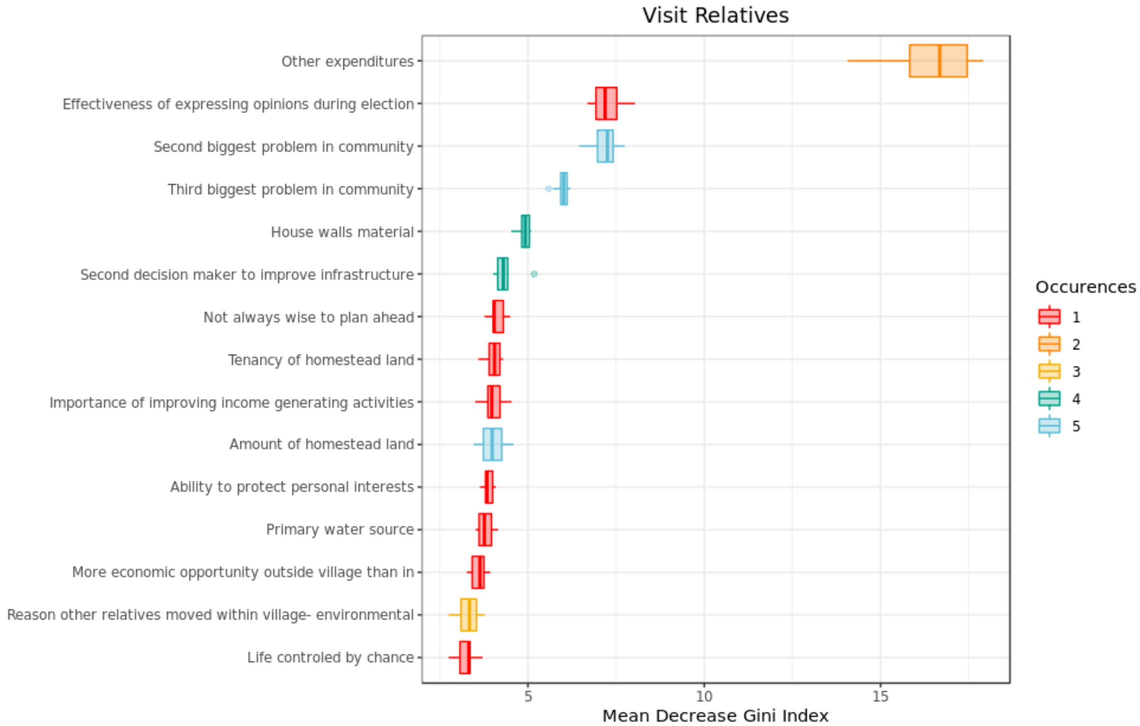


Figure 9: Top 15 variables of importance (fully defined in **Table 2**) identified by random forest models from top to bottom by mean decrease in Gini Index for environmental migration (a.), migration for education (b.), migration for health (c.), migration for trade (d.), and migration to visit relatives (e.). Colors represent how many times a specific variable was in the top 15 most important variables for another model.

In addition, **Table 2** shows salient variables from all model results grouped into higher level categories of variables related to migration ("Migration"), livelihood and wealth ("Livelihood"), community level variables ("Community"), reporter infrastructural support ("Infrastructural Support"), level of trust in others including community and government ("Trust"), personal and household level demographics ("Personal"), and perceptions of locus of control ("Control"). These categories are useful to begin to identify differences in the drivers between the models of migration. The table includes the high-level category, the actual survey question that corresponds to the variable, the variable name indicated on **Figure 9,** and the models in which the variable appears.

Table 2: Variables of importance identified by random forest models of Survey 1 and the original survey questions. Top 15 variables out of more than 1,500 variables total are highlighted for each model. The "Models" column identifies the outcome variable type of migration for which the variable was salient. Variables are categorized as "migration", "community", "control", "livelihood", "personal", "infrastructural support", and "trust" based on key variable themes.

Category	Survey Question	Variable Name	Models
Migration	Have any of your other relatives not living with you now ever moved their whole household temporarily to another place within this village because of an environmental event?	Other relatives moved within village-environmental	Environmental
Migration	Thinking of the event that caused your family to move as you have just said, what was it? (1988 flood, Bhola cyclone, Sidr cyclone, Aila cyclone, other)	Reason other relatives moved within village-environmental	Environmental, Health, Visit relatives
Migration	Have any of your other relatives not living with you now ever moved their whole household temporarily to another village near here because of an environmental event?	Other relatives moved to another village-environmental	Environmental
Migration	Thinking of the event that caused your family to move as you have just said, what was it? (1988 flood, Bhola cyclone, Sidr cyclone, Aila cyclone, other)	Reason other relatives moved to another village-environmental	Environmental
Migration	Have any of your other relatives not living with you now ever moved their whole household permanently to another place because they could not make a livelihood here?	Reason other relatives moved-livelihood	Environmental
Community	As a community member here, what problems do you face regularly? In other words, what are the top 3 problems of this community for you?	Biggest problem in community	Environmental

Community	As a community member here, what problems do you face regularly? In other words, what are the top 3 problems of this community for you?	Second biggest problem in community	All
Community	As a community member here, what problems do you face regularly? In other words, what are the top 3 problems of this community for you?	Third biggest problem in community	All
Community	Imagine that the village receives funds to invest in improving infrastructure in the village. A decision needs to be made about how the funds should be spent. Who would play the biggest role in resolving the dispute?	First decision maker to improve infrastructure	Health, Trade
Community	Who would play the second biggest role?	Second decision maker to improve infrastructure	Environmental, Health, Trade, Visit relatives
Community	What is the minimum level of education that a man can have in your village?	Minimum level of education for men	Health
Community	What is the minimum level of education that a woman can have in your village?	Minimum level of education for women	Health
Community	Who maintains the water source now?	Who maintains primary water source	Health
Community	What do you think should be implemented/improved to help you addressing your future needs related to disaster? Income generating activities.	Importance of improving income generating activities	Visit relatives
Community	There is more economic opportunity outside my village than in it	More economic opportunity outside community than in	Visit relatives
Control	When decisions are made on issues that affect all villagers, do you feel that you are influential in determining the outcome?	Influence in decision making	Health
Control	If you had concerns about how things were going in your village, tell me whether or not you think these things would	Effectiveness of expressing opinions during elections	Visit relatives

	help- Express opinions in elections.		
Control	It's not always wise for me to plan too far ahead because many things turn out to be a matter of good or bad fortune	Not always wise to plan ahead	Visit relatives
Control	I am usually able to protect my personal interests (I can usually look after what is important to me)	Ability to protect personal interests	Visit relatives
Control	To a great extent my life is controlled by accidental/chance happenings	Life controlled by chance	Visit relatives
Livelihood	Residential house wall construction material	House wall material	Environmental, Health, Trade, Visit relatives
Livelihood	Residential house roof construction material	House roof material	Education
Livelihood	Land (homestead) in decimal	Amount of homestead land	All
Livelihood	Tenancy (homestead)	Tenancy of homestead land	Visit relatives
Livelihood	Expenses per year, clothing	Expenditure on clothing	Trade
Livelihood	Expenses per year, health	Expenditure on health	Health, Trade
Livelihood	Please let me know how much you expend daily for giving food to your family members?	Expenditure on food	Trade
Livelihood	Expenses per year, education	Expenditure on education	Education
Livelihood	Expenses per year, other	Other expenses	Health, Visit relatives
Livelihood	How much did you make in taka per month?	Monthly income	Environmental, Trade
Livelihood	Land (agriculture) in decimal	Amount of agricultural land	Education
Livelihood	Tenancy (agricultural)	Tenancy of agricultural land	Education
Livelihood	Source of Income, location	Location of livelihood	Trade
Personal	Year of birth	Birth year	Environmental, Education, Health, Trade
Personal	Head of household level of education	Household head education level	Education

Personal	Level of education, university	Family members attended university	Education
Personal	Level of education, college	Family members attended college	Education
Personal	Lives in the household- No	Family members living outside of household	Education
Personal	Travel time to source (in minutes)	Distance to primary water source	Environmental, Trade
Personal	What are the drinking water sources here?	Primary water source	Trade
Infrastructural Support	Number of minutes it takes to go to cyclone shelter on foot in day light?	Travel time to cyclone shelter	Education
Infrastructural Support	How far is your home from a cyclone shelter? (in km)	Distance to cyclone shelter	Education
Trust	If you suddenly needed a small amount of money, enough to pay for expenses for your household for one week, is there at least one person from the following groups that you could turn to who would be willing to provide this money?- Friends/ neighbors	Money borrowed from friends	Trade
Trust	If you suddenly needed a small amount of money, enough to pay for expenses for your household for one week, is there at least one person from the following groups that you could turn to who would be willing to provide this money?- Members of my extended family/relatives	Money borrowed from family	Trade

Figure 10 shows the results of the variable importance assessment from the random forest model of Survey 2. Again, the top 15 important variables are shown in descending importance from top to bottom. **Table 3** presents salient variables by variable name in **Figure 10** and the corresponding question from the original Survey 2.

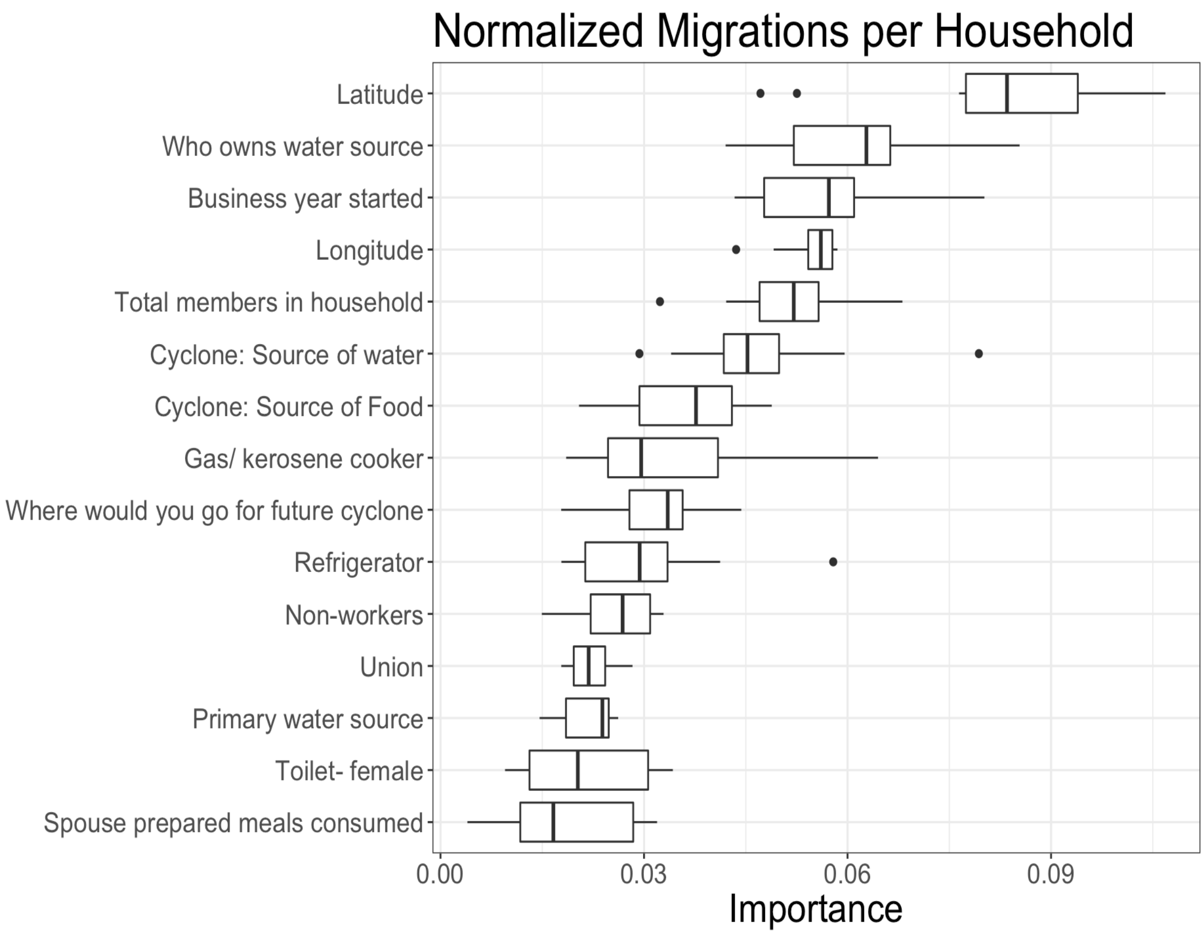


Figure 10: Top 15 variables of importance in Survey 2 identified by random forest models of total household migrations normalized by person-years.

