The state-of-the-art on

Drought displacement modelling

iDMC internal displacement monitoring centre

Acknowledgements

Authors: José María Tárraga Habas – UVEG,
Eva Sevillano Marco – UVEG,
Maria Teresa Miranda – IDMC

Acknowledgements: Universitat De València (UVEG) (Gustau Camps-Valls), UNHCR Innovation Service (Catherine Schneider, Rebeca Moreno Jimenez, Sofia Kyriazi), IDMC (Bina Desai, Sylvain Ponserre)

This research was made possible thanks to the generous support of the Federal Foreign Office of Germany

Cover photo: *Women in Djibo, Burkina Faso, filling 20-litre watercans. © Jacques Bouda / NRC, April 2022.*

Federal Republic of Germany
Foreign Office

Children fetch water from the outskirts of Savo displacement camp, in Djugu Territory, Ituri Province, the Democratic Republic of the Congo. © Hugh Kinsella Cunningham / NRC, April 2022.

Table of contents

6 Abstract

8 Key messages

12 Introduction

14 Drought and internal displacement from the modelling perspective

Primary goals of drought mobility models 15

Hypothesis testing 15

Scenario simulation 15

Time-Series forecasting 16

Potential applications of the models reviewed 16

Support of operational actors with displacement forecasting 17

Projection intentions 17

Stated policy applications 18

19 Existing drought mobility models: strengths and weaknesses

Logistic regression analysis 20

Econometric models 20

Gravity models 21

Radiation models 22

Fixed effects linear regression models 22

Complex systems modelling: agent-based, system dynamics models, Bayesian networks 22

Agent-based models (ABM) 22

System dynamics modelling (SD) 23

Bayesian networks (BN) 24

Other models 25

Tree-based ensembles 25

Exposure modelling using K-means clustering 25

26 Main challenges and data gaps to model internal displacement triggered by drought

Lack of monitoring of drought impacts 26

Fulfilling the sufficiency assumption 27

Lack of data on mobility patterns and internal displacement 27

Lack of monitoring of affected and displaced populations 28

Accounting for complex dynamics 28

Accounting for accumulated impacts of drought 29

Identifying displacement tipping points 29

Quantifying the importance of multiple drivers of displacement 29

Identifying correct attributions 29

Fitting the right conceptual framework 30

Biases in the collection of drought impacts and displacement data 30

Biases in the integration and aggregation of displacement data 30

Estimating uncertainty in models 31

32 Mobility datasets used in drought modelling

Survey data (microdata) 33

Census data 34

Net migration estimates 34

Time-series data 35

Openess of mobility data 35

Drought and internal displacement, why do we need this data? 36

39 Covariates or driving factors used to model drought related human mobility

A boy walks over a bridge in a dry riverbed in North-eastern Syria. ©Tareq Mnadili/ NRC, June 2021.

Availability of covariates 41

42 Main data gaps identified

44 Recommendations to improve drought displacement modelling

45 Improving internal displacement timeseries data

Implementing best practices based on key datasets (UNHCR PRMN Somalia IOM DMT site assessments in Ethiopia) 46

Supporting the collection of even spatiotemporal sampling 46

Improving the accessibility to flow data (new displacements) and total population counts (IDP stock data) 46

Implementing multicausal drivers for displacement reasons 47

Supporting the dissemination of open data, among data collectors of drought impacts and displacement datasets 47

Supporting the data collection of comparable data between countries and organizations 47

48 Improving modelling methodologies: ways forward

Using innovative modelling techniques 48

Better assess models using explainable AI (XAI) approaches 49

Exploring causal discovery and causal Inference techniques 49

Exploiting available data sources and state-of-the-art drought indexes 50

Implementing model evaluation or peer review systems 50

Model access and reusability 51

Uncertainties and bias assessment 52

Interdisciplinary and multisectoral collaboration 53

54 Endnotes

Abstract

This report aims to provide a support tool for the humanitarian and development community to guide their understanding of existing data-driven methodologies in forecasting or comprehending the risk of future drought displacement episodes. A literature review of 42 academic articles and open access documentation about models used in the humanitarian sector was conducted to synthesize the state-of-the-art of drought displacement modelling. Its purpose is (1) to document the strengths and weaknesses of current drought modelling approaches in which ten different modelling approaches have been identified, (2) to show the current state of data availability and existing data gaps in order to improve data collection for mobility variables and confounders of drought-related human mobility, (3) to expose current challenges that must be overcome from the modelling standpoint to build robust and reliable models and (4) understanding model limitations and constraints in terms of fit-for-purpose and application requirements, specifically through a deep analysis of the suitability of models to support evidence-based policymaking. Finally, the report highlights some suggestions for improving current modelling methodologies. The complete catalogue of the reviewed modelling approaches is included in *Appendix I*.

Naim Hamamdeh irrigates his olive trees next to hishome in Umm Fagarah, in the southern West Bank'sMasafer Yatta area. © NRC, November 2021.

Key messages

1

Drought displacement modelling is a budding field where major advances can be made.

A literature review of 42 research articles, conference papers, technical reports and websites was conducted. Of these, only two models explicitly focused on drought-related internal displacement.

2

Understanding how drought affects different patterns of human mobility is a research priority.

Drought displacement modelling should aim to address context specific situations. Not all types of human mobility arise from the same drivers or for the same reasons.

The quantification of drivers would allow for simulation models that can determine the links and influence between components with a certain level of confidence.

3

The question of which specific tipping points trigger internal displacement is still unanswered.

It requires a better understanding of how and when environmental stressors caused by drought surpass critical socioecological thresholds.

4

Determining windows in which the different variables affect drought displacement is essential to refine results.

For example, determining the time-lag when rainfall deficit starts affecting displacement is important for crisis prevention applications.

5

The impacts of drought should be understood less as a natural hazard and more as a social phenomenon.

Demographic and socioeconomic factors have been shown to play a more important role than environmental factors or climate shocks in shaping displacement outcomes. Mitigating risk and building resilience can significantly reduce impacts.

