Global Environmental Change 88 (2024) 102920

Contents lists available at ScienceDirect

Global Environmental Change

journal homepage: www.elsevier.com/locate/gloenvcha

Scenario projections of South Asian migration patterns amidst environmental and socioeconomic change

Sophie de Bruin [a,*], Jannis Hoch [b], Jens de Bruijn [a,c], Kathleen Hermans [d], Amina Maharjan [e], Matti Kummu [f], Jasper van Vliet [a]

[a] Institute for Environmental Studies, VU University, De Boelelaan 1087, 1081HV Amsterdam, the Netherlands
[b] Fathom, Bristol, UK
[c] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[d] Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Theodor-Lieser-Str. 2, 06120 Halle (Saale), Germany
[e] International Centre for Integrated Mountain Development (ICIMOD), Kathmandu, Nepal
[f] Water and Development Researchakno Group, Aalto University, Finland

ARTICLE INFO

Keywords:
Environmental change
Climate change
Migration
South Asia
Machine-learning
Scenario projections

ABSTRACT

Projecting migration is challenging, due to the context-specific and discontinuous relations between migration and the socioeconomic and environmental conditions that drive this process. Here, we investigate the usefulness of Machine Learning (ML) Random Forest (RF) models to develop three net migration scenarios in South Asia by 2050 based on historical patterns (2001–2019). The model for the direction of net migration reaches an accuracy of 75%, while the model for the magnitude of migration in percentage reaches an R^2 value of 0.44. The variable importance is similar for both models: temperature and built-up land are of primary importance for explaining net migration, aligning with previous research. In all scenarios we find hotspots of in-migration North-western India and hotspots of out-migration in eastern and northern India, parts of Nepal and Sri Lanka, but with disparities across scenarios in other areas. These disparities underscore the challenge of obtaining consistent results from different approaches, which complicates drawing firm conclusions about future migration trajectories. We argue that the application of multi-model approaches is a useful avenue to project future migration dynamics, and to gain insights into the uncertainty and range of plausible outcomes of these processes.

1. Introduction

Human migration is intrinsically related to societal change and development. People leave their place of origin for multiple reasons, including better economic or educational opportunities elsewhere, family matters or escaping conflict or persecution (IOM, 2022). Increasingly, environmental change is understood to shape migration patterns via impacts on agriculture and the habitability of regions (Adger et al., 2015; Horton et al., 2021). Yet, the identification of causal linkages between environmental change and migration is challenging or even impossible, due to the intersecting and context-specific nature of factors affecting migration decisions (Boas et al., 2019; Cattaneo et al., 2019; Hermans and McLeman, 2021). Yet, understanding migration dynamics in an era of climate change is both scientifically pertinent and societally relevant, especially in a world where views and voices in national and global migration debates are often ideologically driven (Boas et al., 2019; de Haas, 2010a).

The consequences of environmental change, especially climate variability impacts, on mobility and immobility have been investigated extensively in the past two decades (McLeman et al., 2021; Piguet, 2022). Specifically, there is a growing recognition that climate-related migration is multi-causal and context-specific (Adger et al., 2024; Cattaneo et al., 2019; Hunter and Simon, 2023). Empirical studies based on interview, survey, or census data have provided an overview of common drivers (Afifi, 2011; Hermans-Neumann et al., 2017; van der Geest, 2011). Evidence strongly suggests that environmental conditions have a greater impact on migration within countries as compared to international migration (Cissé et al., 2022; Cundill et al., 2021). However, while some quantitative research on the importance of environmental conditions to international migration dynamics is available, (Abel et al., 2019; Beine and Parsons, 2015; Falco et al., 2019; Grecequet et al., 2017), this type of research within countries is largely absent.

* Corresponding author.
E-mail address: sophiepieternel@gmail.com (S. de Bruin).

https://doi.org/10.1016/j.gloenvcha.2024.102920
Received 23 February 2024; Received in revised form 14 August 2024; Accepted 27 August 2024
Available online 2 September 2024

The lack of subnational longitudinal migration data has limited progress on large-scale quantitative subnational migration research (Piguet, 2022). In line with this, little research has been published on developing quantitative multi-country migration scenario projections on a subnational level that encompass the potential impact of environmental change (Beyer et al., 2023; McLeman, 2013; Oakes et al., 2023). Developing these scenarios is complicated by several factors. There are multiple directions of influence of climate change impacts on migration. Impacts such as prolonged droughts and rising temperatures can both reduce and increase migration, depending on the specific socioeconomic conditions of the region and the experiences and resources of those being affected (Dallmann and Millock, 2017; Mueller et al., 2020; Mueller et al., 2014). Additionally, the availability of subnational socioeconomic scenario projection data are limited, especially within the framework of climate change (Buhaug and Vestby, 2019). Some agent-based and integrated modelling studies exist of future human migration which focus on individual countries (Thober et al., 2018). These studies use national demographic survey data, reducing their comparability across countries. Furthermore, the two Groundswell reports (Clement et al., 2021; Rigaud et al., 2018) and the African Shifts report (Amakrane, 2023) develop various scenarios by deploying a gravity model with and without hydroclimatic and agricultural variables for Sub-Saharan Africa, South Asia, and Latin America. By showing the difference between the scenario with and without the hydroclimatic and agricultural variables, the studies obtain the number of internal migrants that could be attributed to hydroclimatic conditions. Gravity models face criticism when used for developing scenario projections of migration, because they are unable to adequately account for the changes in-migration patterns over time (Beyer et al., 2022). Furthermore, gravity models cannot handle the discontinues impacts of drivers (Robinson and Dilkina, 2018).

The main objective of this article is to better understand the usefulness of a new approach to developing migration scenarios in the context of climate change: Machine Learning (ML) Random Forest (RF) models. RF approaches can combine multiple input data and include discontinuous relations (Robinson and Dilkina, 2018), making them potentially well-suited for understanding and modelling migration. However, RF approaches have not been used before to develop migration scenarios. To further explore the potential of RF approaches, we employ two different models: a RF classification model and a RF regression model. The former projects the direction of net migration per region, while the latter projects the magnitude of the net migration. We test these RF approaches in for South Asia. To do so, we first train both models to explain historical net migration patterns based on known drivers for migration, using a novel high-resolution large-scale migration dataset (Niva et al., 2023a). We interpret the result of this training as the capability of our models to explain net migration patterns, which provides an indication of their capacity to also explain future patterns. We project net migration in the year 2050 under different socioeconomic and environmental change scenarios using both trained models. Results are analysed mainly in the context of the usability of the two RF approaches for envisioning migration scenarios. By doing so, we can discern the insights they offer, facilitating a more robust evaluation and interpretation of approaches to developing migration scenarios.

2. Data & methods

2.1. Case study region

We focus on South Asia, including Bangladesh, Bhutan, India, Nepal, Pakistan, and Sri Lanka, following the delineation of this region by the World Bank. World Bank regions represent relatively homogenous socioeconomic regions, making them appropriate for assessing migration dynamics and interpreting model results with the same model. South Asia was selected for two reasons. The region was selected for two reasons. First, the region is characterised by a high dependency on climate-sensitive livelihoods, mostly in agriculture (Tucker et al., 2015). Second, the death and birth rates are well documented on the subnational level, compared to large parts of the Middle East region and the African continent (Niva et al., 2023a). South Asian net migration data can therefore be regarded as more accurate than other regions with a high dependency on climate-sensitive livelihoods. We excluded Afghanistan ex-ante from the analysis. In this country, 3.9 million refugees were repatriated to Afghanistan between 2002 and 2015 (UNHCR, 2022). Overall, 5.3 million people returned to Afghanistan since 2002 until the Taliban took over control in 2021 (UNHCR, 2023). These exceptional conditions make the country unsuitable for assessing general migration dynamics driven by socioeconomic and environmental changes.

As spatial units we employed the Global Administrative Areas (GADM) level 2 (Global Administrative Areas, 2022), which represents sub-divisions of provinces or districts. Level 2 areas which were smaller than 100 km^2 or inhabited by less than 500 people in the year 2001 were merged with the adjacent area in the same province with which it shared the longest border. This avoids a potentially large influence of small population numbers on the model. Table 1 provides the demographic characteristics of these regions per country.

2.2. Data description

2.2.1. Target data

We used the global net migration dataset developed by Niva et al. (2023a) as target data, which is based on annual harmonised subnational data for births and deaths for the 2001–2019 period. This subnational data was downscaled to a 5 arc-minute resolution globally and then combined with observed changes in population numbers (WorldPop, 2021). Net migration is defined as the difference between the total population change and the natural population change (births minus deaths) per year, reflecting the net number of people moving into or out of a given area. Simulated data was validated against reported detailed net migration data of various countries, indicating a good performance of the downscaling method (Niva et al., 2023a). The final dataset provides information on the net number of people moving in or out of a grid cell per year. Following this approach, migration entails both national and international migration, and the two cannot be separated. Consequently, the case study region for this study is not a closed system, and the sum of all migration can be higher or lower than 0.

We pre-processed the target migration data to facilitate the setup of our RF models. For the Random forest regression (RFR), the net percentage of the population per administrative area that migrates was calculated by dividing the total number of net migrants by the total population of that area. We used the WorldPop population data (WorldPop, 2021) as used by Niva et al. (2023a). For the Random forest classification (RFC) model, an area was classified as 0 when the area faced net out-migration in a particular year, and as 1 when the area faced net in-migration in a particular year. In seven areas we observed

Table 1
Demographic characteristics of the countries in the study region.