Table 3: Top 15 variables of importance (out of approximately 1,100 total variables) identified by random forest model of migration in Survey 2 and the original survey questions.

Variable Name	Survey Question
Latitude	Household latitude
Who owns water source	Water Sources: Who owns?
Business year started	Business: year started
Longitude	Household longitude
Total members in household	Household: total number of members
Cyclone: source of water	What was your principle source of water during the last cyclone?
Cyclone: Source of food	What was your principle source of food during the last cyclone?
Gas/ kerosene cooker	House Services: Kerosene/gas cooker
Where would you go for future cyclone	Where would you go if there was a future cyclone?
Refrigerator	House Services: Refrigerator
Non-workers	Total number of non-workers in household
Union	Household union
Primary water source	Primary water source
Toilet- female	What kind of toilet facility do female household members use?
Spouse prepared meals consumed	Has household consumed prepared meals? If yes, who? Spouse.

4.1.2 Predictive accuracy

For Survey 1, predictive accuracy was assessed by percent error on predicting the test data. The percent test error was calculated from the percentage of total predictions that the random forest predicted incorrectly (i.e. the model predicted migration when, in fact, the household did not report a migration or vice versa). A lower value of percent test error indicates that a model performs better at predicting test data than a model with a higher value of percent test error. **Figure 11** shows the percent test errors for each of the five types of migration assessed in Survey 1.

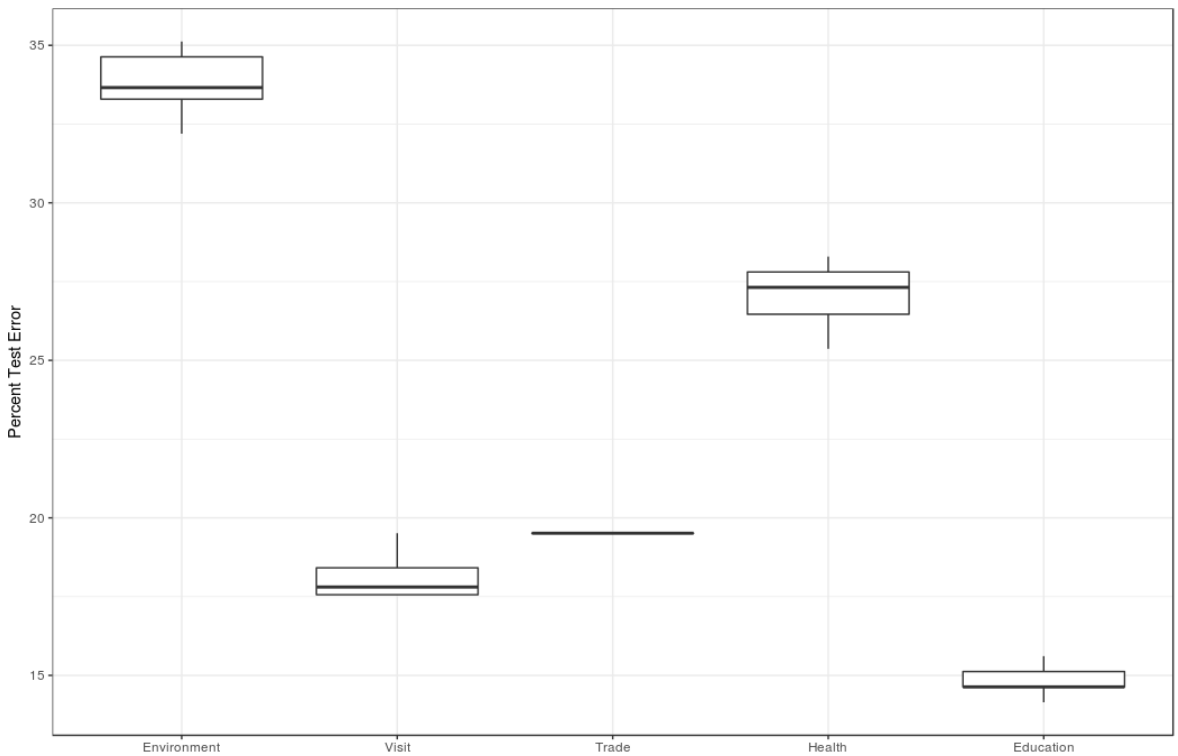


Figure 11: Percent test errors for each random forest model of migration assessed in Survey 1. Test errors are calculated based on predictions of test data from models fitted with training data. The figure shows that the model of migration for education has the lowest test error, while the model of environmental migration has the highest percent test error. These differences represent that random forests' predictive abilities vary based on outcome variables and underlying patterns in data.

4.2 Results of Objective 2

4.2.1 Kaplan-Meier estimates

The results of the Kaplan-Meier Estimate of the survival function for the complete time-event data for internal migration are shown in **Figure 12**.

The Kaplan-Meier method can also be used to quickly view the effects of factored variables by splitting the data and fitting two separate survival models. The results of this comparison can be seen in **Figure 13** and **Figure 14**. **Figure 13** shows the estimates of the survival function for answers of "Yes" and "No" to the question "Do you own a refrigerator?". Similarly, **Figure 14** compares the survival functions for respondents who did and did not own a gas cooker.

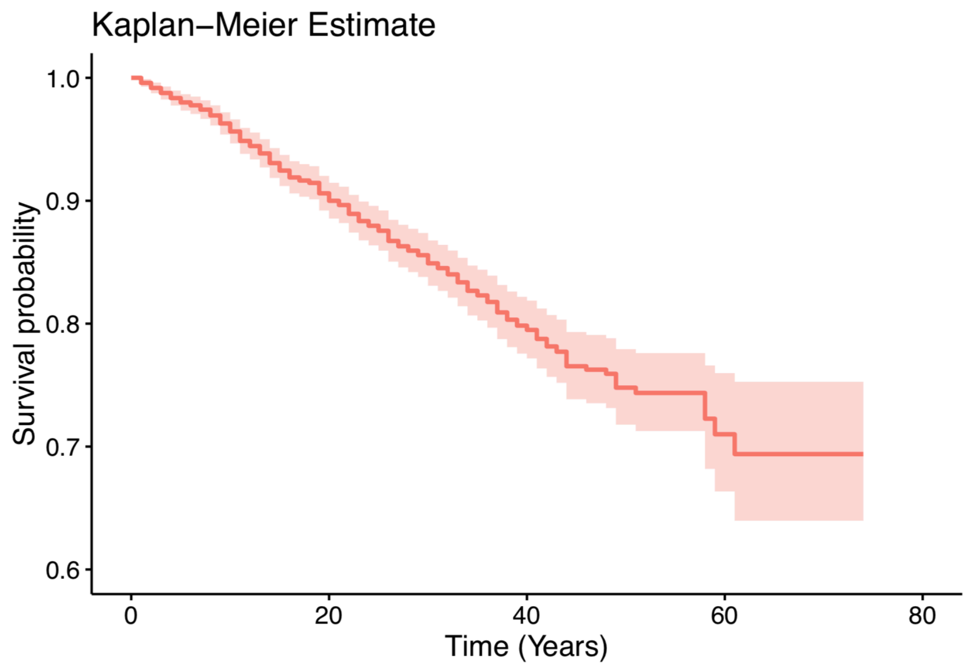


Figure 12: Kaplan-Meier estimate of survival probability, representing the probability of the head of household having not yet taken a first migration trip, over time (person-years)

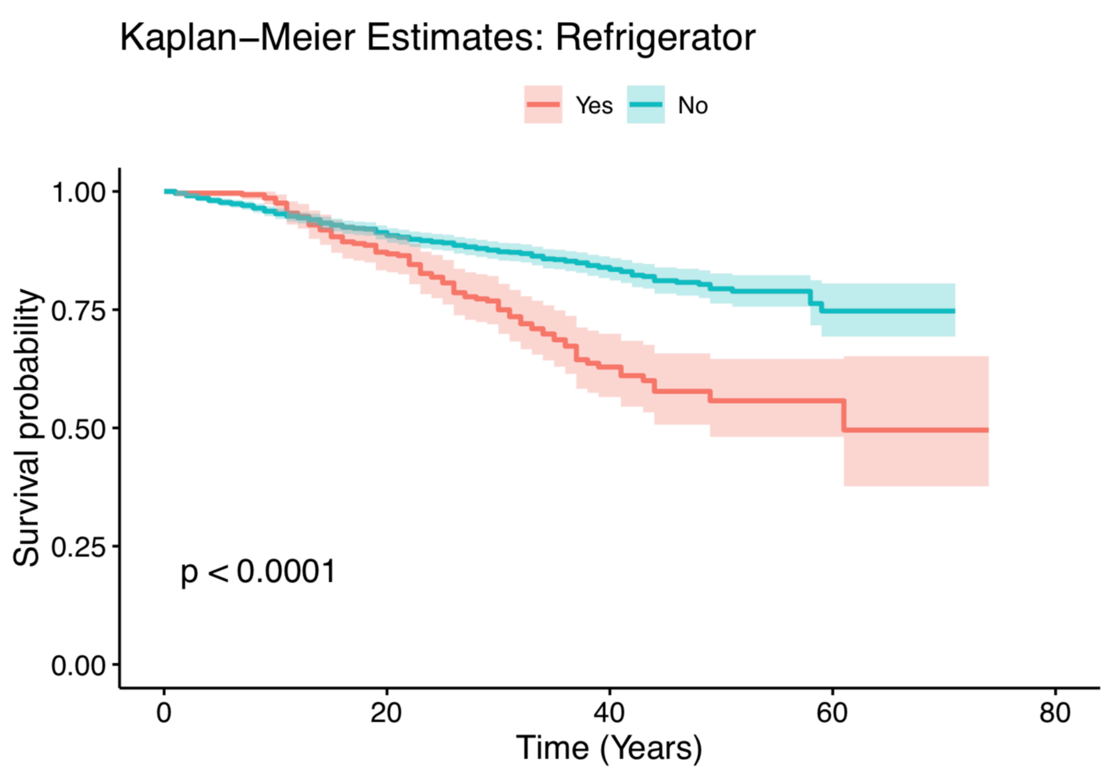


Figure 13: Kaplan-Meier estimates of survival probability over time (person-years) for households reporting “Yes” and “No” to owning a refrigerator.