6

Developing accessible and reusable models is the only way to ensure collective progress and return on investment.

Donors, academics, and humanitarian and development organisations should collaborate to catalogue and promote transparent models that support operational decision-making. We also recommend introducing standards that guarantee the reliability, comparability and re-usability of models, the latter of which will save time and money.

Challenges to model drought displacement – key messages:

- Current limitations and challenges in modelling drought displacement consist, in summary, of three main factors: a lack of data access and availability; the need to improve existing methodologies for modelling drought-related mobility; and the current lack of understanding of the processes that lead to drought displacement in context-specific situations.

- Mobility dynamics are intricate. Non-linear relationships, socio-ecological feedback effects, autoregressive effects, interactions between variables, non-stationarities and adaptive responses are to be expected in mobility dynamics. In this regard, describing temporal patterns of human mobility is inherently more challenging than characterizing spatial patterns.

- Multisectoral information about pre-drought conditions is highly relevant to understanding the impacts of drought.

- The affected population includes all those whose lives have been exposed to the drought hazard. Establishing reliable figures has a direct effect when addressing exposure mapping, populations at risk of internal displacement, and projections of internal displacement.

Key challenges from the modelling perspective

- Logistic regression is often used to analyse panel data that targets the factors that affect or do not affect mobility (hypothesis testing). It is used because the models are simple and interpretable and work with binary data. These models, however, are not good at dealing with non-linear relationships, and multicollinearity and confounding effects must be accounted for. They can be used for gathering evidence on the factors that affect drought displacement and support decision-making.

- Econometric models are used if one is interested in understanding and predicting where and why people move to certain regions. Gravity models, for example, are highly flexible and interpretable. They can be used to infer the reasons of displacement to different regions in the interests of preparedness, to understand the pressures at the points of destination, and particular policy formulations. Implementing them to predict and project patterns of displacement in time, however, must be done with extreme caution.

- Complex system and simulation modelling, such as system dynamics, agent-based models and Bayesian networks, are used if one is interested in explicitly understanding the complex relationships involved and performing interventions on its components to outline possible hypothetical scenarios of drought displacement, rather than to make accurate predictions. These models, however, require that the underlying theory be precise and a great amount of potentially unavailable data for calibration. It must be noted that in modelling a complex system, a simple conceptual approach will be unable to fully characterize the system. Models that are too complex, however, could be difficult to calibrate, validate or interpret and, without the necessary data, it will be impossible to do so.

- Some of the extrapolations from the models reviewed were conducted without taking into consideration out-of-sample data (i.e., data that has not been used to train the model). This is a crucial step in model validation which compromises models' reliability.

- The lack of use of other advanced machine-learning algorithms, eXplainable AI, or inference with causal AI in the reviewed literature indicate this is a young field where many advances can be achieved.

- The combination of different approaches can help deliver quality operational products. This includes exploiting different strengths of the applicable models and combining advanced machine-learning algorithms, eXplainable AI, or inference with causal AI with traditional methods.

- One must consider that models are as good as the data they are trained on, and their computed errors are usually based on how well the model replicates the observed data. For this reason, uncertainty estimations of models are heavily related to the bias in available data.

Key messages from the data perspective:

We must bear in mind that the usefulness of any data-driven methodology depends ultimately on the specific problem it tries to solve and the nature of the data. Datasets with the potential to characterize drought displacement, however, often do not exist, are not openly accessible, are incomplete or present high degrees of uncertainty. For the moment, only in situ measurements such as the International Organization for Migration's Displacement Tracking Matrix (IOM DTM) Ethiopia, the UN Refugee Agency (UNHCR) Somalia Population Movement Tracker (PMT), and the Protection and Return Monitoring Network (PRMN) data are available as reliable data sources of drought internal displacement for modelling purposes. Donors and operational actors could play a role promoting and disseminating such good practices and scaling them up.

- Improving the quality of existing models will require significant investments in data collection, including increasing efforts dedicated to the collection of data on the interactions of households within communities, villages and district-level systems.

- As counterintuitive as it may seem, monitoring areas in which drought impacts are less severe is equally important from the modelling standpoint. Data collection in less affected areas allows for a comparison of affected regions and the identification of the specific factors (climatic and structural) that are driving displacement and at what magnitude.

- Both long-term and high spatiotemporal resolution data are required to fully characterize slow-onset displacement dynamics. Longitudinal (long-term) geospatial data is needed to track the slow accumulation of environmental and societal changes leading to displacement. Consistent records of micro-level data on drought impacts, such as households' capital losses (crop yields, livestock deaths or any other related loss), is not available.

- Existing internal displacement data presents different biases and uncertainties which should be considered. After all, models extract their conclusions from the data they are trained on, and biased data can skew conclusions or produce inaccurate predictions. We identify representation or selection bias, location bias, historical bias, interviewer bias (this could also be related to language bias), response bias and aggregation bias as principal limitations in currently available human mobility data.

- We strongly recommend that data collection methodologies present comparable and harmonizable data between countries. Model and empirical evidence intercomparison could permit the establishment of common frameworks of data and model evaluation.

- Most of the reviewed models use climate, demographic and socioeconomic data as the main drivers for modelling drought displacement. A smaller number of the models, however, account for an explicit agricultural or livestock pathway to drought, or for environmental stressors, such as vegetation indexes. In addition, there are also political indicators and data on violent conflict, land degradation variables, and geographical or migration related data, such as migration networks. All of these have been found to have an impact on modelling displacement across the literature.

- Earth observation guarantees a globally consistent and continuous record of climatic and environmental data and human processes monitoring. Advances in exploiting these datasets are at the frontier of a better understanding of drought displacement drivers.

- We recommend that the elaboration of all databases not containing sensitive information compromising the wellbeing of affected populations should follow FAIR (findable, accessible, interoperable, reusable) principles. By reaching such standards, it can accelerate innovation and maximize research impacts to help develop tools that can better assist affected populations and prevent and reduce future displacements.