Country	Population 2015 X 1 000	Number of admin level 2 areas included	Average population*/admin level 2 area X 1 000
Bhutan	740	133	6
Bangladesh	152 513	65	2 346
India	1 284 631	664	1 935
Nepal	26 813	14	1 915
Pakistan	193 696	33	5 870
Sri Lanka	19 607	227	86

* **Population numbers as used by** Niva et al. (2023b) **from**WorldPop (2021).

net migration values exceeding 20 %, which are likely errors following the skewed pattern (e.g., one year facing net out-migration of 20–40 % followed by net in-migration in a following year of 20–40 %). Therefore, we decided to remove these regions from our analysis. Fig. 1A presents the average yearly net migration ratio (%) for the reference period. 1B presents the average of net in (1) and net out (0) migration per year for the entire reference period 2001–2019.

2.2.2. Independent variables

We used a combination of socioeconomic and environmental variables as the independent variables which are further detailed in Table 2. The variables were selected based on three criteria. First, we selected variables that are potentially relevant for explaining migration according to previous research. Second, we selected those variables for which consistent historical and future data for the whole study region is available. Third, we analysed multi-collinearity, and removed variables that are too strongly correlated with each other (Chan et al., 2022). Although RF-models can handle (multi)collinearity in terms of prediction, having highly correlated predictors can affect the interpretability of the variable importance. It might assign relatively lower importance to one of the correlated variables compared to what it would if the variables were not highly correlated (Chan et al., 2022). For more details on (multi)collinearity between the selected variables and cut-off values, see Fig. S1 and Table S1 on (multi)collinearity in the Supplementary Material.

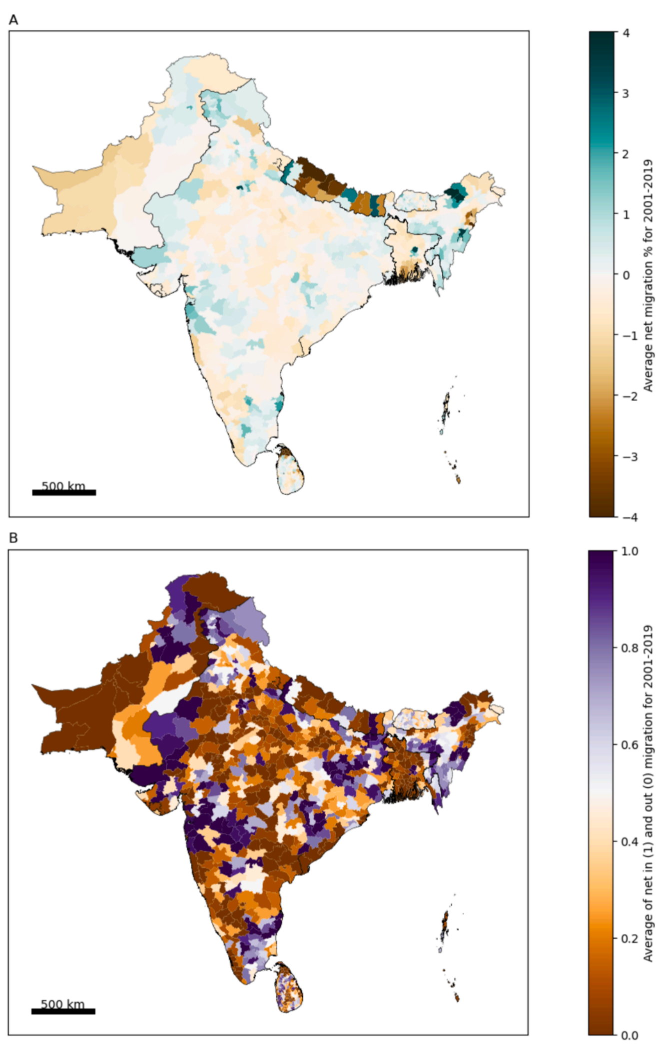

Fig. 1. (A) The average migration ratio for the reference period 2001–2019. (B) The average of net in (1) and net out (0) migration for the reference period 2001–2019.

Table 2
Characteristics of the socioeconomic and environmental annual variables.

Variable	Brief description	Unit	Source	Original resolution
Built-up area	The average built-up area.	Km^2/cell	Wolff et al. (2018)	9.25 × 9.25 km
Conflict	Armed conflict events with over 10 deaths.	Number of events	UCDP Georeferenced Event Dataset V23.1 (Davies et al., 2023; Sundberg and Melander, 2013)	Georeferenced events
Education	Average total years of schooling per person.	Years of schooling	Smits and Permanyer (2019), harmonised and rasterised using methods by Kummu et al (2018).	Admin 1 level, except for Sri Lanka, which is country-based.
GNI per cap	Gross domestic income per capita.	US$2017	Smits and Permanyer (2019) harmonised and rasterised using methods by Kummu et al (2018) and further downscaled to admin 2 level.	5 arcminutes
Inequality (GINI)	Income inequality between people, following the Gini index.	0–1 index	Solt (2020) downscaled using subnational GINI data	Admin 1 level, except for Sri Lanka, which is country-based.
Dry spell	Number of days in the longest period without significant precipitation of at least 1 mm.	Number of days	World Bank Group (2023)	0.5x0.5 degrees
Fluvial flood Volume	Volume of the largest flooding event in an area.	M^3	Sutanudjaja et al. (2018)	5 arcminutes
Precipitation	Total precipitation.	Millimetres	World Bank Group (2023)	0.5x0.5 degrees
Temperature	Average temperature.	Degree C°	IMAGE model (Doelman et al., 2018; Stehfest et al., 2014), 2023 update.	0.5x0.5 degrees
Actual crop yield	The actual simulated total crop yield.	Ton/km2	IMAGE model (Doelman et al., 2018; Stehfest et al., 2014), 2023 update.	5 arcminutes

The socioeconomic variables included are built-up land, conflict, education, Gross National Income (GNI) per capita and economic inequality (GINI). Built-up land was included as a proxy for the level of urbanicity and the access to services, conditions generally attracting migrants (Selod and Shilpi, 2021). Conflict was included since it can lead to forced displacement (Braithwaite et al., 2019), although it is not clear to what extent conflict events are a factor of importance to migration over large regions and long-time spans. Education levels affect migration following the opportunities that come with being educated (de Haas, 2010b; Neumann and Hermans, 2017). Income levels, in this study GNI per capita, are an important requirement for migration (Neumann and Hermans, 2017; Niva et al., 2021). The role of economic inequality in-migration dynamics is equivocal. However, earlier research has showed that inequality is often high in fast-growing cities what would imply that it could be a factor explaining rural–urban migration (Østby, 2016).

The environmental variables included in this study are the length of dry spells, flood volume, precipitation, temperature, and actual crop yields. Dry spells have been included since these have been associated with out-migration (Carrico and Donato, 2019). Flood volume was included because floods can affect agricultural yields, as well as habitability of an area affecting migration dynamics (Horton et al., 2021; Mueller et al., 2014). Precipitation and temperature have been found to affect migration mainly via agricultural productivity, although often to a limited extent (Bohra-Mishra et al., 2014; Cattaneo and Peri, 2016; Mueller et al., 2014). Finally, we included actual crop yields to account for the large-scale trends of productivity of an area. Productive areas might indicate that these areas are attractive to rural-rural migrants (Hathie et al., 2015). At the same time, people living in productive areas might have the financial means required to migrate (Groth et al., 2020).

From most variables, we calculated the yearly varying average value per administrative area. For conflict, the total number of conflict events per year per area was used. For flooding, the maximum annual fluvial flood volume per area was taken, representing the largest flood event for that year following exceeding river discharge following excess rainfall after accounting for soil infiltration and evapotranspiration. For all runs, a one-year time lag was applied for all variables except for conflict events and flood volume, assuming that those trigger migration directly. For more context regarding the socioeconomic and environmental variables regarding the minimal, maximal, average values and standard deviation, see Table S2 in the Supplementary Material.

2.2.3. Projection data

Three SSP-RCP combinations were employed to reflect a range of socioeconomic and climate developments: SSP1 with RCP2.6, SSP2 with RCP4.5, and SSP3 with RCP7.0. See Table 3 for a brief description of these scenarios. The SSP-RCP scenario projection data are consistent with the historical data. For the environmental data, CMIP6-based ISIMIP-3 protocol projections were used for RCP2.6, RCP4.5 and RCP7.0, except for flood volume since this data was not available in CMIP6. Therefore, we use CMIP5-based ISIMIP-3 protocol data for RCP2.6, RCP4.5 and RCP6.0 for flood volume. Although CMIP5 and CMIP6 are different in terms of variables, resolution, and scenarios, the phases are consistent in terms of simulation protocols and model evaluation standards (Tebaldi et al., 2021). It is therefore justified to use data from both phases in one analysis. For more scenario details per variable see Table S3 in the Supplementary Material. Conflict events were not included in any scenario because projections of conflict risk are not available for the region. While projections of conflict are lacking, we incorporate historical conflict events nonetheless, to gain insights into the importance of these events in shaping the migration patterns.

Table 3
Summary of the scenario narrative of the SSP-RCP combinations used in this study based on O'Neill et al. (2016) and Kriegler et al. (2012).