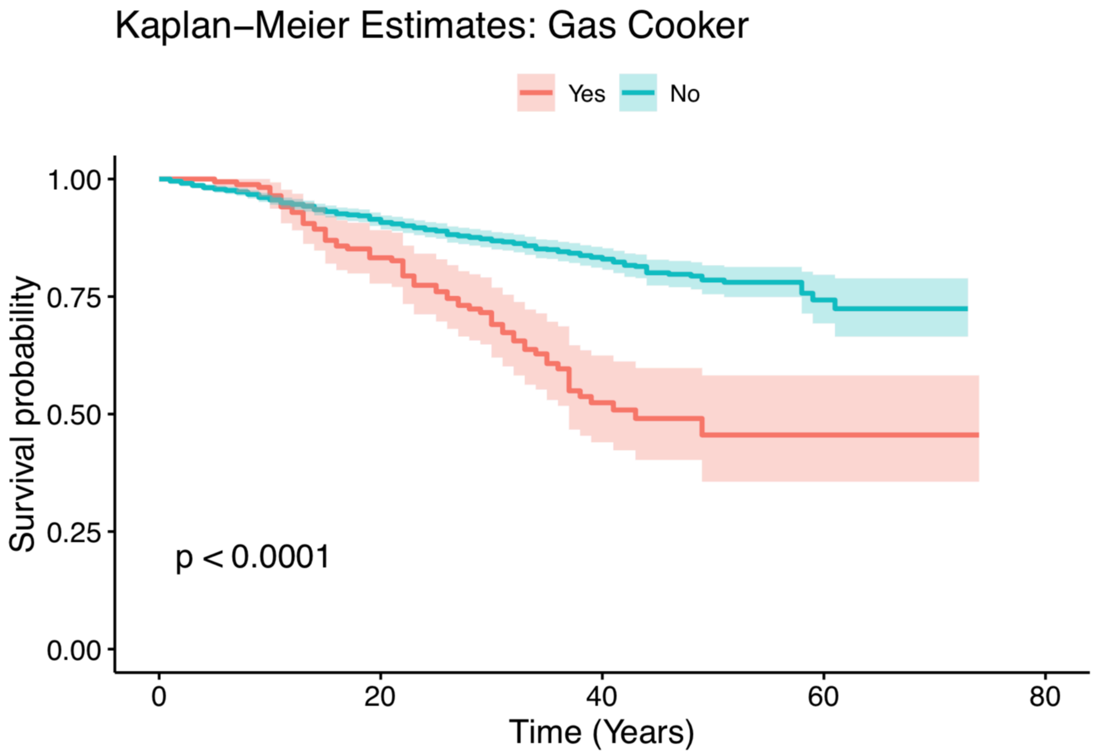


Figure 14: Kaplan-Meier estimates of survival probability over time (person-years) for households reporting "Yes" and "No" to owning a kerosene or gas cooker.

4.2.2 Cox proportional hazards models

Univariate Cox proportional hazards models were fit for each of the salient variables in **Table 3** identified by the random forest model. For each univariate model, **Table 4** shows the estimated value of the coefficient "Beta", the estimated hazard ratio (HR) and 95% confidence interval boundaries, the R^2 value, and the p-value.

The hazard ratio describes how a covariate impacts the hazard (whether it has a positive or negative effect) (Harrell 2015). The hazard ratio for a covariate is calculated by computing the ratio of the hazard for that covariate over the baseline hazard. Therefore, a hazard ratio of 1 indicates that the covariate has no effect on the hazard. A hazard ratio less than 1 means that the covariate reduces the hazard of an event, and a hazard ratio greater than 1 means that the covariate increases the hazard from the baseline. Hazard ratios for the univariate models are indicated in **Table 4**.

In **Table 4**, the R^2 value reported is a generalized R^2, estimated by

$$1 - exp(\chi_{LR}/n)$$

where χ_{LR} is the chi-square statistic for the likelihood ratio of the overall model, and n is the total number of observations.

Table 4: Results of univariate Cox proportional hazards models with each salient variable identified by the random forest models. For each univariate model, the fitted coefficient Beta is presented, along with the hazard ratio (HR) and 95% confidence intervals for HR, the generalized R^2, and p-value.

Variable	Beta	HR (95% CI for HR)	R^2	P-value
Latitude	-0.0012	0.999 (0.994-1)	0.000181	0.653
Business: year started	-0.000406	1 (0.999-1)	0.026	1.1e-10
Longitude	-0.000651	0.999 (0.995-1)	7.14e-05	0.75
Total members in household	-0.129	0.879 (0.831-0.93)	0.0133	7.45e-06
Kerosene/ gas cooker	-1.02	0.36 (0.274-0.472)	0.0256	1.94e-13
Refrigerator	-0.801	0.449 (0.351-0.574)	0.0211	1.71e-19
Cyclone: Source of water	0.902	2.22 (0.223-27.2)	0.0164	4.14e-05
Cyclone: Source of Food	0.232	-0.242 (0.874-1.82)	0.016	9.43e-05
Where would you go for future cyclone	-13.5	0.146 (0.853-1.57)	0.0111	0.000221
Who owns water source	-0.357	0.0357 (0.378-1.3)	0.00405	0.292
Spouse prepared meals consumed	1.14	3.11 (1.6-6.05)	0.00474	0.000801
Union	-0.0098	0.99 (0.986-0.994)	0.0131	3.38e-06
Non-workers	-0.0909	0.913 (0.859-0.971)	0.00532	0.00356
Toilet- female	0.226	0.646 (0.934-1.68)	0.0141	0.000154
Primary water source	0.698	NA	0.00889	0.0988

From these univariate models, it is apparent that the variables “Latitude”, “Longitude”, “Who owns water source”, and “Primary water source” are not significant. Furthermore, the uncertainty is so large, as shown by the range of the confidence intervals for the variables related to the most recent cyclone and female toilet, that they are omitted from the continued analysis. Because the 95% confidence intervals for these variables cross the hazard ratio value of 1, it cannot be reliably assumed that these variables will impact the survival function.

Next, a series of nested Cox proportional hazards models were developed with the remaining variables by starting with a univariate model and systematically adding an additional significant covariate to the model. R^2 values for these models are reported in **Table 5**. Analysis of variance (ANOVA) tests show that each subsequent model with an additional variable is significantly different from the previous model. The hazard ratios for the covariates of the complete model are shown in **Figure 15**.

Table 5: Nested Cox proportional hazards models of increasing complexity and generalized R^2.

Model	$\mathbf{R^2}$
Business	0.026
Business + Household members	0.047
Business + Household members + Refrigerator	0.057
Business + Household members + Refrigerator + Stove	0.06
Business + Household members + Refrigerator + Stove + Union	0.063
Business + Household members + Refrigerator + Stove + Union + Non-workers	0.066
Business + Household members + Refrigerator + Stove + Union + Non-workers + Spouse prepared meals	0.069

Finally, from the output of the complete Cox proportional hazard model, the survival function was estimated. A comparison of the survival model from the Kaplan-Meier estimator and the Cox proportional hazards model can be seen in **Figure 16**. **Figure 16** indicates that there is good agreement between the two models in their estimates of the overall survival function.

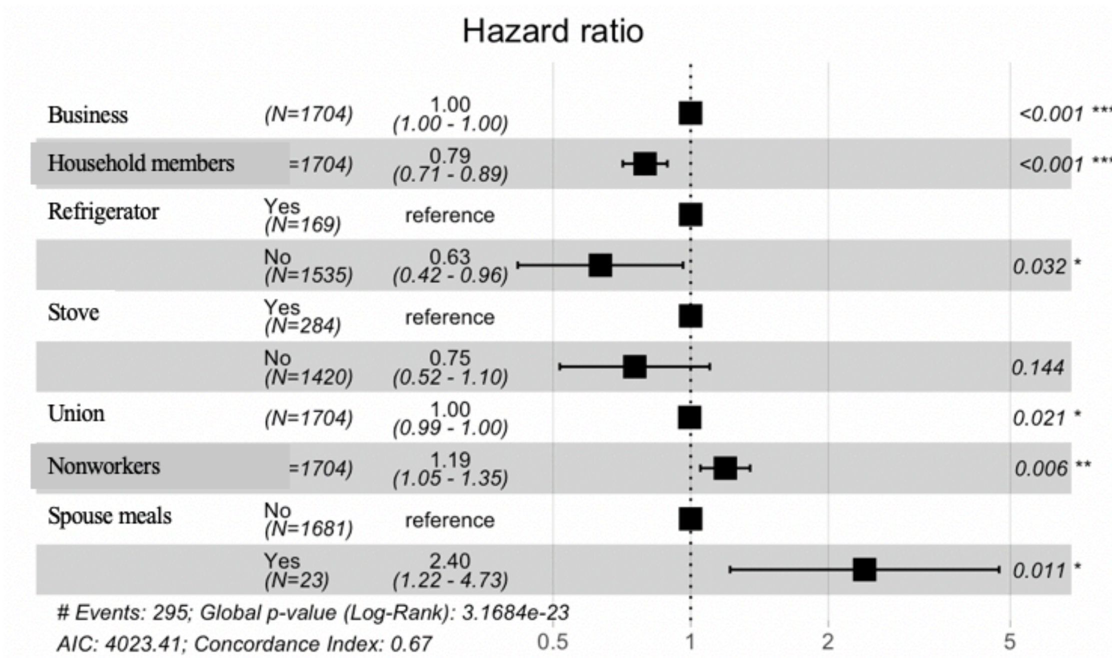


Figure 15: Hazard ratios for the final Cox-Proportional hazards model. A hazard ratio greater than 1 (to the right of the dashed line), indicates that the variable increases mobility, while a hazard ratio less than 1 (to the left of the dashed line) indicates that the variable decreases mobility.

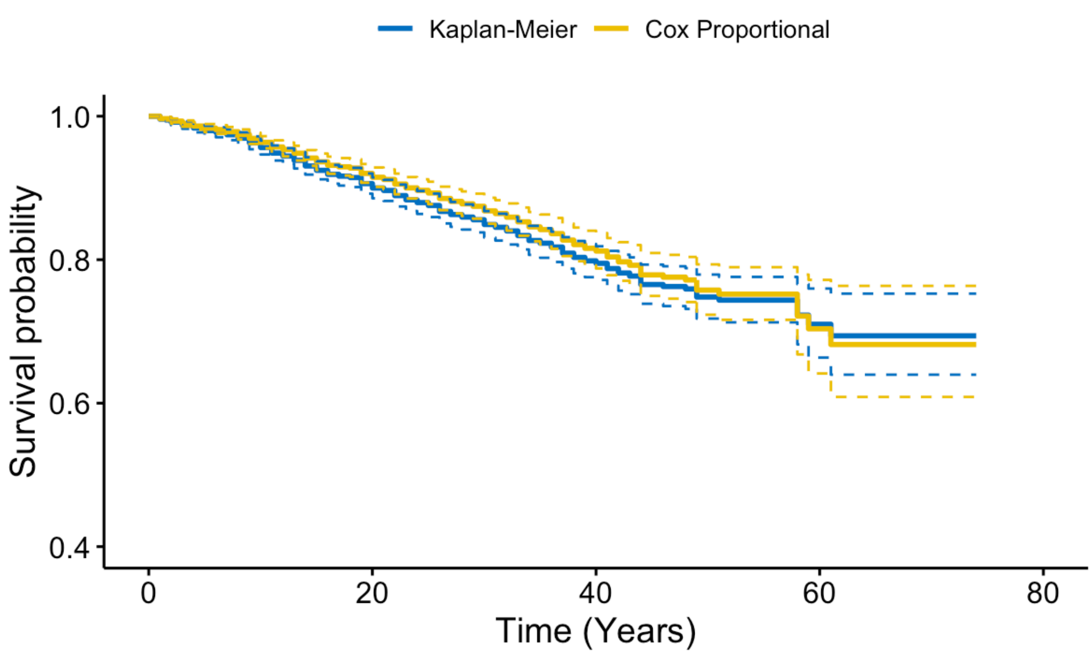


Figure 16: Comparison of Kaplan-Meier estimation and Cox proportional hazards estimation of survival probability over time.