Introduction

The direct and indirect impacts of drought pose major challenges for populations depending primarily on natural resources. They can generate major disruptions in people's livelihoods, food security, the economy, and ecosystems[2]. Forced displacement triggered by drought occurs when the direct or indirect impacts of drought push communities to critical thresholds and erodes their traditional coping strategies (such as mobility), making livelihoods unsustainable or unviable. For that reason, drought can be a driver or an amplifier of internal displacement and other forms of human mobility when it results in increasing food insecurity, the erosion of livelihoods systems, livestock loss or damage, or economic loss, or when it results in resource depletion, or inaccessible water or pastureland[3]. Drought can also increase the risk or exposure to violence and insecurity, leading to forced displacement[4].

IDMC reported about 1.8 million internal displacements resulting from drought conditions in Ethiopia and Somalia between 2017 and 2020[5]. IOM estimated in 2022 that around 1.4 million displacements could happen as result of drought in Somalia alone[6]. Data on internal displacement resulting from drought conditions, however, often underestimates the problem, and available data is sparse.

Drought and its impacts are among the most difficult hazards to monitor, more so than sudden-onset disaster events. Droughts are episodic (time-limited events), complex and multifaceted phenomena that result from intricate interactions between natural processes and human activities. They occur when there is an extreme lack of water compared with normal circumstances[7]. For that reason, droughts unfold at different spatial and temporal scales (months, years, and even decades) and can result from different cascading events. Understanding the characteristics and the development of drought events and their societal impacts is crucial to anticipating and reducing the risk of their negative impacts. It is also essential to improving emergency response and drought risk reduction and preparedness actions. Only with a thorough comprehension of drought displacement drivers and triggers can effective assistance strategies, prevention measures and future drought displacement risk assessments be implemented, and the impacts of drought ameliorated.

Tackling drought mobility, and particularly drought displacement, from a data-driven perspective is still at an early stage. More advanced methodologies have only been applied in this area in the last decade. One of the main findings of this report is that internal displacement is underrepresented by current modelling approaches. Recent advances in data acquisition methods related to displacement monitoring, Earth observation from satellite imagery and the appearance of alternative data sources[i], have boosted the potential and development of novel techniques to quantify the impacts of extreme events and to model drought displacement. These models and methods, however, have not been systematically compared.

The goal of this report is to present an overview of the state-of-the-art of drought displacement modelling based on a literature review of drought displacement models. To provide the full picture, other climate or environmental proxies of drought where also included. For this purpose, we compiled and organized the information in a catalogue that lists all the models found in the literature. Overall, 42 research articles, conference papers, technical reports, and websites (model documentation or code repository) were reviewed. Two of these articles represent the humanitarian sector. The review was conducted in December 2021 via desktop research. The articles reviewed cover the period from 2001 to 2021. The catalogue describes models that were developed using different methodologies (around eight model types) and spatial resolutions[ii] in 23 countries[iii] and regions, such as sub-Saharan Africa and the Horn of Africa, and worldwide projections. Different forms of human mobility were covered by the literature review, including internal and international migration and internal displacement. Expanding the research to different mobility types allowed us to increase the number of articles reviewed, as limited open documentation was found regarding models used in operational settings. All the publications examined are presented in *Appendix I*.

This catalogue also contains general information about the reviewed models, including their intended use, a description of mobility and covariates[iv] (external variables that can be used to explain human mobility), data usage, model development and model evaluation specifications, benchmarking weakness and strengths, operational readiness, and possible ethical considerations. These categories were established using as a reference the Peer Review Framework for Predictive Analytics in Humanitarian Response published by the OCHA Center for Humanitarian Data[1], and other relevant considerations for drought displacement.

This report aims to document the main data inputs used in the models reviewed because of the importance of reporting the current availability of drought mobility datasets and specifying how these datasets can be exploited. That also allows us to identify current data gaps that limit the application of the models and help improve the data collection methodologies of key drought-displacement indicators.

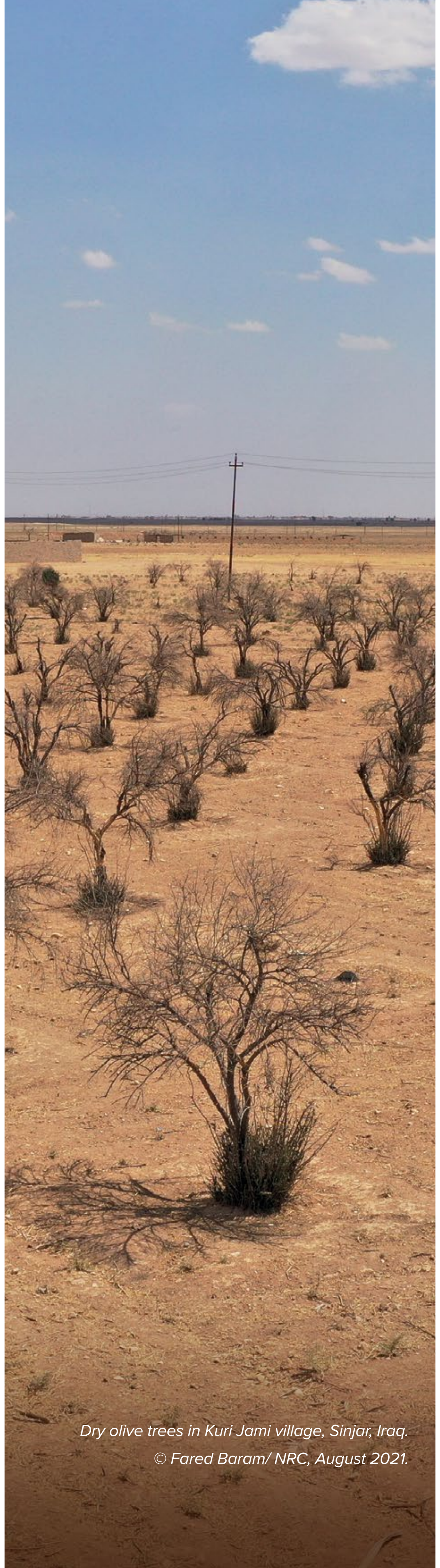
This report, however, should be read with the following disclaimer: The literature review findings are presented in a descriptive manner based on the authors' analysis. It does not represent an exhaustive description of all the existing models. We based our analysis on articles that focus on drought (or proxies of drought) as a trigger of human mobility. The results of the analysis and additional remarks on drought displacement from the modelling perspective are presented in the subsequent sections.

