# Machine Learning to Identify Drivers of Internal Migration in Coastal Bangladesh

Kelsea Best[1], Jonathan Gilligan[1], Hiba Baroud[2], Brooke Ackerly[3], Bishawjit Mallick[4]
[1]Department of Earth & Environmental Science, Vanderbilt University, Nashville, TN
[2]Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, TN
[3]Department of Political Science, Vanderbilt University, Nashville, TN
[4]Technical University Dresden, Dresden, Germany

VANDERBILT UNIVERSITY

## INTRODUCTION

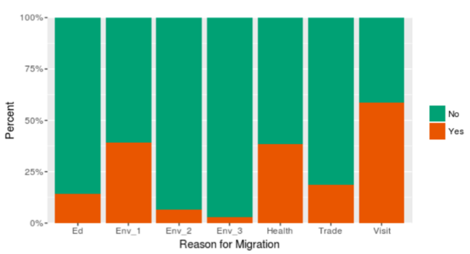
- As climate change increases pressure on vulnerable communities, migration is one adaptation strategy.
- The decision to migrate is highly complex, and is influenced by economic, social, and environmental drivers.
- Bangladeshi communities have long adapted to a dynamic and challenging natural environment. Seasonal migration and livelihood diversification are important adaptation methods, especially rural to urban migration.
- This work addresses a gap in current research by beginning to investigate how different "push" and "pull" drivers of migration might have different variables that contribute to the ultimate decision to move or stay.

## METHODOLOGY

Process data and impute missing values → Test and tune models for prediction → Assess random forest variable importance

## DATA AND DATA PROCESSING

- Data was collected by household interviews throughout 40 communities in the southwest area of Bangladesh from 2012 to 2014.
- Multiple imputation with random forest was conducted to address data missing completely at random (*m = 10*).
- Dummy variables were created for each question that included a categorical or ordered response type.
- Random forest (ensemble of decision trees) was selected after assessing predictive ability of different models on complete cases.
- Outcome variables are binary indicators of migration for environmental reasons, education, health, trade, or to visit relatives.
- Random forest models were run for each complete dataset and for each outcome variable (50 models total).




## MODELS TESTED FOR PREDICTION

- Each model was trained on a sample of 80% of complete cases of data, and tested on the remaining 20% to test predictive accuracy.
- Model parameters were tuned using cross validation.

**Prediction errors (percent) for each model.**

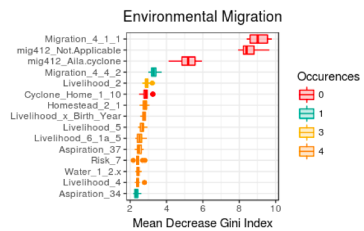
| Model | Env_1 | Ed | Health | Trade | Visit |
|---|---|---|---|---|---|
| Logistic | 47.1 | 44.9 | 44.1 | 43.4 | 42.6 |
| SVM | 36.0** | 16.2** | 36.0** | 19.9** | 41.2** |
| Random Forest | 35.5 | 14.7 | 33.1 | 19.9** | 33.8 |

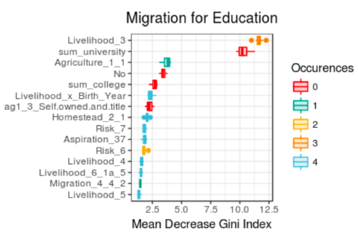
**Indicates that the model was unable to predict both "Yes" and "No" values, but predicted only one or the other, resulting in class errors of 0% and 100%.

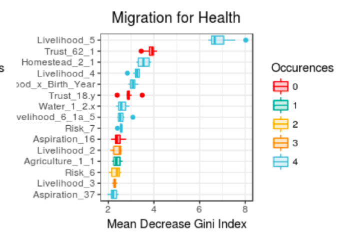
## VARIABLE IMPORTANCE IN RANDOM FOREST MODELS

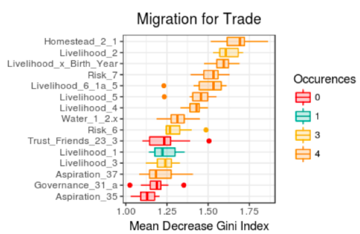
- **Gini Index** is used to measure impurity when building decision trees
- Mean decrease in Gini Index is used to identify top 15 variables of importance in random forest models of each type of migration.