Chapter 5

Discussion

5.1 Discussion of Objective 1

5.1.1 Insights from Survey 1

The analysis of variable importance from random forest models reveals similarities and differences between the variables associated with different types of migration in Survey 1: environmental migration, migration for education, migration for trade, migration for health, and migration to visit relatives. For all five of the models, possible proxies for wealth or socio-economic status such the amount of homestead land owned were among the most important variables that influence the migration outcome variable. The material of the respondent's home was important in models of four of the five types of migration. Previous research has indicated that livelihood and economic opportunity can greatly motivate or limit mobility (Adger et al. 2015; Bennett et al. 2011). Perceived issues in the community were also important across all of the models ("Biggest problem in community", "Second biggest problem in community", and "Third biggest problem in community"), suggesting that perceptions and satisfaction with one's home are also importantly associated with migration, regardless of the dominant motivation. Birth year of the head of household was also important for many of the models, suggesting that age is likely an important personal factor associated with migration decisions.

Assessing the individual models more closely provides additional insights into the differences between the models of migration as a result of varying self-reported motivations. **Figure 17** shows high level differences between models based on the frequency of important variables categorized by theme.

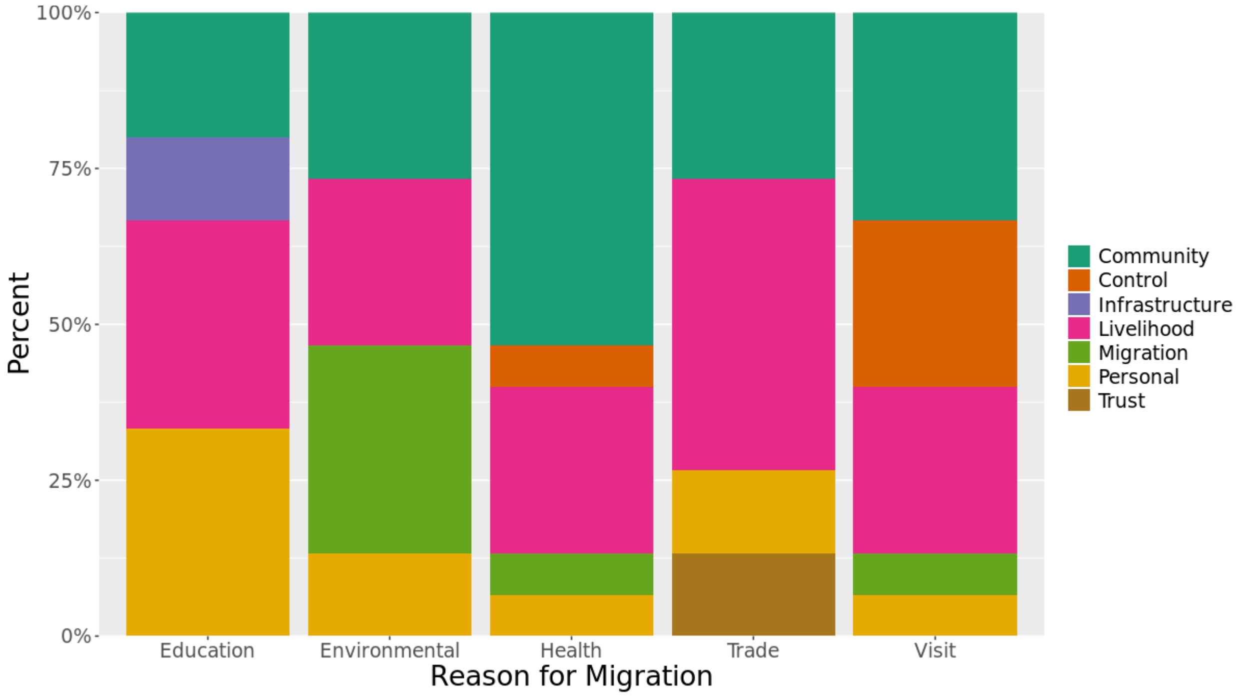


Figure 17: Overview of categories of variables present in top 15 variables of importance for models of migration assessed with Survey 1. The legend shows the colors corresponding to the categories of variables.

From this analysis, it is apparent that the response variable for environmental migration is uniquely associated with knowing others who have also migrated for environmental reasons, as the first two most important variables reflect this. Past research has shown that the barrier to migrate can be significantly lowered by potential migrants having social connections with others who have migrated in the past (Black et al. 2011; Haug 2008; Hunter et al. 2013). It is also noteworthy that there were not any explicitly environmental variables amongst the most important variables for environmental migration. This reinforces the common understanding that even when environmental pressures impact migration, they are rarely the only driver (Obokata et al. 2014).

In contrast to environmental migration, the model of migration for education is uniquely influenced by variables that relate to household education level, such as the annual household expenditures on education and the number of household members who attended college and university. In addition to these variables related to education, the models have identified variables related to socioeconomic status and access to infrastructure, such as tenancy of agricultural land and distance to a cyclone shelter as important variables.

Other nuances in the important variables in the models of migration for health care, migration for trade, and migration to visit relatives further demonstrate random forest's ability to identify nuances and complex relationships in data. For instance, migration for trade or commerce is highly impacted by the location of the respondent's primary source of livelihood. Migration for health is uniquely influenced by factors that reflect community level conditions, such as the minimum level of education that a community member would be able to obtain. Migration to visit relatives is uniquely influenced highly by variables related to locus of control such as the respondent's faith in their ability to plan ahead and ability to express themselves during elections.

The results of the percent test error for the random forest models for each type of migration may also be telling. Migration for education has the lowest percent test error, suggesting that this type of migration is easier for a model to predict. Environmental migration, in contrast, has the highest percent test error, followed by migration for health care. There seems to be a clear divide between models to predict migration for education, migration to visit relatives, and migration for trade performing relatively well, while models for environmental migration and migration for health care performed more poorly. One possible explanation is that environmental events or health challenges are "push" factors, or negative conditions in a community that may push someone to move to another location. Education, visiting relatives, and economic opportunities, however, are "pull" factors, or positive conditions in a destination location that draw someone to move (Amrith 2013; Hassani-Mahmooei 2012). It would seem, therefore, that it is more difficult to predict migration driven by "push" factors, perhaps because they are things that are less likely to be planned for, but are rather events that happen to a potential migrant.

5.1.2 Insights from Survey 2

Results from the analysis of Survey 2 also demonstrate random forests' ability to provide insights into the important variables associated with migration in Bangladesh, this time where the outcome variable is continuous rather than dichotomous. Here, the first and fourth most important variables are latitude and longitude, which are important even when controlling for the survey community. This suggests that latitude and longitude may be representing environmental conditions that vary spatially. For example, it is known that soil salinity varies strongly on a north-south gradient, and salinity has been shown to be an important factor influencing migration in Bangladesh (Chen and Mueller 2018). The second most important variable reflects ownership of the household's primary water source, and primary water source is also important, which could further support the idea that latitude is representing challenges related to water quality and salinity or that water quality issues are important for migration.

In addition to latitude and longitude, several variables that suggest socio-economic status are important in the model. For example, the year a business was started as well as owning a refrigerator or a gas cooker are all proxies for socioeconomic level. Toilet facilities used by female members of the household, and whether or not the spouse of the household head consumes prepared meals may also reflect socioeconomic status. Toilet facilities specifically may impact health for both women and children in a household (Amin et al. 2010; Hong et al. 2006). In addition, it has been established that there are gender disparities in food security in Bangladesh, the extent of which may reflect gender empowerment, education, household wealth, and household employment (Sraboni et al. 2014). Much like the results from Survey 1, this supports the common understanding that livelihood and economic opportunity can greatly impact mobility (Adger et al. 2015; Bennett et al. 2011). These results from Survey 2 suggest that household composition is also important to mobility. Total members in the household as well as the number of non-workers in the household are both important, even when controlling for these things explicitly by normalizing number of migrations by person-years.

Finally, several variables related to the most recent cyclone were important in this analysis, including sources of food and water during the latest cyclone, as well as where a household would go in the event of a future cyclone. As previously mentioned, cyclones have been studied as a possible driver of migration in Bangladesh (Lu et al. 2013; Mallick and Vogt 2014). As one example, Mallick and Vogt found that that male household members migrated towards cities in order to access livelihood opportunities after the cessation of emergency aid after the 2009 cyclone Aila (Mallick and Vogt 2014).

5.1.3 Application of random forests to migration

Random forest models are promising tools for predicting migration from a large collection of covariates. Assessed by the accuracy of out-of-sample predictions, random forests outperformed logistic regressions and SVM models. This is likely due to the fact that tree-based models allow us to identify nonlinear, nonmonotonic, and even multimodal interactions between variables that cannot be effectively modeled with generalized linear models or SVMs. As previously mentioned, predictive ability is critical for informing future climate policy and adaptation strategies that aim to address migration (Ahsan et al. 2011; Black et al. 2011; Biermann et al. 2010; Stern 2006). Where prediction might be more important than understanding underlying drivers, especially when providing information to policymakers, random forest should be explored further as a possible tool. This insight suggests that as modelers and researchers continue to work to improve their ability to predict migration, random forest models should be included in future analysis. However, modelers should continue to develop more sophisticated methods, as these results show that even the random forest model still has limited predictive accuracy. It is possible that the limited predictive power of the random forest models was because the model did not explicitly include push factors (such as cyclones or illnesses) in models of migrations driven by push factors, such as migration for environmental reasons and for health. The inclusion of push factors in a future analysis could improve model predictive accuracy.

This work also demonstrates that random forest models can help researchers identify salient variables from large social surveys when studying migration. This is especially useful when dealing with large, complex datasets from social surveys, where it can be challenging to decide which variables are worthwhile for further investigation. For both the case of categorical outcome variables and a continuous outcome, random forest models were able to identify the most important predictors of migration from an original set of more than 1,200 total predictors. From these top important variables, I was able to provide insights into the underlying patterns in the datasets and thus identify nuances in the drivers of different kinds of migration in southwestern Bangladeshi communities.

5.2 Discussion of Objective 2

5.2.1 Variable impact on mobility

This work also helps to illuminate how salient variables related to location, livelihood, and family structure might impact a household's risk of internal migration in coastal Bangladeshi communities. The univariate Cox proportional hazards models outlined in **Table 3** demonstrate that the number of members in a household, the year a business is started, whether or not the household owns a refrigerator, and whether or not the household owns a gas cooker. Latitude, longitude, and variables related to the most recent cyclone were not significant predictors of the hazard function or reflected too much uncertainty to be reliable.

It is especially surprising that latitude and longitude were not significant covariates, because they were the first and third most important variables identified by the previous work using random forest algorithms. It was thought that latitude especially would be significant, because there is a clear gradient of increasing soil salinity from north to south in Bangladesh, and previous studies have suggested that soil salinity is important for driving migration in Bangladesh (Chen and Mueller 2018). It is possible that the random forest algorithm is able to identify patterns in the latitude and longitude data that are more complex than the linear relationship that the Cox proportional hazards model assumes. For example, the random forest algorithm would be able to identify geographic clusters of migration, which the Cox proportional hazards model would not.