i Including national statistics agencies and NGO surveys, but also VGI (volunteer geographic information), crowd-data, social media in near real time, ARD (analysis ready data) and remote sensing products.

ii Example: Village level, individual level and regional projections.

iii Bangladesh, Burkina Faso, Colombia, Ecuador, Ethiopia, Ghana, India, Iran, Ivory Coast, Kenya, Mali, Mauritania, Mexico, Niger, Nigeria, Pakistan, Senegal, Somalia, South Africa, Tanzania, Thailand, Uganda and, Viet Nam.

iv Covariates are predictive or explanatory variables of a depended variable.



Dry olive trees in Kuri Jami village, Sinjar, Iraq.
© Fared Baram/ NRC, August 2021.

Drought and internal displacement from the modelling perspective

Models or computer simulations can be used as tools to improve the understanding and predictability of drought. They can also allow for investigation into factors and dynamics that can impact the severity or scale of droughts and associated forced displacement at different spaciotemporal scales and using different hypothetical scenarios (e.g., under different climate change or other what-if scenarios). For that, systematic data related to drought and its impacts is necessary, covering different temporal, environmental, economic, and societal dimensions[8]. Long-term capture of comprehensive data is, for that reason, essential to ensuring the accuracy of model predictions, as the model outcomes depend on the data inputs provided. Data on drought impacts is also crucial for the development of early warning systems, as well as drought prevention, reduction, mitigation and adaptation measures.

The research interest in the linkage of environmental change, drought and human mobility has increased significantly in recent decades. Accurately quantifying the effect of specific drivers on human mobility and drought displacement, however, is still a great challenge because of the complexity of the scientific problem and the shortage and quality of the existing data.

Drought-affected mobility is a multicausal and multidimensional phenomenon where no driver can be determined as the single reason for displacement. Its dynamics emerge instead from a combination of socioeconomic, political and environmental drivers that interact at different spatiotemporal scales in non-linear ways[9]. Drought crisis events also involve intrinsic complexities, such as the duration of the hazard, uncertain onset times, the influence of compound effects (e.g., land degradation over the years) and climate patterns (e.g., El Niño/Southern Oscillation). There is not a single definition of drought. Researchers have defined several types, such as agricultural, hydrological, meteorological, or socioeconomic drought. The literature also provides different indicators that describe drought conditions and indices that are used to describe drought severity through numerical representations[7,10,11].

The following sections, describe the reviewed models, organized by purpose and potential applications. In section 5 we present the different types of documented models, highlighting some of the strengths and weaknesses of the different methodological approaches. We also present an overview of the challenges and data gaps in modelling drought mobility.

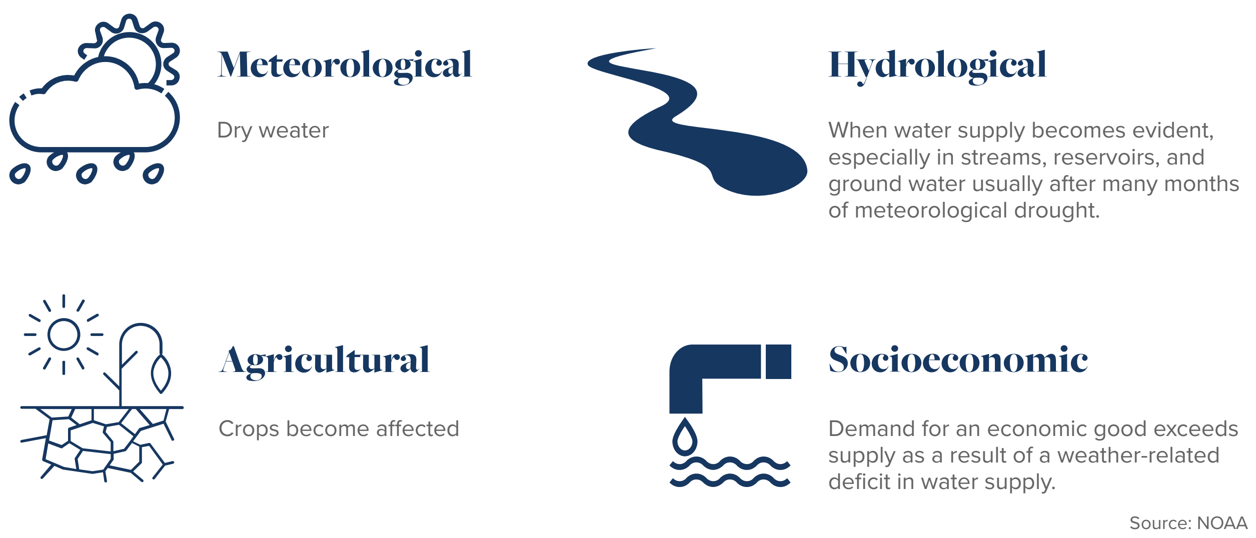


Figure 1: Different types of droughts described in the literature

Primary goals of drought mobility models

According to the results of the literature review, most of the modelling approaches have been developed as hypothesis testing tools intended to analyse drought-mobility drivers. The reviewed models aim to support different applications, such as policy design, the projection of human mobility estimates, the identification of drivers of displacement and the support of operational actors. They are also intended to inform future data collection efforts (See Figure 2).