SSP-RCP	Scenario description
SSP1 – RCP2.6 Sustainable Development – Low Emissions	This scenario envisions a future where the world follows a sustainable development pathway with low greenhouse gas emissions. It assumes that society places a strong emphasis on environmental sustainability, energy efficiency, and the reduction of carbon emissions. Challenges for mitigation and adaptation are low. Under RCP 2.6, total radiative forcing increases to 3.0 W m − 2 until mid-century before a decline begins. The goal is to limit global warming to well below 2 degrees Celsius above pre-industrial levels.
SSP2 – RCP4.5 Middle of the Road	This scenario represents a middle-of-the-road development pathway, with moderate greenhouse gas emissions without fundamental breakthroughs. It anticipates a world where efforts to mitigate climate change are moderate, with a focus on balancing economic growth and environmental concerns.
SSP3 – RCP7.0 Regional Rivalry – High Emissions	In this scenario, a future is envisioned where there is a lack of global cooperation, leading to regional rivalries and fragmentation of efforts. Greenhouse gas emissions are relatively high, resulting in substantial global warming. This pathway highlights the potential consequences of limited international collaboration.

2.3. Set-up of the Random Forest models

To evaluate the importance of the socioeconomic and environmental variables to the magnitude and direction of migration, we employed an RFR model and an RFC model. For our analysis we used the existing CoPro framework, which is designed to apply machine-learning approaches for projections, used before to develop conflict risk projections (modified from Hoch et al., 2021a). Within the CoPro framework, the Scikit-learn library is used to implement the RFR and the RFC algorithms (Pedregosa et al., 2011). Both algorithms are trained with 20 years of data (2001–2019) to quantify the historical relation between the independent variables and net migration, the target data.

To explore the explanatory power of the two approaches, we first trained the RFR and RFC for each country separately. For our final projections in the entire region, we included only those countries where historical migration could be explained by the model. Moreover, by first assessing the individual countries, we could also examine to what extent the variable importance was comparable between the individual countries for the two approaches.

For projecting future migration patterns, we aggregated those countries where migration could be (partly) explained by the RFR and RFC approach. This aggregation excludes Bhutan and Pakistan (see results and discussion). The RFR and the RFC were subsequently trained over this combined study region to develop the out-of-sample scenario projections.

For each year in the reference period, values were extracted from the independent variables for each administrative area. In total, this yields data points equal to the number of administrative areas times the number of years. Subsequently, for both the regression and the classification run, 100 RF trees were initialised to capture the variance in the data without overfitting. For each tree, 70 % of the data points were randomly drawn to train the model, and the remaining 30 % were used for validation using several evaluation metrics (Skicit-learn, 2023; Ting, 2011):

- **Accuracy**: reflects the fraction of correct classifications [0–1] – higher represents more correct classifications.
- **Recall**: presents the total number of true positives out of the total number of positives [0–1] – higher represents more correct classifications.
- **Precision:** the number of true positives out of the total number of positives predicted, including the false positives [0–1] – higher represents more correct classifications.
- **The ROC AUC** (Receiver Operating Characteristic – Area Under the Curve): this value quantifies the overall ability of a binary classification model to distinguish between the positive and negative classes [0 – 1] – higher represents more correct classifications.
- **R^2**: For the regressions runs the R2 is presented. This is a statistic that measures the proportion of the variance in the target variable explained by the independent variables in the model. It provides a

value between 0 and 1, a higher R2 indicates a better fit of the model to the data.

We calculated the variable importance based on the same training runs for both models. Variable importance refers to the relative importance of each input variable in explaining migration, the target variable. In total, the variable importance of the combined independent variables is 1. Quantifying the importance of the separate variables for the model predictions enhances our understanding of the importance of the socioeconomic and environmental conditions to net migration. Although variable importance provides the magnitude of influence it does not provide the direction of influence, since the direction can go both ways within the same model, depending on the other variables. This property is both a strength and a weakness of RF models since it makes them flexible but also harder to interpret.

From the end of the reference period (2019) until 2050, we made annual out-of-sample projections. The projections are made for the combined region of countries where a share of the historical migration could be explained in both the RFR and the RFC. To maintain the internal consistency of each projection pathway, this was done for each selected SSP-RCP combination. For the RFR, the projections represent the average net in or out-migration ratio in 2050 for the 100 trees in the model. For the RFC, the projections represent the average probability of net in-migration in 2050 for the 100 trees in the model. For both the RFR and RFC we compared the 2050 values with the average net migration value per region for the reference period to indicate where future migration might deviate from historical patterns. The last step involved a comparison of the scenario projections based on the RFR and the RFC approach to check for consistency. This was done by comparing the projected net in- and net out-migration. For the RFC approach, net in-migration was defined as a probability of net in-migration of over 0.5, while for the RFR net in-migration was defined as a positive migration percentage.

3. Results

3.1. Model validation

For the RFR run, which aims to predict the annual net migration as a percentage of the population, the R^2 varies markedly between the countries. For India, the explained variance is low with 0.17. The R^2 for Bangladesh, Nepal and Sri Lanka are relatively high, respectively 0.59, 0.91 and 0.53. For Bhutan and Pakistan, migration could not be explained, as indicated by the negative R^2 values in Table 4. These countries are therefore excluded from the runs to make scenario projections of the combined region (see Discussion for further interpretation of these results). The explained variance of the combined region, Bangladesh, India, Nepal and Sri Lanka is 0.44.

For the RFC run, which aims to predict binary in (1) or out (0) migration, the overall model performance is good, as indicated by ROC-AUC scores of above 0.8, except for Bhutan (Table 4). For Bhutan, historical migration cannot be explained with the RFC approach. Overall accuracy of the other countries— the fraction of correct classifications—is reasonable to good. Mean precision – the ability of the RFC not to label an observation as in-migration that is out-migration – is also reasonable to good. The recall scores are lower for all areas, indicating a restricted ability of the classifier to find all positive observations. Migration in the combined region of Bangladesh, India, Nepal and Sri Lanka can be explained relatively well, as indicated by an overall accuracy of 0.75.

Table 4
Mean evaluation metrics for RFC and RFR run.

Country	RF Classifier				RF Regression
	ROC AUC	Accuracy	Precision	Recall	R^2
Bangladesh	0.88	0.88	0.80	0.53	0.59
Bhutan	0.51	0.50	0.51	0.53	−0.19
India	0.82	0.75	0.74	0.66	0.17
Nepal	0.98	0.93	0.95	0.86	0.91
Pakistan	0.90	0.83	0.80	0.72	−0.25
Sri Lanka	0.80	0.73	0.67	0.56	0.53
South Asia (excl. Bhutan and Pakistan)	0.82	0.75	0.72	0.61	0.44

3.2. Predictors of net migration

The variable importance in the RFR and RFC run per country are rather comparable (Fig. 2). The average annual temperature is the most important variable in both analyses. Built-up land, yields and annual precipitation rates also have a relatively high importance across the countries and for the combined region in both runs. Conflict is not important to explain net migration in the reference period. The importance of the variables in the two models is fairly comparable between countries for many, though not all, variables. For the variable importance per country see Fig. S2 in the Supplementary Material. Especially in Nepal there are major differences in the importance of the same variable between runs. The importance of some variables, including inequality and temperature, differ considerably among countries.

3.3. Scenario projections

Based on the historical relations learned in the RFC and RFR run, projections with data from three SSP-RCP combinations are made for the combined South Asian regions of Bangladesh, India, Nepal and Sri Lanka.

The RFR projections, which yield the magnitude of net in- or out-migration in 2050, show a scattered image across the region (Fig. 3a, c, e). Some regions are projected to face net out-migration in all SSP-RCP combinations, including western Nepal, north-eastern India, and large parts of northern and southern Sri Lanka and coastal zones of Bangladesh. At the same time, net in-migration is projected in all scenarios for the coastal areas of western India, South-East India and large parts of inland Bangladesh, including the capital region of Dhaka. The magnitudes of net in- or out-migration are in general more moderate for SSP1-RCP26, as compared to the other two scenarios. The SSP2-RCP45 scenario projection is most outspoken in terms of differences in net in- and out-migration. Most areas are projected to face net in-migration in the SSP3-RCP70 scenario, 474 out of the 969 areas. This number is 286 for the SSP1-RCP70 and 273 for the SSP2-RCP45 scenario. However, the average net migration over all administrative areas is more similar among the scenarios, with a weighted average of 0.1 % in SSP3-RCP70, while this is −0.1 % for the SSP1-RCP70 and −0.5 % for the SSP2-RCP45 scenario. There are some notable differences between the scenario projections and the reference period. Especially in northern Sri Lanka, northern Nepal, north-east India and parts of central-western India, the differences between the scenario projections and the reference period are substantial (Fig. 3b, d, f).