The overall Kaplan-Meier estimator of survival probability (**Figure 12**) shows that the survival probability decreases between 1 and slightly less than 0.7 over a range of time from 0 to approximately 75 years. Additional Kaplan-Meier estimators (**Figures 13, 14**) indicate that families that own a refrigerator or own a gas cooker are at a significantly higher risk over time of making a first internal trip. This result is not surprising, as previous work demonstrates that livelihood factors are extremely important for a household's ability to migrate (Bennett and Beddington 2011; Islam et al. 2017; Qin et al. 2010). Wealthier households (such as those that could afford to have a refrigerator or a gas cooker) are commonly able to move more frequently, while poorer households might be "trapped" and immobile (Penning-Rowsell et al. 2013).

From the variables that were significant in the univariate models, a nested multi-variate Cox proportional hazards model was developed. The best performing model was the complete model with year business was started, whether or not the household owned a refrigerator, whether or not the household owned a gas cooker, total members in the household, union, and whether or not the spouse of the household head consumes prepared meals. This final model had a generalized R^2 value of 0.069 (**Table 4**). It is possible that this value of R^2 is so low because, again, the covariates are unlikely to follow a simple linear relationship assumed by the Cox proportional hazards model.

Despite the low value of R^2, the multi-variate Cox proportional hazards model is useful in beginning to understand how these variables influence the underlying risk of migrating. The values of hazard ratios shown in **Figure 15** quantify these impacts. The hazard ratios to the left of the dotted line in the figure show the variables have a negative impact on the overall risk of migration. This means that these variables decrease the underlying hazard. These variables include total household members, not owning a refrigerator, and not owning a gas or kerosene cooker. Hazard ratios that fall to the right of the dotted line in **Figure 15** show variables that have a positive impact on migration, meaning they increase the underlying hazard of migration. These variables are the number of non-workers in the household and that the spouse of the household head consumes prepared meals.

It has been established that there are gender disparities in food security in Bangladesh, the extent of which may reflect gender empowerment, education, household wealth, and household employment (Sraboni et al. 2014). Similar to the ownership of a refrigerator or a gas/ kerosene stove, whether or not a spouse consumes prepared meals may reflect household socioeconomic status. These results are therefore consistent with the results related to owning a refrigerator or gas stove, as they suggest that a higher level of household wealth will contribute to an increased hazard of taking a first migration trip. These results are possibly insightful on another level if spousal food security is a representation of female empowerment, as Sraboni et al. suggest (2014). In this interpretation, higher female empowerment would contribute to increased risk of migration by a male head of household. Previous work has suggested that male labor migration in Bangladesh may contribute to increased female empowerment in decision making (IOM), but there is not a clear way to describe the connection between female empowerment and household head migration in the opposite causal direction.

These results show that the total number of members of a household has a negative impact on migration, while number of non-workers in the household seems to increase migration by the household head. This, therefore, provides evidence that complicates the hypothesis posited from just the random forest results that a larger family may increase pressure on a head of household to migrate in order to support the family. It is possible that this is true in the case of non-workers, reinforcing the importance of remittances that migratory members of a household can send home to support their families (Massey 1990; Edwards and Ureta 2003). A household with a higher number of non-workers to support may be more dependent on remittances from a migratory head of household. Because this analysis uses data collected about the household after the migration has occurred, these results could also suggest that a household with a successful migrant returning remittances can afford to have more non-workers in the household. This analysis cannot conclude which causal direction is correct. However, it seems that larger households may also create an anchoring effect that keeps the head of household from migrating, perhaps because migrating from the household, even temporarily, would leave the household more vulnerable and economically stressed. This suggests that household size has a complex effect on probabilities of migration which reflects household livelihood capacity as well as vulnerability.

Conclusions

This work confirms that random forests were useful to identify salient variables in both ethnosurveys related to migration in Bangladesh, and they were able to identify nuances between different forms of migration. Additionally, random forests were shown to have superior predictive ability to other algorithms, including simple logistic regression.

In relation to **Objective 2**, the results were slightly more mixed. Though survival analysis was useful to identify the directionality of factors contributing to migration, the low values of R^2 raise questions as to the reliability of the analysis. This work did show that variables related to livelihood, including whether or not a household owned a refrigerator and a gas stove, did positively impact migration. As discussed, this result supports previous work which demonstrates that livelihood factors are extremely important for a household's ability to migrate (Bennett and Beddington 2011; Islam et al. 2017; Qin et al. 2010). Wealthier households (such as those that could afford to have a refrigerator or a gas cooker) are commonly able to move more frequently, while poorer households might be "trapped" and immobile (Penning-Rowsell et al. 2013).

Overall, this analysis reinforces that the decision to migrate, for any motivation, is highly complex. Because of the complexity of the problem, random forest models can be useful tools for researchers studying migration, especially environmental migration, where the theory is not clearly established or varies from one place (or context) to another. One downside of the random forest models, however, is that though they can quantify variable importance, they do not provide insights into the directionality or scale of individual predictors on the outcome variable. For example, from random forest outputs, it is shown that latitude is important to predicting normalized total number of household internal migration trips, but there is not a simple relationship, such as greater rates of migration at higher latitudes, and the more complex relationship discovered by the random forest models is not available for easy inspection. A combination of theory and traditional regression methods may be more appropriate, once important variables are identified, in identifying more directly how those variables impact migration. In this way, random forest models are not the final answer to assessing or modeling environmentally induced migration, but can serve as a first step for researchers to provide insights into their datasets, inform hypotheses, or support theories.

To begin to address this need, this work also demonstrates that survival analysis can be a useful tool for studying and quantifying the risks of migration over time. Because migration is highly complex, non-parametric and semi- parametric methods are useful, as the underlying hazard is unlikely to follow a known functional form. Non-parametric Kaplan-Meier estimators can provide insights into overall survival and hazard functions, while semi-parametric Cox proportional hazards models can provide specific insight into how covariates of interest impact the hazard. Here, the impact of covariates on risk is of higher interest than the underlying form of the hazard function, as these insights are more valuable for identifying risk factors of migration.

This analysis provides insights into migration dynamics, but it does not begin to accurately quantify migration risks. Assumptions of linearity and challenges in identifying important variables contribute to low predictive ability. Low values of R^2 could also be in part due to the high degree of censored versus uncensored data used in this analysis. It is possible that the data is zero-inflated, and that different processes are influencing the households that never move and those who do move. As this work advances, it will be important to shift from an exploratory analysis to one that focuses on prediction. This will be important in providing rigorous information to decisionmakers and stakeholders about how future socioeconomic pressures and environmental change will impact migration numbers on a more aggregate level.

Future work should continue to develop modeling methods that are able to assess and explain the complex relationship between environmental, economic, and political factors and how they contribute to migration decisions. Predictive accuracy should remain a priority for future research, but this will require a fundamental understanding of the systems at play to best inform model development and selection. Therefore, future work should also seek to understand feedbacks between migration decisions, the environment, and demographics (Gray 2014; Hugo 2011; Shayegh 2017).

In addition, studying complex, multifaceted systems such as human migration and climate change requires interdisciplinary research teams and an openness to collaboration (Speelman et al. 2017). Future efforts should incorporate multidisciplinary teams of social scientists, environmental scientists, psychologists, and data scientists to attempt to understand the factors that influence human migration simultaneously, as a system, rather than in isolation. In this way, all available tools, including machine learning algorithms like random forest models, can be applied to understanding the complex and important phenomenon of environmental migration.

References

Abel, G. J., Brottrager, M., Crespo Cuaresma, J., & Muttarak, R. (2019). Climate, conflict and forced migration. *Global Environmental Change*, *54*, 239–249. https://doi.org/10.1016/j.gloenvcha.2018.12.003

Ackerly, B. A., Anam, M. M., & Gilligan, J. (2015). Environment, political economies and livelihood change. In B. Mallick & B. Etzold (Eds.), *Environment, Migration and Adaptation: Evidence and Politics of Climate Change in Bangladesh*. Dhaka, Bangladesh: AH Development Publishing House (AHDPH). Retrieved from http://eprints.qut.edu.au/84192/

Ackerly, B.A., Anam, M. M., Gilligan, J. & Goodbred, S. (2017). Climate and community: The human rights, livelihood, and migration impacts of climate change, in Dimitra Manou, D., Baldwin, A., Cubie, D., Mihr, A., & Thorp, T. M. (Eds.). *Climate change, migration and human rights: Law and policy perspectives*. Milton Park, Abingdon, Oxon; New York, NY: Routledge.

Adams, H., & Kay, S. (2019). Migration as a human affair: Integrating individual stress thresholds into quantitative models of climate migration. *Environmental Science & Policy*, *93*, 129–138. https://doi.org/10.1016/j.envsci.2018.10.015

Adger, W. N., Arnell, N. W., Black, R., Dercon, S., Geddes, A., & Thomas, D. S. G. (2015). Focus on environmental risks and migration: causes and consequences. *Environmental Research Letters*, *10*(6), 060201. https://doi.org/10.1088/1748-9326/10/6/060201

Afsar, R. (2003). Internal migration and the development nexus: the case of Bangladesh. In *Regional Conference on Migration, Development and Pro-Poor Policy Choices in Asia* (pp. 22–24).

Ahmed, N., Islam, M. N., Hasan, M. F., Motahar, T., & Sujauddin, M. (2018). Understanding the political ecology of forced migration and deforestation through a multi-algorithm classification approach: The case of Rohingya displacement in the southeastern border region of Bangladesh. *Geology, Ecology, and Landscapes*, *0*(0), 1–13. https://doi.org/10.1080/24749508.2018.1558025

Ahsan, R., Karuppannan, S., & Kellett, J. (2011). Climate Migration and Urban Planning System: A Study of Bangladesh. *Environmental Justice*, *4*(3), 163–170. https://doi.org/10.1089/env.2011.0005

Alam, G. M. M., Alam, K., & Mushtaq, S. (2017). Climate change perceptions and local adaptation strategies of hazard-prone rural households in Bangladesh. *Climate Risk Management*, *17*, 52–63. https://doi.org/10.1016/j.crm.2017.06.006

Alauddin, M., & Sharma, B. R. (2013). Inter-district rice water productivity differences in Bangladesh: An empirical exploration and implications. *Ecological Economics*, *93*, 210–218. https://doi.org/10.1016/j.ecolecon.2013.05.015

Amin, R., Shah, N. M., & Becker, S. (2010). Socioeconomic factors differentiating maternal and child health-seeking behavior in rural Bangladesh: A cross-sectional analysis. *International Journal for Equity in Health*, *9*(1), 9. https://doi.org/10.1186/1475-9276-9-9

Amrith, S. S. (2013). *Crossing the Bay of Bengal: the furies of nature and the fortunes of migrants*. Cambridge, Mass. London, England: Harvard University Press.