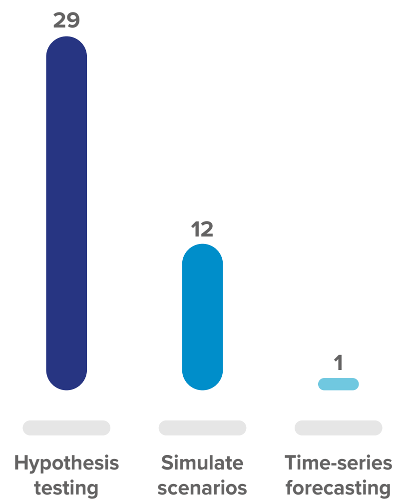


Figure 2: Primary goals of models reviewed

Twelve out of the 42 models reviewed were developed to simulate 'what-if' scenarios. The goal is exploring outcomes under different drought impacts or climate pathways and understanding the underlying dynamical processes that result in displacement. Only one of the reviewed modelling approaches aimed to provide time-series forecasting of drought-related displacement as its main objective.

Hypothesis testing

Most of the resources reviewed (29/42) aimed to test whether climate, environmental, socioeconomic, or other factors have an influence on human mobility, and to quantify this evidence to improve our understanding of such phenomena. For that purpose, statistical regression was the most common modelling method used. Regression analysis[v] mainly works by finding relationship patterns in data, with or without strong prior assumptions. Its goal is to analyse how effectively a set of explanatory variables can predict or explain human mobility outcomes. Once a goodness-of-fit measure (R-squared, Chi-squared, MSE, MAE) representing model error has been obtained, one can assess how reliably the model explains the targeted human mobility data with the accounted factors and assumptions. The lack of time-sensitive drought impact or displacement data limits its applicability, however.

Since the main goal of these approaches is to determine the impact of specific factors on human mobility, most reviewed models allow us to inspect their learned parameters calibrated to best explain the mobility data (model fit). A usual case is that these parameters are coefficients assigned to each variable and can be interpreted as "strength factors" measuring how much they contribute to the model's predictions. By measuring its coefficients, a regression model provides evidence of the influence of the considered factors on displacement (e.g., does rainfall variability increase or decrease migration?). The main algorithms used for hypothesis testing are logistic regression, econometric models (gravity, radiation and fixed linear regression), tree-based ensembles and K-means clustering (which is instead a classification algorithm). These will be discussed further in this section.

Scenario simulation

These approaches aim to explore the outcomes of plausible responses to drought under hypothetical scenarios. After the behaviour of the target system is correctly replicated, the model is then used to test how the evolution from initial conditions, changing relationships or future events would affect human mobility. This can be achieved by running a computer simulation or by extrapolating the results from a regression model. The main task in computer simulations is

v Regression analysis is widely used for prediction, forecasting and quantifying the connection between a target variable and its predictors. For example, it allows the quantification of the connection between birth rate and poverty levels (the typical X vs. Y variable relationship in its simplest version) and eventually predict poverty levels by using birth rate. A regression analysis provides quantifiable metrics representing how well the model fits the data (goodness-of-fit), like the R-squared statistics (R-squared for example measures "how much" of the variability of the mobility data the regression model was able to explain on a convenient 0 - 100% scale).

to conceptualize the system with causal rules and quantify the interactions between its components. Complex system models such as Agent-based models, Dynamic System modelling and Bayesian networks are the main representatives, allowing one to model the complex interactions of human mobility. Other approaches aim instead to project aggregated mobility flows relying on the extrapolations of the model under different development or global CO2 emission pathways. Such is the case with approaches like Groundswell's gravity model[12]. Both approaches can be used for policy design by applying what-if scenarios and verifying conceptual frameworks or theories through the extraction of qualitative knowledge and even predictions. This can be done if the models are properly validated and calibrated with real data.

Time-series forecasting

Time-series analysis models are used to make predictions based on historical time-stamped data. They depend on the amount of available historical data, possibly more so than other types of modelling approaches. One of the assumptions of time-series forecasting is that some aspects of past patterns will persist into the future. A wide range of methods have been developed for the forecasting of time series, with the three most common approaches being statistical methods (ex. ARIMA), machine learning (e.g., Bayesian Neural Network) and the use of hybrid methods.

The only reviewed model that aimed for explicit forecasting of displacement flows is the UNHCR and UN Global Pulse's Project Jetson[13]. Project Jetson represents one of the first explorations of time series forecasting by the humanitarian sector. It focuses on explicit time-series forecasting to understand the nexus between climate, displacement and violent conflict. The project provides open access to technical documentation, the source code of the model, and a visualization interphase.

Project Jetson was launched by the UNHCR Innovation Service as a machine-learning-based project aimed at forecasting the number of internally displaced people within Somalia and the number of refugees along the Somali-Ethiopian border. It was also implemented to discover, understand and measure the factors that cause and exacerbate ongoing displacement in Somalia and people's decision to flee. Humanitarian organizations can potentially use it to improve response efforts and relief services for displaced populations within Somalia.

In comparison with other modelling approaches, Project Jetson projections aim to forecast displacements up to three months in advance. This analysis is complemented with a geospatial model using a gravity model approach with a horizon longer than three months[13]. Jetson's predictions are based on available data and are subject to its inherent biases and uncertainties. Project Jetson was never operationalized (e.g., used to inform decision making in UNHCR field operations), and predictions were retroactively compared with actual data collected in Somalia and Ethiopia. This ensured that there was still accountability for displaced populations. Technical tools were not implemented without extensive algorithmic and human rights due diligence around their limitations and impact. A limitation of this approach is that current time-series forecasting models are not robust to changing causal relationships in a system (e.g., the effect of an external unconsidered factor) and to emerging patterns of mobility that are not already present in the data upon which the model was trained. This type of model, however, could be retrained on new data and corrected for previous failures.

Potential applications of the models reviewed

Most of the model's reviewed were designed to serve one or several of five main purposes:

- Exploring the potential impact of policies or interventions aiming to prepare for or reduce migration[14 15]. These models are grouped under the category of policy design.
- Exploring future scenarios of mobility associated with preparedness or climate change scenarios[16, 12]. These models are grouped under the category projection intentions.
- To identify drivers of migration[17].
- To support operational actors by helping to understand the factors that cause and exacerbate ongoing displacement[13].
- And to inform future data collection efforts[18].

In the sections below we introduce some potential usages and limitations regarding how the models can be used in forecasting applications and in the projection of future scenarios of displacement or migration. Finally, we will discuss some applications for policy formulations.