The RFC projections show a rather uniform net migration pattern across the region for all three scenario projections (Fig. 4a, c, e). For the majority of the regions, the projected probability of net in-migration is below 0.5, indicating that these regions are mostly characterised by socioeconomic and environmental conditions historically associated with out-migration. In SSP1-RCP26 64 out of 969 regions are projected to face net in-migration, while this is 69 for SSP2-RCP4.5 and 289 for SSP3-RCP70. In all scenarios, the North-West of India is projected to have the highest probability of net in-migration. The differences between the SSP-RCP runs are notable, with SSP3-RCP70 facing the lowest level of out migration. Comparing the scenario projections to the average of the reference period, the picture is scattered though. Changes in net migration are considerable. Major parts of Nepal, the north and

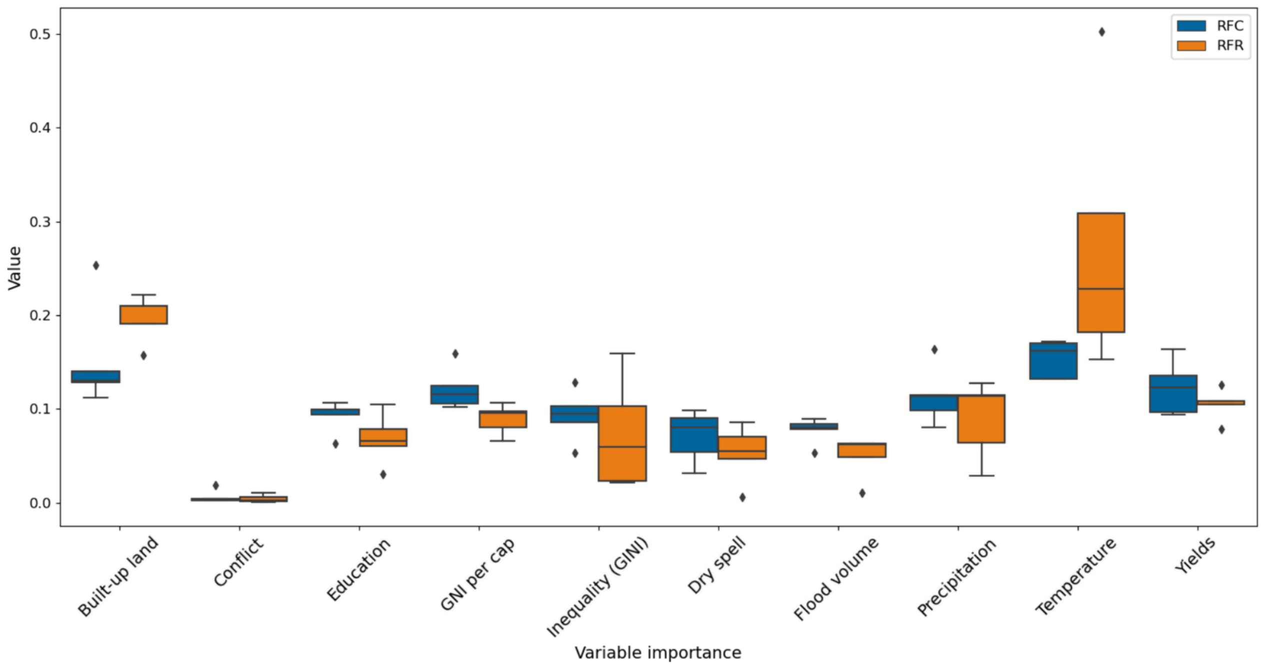

Fig. 2. Distribution of the variable importance based on the initial trees for the RFR and the RFC run for the separate country runs. The boxes represent the interquartile range, the line in the box is the average value, the whiskers maximum is 1.5 times the interquartile range, and the red dots are data outliers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

central parts of Sri Lanka, most of inland Bangladesh, central and southern India are projected to have a probability of net in-migration below 0.5, indicating a higher chance on out-migration then in-migration. Since the average share of net out-migration in the reference migration was lower than the projected probability of net in-migration by 2050, these regions face a positive difference when comparing the reference period with the projection period (Fig. 4b, d, f).

The agreement between the RFR and RFC approaches is moderate (Fig. 5). Areas are marked green if the RFR ratio is positive and the RFC probability is above 0.5, or if the RFR ratio is negative and the RFR probability is below 0.5. For SSP1-RCP26, 72 % of the areas overlap in terms of direction, 73 % for SSP2-RCP45 and 61 % for SSP3-RCP70. The correlation coefficient between the projections is 0.2, 0.22 and 0.27($p < 0.01$) for respectively SSP1-RCP26, SSP2-RCP45 and SSP3-RCP70. The approaches are thus weakly in agreement.

4. Discussion

4.1. Interpretation of the results

We use two different models to explain historical migration patterns in South Asia and find largely similar patterns in variable importance for both models. The evaluation results of both models show that the degree to which net migration can historically be explained differs between countries and between the two approaches. An ROC of 0.83 for the RFC model and the R^2 of 0.44 for the RFR model indicate that both are able to explain a considerable share of historical migration. At the same time, these evaluation metrics also indicate that a significant part of this process is either explained by variables that are not included or that the process is, to some extent, intrinsically uncertain. For example, societal and political sentiment towards migrants or migration and refugee policies (Hatam, 2019), could not be included due to the absence of suitable subnational data. Also, perceptions of risks and opportunities in the place of origin or destination can affect migration decisions, which might not always reflect the objective reality (Fischer et al., 1997). Furthermore, factors such as place attachment, people's aspirations, and cultural preferences contribute to the complexity of understanding migration, rendering it often an irrational and unpredictable decision (de Haas, 2021).

The importance of different variables largely confirms the existing understanding of processes driving migration. In both approaches, the annual average temperature is the most important variable, followed by built-up land. This implies that rising temperatures and a rising proportion of built-up land could change migration patterns considerably. Existing literature has also showed that temperature can affect migration in various ways. For example Cattaneo and Peri (2016) show that higher temperatures can increase out-migration levels in middle-income countries by lowering agricultural productivity. Mueller et al. (2014) conclude that heat stress consistently increases the long-term migration of men in Pakistan, driven by a negative effect on farm and non-farm income. Another explanation for this study could be that in the reference period, people tended to move from the colder, more inaccessible mountainous areas in Nepal and India to warmer, more fertile and urban areas (Biella et al., 2022; Maharjan et al., 2020). This could explain why net in-migration is projected in most areas in the RFR projection of SSP3-RCP70, since temperatures rise quicker and built-up land grows faster compared to SSP1-RCP26 and SSP2-RCP45. This possibly makes areas more attractive in the projections based on the obtained historical relations, although it is questionable to what extent this historical relation will be a predictor for future dynamics. In the RFR SSP3-RCP70 projections, least out-migration is projected in the colder climates of the Mountainous North of India and Nepal, compared to the other two scenarios. This dynamic is not captured in the RFC scenarios, since here out-migration is more probable in the colder mountainous regions in all scenarios.

Contrary to existing insights (Neumann and Hermans, 2017; Niva et al., 2021), we did not find income to be of primary importance for explaining migration. This observation might be explained by the fact that we assess net migration, rather than absolute flows of in- and out-migration While high income regions are potentially attractive to migrants following the perceived economic opportunities while at the same time out-migration is also found to increase with higher incomes due to associated migration costs (Clemens, 2020; de Haas, 2021; Groth et al., 2020). Yet, this might not be visible in the net migration numbers. The

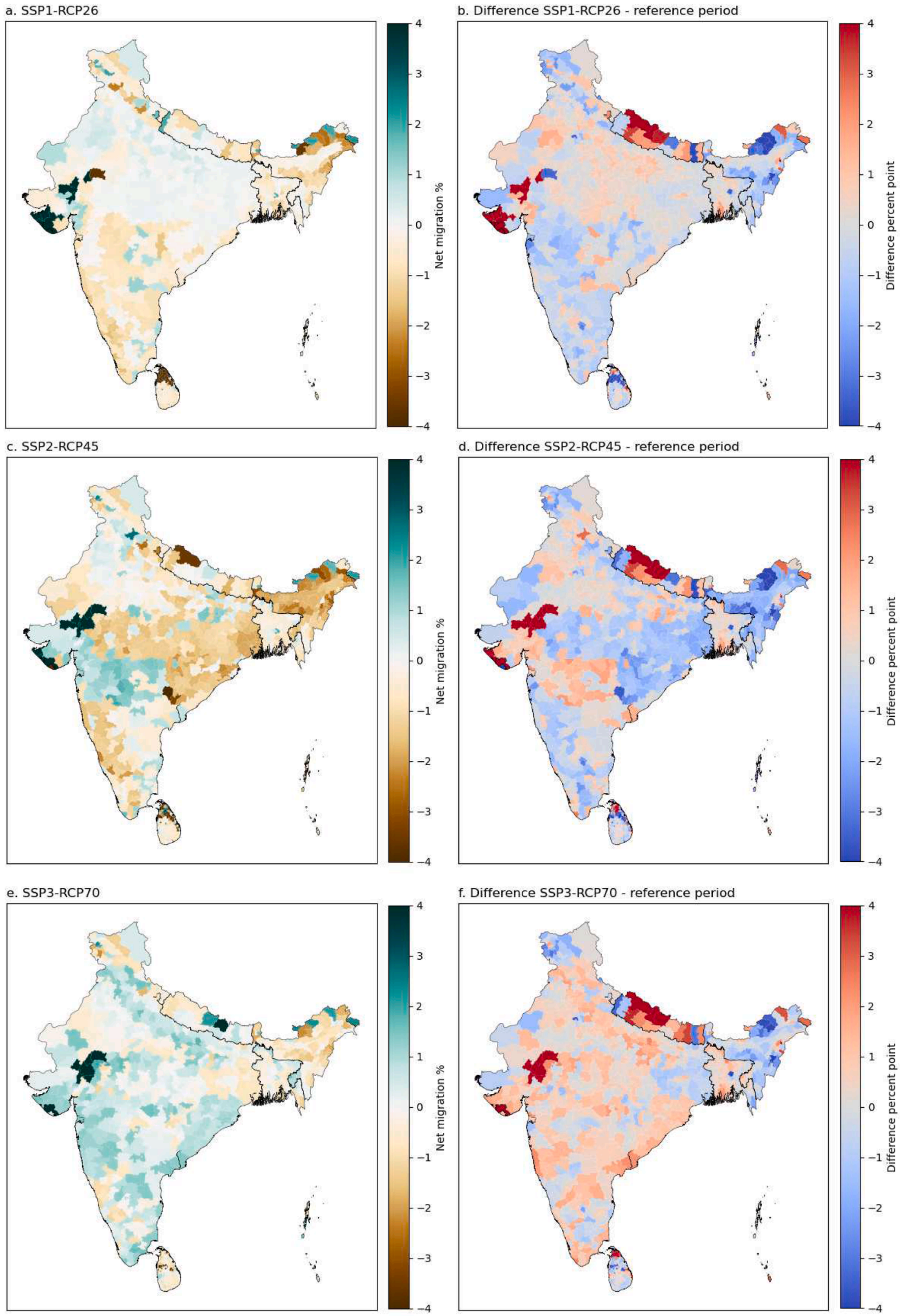

Fig. 3. Left maps: RFR scenario projections of the magnitude of net migration in percentage. Right maps: The absolute difference between the projected 2050 net migration percentage and the average net migration percentage over the reference period 2001–2019.