Ansell, J. I., & Philipps, M. J. (1997). Practical aspects of modelling of repairable systems data using proportional hazards models. *Reliability Engineering & System Safety*, *58*(2), 165–171. https://doi.org/10.1016/S0951-8320(97)00026-4

Auerbach, L. W., Goodbred Jr, S. L., Mondal, D. R., Wilson, C. A., Ahmed, K. R., Roy, K., … Ackerly, B. A. (2015). Flood risk of natural and embanked landscapes on the Ganges–Brahmaputra tidal delta plain. *Nature Climate Change*, *5*(2), 153–157. https://doi.org/10.1038/nclimate2472

Azad, A. K., Jensen, K. R., & Lin, C. K. (2009). Coastal Aquaculture Development in Bangladesh: Unsustainable and Sustainable Experiences. *Environmental Management*, *44*(4), 800–809. https://doi.org/10.1007/s00267-009-9356-y

Bailey, A. J. (1993). Migration history, migration behavior and selectivity. *The Annals of Regional Science*, *27*(4), 315–326. https://doi.org/10.1007/BF01583571

Barabadi, A., Barabady, J., & Markeset, T. (2011). Maintainability analysis considering time-dependent and time-independent covariates. *Reliability Engineering & System Safety*, *96*(1), 210–217. https://doi.org/10.1016/j.ress.2010.08.007

Barabadi, A., Barabady, J., & Markeset, T. (2014). Application of reliability models with covariates in spare part prediction and optimization – A case study. *Reliability Engineering & System Safety*, *123*, 1–7. https://doi.org/10.1016/j.ress.2013.09.012

Barbieri, A. F., & Carr, D. L. (2005). Gender-specific out-migration, deforestation and urbanization in the Ecuadorian Amazon. *Global and Planetary Change*, *47*(2), 99–110. https://doi.org/10.1016/j.gloplacha.2004.10.005

Barker, K., & Baroud, H. (2014). Proportional hazards models of infrastructure system recovery. *Reliability Engineering & System Safety*, *124*, 201–206. https://doi.org/10.1016/j.ress.2013.12.004

Barrios, S., Bertinelli, L., & Strobl, E. (2006). Climatic change and rural–urban migration: The case of sub-Saharan Africa. *Journal of Urban Economics*, *60*(3), 357–371. https://doi.org/10.1016/j.jue.2006.04.005

Bell, A. R. (2017). Informing decisions in agent-based models — A mobile update. *Environmental Modelling & Software*, *93*, 310–321. https://doi.org/10.1016/j.envsoft.2017.03.028

Bell, A., Ward, P., Tamal, Md. E. H., & Killilea, M. (2019). Assessing recall bias and measurement error in high-frequency social data collection for human-environment research. *Population and Environment*, *40*(3), 325–345. https://doi.org/10.1007/s11111-019-0314-1

Bennett, G., Thomas, S. M., & Beddington, J. R. (2011). Migration as adaptation. *Nature*, *478*, 447–449.

Benneyworth, L., Gilligan, J., Ayers, J. C., Goodbred, S., George, G., Carrico, A., … Piya, B. (2016). Drinking water insecurity: water quality and access in coastal south-western Bangladesh. *International Journal of Environmental Health Research*, *26*(5–6), 508–524. https://doi.org/10.1080/09603123.2016.1194383

Bhatta, G. D., Aggarwal, P. K., Poudel, S., & Belgrave, D. A. (2016). Climate-induced migration in South Asia: Migration decisions and the gender dimensions of adverse climatic events. *Journal of Rural and Community Development*, *10*(4).

Bhuiyan, M. A. H., Islam, S. M. D.-U., & Azam, G. (2017). Exploring impacts and livelihood vulnerability of riverbank erosion hazard among rural household along the river Padma of Bangladesh. *Environmental Systems Research*, *6*(1). https://doi.org/10.1186/s40068-017-0102-9

Biermann, F., & Boas, I. (2010). *Preparing for a Warmer World: Towards a Global Governance System to Protect Climate Refugees* (Vol. 10). https://doi.org/10.1162/glep.2010.10.1.60

Bilsborrow, R. E., & Henry, S. J. F. (2012). The use of survey data to study migration–environment relationships in developing countries: alternative approaches to data collection. *Population and Environment*, *34*(1), 113–141. https://doi.org/10.1007/s11111-012-0177-1

Black, R., Adger, W. N., Arnell, N. W., Dercon, S., Geddes, A., & Thomas, D. (2011). The effect of environmental change on human migration. *Global Environmental Change*, *21*, S3–S11. https://doi.org/10.1016/j.gloenvcha.2011.10.001

Black, R., Arnell, N. W., Adger, W. N., Thomas, D., & Geddes, A. (2013). Migration, immobility and displacement outcomes following extreme events. *Environmental Science & Policy*, *27*, S32–S43. https://doi.org/10.1016/j.envsci.2012.09.001

Black, R., Kniveton, D., & Schmidt-Verkerk, K. (2011). Migration and Climate Change: Towards an Integrated Assessment of Sensitivity. *Environment and Planning A*, *43*(2), 431–450. https://doi.org/10.1068/a43154

Black, R., Kniveton, D., Skeldon, R., Coppard, D., Murata, A., & Schmidt-Verkerk, K. (2008). Demographics and climate change: Future trends and their policy implications for migration. *Development Research Centre on Migration, Globalisation and Poverty. Brighton: University of Sussex*.

Black, R., Natali, C., & Skinner, J. (2005). *Migration and inequality*. World Bank Washington, DC.

Boas, I. (2017). Environmental change and human mobility in the digital age. *Geoforum*, *85*, 153–156. https://doi.org/10.1016/j.geoforum.2017.07.022

Boas, I., Dahm, R., & Wrathall, D. (n.d.). Grounding Big Data on Climate-Induced Human Mobility. *Geographical Review*, *0*(0). https://doi.org/10.1111/gere.12355

Bohra-Mishra, P., Oppenheimer, M., Cai, R., Feng, S., & Licker, R. (2017). Climate variability and migration in the Philippines. *Population and Environment*, *38*(3), 286–308. https://doi.org/10.1007/s11111-016-0263-x

Bohra-Mishra, P., Oppenheimer, M., & Hsiang, S. M. (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(27), 9780–9785. https://doi.org/10.1073/pnas.1317166111

Brammer, H. (2014). Bangladesh's dynamic coastal regions and sea-level rise. *Climate Risk Management*, *1*, 51–62. https://doi.org/10.1016/j.crm.2013.10.001

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, *82*(5), 1671–1748. https://doi.org/10.3982/ECTA10489

Bull, K., & Spiegelhalter, D. J. (1997). Survival analysis in observational studies. *Statistics in Medicine*, *16*(9), 1041–1074.

Buolawmini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proc. Machine Learning Res. 81*, 77-91. http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline

Burrows, K., & Kinney, P. L. (2016). Exploring the Climate Change, Migration and Conflict Nexus. *International Journal of Environmental Research and Public Health*, *13*(4). https://doi.org/10.3390/ijerph13040443

Cai, R., Feng, S., Oppenheimer, M., & Pytlikova, M. (2016). Climate variability and international migration: The importance of the agricultural linkage. *Journal of Environmental Economics and Management*, *79*, 135–151. https://doi.org/10.1016/j.jeem.2016.06.005

Cai, R., & Oppenheimer, M. (2013). An Agent-based model of climate-induced agricultural labor migration. In *2013 Annual meeting, August* (pp. 4–6).
Call, M. A., Gray, C., Yunus, M., & Emch, M. (2017). Disruption, not displacement: Environmental variability and temporary migration in Bangladesh. *Global Environmental Change*, *46*, 157–165. https://doi.org/10.1016/j.gloenvcha.2017.08.008
Cattaneo, C., & Peri, G. (2016). The migration response to increasing temperatures. *Journal of Development Economics*, *122*, 127–146. https://doi.org/10.1016/j.jdeveco.2016.05.004
Chen, J., & Mueller, V. (2018). Coastal climate change, soil salinity and human migration in Bangladesh. *Nature Climate Change*. https://doi.org/10.1038/s41558-018-0313-8
Coffey, D., Papp, J., & Spears, D. (2015). Short-Term Labor Migration from Rural North India: Evidence from New Survey Data. *Population Research and Policy Review*, *34*(3), 361–380. https://doi.org/10.1007/s11113-014-9349-2
Crowley, J., & Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association*, *72*(357), 27–36. https://doi.org/10.1080/01621459.1977.10479903
Cutler, F. original by L. B. and A., & Wiener, R. port by A. L. and M. (2018). randomForest: Breiman and Cutler's Random Forests for Classification and Regression (Version 4.6-14). Retrieved from https://CRAN.R-project.org/package=randomForest
Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1
Davis, K. F., Bhattachan, A., D'Odorico, P., & Suweis, S. (2018). A universal model for predicting human migration under climate change: Examining future sea level rise in Bangladesh. *Environmental Research Letters*, *13*(6), 064030. https://doi.org/10.1088/1748-9326/aac4d4
DeAngelis, D. L., & Diaz, S. G. (2019). Decision-Making in Agent-Based Modeling: A Current Review and Future Prospectus. *Frontiers in Ecology and Evolution*, *6*. https://doi.org/10.3389/fevo.2018.00237
De Jong, G. F. (2000). Expectations, gender, and norms in migration decision-making. *Population Studies*, *54*(3), 307–319. https://doi.org/10.1080/713779089
Dewan, A. M., Islam, M. M., Kumamoto, T., & Nishigaki, M. (2007). Evaluating Flood Hazard for Land-Use Planning in Greater Dhaka of Bangladesh Using Remote Sensing and GIS Techniques. *Water Resources Management*, *21*(9), 1601–1612. https://doi.org/10.1007/s11269-006-9116-1
Dillon, A., Mueller, V., & Salau, S. (2011). Migratory Responses to Agricultural Risk in Northern Nigeria. *American Journal of Agricultural Economics*, *93*(4), 1048–1061. https://doi.org/10.1093/ajae/aar033
Donato, K. M., & Massey, D. S. (2016). Twenty-First-Century Globalization and Illegal Migration. *The ANNALS of the American Academy of Political and Social Science*, *666*(1), 7–26. https://doi.org/10.1177/0002716216653563
Donato, K. M., Massey, D. S., Donato, K. M., Carrico, A. R., Sisk, B., & Piya, B. (2016). Different but the Same: How Legal Status Affects International Migration from Bangladesh. *The ANNALS of the American Academy of Political and Social Science*, *666*(1), 203–218. https://doi.org/10.1177/0002716216650843