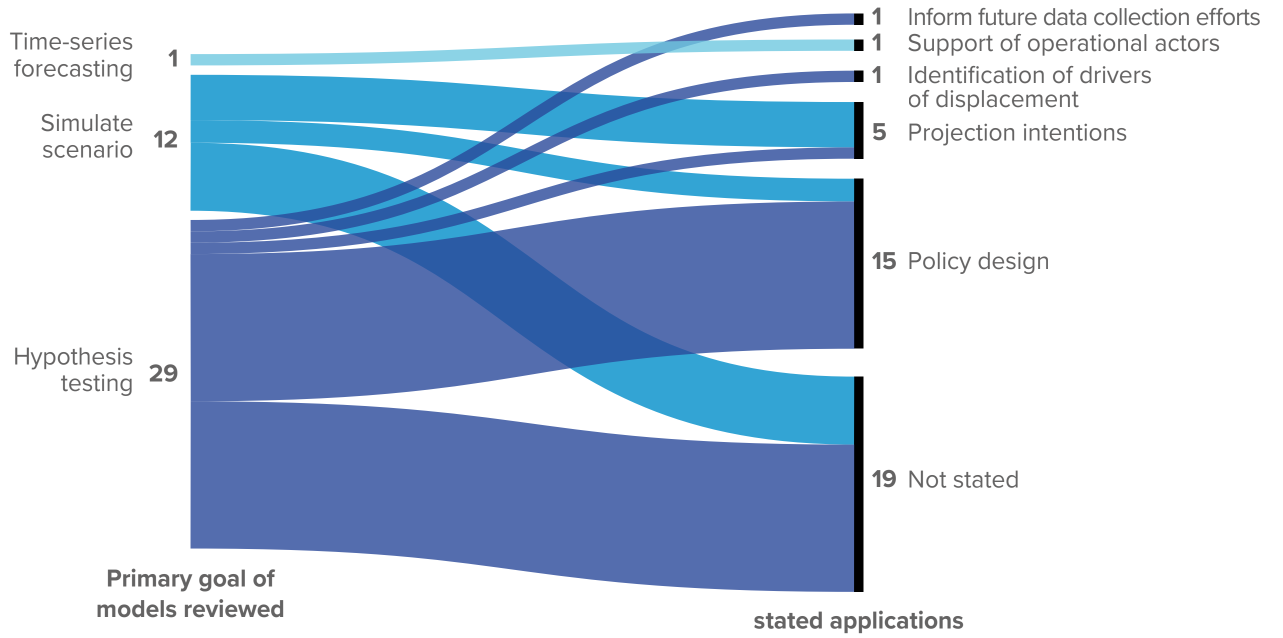


Figure 3: Goals of reviewed drought mobility models and stated applications

Support of operational actors with displacement forecasting

Time series forecasting initiatives provide one with the ability to predict future displacement magnitudes and trends on short-time frames. This is very useful for supporting operational actors in the field[19]. Accurate forecasts allow for strategic and effective short- and middle- term planning and crisis prevention when it comes to droughts. Given the multiple uncertainties involved in data acquisition methods and modelling approaches and the fact that future scenarios are fundamentally uncertain, quantitative displacement projections (long-term) and forecasting (short-term) present tremendous challenges. Project JETSON, for example, in aiming to forecast drought displacement in Somalia with machine-learning, always categorises its efforts as "experiments," not as readily operational products[13].

The success of an effective forecast depends on the existence of clean and well time-stamped data in which the historical trends and patterns can be identified and on data biases present in data collection and data preprocessing. One can always expect errors and uncertainties in forecasting human mobility: the future is fundamentally uncertain and precise predictions on how the socio-ecological interactions will evolve are very difficult to obtain[20]. As a result, point precision predictions of displaced populations in time are currently unachievable. Intervals of plausible magnitudes could be obtained instead. Since the field is young and still maturing, the feasibility of a reliable and operational forecasting product for drought displacement is also yet to be explored by modellers and practitioners.

Following this line, a practical implementation must deal with how to treat uncertainty in forecasting. This is related to the acceptable margin of error for the humanitarian situation if the forecasting fails. For that reason, being explicit and transparent about uncertainty is crucial in any application. Two critical situations must be addressed: Who would be affected if there are false positives (the forecasting falsely detects displacement) and who would be affected if there are false negatives (the forecasting did not detect displacement)? The humanitarian and scientific community should thoroughly discuss these issues if an operational product is implemented.

Projection intentions

Projections are long-term predictions that describe hypothetical conditions in the future (estimated conditions) based on assumed or expected circumstances. Forecasts, by contrast, are short-term, time-based predictions that rely on time series data. Given the climate emergency, accurate projections could be extremely useful to the humanitarian sector in helping to prepare for the conditions future droughts will impose on vulnerable populations around the globe.

Projection intentions can be used to explore how some environmental, climatic and/or conflict drivers could affect internal displacement, migration or the mobility of pastoralists. Projections are a common type of application in modelling, but we recommend caution in the use and communication of long-term projections' modelling outcomes. Macro-level research on the possible crisis triggered by climate change is useful for the preparedness of humanitarian organizations and governments. We must, however, bear in mind that future scenarios are fundamentally uncertain, as many unexpected circumstances that are not accounted for in models can unfold. After all, statistical models learn from past data, and initial conditions might dramatically change outside the conditions in which the model was trained. Recent studies have also emerged that point out how some of the current methodologies used for projecting international migration may not be apt for describing temporal patterns of migration[21]. Using projections resulting from these models should be done with caution.

Some of the extrapolations from the reviewed models were conducted without taking into consideration out-of-sample data (i.e., data that has not been used to train the model). This is a crucial step in model validation, as a subset of the data should be kept from the model training to later test the reliability of the model's predictions. This step gives information about the model's ability to generalize or describe reality. Statistically significant correlations derived from in-sample data often constitute poor predictors for new data[22].