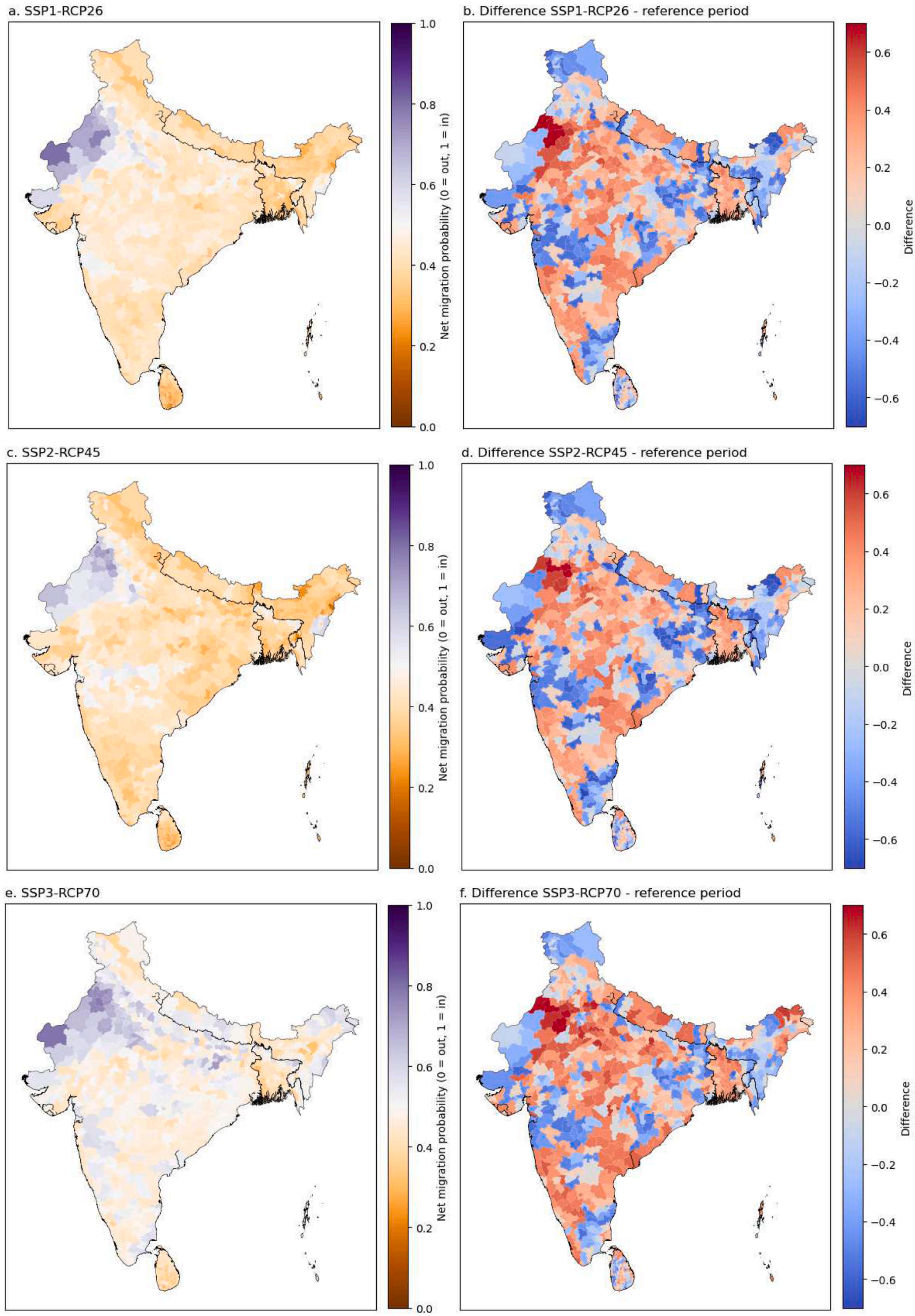

Fig. 4. Left maps: RFC scenario projections of the probability of net in-migration. Right maps: The absolute difference between the projected 2050 probability of net in- (1) to out- (0) migration and the average net in- (1) to out- (0) migration over the reference period.

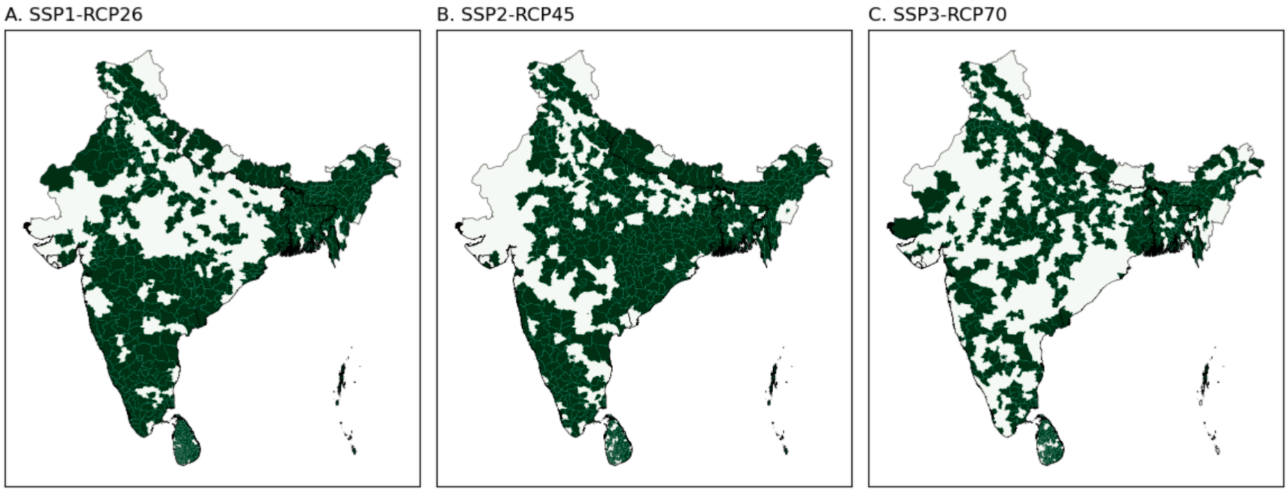

Fig. 5. Overlap in scenario projection direction for (A) SSP1-RCP26, (B) SSP2-RCP45, (C) and SSP3-RCP70.

importance of built-up area can be explained by the fact that South Asia is already for decades characterised by rural to urban migration. This explanatory variable also represents the search for economic opportunities, and its' importance is apparent in both modelling approaches (Maharjan et al., 2020; Selod and Shilpi, 2021). The other variables, both socioeconomic and environmental, have a medium influence, consistent with prior findings as underlying the design of our model (Bohra-Mishra et al., 2014; de Haas 2010b).

Conflict stands out as a variable of little importance in explaining migration, despite the fact that conflict has affected displacement in several of the regions analysed in the reference period, including the long-lasting civil wars in Sri Lanka (1983–2009) and Nepal (1996–2006). Possibly, this reflects the fact that in both countries most people were displaced within their own region (Singh et al., 2007; Steele, 2019). Alternatively, the number of people displaced was limited compared to the total number of migrants in the reference period. Due to a lack of future scenarios on conflicts, this variable was not included in our projections, and the low variable importance suggests this decision will have little effect on our projections.

For Pakistan, no relation was detected between the sample and target data in the training of the RFR model. This could possibly be explained by the fact that the average area size is larger than in other countries, 27,411 km^2, including almost 6 M inhabitants on average. This obscures local differences between more rural and urban areas, as well as differences between socioeconomic and environmental conditions. For Bhutan, no relation between the sample and target could be detected in neither the RFR nor the RFC run. This could be due to the relatively small average administrative area size, 294 km^2, combined with the low total population per administrative area, which makes the analysis more sensitive to local migration drivers, such as labour demand for the construction of a new neighbourhood. An additional explanation could be that the quality of the migration data is limited.

When making out-of-sample predictions for the scenario projections we assume that the historically identified relationship remains stable over time. This is a limitation shared in previous research on conflict projections (Bowlsby et al., 2019; Hoch et al., 2021b). This assumption overlooks the dynamic nature of socioeconomic and environmental conditions, thus potentially limiting the accuracy of long-term migration projections. Climate change impacts in South Asia are expected to worsen, especially after 2050 (Shaw et al., 2022), making it harder to sustain agricultural-dependent livelihoods. Especially temperatures are projected to rise considerably in many places in South Asia, which will affect the habitability of these places (Lenton et al., 2023; Xu et al., 2020). It is not straightforward how this will affect migration, since currently popular migrant destinations, such as Mumbai or Delhi, are projected to be exposed to unliveable temperatures. Still, extremely high temperatures can compromise living conditions in such a way (Xu et al., 2020), that liveability can only be safeguarded when air-conditioned rooms and greenhouses are available to a wide public. Additionally, climate change impacts such as sea level rise and changing monsoon dynamics might affect migration in unprecedented ways (Dallmann and Millock, 2017; Hauer et al., 2020).