Dustmann, C., & Okatenko, A. (2014). Out-migration, wealth constraints, and the quality of local amenities. *Journal of Development Economics*, *110*, 52–63. https://doi.org/10.1016/j.jdeveco.2014.05.008
Edwards, A. C., & Ureta, M. (2003). International migration, remittances, and schooling: evidence from El Salvador. *Journal of Development Economics*, *72*(2), 429–461. https://doi.org/10.1016/S0304-3878(03)00115-9
Farris, S. R. (2010). Interregional Migration: The challenge for gender and development. *Development*, *53*(1), 98–104. https://doi.org/10.1057/dev.2009.89
Feng, S., Krueger, A. B., & Oppenheimer, M. (2010). Linkages among climate change, crop yields and Mexico-US cross-border migration. *Proceedings of the National Academy of Sciences*, *107*(32), 14257–14262. https://doi.org/10.1073/pnas.1002632107
Findley, S. E. (1994). Does drought increase migration? A study of migration from rural Mali during the 1983-1985 drought. *The International Migration Review*, *28*(3), 539–553.
Fischer, J., Gardner, T. A., Bennett, E. M., Balvanera, P., Biggs, R., Carpenter, S., … Tenhunen, J. (2015). Advancing sustainability through mainstreaming a social–ecological systems perspective. *Current Opinion in Environmental Sustainability*, *14*, 144–149. https://doi.org/10.1016/j.cosust.2015.06.002
Fussell, E., Hunter, L. M., & Gray, C. L. (2014). Measuring the environmental dimensions of human migration: The demographer's toolkit. *Global Environmental Change*, *28*, 182–191. https://doi.org/10.1016/j.gloenvcha.2014.07.001
Fussell, E., Sastry, N., & VanLandingham, M. (2010). Race, socioeconomic status, and return migration to New Orleans after Hurricane Katrina. *Population and Environment*, *31*(1–3), 20–42. https://doi.org/10.1007/s11111-009-0092-2
Galvin, K. A. (2009). Transitions: Pastoralists Living with Change. *Annual Review of Anthropology*, *38*(1), 185–198. https://doi.org/10.1146/annurev-anthro-091908-164442
Geddes, A., Adger, W. N., Arnell, N. W., Black, R., & Thomas, D. S. G. (2012). Migration, Environmental Change, and the 'Challenges of Governance.' *Environment and Planning C: Government and Policy*, *30*(6), 951–967. https://doi.org/10.1068/c3006ed
Gemenne, F. (2011). Why the numbers don't add up: A review of estimates and predictions of people displaced by environmental changes. *Global Environmental Change*, *21*, S41–S49. https://doi.org/10.1016/j.gloenvcha.2011.09.005
Gioli, G., Khan, T., Bisht, S., & Scheffran, J. (2014). Migration as an Adaptation Strategy and its Gendered Implications: A Case Study From the Upper Indus Basin. *Mountain Research and Development*, *34*(3), 255–265. https://doi.org/10.1659/MRD-JOURNAL-D-13-00089.1
Gray, C., Frankenberg, E., Gillespie, T., Sumantri, C., & Thomas, D. (2014). Studying Displacement After a Disaster Using Large Scale Survey Methods: Sumatra After the 2004 Tsunami. *Annals of the Association of American Geographers. Association of American Geographers*, *104*(3), 594–612. https://doi.org/10.1080/00045608.2014.892351
Gray, C. L., & Mueller, V. (2012a). Natural disasters and population mobility in Bangladesh. *Proceedings of the National Academy of Sciences*, *109*(16), 6000–6005. https://doi.org/10.1073/pnas.1115944109
Gray, C., & Mueller, V. (2012b). Drought and Population Mobility in Rural Ethiopia. *World Development*, *40*(1), 134–145. https://doi.org/10.1016/j.worlddev.2011.05.023
Gray, C., & Wise, E. (2016). Country-specific effects of climate variability on human migration. *Climatic Change*, *135*(3–4), 555–568. https://doi.org/10.1007/s10584-015-1592-y

Gray, Clark L. (2011). Soil quality and human migration in Kenya and Uganda. *Global Environmental Change*, *21*(2), 421–430. https://doi.org/10.1016/j.gloenvcha.2011.02.004
Gray, Clark L., & Bilsborrow, R. E. (2014). Consequences of out-migration for land use in rural Ecuador. *Land Use Policy*, *36*, 182–191. https://doi.org/10.1016/j.landusepol.2013.07.006
Greiner, C., & Sakdapolrak, P. (2013). Rural–urban migration, agrarian change, and the environment in Kenya: a critical review of the literature. *Population and Environment*, *34*(4), 524–553. https://doi.org/10.1007/s11111-012-0178-0
Hallegatte, S. (2012). A framework to investigate the economic growth impact of sea level rise. *Environmental Research Letters*, *7*(1), 015604. https://doi.org/10.1088/1748-9326/7/1/015604
Harrell, F. E. (2015). *Regression Modeling Strategies*. https://doi.org/10.1007/978-3-319-19425-7
Hassani-Mahmooei, B., & Parris, B. W. (2012). Climate change and internal migration patterns in Bangladesh: an agent-based model. *Environment and Development Economics*, *17*(06), 763–780. https://doi.org/10.1017/S1355770X12000290
Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction.* 2nd ed. New York: Springer.
Hauer, M. E. (2017). Migration induced by sea-level rise could reshape the US population landscape. *Nature Climate Change*, *7*(5), 321–325. https://doi.org/10.1038/nclimate3271
Haug, S. (2008). Migration Networks and Migration Decision-Making. *Journal of Ethnic and Migration Studies*, *34*(4), 585–605. https://doi.org/10.1080/13691830801961605
Henry, S., Boyle, P., & Lambin, E. F. (2003). Modelling inter-provincial migration in Burkina Faso, West Africa: the role of socio-demographic and environmental factors. *Applied Geography*, *23*(2–3), 115–136. https://doi.org/10.1016/j.apgeog.2002.08.001
Henry, S., Schoumaker, B., & Beauchemin, C. (2004). The Impact of Rainfall on the First Out-Migration: A Multi-level Event-History Analysis in Burkina Faso. *Population and Environment*, *25*(5), 423–460. https://doi.org/10.1023/B:POEN.0000036928.17696.e8
Higgins, S. A., Overeem, I., Steckler, M. S., Syvitski, J. P. M., Seeber, L., & Akhter, S. H. (2014). InSAR measurements of compaction and subsidence in the Ganges-Brahmaputra Delta, Bangladesh. *Journal of Geophysical Research: Earth Surface*, *119*(8), 1768–1781. https://doi.org/10.1002/2014JF003117
Hindman, M. (2015). Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 48–62. https://doi.org/10.1177/0002716215570279
Hino, M., Field, C. B., & Mach, K. J. (2017). Managed retreat as a response to natural hazard risk. *Nature Climate Change*, *7*(5), 364–370. https://doi.org/10.1038/nclimate3252
Hong, R., Banta, J. E., & Betancourt, J. A. (2006). Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. *International Journal for Equity in Health*, *5*(1), 15. https://doi.org/10.1186/1475-9276-5-15
Hosmer, D. W. Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.
Hugo, G. (1996). Environmental Concerns and International Migration. *The International Migration Review*, *30*(1), 105–131. https://doi.org/10.2307/2547462
Hugo, G. (2011). Future demographic change and its interactions with migration and climate change. *Global Environmental Change*, *21*, S21–S33. https://doi.org/10.1016/j.gloenvcha.2011.09.008

Hunter, L. M. (2005). Migration and Environmental Hazards. *Population and Environment*, *26*(4), 273–302. https://doi.org/10.1007/s11111-005-3343-x
Hunter, L. M., Murray, S., & Riosmena, F. (2013). Rainfall Patterns and U.S. Migration from Rural Mexico. *The International Migration Review*, *47*(4), 874–909. https://doi.org/10.1111/imre.12051
Hunter, L. M., Nawrotzki, R., Leyk, S., Laurin, G. J. M., Twine, W., Collinson, M., & Erasmus, B. (2014). Rural Outmigration, Natural Capital, and Livelihoods in South Africa. *Population, Space and Place*, *20*(5), 402–420. https://doi.org/10.1002/psp.1776
Hunter, L. M., Luna, J. K., & Norton, R. M. (2015). Environmental Dimensions of Migration. *Annual Review of Sociology*, *41*(1), 377–397. https://doi.org/10.1146/annurev-soc-073014-112223
IOM. (n.d.). Gender and Labour Migration in Asia - | IOM Online Bookstore. Retrieved February 28, 2019, from https://publications.iom.int/books/gender-and-labour-migration-asia
IPCC - SR15. (n.d.). Retrieved October 24, 2018, from http://www.ipcc.ch/report/sr15/
Islam, S. (2014). *Confronting the Blue Revolution: Industrial Aquaculture and Sustainability in the Global South*.
Islam, M. M., & Sado, K. (2000). Development of flood hazard maps of Bangladesh using NOAA-AVHRR images with GIS. *Hydrological Sciences Journal*, *45*(3), 337–355. https://doi.org/10.1080/02626660009492334
Islam, M. Rezaul. (2017). Climate Change, Natural Disasters and Socioeconomic Livelihood Vulnerabilities: Migration Decision Among the Char Land People in Bangladesh. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1563-y
Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U., & Bonin, H. (2010). Direct Evidence on Risk Attitudes and Migration. *Review of Economics and Statistics*, *92*(3), 684–689. https://doi.org/10.1162/REST_a_00020
Joarder, M. A. M., & Miller, P. W. (2013). Factors affecting whether environmental migration is temporary or permanent: Evidence from Bangladesh. *Global Environmental Change*, *23*(6), 1511–1524. https://doi.org/10.1016/j.gloenvcha.2013.07.026
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415
Kartiki, K. (2011). Climate change and migration: a case study from rural Bangladesh. *Gender & Development*, *19*(1), 23–38. https://doi.org/10.1080/13552074.2011.554017
Khandker, S. R. (2012). Seasonality of income and poverty in Bangladesh. *Journal of Development Economics*, *97*(2), 244–256. https://doi.org/10.1016/j.jdeveco.2011.05.001
Klabunde, A., & Willekens, F. (2016). Decision-Making in Agent-Based Models of Migration: State of the Art and Challenges. *European Journal of Population*, *32*(1), 73–97. https://doi.org/10.1007/s10680-015-9362-0
Kniveton, D., Smith, C., & Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change*, *21*, S34–S40. https://doi.org/10.1016/j.gloenvcha.2011.09.006
Koubi, V., Spilker, G., Schaffer, L., & Bernauer, T. (2016). Environmental Stressors and Migration: Evidence from Vietnam. *World Development*, *79*, 197–210. https://doi.org/10.1016/j.worlddev.2015.11.016

L. Perch-Nielsen, S., B. Bättig, M., & Imboden, D. (2008). Exploring the link between climate change and migration. *Climatic Change*, *91*(3–4), 375–393. https://doi.org/10.1007/s10584-008-9416-y

Lagakos, D., Mobarak, A. M., & Waugh, M. (2018). *The Welfare Effects of Encouraging Rural-Urban Migration* (No. w24193). https://doi.org/10.3386/w24193

Liu, Y., & De, A. (2015). Multiple Imputation by Fully Conditional Specification for Dealing with Missing Data in a Large Epidemiologic Study. *International Journal of Statistics in Medical Research*, *4*(3), 287–295. https://doi.org/10.6000/1929-6029.2015.04.03.7

Lu, X., Wrathall, D. J., Sundsøy, P. R., Nadiruzzaman, M., Wetter, E., Iqbal, A., … Bengtsson, L. (2016). Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environmental Change*, *38*, 1–7. https://doi.org/10.1016/j.gloenvcha.2016.02.002

Mallick, B. & Etzold. B. (Eds.) (2015). *Environment, Migration and Adaptation: Evidence and Politics of Climate Change in Bangladesh*. Dhaka, Bangladesh: AH Development Publishing House (AHDPH).