One must also take into account that the model assumptions used in the modelling process are also extrapolated into the projections' results. For example, if the model assumes a linear relationship between the lack of rainfall and forced displacement (as most reviewed models do), future rainfall deviations scenarios will be tied to the observed displacement magnitudes in a direct proportionality relation. This could sound like a reasonable hypothesis. It does not, however, account for possible adaptation strategies or unconsidered interacting factors. This has been extensively criticized by application domain-experts as it usually leads to projections of alarming levels of mass migration without rigorous evidence to support it[23].

Despite the desirable applications of displacement magnitude forecasting and displacement projections, predictive analytics could also be employed for other operational approaches in the field. Such is the case of financial-based forecasting aimed at predicting the resources needed to resolve a current situation. It is also the case for other non-forecasting approaches, such as early warning systems, hotspot and vulnerability assessments, and exposure mapping of populations at risk of internal displacement associated with drought.

Stated policy applications

Transitioning to knowledge-transfer and decision-support activities, several articles seek to understand and simulate mobility dynamics in order to develop future scenarios or to point out targets for intervention in the interest of policy preparedness. Most of the reviewed models aim to use simulation modelling approaches to offer policy design tools for interventions under different scenarios or to understand the intricate livelihood dynamics of pastoralists[24]. For example, a Bayesian network model in Ethiopia was used as a discussion, communication and learning tool to investigate stakeholders' perception about the effect of precipitation and soil degradation on human-mobility[25]. Hypothesis testing approaches, on the other hand, measure the effects of different variables and are set to inform targeted intervention based on specific factors or by outlining vulnerable areas susceptible to drought-related risks. For more information on stated policy applications, we refer the reader to section *5.4 Intended Policy Advising Objective*, from the model catalogue in *Appendix I*.

Existing drought mobility models: strengths and weaknesses

Most reviewed resources are scientific papers (36/42), and only one resource had the model source code hosted on an open access code hosting platform. Finding literature or repositories of drought displacement models is challenging, particularly in operational settings such as the humanitarian sector, because their technical information is often not accessible. This is a major caveat in knowledge transfer and a big barrier to the development of actionable science in the domain.

The following sections provide an overview of the different types of models used in the literature, as well as the strengths and weaknesses specific to particular methods. Note that the reviewed methodologies do not constitute an exhaustive description of all the existing models, but rather a representation of current approaches. The reviewed models have different methodologies, cover different regions of interest, and have different objectives, limiting a possible comparison of performance or strengths across modelling approaches. Depending on the data availability and the research questions of a particular experiment or project, different types of models or methodologies could be applied to understand or predict the dynamics that trigger displacement or human mobility, or that could support the exploration of different scenarios of displacement associated with drought. We must bear in mind that the usefulness of any data-driven methodology depends ultimately on the specific problem it tries to solve and the

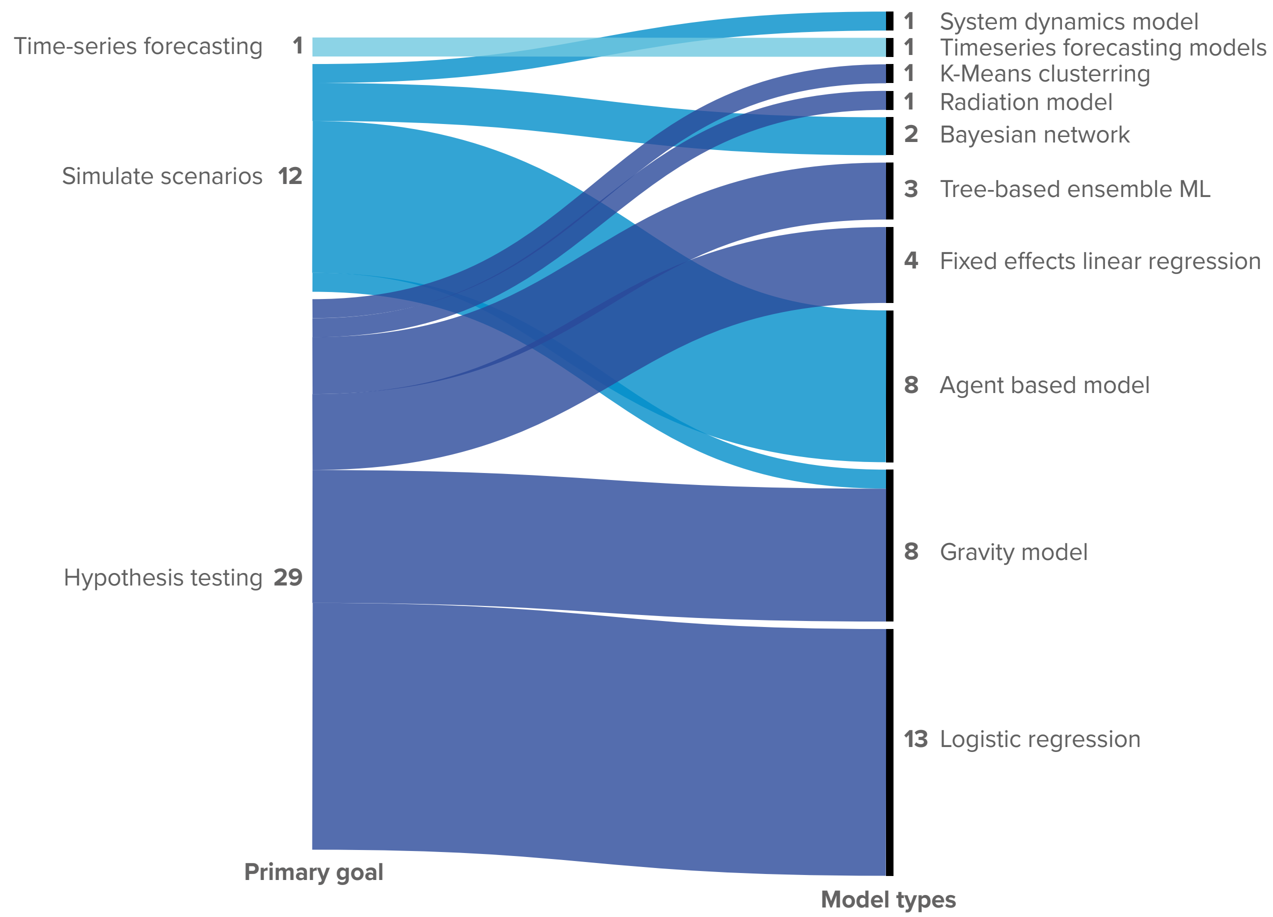


Figure 4: Illustration of the frequency of models reviewed by goal and type of model

nature of the data. The main modelling methodologies, with the associated strengths and weaknesses for each model, are summarized in *Annex I*.