The projections for the three SSP-RCP combinations provide a mixed picture. In both the RFR and the RFC projections, we find hotspots of net in-migration in North-western India. North-western India is also in the reference period mostly characterised by net in-migration. However, the extent of in-migration along all scenarios decreases, albeit in different magnitudes, suggesting a decrease in overall net in-migration by 2050. Net out-migration regions across the RFR and the RFC projections are found in most parts of eastern and northern India, and large parts of Sri Lanka. For eastern and northern India, these projections differ from the observed trend in the reference period, in which those regions faced, on average, in-migration. These are the mountainous parts of the country which are rich in cultural heritage, had limited accessibility, and people have a strong place attachment. These characteristics have resulted in reduced out-migration, a trend that could change in a future with more urban development, higher education attainment and cultural connections with wider India. Consequently, the projected difference in migration dynamics when comparing the various SSP-RCP combinations by 2050 with the reference period is plausible. For the South and North of Sri Lanka, net out-migration is projected in all SSP-RCP combinations, while the pattern of net in- or out-migration was heterogenous in the reference period. North Sri Lanka is a dry zone and Jaffna in North-East has been the centre of decades long civil war resulting in isolation and missed developmental opportunities. Lastly, the direction of future migration in most parts of Bangladesh and the remaining parts of India differ between the two approaches and between the SSP-RCP combinations.

4.2. *Challenges related to projecting migration*

In this study we present two Random Forest-based approaches to explain and project migration scenarios for South Asia. The approach of this study is innovative and constraining at the same time. It is innovative, because projecting migration using RF models is novel to the field, although machine-learning approaches have been applied to understand historical migration patterns (Aoga et al., 2024; McLeman et al., 2022; Molina et al., 2023). It is constraining, because there are limitations to this approach, which are discussed in the comparison of both methods below.

The three SSP-RCP scenarios analysed using both methods provide insight in where net in- or out-migration is projected based on the identified historical relations. By developing scenarios for direction (based on the RFC model), and magnitude of migration (based on the RFR model), we can compare the two approaches in terms of variable importance and scenario projections. The two approaches demonstrated reasonably good predictive performances individually and similar variable importance (Fig. 2) for the combined region which implies at first sight comparable underlying processes. However, the migration projections show notable disparities between the scenarios, particularly evident in the SSP3-RCP70 scenario (Fig. 5). The variation in these projections highlights the need to investigate the underlying factors responsible for these discrepancies. The differences in outcomes between both models align with a wider academic concern about analyses and projections based on single model applications (Gould et al., 2023).

The discrepancy between the RFR and RFC model is particularly interesting since both models are driven by the same set of input data and the feature importance of these input data is also well aligned. This could lead to the intuitive expectation that errors would also be correlated, as the same unobserved variables are missing from both models. Instead, the large discrepancy and low correlation suggests that uncertainties in both models are at best weakly related. This observation complicates articulating firm conclusions about future migration patterns. We see the explicit analysis of this uncertainty as a strength of this study, rather than a weakness. When only the results of either the RFR or RFC model would be presented, this uncertainty would remain hidden, and the reasonable validation scores could suggest a higher certainty to either of these.

While our chosen methodological approaches excel in detecting complex, nonlinear, and discontinuous relations between independent and dependent variables, their inherent black box nature restricts transparency. Consequently, the study's contribution to migration theory remains confined to the confirmation of the importance of the variables identified in advance, but we cannot provide further insights into their direction or role in shaping migration patterns, as also concluded by Best et al. (2022). A related limitation of the RFR model is that it cannot predict results outside the scope of the training data (i.e., higher net in-migration or out-migration than observed during the training period). This contrasts with standard regression, which permits extrapolation, enabling predictions beyond the range of observed training data, and limiting the ability of our models to deal with unseen and more extreme socioeconomic and environmental conditions that are expected in South Asia (Shaw et al., 2022).

The value of developing the migration scenarios presented in this study remains significant, despite, and to some extent because of the methodological challenges. Employing multiple methodological approaches can enhance the comprehensiveness of an analysis. While the resulting projections vary, the disparities highlight the complexity and the multifaceted nature of migration. The disparities between the two approaches' projections may raise questions about consistency. However, the contrasting outcomes emphasise the intricacy of migration analysis, and this would remain hidden when presenting only one approach. The main value of this study lies in exploring two methodological approaches, to provide a more nuanced, comprehensive, and critical assessment of migration scenarios. This is not solely valuable in terms of academic interest; this study also provides foundational knowledge for informed discussion in decision-making arenas. This knowledge can contribute to the development of a more comprehensive understanding of what can and what cannot be known about future migration patterns. For example, the influential Groundswell reports (Clement et al., 2021; Rigaud et al., 2018) use one (gravity modelling) approach to project climate-related migration and one evaluation metric. Our study suggests that applying a multi-model approach could provide a more nuanced and complete view on possible migration futures.

5. Concluding remark

Our study has investigated the usefulness of two machine learning approaches to developing net migration scenario projections for different socioeconomic and environmental future trajectories in South Asia. Both approaches achieve a reasonable accuracy for the training period for the major parts of the region, and further analysis shows rather consistent patterns of variable importance. We found that temperature was the most important variable to explain net migration, followed by built-up land and yields. These conditions are expected to change considerably following demographic change and climate change impacts, implying changing migration patterns as well. Besides, we show that with the same data but different methodological approaches scenario projections can differ considerably. While we find overlapping areas of in-migration North-western India and areas projected to face out-migration in eastern and northern India, large parts of Nepal and Sri Lanka, there are also substantial disparities in other areas between the two approaches. This indicates that not one approach to modelling future migration patterns on its own should be considered as complete or reliable. With this study, we take the underdeveloped field of developing net migration scenario projections a step forward.

CRediT authorship contribution statement

Sophie Pieternel de Bruin: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Jannis Hoch:** Writing – review & editing, Software, Methodology. **Jens de Bruijn:** Writing – review & editing, Software. **Kathleen Hermans:** Writing – review & editing, Conceptualization. **Amina Maharjan:** Writing – review & editing, Conceptualization. **Matti Kummu:** Writing – review & editing, Data curation. **Jasper van Vliet:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The open-access and open-source model code used to perform the simulations can be found on GitHub: https://github.com/JannisHoch/copro/tree/Final_migration_run Data available upon request.

Acknowledgements

We acknowledge the contributions of Jonathan Doelman (PBL Netherlands Environmental Assessment Agency), Bep Schrammeijer (Vrije Universiteit Amsterdam) and Niko Wanders (Universiteit Utrecht) for providing data and code snippets.

SdB and JvV were supported by the Netherlands Organization for Scientific Research NWO in the form of a VIDI grant (Grant No VI. Vidi.198.008). MK was supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (SOS.aquaterra, grant agreement No. 819202) and Academy of Finland funded project TREFORM (grant no. 339834).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloenvcha.2024.102920.

References

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

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. Environ. Res. Lett. 10 (6), 060201 https://doi.org/10.1088/1748-9326/10/6/060201.

Adger, W.N., Fransen, S., Safra de Campos, R., Clark, W.C., 2024. Migration and sustainable development. Proc. Natl. Acad. Sci. 121 (3), e2206193121 https://doi.org/10.1073/pnas.2206193121.

Afifi, T., 2011. Economic or environmental migration? The push factors in Niger. Int. Migr. 49, e95–e124. https://doi.org/10.1111/j.1468-2435.2010.00644.x.

Amakrane, K. R., Sarah; Simpson, Nicholas P.; de Sherbinin, Alex; Linekar, Jane; Horwood, Chris; Jones, Bryan; Cottier, Fabien; Adamo, Susana; Mills, Briar; Yetman, Greg; Chai-Onn, Tricia; Squires, John; Schewe, Jacob; Frouws, Bram; Forin, Roberto, 2023. African Shifts: The Africa Climate Mobility Report, Addressing Climate-Forced Migration & Displacement. https://africa.climatemobility.org/report.

Aoga, J.O.R., Bae, J., Veljanoska, S., Nijssen, S., Schaus, P., 2024. Impact of weather factors on migration intention using machine learning algorithms. Oper. Res. For. 5 (1), 8. https://doi.org/10.1007/s43069-023-00271-y.

Beine, M., Parsons, C., 2015. Climatic factors as determinants of international migration. Scand. J. Econ. 117 (2), 723–767. https://doi.org/10.1111/sjoe.12098.

Best, K., Gilligan, J., Baroud, H., Carrico, A., Donato, K., Mallick, B., 2022. Applying machine learning to social datasets: a study of migration in southwestern Bangladesh using random forests. Reg. Environ. Chang. 22 (2), 52. https://doi.org/10.1007/s10113-022-01915-1.

Beyer, R.M., Schewe, J., Lotze-Campen, H., 2022. Gravity models do not explain, and cannot predict, international migration dynamics. Humanities and Social Sciences Communications 9 (1), 56. https://doi.org/10.1057/s41599-022-01067-x.

Beyer, R.M., Schewe, J., Abel, G.J., 2023. Modeling climate migration: dead ends and new avenues [Hypothesis and Theory]. Front. Clim. 5 https://doi.org/10.3389/fclim.2023.1212649.

Biella, R., Hoffmann, R., Upadhyay, H., 2022. Climate, agriculture, and migration: exploring the vulnerability and outmigration nexus in the Indian Himalayan Region. Mt. Res. Dev. 42 (2) https://doi.org/10.1659/MRD-JOURNAL-D-21-00058.1.