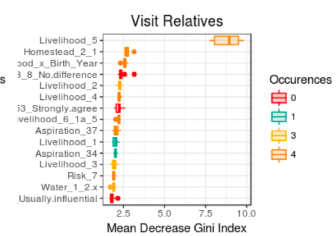
$$Gini = 1 - \sum_j p(j)^2$$




## INSIGHTS AND CONCLUSIONS

| Type of Migration | Key Variables |
|---|---|
| **All models** | • Livelihood (annual expenditures, daily expenditures, amount of homestead land owned, monthly income)<br>• Travel time to primary water source<br>• Distance to cyclone shelter<br>• Birth Year |
| **Environmental** | • Knowing others who have migrated<br>• Environmental damage to home and livelihood |
| **Education** | • Annual expenditure and importance of education<br>• Members of household completed university or college |
| **Health** | • Community efforts to dig a pond for drinking water<br>• Feeling unsafe traveling to work |
| **Trade** | • Borrowing money from friends<br>• Fewer unique variables, higher variance |
| **Visit relatives** | • Low locus of control in community<br>• Strong belief that more economic opportunities exist outside of community |

**Conclusions:**

- Environmental migration is uniquely influenced by knowledge of others who have migrated and economic impacts of environmental events.
- Livelihood variables, age, social status, and proximity to water sources are highly important for all forms of migration.
- Imputing missing data significantly impacts results, demonstrating importance.

## FUTURE WORK

- Repeat analysis with Bangladesh Environment and Migration Survey (BEMS).
- Develop survival models from survey data to quantify probabilities of migration.
- Use insights from statistical analysis to inform agent-based model to test scenarios and adaptation strategies.
- Utilize multidisciplinary teams to understand the factors that influence human migration as a system, rather than in isolation.

## REFERENCES

1. Ackerly, Brooke A., Md Mujibul Anam, and Jonathan Gilligan. 2015. "Environment, Political Economies and Livelihood Change." In Environment, Migration and Adaptation: Evidence and Politics of Climate Change in Bangladesh, edited by Bishawjit Mallick and Benjamin Etzold. Dhaka, Bangladesh: AH Development Publishing House (AHDPH). http://eprints.qut.edu.au/84192/.
2. Alam, G.M. Monirul, Khorshed Alam, and Shahbaz Mushtaq. 2017. "Climate Change Perceptions andLocal Adaptation Strategies of Hazard-Prone Rural Households in Bangladesh. "Climate Risk Management17: 52–63. doi:10.1016/j.crm.2017.06.006.
3. Black, Richard, W. Neil Adger, Nigel W. Arnell, Stefan Dercon, Andrew Geddes, and David Thomas. 2011."The Effect of Environmental Change on Human Migration." Global Environmental Change 21 (December): S3–S11. doi:10.1016/j.gloenvcha.2011.10.001
4. Call, Maia A., Clark Gray, Mohammad Yunus, and Michael Emch. 2017. "Disruption, Not Displacement: Environmental Variability and Temporary Migration in Bangladesh." Global Environmental Change 46 (September): 157–65. doi:10.1016/j.gloenvcha.2017.08.008.
5. Islam, M. Rezaul. 2017. "Climate Change, Natural Disasters and Socioeconomic Livelihood Vulnerabilities: Migration Decision Among the Char Land People in Bangladesh." Social Indicators Research, January. doi:10.1007/s11205-017-1563-y.
6. L. Perch-Nielsen, Sabine, Michale B. Battig, and Dieter Imboden. 2008. "Exploring the Link Between Climate Change and Migration." Climatic Change 91 (3-4): 375–93. doi:10.1007/s10584-008-9416-y.