Mallick, B., & Vogt, J. (2014). Population displacement after cyclone and its consequences: empirical evidence from coastal Bangladesh. *Natural Hazards*, *73*(2), 191–212. https://doi.org/10.1007/s11069-013-0803-y

Marchiori, L., Maystadt, J.-F., & Schumacher, I. (2012). The impact of weather anomalies on migration in sub-Saharan Africa. *Journal of Environmental Economics and Management*, *63*(3), 355–374. https://doi.org/10.1016/j.jeem.2012.02.001

Martin, M., Billah, M., Siddiqui, T., Abrar, C., Black, R., & Kniveton, D. (2014). Climate-related migration in rural Bangladesh: a behavioural model. *Population and Environment*, *36*(1), 85–110. https://doi.org/10.1007/s11111-014-0207-2

Mason, W., Vaughan, J. W., & Wallach, H. (2014). Computational social science and social computing. *Machine Learning*, *95*(3), 257–260. https://doi.org/10.1007/s10994-013-5426-8

Massey, D. S. (1990). Social Structure, Household Strategies, and the Cumulative Causation of Migration. *Population Index*, *56*(1), 3–26. https://doi.org/10.2307/3644186

Massey, D. S., Axinn, W. G., & Ghimire, D. J. (2010). Environmental Change and Out-Migration: Evidence from Nepal. *Population and Environment*, *32*(2), 109–136. https://doi.org/10.1007/s11111-010-0119-8

McGranahan, G., Balk, D., & Anderson, B. (2007). The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization*, *19*(1), 17–37. https://doi.org/10.1177/0956247807076960

McLeman, R. A., & Hunter, L. M. (2010). Migration in the context of vulnerability and adaptation to climate change: insights from analogues. *Wiley Interdisciplinary Reviews. Climate Change*, *1*(3), 450–461. https://doi.org/10.1002/wcc.51

McLeman, R., & Smit, B. (2006). Migration as an Adaptation to Climate Change. *Climatic Change*, *76*(1), 31–53. https://doi.org/10.1007/s10584-005-9000-7

McLeman, Robert. (2013). Developments in modelling of climate change-related migration. *Climatic Change*, *117*(3), 599–611. https://doi.org/10.1007/s10584-012-0578-2

Miletto, M., Caretta, M. A., Burchi, F. M., & Zanlucchi, G. (2017). *Migration and its interdependencies with water scarcity, gender and youth employment*. UNESCO Publishing.

Moon, B. (1995). Paradigms in migration research: Exploring "moorings" as a schema. *Progress in Human Geography*, *19*(4), 504–524. https://doi.org/10.1177/030913259501900404

Mueller, V., Gray, C., & Kosec, K. (2014). Heat stress increases long-term human migration in rural Pakistan. *Nature Climate Change*, *4*(3), 182–185. https://doi.org/10.1038/nclimate2103

Myers, N. (2002). Environmental refugees: a growing phenomenon of the 21st century. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *357*(1420), 609–613. https://doi.org/10.1098/rstb.2001.0953

Neumann, K., & Hilderink, H. (2015). Opportunities and Challenges for Investigating the Environment-Migration Nexus. *Human Ecology*, *43*(2), 309–322. https://doi.org/10.1007/s10745-015-9733-5

Nicholls, R. J., & Goodbred, S. L. (2004). Towards integrated assessment of the Ganges–Brahmaputra delta. In *Proceedings of 5th International Conference on Asian Marine Geology, and 1st Annual Meeting of IGCP475 DeltaMAP and APN Mega-Deltas* (p. 9).

Nicholls, R. J., Wong, P. P., Burkett, V., Codignotto, J., Hay, J., McLean, R., … Arblaster, J. (2007). Coastal systems and low-lying areas.

Nicholls, R. J., Wong, P. P., Burkett, V., Woodroffe, C. D., & Hay, J. (2008). Climate change and coastal vulnerability assessment: scenarios for integrated assessment. *Sustainability Science*, *3*(1), 89–102. https://doi.org/10.1007/s11625-008-0050-4

Obokata, R., Veronis, L., & McLeman, R. (2014). Empirical research on international environmental migration: a systematic review. *Population and Environment*, *36*(1), 111–135. https://doi.org/10.1007/s11111-014-0210-7

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, *325*(5939), 419–422. https://doi.org/10.1126/science.1172133

Passalacqua, P., Lanzoni, S., Paola, C., & Rinaldo, A. (2013). Geomorphic signatures of deltaic processes and vegetation: The Ganges-Brahmaputra-Jamuna case study. *Journal of Geophysical Research: Earth Surface*, *118*(3), 1838–1849. https://doi.org/10.1002/jgrf.20128

Paul, B. K. (2005). Evidence against disaster-induced migration: the 2004 tornado in north-central Bangladesh. *Disasters*, *29*(4), 370–385. https://doi.org/10.1111/j.0361-3666.2005.00298.x

Paul, B. G., & Vogl, C. R. (2011). Impacts of shrimp farming in Bangladesh: Challenges and alternatives. *Ocean & Coastal Management*, *54*(3), 201–211. https://doi.org/10.1016/j.ocecoaman.2010.12.001

Penning-Rowsell, E. C., Sultana, P., & Thompson, P. M. (2013). The 'last resort'? Population movement in response to climate-related hazards in Bangladesh. *Environmental Science & Policy*, *27*, S44–S59. https://doi.org/10.1016/j.envsci.2012.03.009

Piguet, E. (2010). Linking climate change, environmental degradation, and migration: a methodological overview. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(4), 517–524. https://doi.org/10.1002/wcc.54

Piguet, E. (2013). From "Primitive Migration" to "Climate Refugees": The Curious Fate of the Natural Environment in Migration Studies. *Annals of the Association of American Geographers*, *103*(1), 148–162. https://doi.org/10.1080/00045608.2012.696233

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, *9*(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, *68*(2), 373–379. https://doi.org/10.1093/biomet/68.2.373

Redehegn, M. A., Sun, D., Eshete, A. M., & Gichuki, C. N. (2019). Development impacts of migration and remittances on migrant-sending communities: Evidence from Ethiopia. *PLOS ONE*, *14*(2), e0210034. https://doi.org/10.1371/journal.pone.0210034

Renaud, F. G., Dun, O., Warner, K., & Bogardi, J. (2011). A Decision Framework for Environmentally Induced Migration: Framework for environmentally induced migration. *International Migration*, *49*, e5–e29. https://doi.org/10.1111/j.1468-2435.2010.00678.x

Reuveny, R. (2007). Climate change-induced migration and violent conflict. *Political Geography*, *26*(6), 656–673. https://doi.org/10.1016/j.polgeo.2007.05.001

Shayegh, S. (2017). Outward migration may alter population dynamics and income inequality. *Nature Climate Change*, *7*(11), 828–832. https://doi.org/10.1038/nclimate3420

Sherbinin, A. de, Castro, M., Gemenne, F., Cernea, M. M., Adamo, S., Fearnside, P. M., … Shi, G. (2011). Preparing for Resettlement Associated with Climate Change. *Science*, *334*(6055), 456–457. https://doi.org/10.1126/science.1208821

Sherbinin, Alex de, Levy, M., Adamo, S., MacManus, K., Yetman, G., Mara, V., … Pistolesi, L. (2012). Migration and risk: net migration in marginal ecosystems and hazardous areas. *Environmental Research Letters*, *7*(4), 045602. https://doi.org/10.1088/1748-9326/7/4/045602

Silveira, J. J., Espindola, A. L., & Penna, T. J. P. (2006). An agent-based model to rural-urban migration analysis. *Physica A: Statistical Mechanics and Its Applications*, *364*, 445–456. https://doi.org/10.1016/j.physa.2005.08.055

Smith, C. D. (2014). Modelling migration futures: development and testing of the Rainfalls Agent-Based Migration Model – Tanzania. *Climate and Development*, *6*(1), 77–91. https://doi.org/10.1080/17565529.2013.872593

Speelman, L. H., Nicholls, R. J., & Dyke, J. (2017). Contemporary migration intentions in the Maldives: the role of environmental and other factors. *Sustainability Science*, *12*(3), 433–451. https://doi.org/10.1007/s11625-016-0410-4

Sraboni, E., Malapit, H. J., Quisumbing, A. R., & Ahmed, A. U. (2014). Women's Empowerment in Agriculture: What Role for Food Security in Bangladesh? World Development, 61, 11–52. https://doi.org/10.1016/j.worlddev.2014.03.025

Stern, N. (2006). the Price of Change. IAEA BULLETIN, 48(2), 25.

Subrahmanian, V. S., & Kumar, S. (2017). Predicting human behavior: The next frontiers. *Science*, *355*(6324), 489–489. https://doi.org/10.1126/science.aam7032

Suckall, N., Fraser, E., Forster, P., & Mkwambisi, D. (2015). Using a migration systems approach to understand the link between climate change and urbanisation in Malawi. *Applied Geography*, *63*, 244–252. https://doi.org/10.1016/j.apgeog.2015.07.004

Suykens, J. A. K., & Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*, *9*(3), 293–300. https://doi.org/10.1023/A:1018628609742

Svart, L. M. (1976). Environmental Preference Migration: A Review. *Geographical Review*, *66*(3), 314–330. https://doi.org/10.2307/213888

Tanner, T., Lewis, D., Wrathall, D., Bronen, R., Cradock-Henry, N., Huq, S., … Thomalla, F. (2015). Livelihood resilience in the face of climate change. *Nature Climate Change*, *5*(1), 23–26. https://doi.org/10.1038/nclimate2431

Teachman, J. D. (1983). Analyzing social processes: Life tables and proportional hazards models. *Social Science Research*, *12*(3), 263–301. https://doi.org/10.1016/0049-089X(83)90015-7
Tessler, Z. D., Vörösmarty, C. J., Grossberg, M., Gladkova, I., Aizenman, H., Syvitski, J. P. M., & Foufoula-Georgiou, E. (2015). Profiling risk and sustainability in coastal deltas of the world. *Science*, *349*(6248), 638–643. https://doi.org/10.1126/science.aab3574
Thiede, B., Gray, C., & Mueller, V. (2016). Climate variability and inter-provincial migration in South America, 1970–2011. *Global Environmental Change*, *41*, 228–240. https://doi.org/10.1016/j.gloenvcha.2016.10.005
Thober, J., Schwarz, N., & Hermans, K. (2018). Agent-based modeling of environment-migration linkages: A review. *Ecology and Society*, *23*(2). https://doi.org/10.5751/ES-10200-230241
Thornton, F., McNamara, K. E., Farbotko, C., Dun, O., Ransan-Cooper, H., Chevalier, E., & Lkhagvasuren, P. (2018). Human mobility and environmental change: a survey of perceptions and policy direction. *Population and Environment*. https://doi.org/10.1007/s11111-018-0309-3
Till, C., Haverkamp, J., White, D., & Bhaduri, B. (2018). Understanding climate-induced migration through computational modeling: A critical overview with guidance for future efforts. *The Journal of Defense Modeling and Simulation*, *15*(4), 415–435. https://doi.org/10.1177/1548512916679038
Uddin, M. N., Saiful Islam, A. K. M., Bala, S. K., Islam, G. M. T., Adhikary, S., Saha, D., … Akter, R. (2019). Mapping of climate vulnerability of the coastal region of Bangladesh using principal component analysis. *Applied Geography*, *102*, 47–57. https://doi.org/10.1016/j.apgeog.2018.12.011
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(1), 1–67. https://doi.org/10.18637/jss.v045.i03
Walsham, M. (2010). Assessing the Evidence: Environment, Climate Change and Migration in Bangladesh. International Organization for Migration.
Warner, K., Hamza, M., Oliver-Smith, A., Renaud, F., & Julca, A. (2010). Climate change, environmental degradation and migration. *Natural Hazards*, *55*(3), 689–715. https://doi.org/10.1007/s11069-009-9419-7
White, M. J., Moreno, L., & Guo, S. (1995). The Interrelation of Fertility and Geographic Mobility in Peru: A Hazards Model Analysis. *International Migration Review*, *29*(2), 492–514. https://doi.org/10.1177/019791839502900207
Willekens, F., Massey, D., Raymer, J., & Beauchemin, C. (2016). International migration under the microscope. *Science*, *352*(6288), 897–899. https://doi.org/10.1126/science.aaf6545
World Bank. (2016). Bangladesh: Growing the Economy through Advances in Agriculture. http://projects-beta.worldbank.org/en/results/2016/10/07/bangladesh-growing-economy-through-advances-in-agriculture
World Bank. (n.d.). Bangladesh | Data. Retrieved February 22, 2019, from https://data.worldbank.org/country/bangladesh