Boas, I., Farbotko, C., Adams, H., Sterly, H., Bush, S., van der Geest, K., Wiegel, H., Ashraf, H., Baldwin, A., Bettini, G., Blondin, S., de Bruijn, M., Durand-Delacre, D., Fröhlich, C., Gioli, G., Guaita, L., Hut, E., Jarawura, F.X., Lamers, M., Lietaer, S., Nash, S.L., Piguet, E., Rothe, D., Sakdapolrak, P., Smith, L., Tripathy Furlong, B., Turhan, E., Warner, J., Zickgraf, C., Black, R., Hulme, M., 2019. Climate migration myths. Nat. Clim. Chang. 9 (12), 901–903. https://doi.org/10.1038/s41558-019-0633-3.

Bohra-Mishra, P., Oppenheimer, M., Hsiang, S.M., 2014. Nonlinear permanent migration response to climatic variations but minimal response to disasters. Proc. Natl. Acad. Sci. 111 (27), 9780–9785. https://doi.org/10.1073/pnas.1317166111.

Bowlsby, D., Chenoweth, E., Hendrix, C., Moyer, J.D., 2019. The future is a moving target: predicting political instability. Br. J. Polit. Sci. 1–13 https://doi.org/10.1017/S0007123418000443.

Braithwaite, A., Salehyan, I., Savun, B., 2019. Refugees, forced migration, and conflict: introduction to the special issue. J. Peace Res. 56 (1), 5–11. https://doi.org/10.1177/0022343318814128.

Buhaug, H., Vestby, J., 2019. On growth projections in the shared socioeconomic pathways. Global Environ. Polit. 19 (4), 118–132. https://doi.org/10.1162/glep_a_00525.

Carrico, A.R., Donato, K., 2019. Extreme weather and migration: evidence from Bangladesh. Popul. Environ. 41 (1), 1–31. https://doi.org/10.1007/s11111-019-00322-9.

Cattaneo, C., Peri, G., 2016. The migration response to increasing temperatures. J. Dev. Econ. 122, 127–146. https://doi.org/10.1016/j.jdeveco.2016.05.004.

Cattaneo, C., Beine, M., Fröhlich, C.J., Kniveton, D., Martinez-Zarzoso, I., Mastrorillo, M., Millock, K., Piguet, E., Schraven, B., 2019. Human migration in the era of climate change. Rev. Environ. Econ. Policy 13 (2), 189–206. https://doi.org/10.1093/reep/rez008.

Chan, J.-Y.-L., Leow, S.M.H., Bea, K.T., Cheng, W.K., Phoong, S.W., Hong, Z.-W., Chen, Y.-L., 2022. Mitigating the multicollinearity problem and its machine learning approach: a review. Mathematics 10 (8), 1283. https://www.mdpi.com/2227-7390/10/8/1283.

Cissé, G., R. McLeman, H. Adams, P. Aldunce, K. Bowen, D. Campbell-Lendrum, S. Clayton, K.L. Ebi, J. Hess, C. Huang, Q. Liu, G. McGregor, J. Semenza, M.C. Tirado. (2022). Health, Wellbeing, and the Changing Structure of Communities. In D. C. R. H.-O. Pörtner, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (Ed.), Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 1041–1170). Cambridge University. 10.1017/9781009325844.009.

Clemens, M.A., 2020. Migration from Developing Countries: Selection, Income Elasticity and Simpson's Paradox. Centro Studi Luca d'Agliano Development Studies Working Paper(465).

Clement, V., Rigaud, K. K., de Sherbinin, A., Jones, B., Adamo, S., Schewe, J., Sadiq, N., & Shabahat, E. (2021). Groundswell Part 2 : Acting on Internal Climate Migration. https://openknowledge.worldbank.org/handle/10986/36248.

Cundill, G., Singh, C., Adger, W.N., Safra de Campos, R., Vincent, K., Tebboth, M., Maharjan, A., 2021. Toward a climate mobilities research agenda: Intersectionality, immobility, and policy responses. Glob. Environ. Chang. 69, 102315 https://doi.org/10.1016/j.gloenvcha.2021.102315.

Dallmann, I., Millock, K., 2017. Climate variability and inter-state migration in India. CESifo Econ. Stud. 63 (4), 560–594. https://doi.org/10.1093/cesifo/ifx014.

Davies, S., Pettersson, T., Öberg, M., 2023. Organized violence 1989–2022 and the return of conflicts between states? J. Peace Res. 60 https://doi.org/10.1177/00223433231185169.

de Haas, H., 2010a. Migration and development: a theoretical perspective. Int. Migr. Rev. 44 (1), 227–264. http://www.jstor.org/stable/20681751.

de Haas, H., 2010b. Migration Transitions: a Theoretical and Empirical Inquiry into the Developmental Drivers of International Migration, Vol. 24. International Migration Institute, University of Oxford.

de Haas, H., 2021. A theory of migration: the aspirations-capabilities framework. Compar. Migr. Stud. 9 (1), 8. https://doi.org/10.1186/s40878-020-00210-4.

Doelman, J.C., Stehfest, E., Tabeau, A., van Meijl, H., Lassaletta, L., Gernaat, D.E., Hermans, K., Harmsen, M., Daioglou, V., Biemans, H., 2018. Exploring SSP land-use dynamics using the IMAGE model: regional and gridded scenarios of land-use change and land-based climate change mitigation. Glob. Environ. Chang. 48, 119–135. https://doi.org/10.1016/j.gloenvcha.2017.11.014.

Falco, C., Galeotti, M., Olper, A., 2019. Climate change and migration: Is agriculture the main channel? Glob. Environ. Chang. 59, 101995 https://doi.org/10.1016/j.gloenvcha.2019.101995.

Fischer, P.A., Martin, R., Straubhaar, T., 1997. Should I Stay or Should I Go? In: Hammar, T., Brochmann, G., Tamas, K., Faist, T. (Eds.), International Migration. Routledge, Immobility and Development.

Global Administrative Areas, 2022. GADM database of Global Administrative Areas, version 4.1 Retrieved 4 June 2023 from www.gadm.org.

Gould, E., Fraser H., Parker, T.H., Nakagawa, S., Griffith, S.C., Vesk, P.A., Fidler, F., Hamilton, D.G., Abbey-Lee, R.N. & Abbott, J.K., e. a. (2023). Same data, different analysts: variation ineffect sizes due to analytical decisions inecology and evolutionary biology. *Preprint at EcoEvoRxiv*. https://doi.org/10.32942/X2GG62.

Grecequet, M., DeWaard, J., Hellmann, J.J., Abel, G.J., 2017. Climate vulnerability and human migration in global perspective. Sustainability 9 (5), 720. https://www.mdpi.com/2071-1050/9/5/720.

Groth, J., Ide, T., Sakdapolrak, P., Kassa, E., Hermans, K., 2020. Deciphering interwoven drivers of environment-related migration – a multisite case study from the Ethiopian highlands. Glob. Environ. Chang. 63, 102094 https://doi.org/10.1016/j.gloenvcha.2020.102094.

Hatam, A.K., 2019. Repatriation of Afghan Refugees from Pakistan: A Legal Analysis (Part-II). Kardan J. Social Sci. Human., 2, 1–18. 10.31841/KJSSH.2021.2.

Hathie, I., Wade, I., Ba, S., Niang, A., Niang, M., Sow, M.M.K., Ndione, M.Y.C., Ba, C.O., 2015. Emploi des jeunes et migration en Afrique de l'Ouest. https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/54153/IDL-54153.pdf.

Hauer, M.E., Fussell, E., Mueller, V., Burkett, M., Call, M., Abel, K., McLeman, R., Wrathall, D., 2020. Sea-level rise and human migration. Nat. Rev. Earth Environ. 1 (1), 28–39. https://doi.org/10.1038/s43017-019-0002-9.

Hermans, K., McLeman, R., 2021. Climate change, drought, land degradation and migration: exploring the linkages. Curr. Opin. Environ. Sustain. 50, 236–244. https://doi.org/10.1016/j.cosust.2021.04.013.

Hermans-Neumann, K., Priess, J., Herold, M., 2017. Human migration, climate variability, and land degradation: hotspots of socio-ecological pressure in Ethiopia. Reg. Environ. Chang. 17 (5), 1479–1492. https://doi.org/10.1007/s10113-017-1108-6.

Hoch, J.M., de Bruin, S., Wanders, N., 2021a. CoPro: a data-driven modelling framework for conflict risk projections. J. Open Source Softw. 6 (58), 2855. https://doi.org/10.21105/joss.02855.

Hoch, J.M., de Bruin, S.P., Buhaug, H., von Uexkull, N., van Beek, R., Wanders, N., 2021b. Projecting armed conflict risk in Africa towards 2050 along the SSP-RCP scenarios: a machine learning approach. Environ. Res. Lett. 16 (12) https://doi.org/10.1088/1748-9326/ac3db2.

Horton, R.M., de Sherbinin, A., Wrathall, D., Oppenheimer, M., 2021. Assessing human habitability and migration. Science 372 (6548), 1279–1283. https://doi.org/10.1126/science.abi8603.

Hunter, L.M., Simon, D.H., 2023. Time to mainstream the environment into migration theory? Int. Migr. Rev. 57 (1), 5–35. https://doi.org/10.1177/01979183221074343.

IOM, 2022. World Migration Report 2022.

Kriegler, E., O'Neill, B.C., Hallegatte, S., Kram, T., Lempert, R.J., Moss, R.H., Wilbanks, T., 2012. The need for and use of socio-economic scenarios for climate change analysis: a new approach based on shared socio-economic pathways. Glob. Environ. Chang. 22 (4), 807–822.

Kummu, M., Taka, M., Guillaume, J.H.A., 2018. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. Scientific Data 5 (1), 180004. https://doi.org/10.1038/sdata.2018.4.

Lenton, T.M., Xu, C., Abrams, J.F., Ghadiali, A., Loriani, S., Sakschewski, B., Zimm, C., Ebi, K.L., Dunn, R.R., Svenning, J.-C., Scheffer, M., 2023. Quantifying the human cost of global warming. Nat. Sustainability. https://doi.org/10.1038/s41893-023-01132-6.

Maharjan, A., Campos, R.S.d., Singh, C., Das, S., Srinivas, A., Bhuiyan, M.R.A., Ishaq, S., Umar, M.A., Dilshad, T., Shrestha, K., Bhadwal, S., Ghosh, T., Suckall, N., Vincent, K., 2020. Migration and household adaptation in climate-sensitive hotspots in South Asia. Curr. Clim. Change Rep. 6, 1–16. https://doi.org/10.1007/s40641-020-00153-z.

McLeman, R., 2013. Developments in modelling of climate change-related migration. Clim. Change 117 (3), 599–611.

McLeman, R., Wrathall, D., Gilmore, E., Thornton, P., Adams, H., Gemenne, F., 2021. Conceptual framing to link climate risk assessments and climate-migration scholarship. Clim. Change 165 (1), 24. https://doi.org/10.1007/s10584-021-03056-6.

McLeman, R., Grieg, C., Heath, G., Robertson, C., 2022. A machine learning analysis of drought and rural population change on the North American Great Plains since the 1970s. Popul. Environ. 43 (4), 500–529. https://doi.org/10.1007/s11111-022-00399-9.

Molina, M.D., Chau, N., Rodewald, A.D., Garip, F., 2023. How to model the weather-migration link: a machine-learning approach to variable selection in the Mexico-U.S. context. J. Ethn. Migr. Stud. 49 (2), 465–491. https://doi.org/10.1080/1369183X.2022.2100549.

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

Mueller, V., Gray, C., Hopping, D., 2020. Climate-Induced migration and unemployment in middle-income Africa. Glob. Environ. Chang. 65, 102183 https://doi.org/10.1016/j.gloenvcha.2020.102183.

Neumann, K., Hermans, F., 2017. What drives human migration in Sahelian Countries? A meta-analysis. Population, Space, Place 23 (1), e1962.

Niva, V., Kallio, M., Muttarak, R., Taka, M., Varis, O., Kummu, M., 2021. Global migration is driven by the complex interplay between environmental and social factors. Environ. Res. Lett. 16 (11), 114019 https://doi.org/10.1088/1748-9326/ac2e86.

Niva, V., Horton, A., Virkki, V., Heino, M., Kallio, M., Kinnunen, P., Abel, G.J., Muttarak, R., Taka, M., Varis, O., Kummu, M., 2023a. Data for: World's human migration patterns in 2000–2019 unveiled by high-resolution data. Zenodo. https://doi.org/10.5281/zenodo.7997134.

Niva, V., Horton, A., Virkki, V., Heino, M., Kosonen, M., Kallio, M., Kinnunen, P., Abel, G. J., Muttarak, R., Taka, M., Varis, O., Kummu, M., 2023b. World's human migration patterns in 2000–2019 unveiled by high-resolution data. Nat. Hum. Behav. https://doi.org/10.1038/s41562-023-01689-4.

O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K., Sanderson, B.M., 2016. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geosci. Model Dev. 9 (9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016.

Oakes, R., Van der Geest, K., Schraven, B., Adaawen, S., Ayeb-Karlsson, S., de Sherbinin, A., Etzold, B., Groth, J., Hermanns, K., Lakeman, S., Nawrotzki, R., Rademacher-Schulz, C., Romankiewic, C., Serraglio, D., Sterly, H., Thalheimer, L., Wiederkehr, C., Williams, D., 2023. A future agenda for research on climate change and human mobility. Int. Migr. 61 (5), 116–125. https://doi.org/10.1111/imig.13169.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12 (85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html.

Piguet, E., 2022. Linking climate change, environmental degradation, and migration: An update after 10 years. WIREs Clim. Change 13 (1), e746.

Rigaud, K. K., de Sherbinin, A., Jones, B., Bergmann, J., Clement, V., Ober, K., Schewe, J., Adamo, S., McCusker, B., & Heuser, S. (2018). Groundswell : Preparing for Internal Climate Migration.

Robinson, C., & Dilkina, B. (2018). A Machine Learning Approach to Modeling Human Migration Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, Menlo Park and San Jose, CA, USA. 10.1145/3209811.3209868.

Selod, H., Shilpi, F., 2021. Rural-urban migration in developing countries: Lessons from the literature. Reg. Sci. Urban Econ. 91, 103713 https://doi.org/10.1016/j.regsciurbeco.2021.103713.

Shaw, R., Y. Luo, T.S. Cheong, S. Abdul Halim, S. Chaturvedi, M. Hashizume, G.E. Insarov, Y. Ishikawa, M. Jafari, A. Kitoh, J. Pulhin, C. Singh, K. Vasant, and Z. Zhang. (2022). Asia. In D. C. R. H.-O. Pörtner, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (Ed.), Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 1457–1579). Cambridge University Press. 10.1017/9781009325844.012.

Singh, S., Sharma, S.P., Mills, E., Poudel, K.C., Jimba, M., 2007. Conflict induced internal displacement in Nepal. Med. Confl. Surviv. 23 (2), 103–110. https://doi.org/10.1080/13623690701248088.

Skicit-learn. (2023). 3.3. Metrics and scoring: quantifying the quality of predictions. Retrieved 6 June from https://scikit-learn.org/dev/modules/model_evaluation.html.

Smits, J., Permanyer, I., 2019. The Subnational Human Development Database. Sci. Data. 10.1038/sdata.2019.386:190038.

Solt, F., 2020. Measuring income inequality across countries and over time: the standardized world income inequality database. Soc. Sci. Q. 101 (3), 1183–1199. https://doi.org/10.1111/ssqu.12795.

Steele, A., 2019. Civilian resettlement patterns in civil war. J. Peace Res. 56 (1), 28–41. https://doi.org/10.1177/0022343318820576.

Stehfest, E., Vuuren, D.v., Kram, T., Bouwman, L., Alkemade, R., Bakkenes, M., Biemans, H., Bouwman, A., Elzen, M.d., Janse, J., Lucas, P., Minnen, J.v., Müller, C., Prins, A.G., 2014. Integrated Assessment of Global Environmental Change with IMAGE 3.0. Model description and policy applications. PBL Netherlands Environmental Assessment Agency.

Sundberg, R., Melander, E., 2013. Introducing the UCDP Georeferenced Event Dataset. J. Peace Res. 50 (4), 523–532. https://doi.org/10.1177/0022343313484347.

Sutanudjaja, E.H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J.H., Drost, N., Van Der Ent, R.J., De Graaf, I.E., Hoch, J.M., De Jong, K., 2018. PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. Geosci. Model Dev. 11 (6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K.B., Hurtt, G., Kriegler, E., Lamarque, J.F., Meehl, G., Moss, 2021. Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. Earth Syst. Dynam. 12 (1), 253–293. https://doi.org/10.5194/esd-12-253-2021.

Thober, J., Schwarz, N., Hermans, K., 2018. Agent-based modeling of environment-migration linkages: a review. Ecol. Soc., 23(2), Article 41. 10.5751/ES-10200-230241.

Ting, K.M., 2011. Encyclopedia of machine learning. Springer. https://doi.org/10.1007/978-0-387-30164-8.

Tucker, J., Daoud, M., Oates, N., Few, R., Conway, D., Mtisi, S., Matheson, S., 2015. Social vulnerability in three high-poverty climate change hot spots: What does the climate change literature tell us? Reg. Environ. Chang. 15 (5), 783–800. https://doi.org/10.1007/s10113-014-0741-6.

UNHCR, 2022, 31 December 2022. Operational Data Portal Country view. Pakistan (Islamic Republic of). The UN Refugee Agency. Retrieved 28 September 2023 from https://data2.unhcr.org/en/country/pak.

UNHCR, 2023. Afghanistan Refugee Crisis Explained. The UN Refugee Agency. Retrieved 5 September 2023 from https://www.unrefugees.org/news/afghanistan-refugee-crisis-explained/#:~:text=Since%202002%2C%20UNHCR%20has%20supported,5.3%20million%20refugees%20to%20Afghanistan.

van der Geest, K., 2011. North-South migration in Ghana: what role for the environment? Int. Migr. 49, e69–e94. https://doi.org/10.1111/j.1468-2435.2010.00645.x.

Wolff, S., Schrammeijer, E.A., Schulp, C.J.E., Verburg, P.H., 2018. Meeting global land restoration and protection targets: What would the world look like in 2050? Glob. Environ. Chang. 52, 259–272. https://doi.org/10.1016/j.gloenvcha.2018.08.002.

World Bank Group, 2023. Climate Change Knowledge Portal. https://climateknowledgeportal.worldbank.org/download-data.

WorldPop, 2021. WorldPop Hub dataset. 10.5258/SOTON/WP00647.

Xu, C., Kohler, T.A., Lenton, T.M., Svenning, J.-C., Scheffer, M., 2020. Future of the human climate niche. Proc. Natl. Acad. Sci. 117 (21), 11350–11355. https://doi.org/10.1073/pnas.1910114117.

Østby, G., 2016. Rural–urban migration, inequality and urban social disorder: Evidence from African and Asian cities. Confl. Manag. Peace Sci. 33 (5), 491–515. https://doi.org/10.1177/0738894215581315.