Ten different methodological approaches used in models were documented during the literature review (See Figure 4).

Logistic regression analysis

As has been discussed, a great proportion of drought displacement literature aims to discover and quantify the effects of climate or other factors in human mobility (i.e., hypothesis testing). A widely used approach in traditional statistics is to use logistic regression[26]. It has a long tradition in statistical analysis because of its simplicity and interpretability. Logistic models differ from linear regression ones in that the former target a binary outcome (migration or non-migration) and can account for categorical variables assessment. This technique assumes linear relationships so it will be effective if data relationships are linearly separable. In the literature, it is used on survey or census data for records of both migrants and non-migrants.

To achieve a successful analysis and extract valid conclusions, two key issues in model development must be considered: confounding effects and multicollinearity. Confounding effects are crucial to quantifying the actual relationship of a specific climate or environmental variable on observed mobility. These variables could invalidate regression results if left out by the model. The implication is that they would not be able to “isolate” the desired effect one wants to measure. For example, one may erroneously conclude that rainfall variability is driving migration. This finding can change, however, when household income is introduced into the analysis. In that case, it can be concluded that rainfall variability only constrains migration for low-income households (see *Alessandrini Alfredo et al., 2021)*[27]. That can have dramatic implications for decision-making and preparedness strategies since the latter result means that efforts should target areas of poverty.

At the same time, the modeller should account for multicollinearity effects. This implies that including redundant or intercorrelated covariates into the model could bias the coefficients or “strength factors” results in assessing the influence of mobility drivers[28]. For example, if one includes GDP and birth rate into the same model (which we know are highly correlated and include similar latent information) the contribution of a driver accounting for poverty in the model may be “split” between both factors and measure a lesser effect than if we only accounted for GDP or birth rate alone. Here we propose including multicollinearity tests in any analysis or using methods in which controlling for this problem is embedded into the algorithm.

Advantages of logistic models are:

- They can jointly measure the association of variables, to account for confounding effects when variables are included in the analysis[29].
- The results or the model are interpretable: coefficients in the model can be seen as “strength factors” of covariates on human mobility.
- They allow for binary targets, meaning that survey and census data can be easily exploited
- Computationally efficient.

However, logistic models:

- Do not work well with data that is not linearly separable, which means that it cannot account for the more complex non-linear relationships present in displacement.
- Small sample sizes can invalidate the coefficient results if too many variables must be included in the study (also known as the curse of dimensionality)
- If covariates are intercorrelated and this is not controlled, then the regression coefficients can be affected and invalidate the results of the study.

Finally, we highly recommended including quality metrics to highlight the model’s limitations and performance. Best practice in validation requires including an overall evaluation of the logistic model by means of statistical tests of individual predictors, goodness-of-fit statistics and an assessment of the predicted probabilities[29]. For these reasons, the success of a logistic regression analysis heavily depends on the modeller’s expertise regarding the specific problem and data availability issues.

Econometric models

Econometric models are the common methodology used to model aggregated migration flows across regions based on regression techniques. In the reviewed literature, econometric models are used to predict and understand the propensity or restriction of aggregate migration flows between different areas and possible drivers. Drawn from economic theories of migration, these models conceptually assume some type of utility maximization criterion: an individual decision to migrate is rationally based on a greater livelihood in other area and agents maximize their utility across the full set of destinations. They are mainly based on the idea of “push” and “pull” factors on migration. Push factors are defined as the drivers that affect people to move, while pull factors refer to the drivers that attract migrants to a specific region. Examples of push factors in literature are drought impacts, rainfall variability and unemployment. Pull factors are economic opportunities at the destination site, language similarity, and the urbanization rate, among others. This is based on the empirical observation that drought conditions push people from their origin and attract people towards nearby areas with greater economic opportunities.

Gravity models

Gravity models are the most common econometric approach found in the literature. They are also quite common in economics and social sciences given their proven efficacy in modelling spatially related flows. They are named after Newton’s law of gravitation because they rely on geographic distance, population, and other factors (e.g., wage differentials) as the attractors of migration flows. One can use gravity models to analyse the spatial patterns of mobility between regions (i.e., which are the most common destinations for regions). They can also be used, by looking at the size of the variables’ adjusted coefficients, to measure how distinct factors affect human mobility.

Gravity models have proven successful when explaining spatial patterns of flows at an aggregated level, between regions or countries, under simple assumptions (*Poot et al., 2016*)[30]. **Methodologies followed by current gravity models, however, have been found to describe spatial, but not temporal, patterns of human mobility**[21]. For that reason, using them to project human mobility flows should be done with caution since that lies outside model capabilities. Projections using gravity models could be highly unreliable.

Examples of the usages of gravity models in literature can be found in the Groundswell Gravity model that aims to project migration flows at grid cell levels in several countries worldwide[12]. They are also used to detect spatial patterns of interprovincial migration flows in Iran in the presence of annual rises in temperature and drops in precipitation by *Shiva and Molana, 2018*[31].

The main strengths of gravity models are:

- They are highly interpretable, meaning that applications intended to analyse spatial patterns of displacement across regions and probe the influence of the specific accounted drivers may find gravity models useful.
- Gravity models allow one to engineer a combination of fixed effect factors into the equations[32]. These fixed effects are introduced by constant parameters aiming to capture unobservable country-level or region-level factors without explicitly defining them (e.g., state migration policy or any other structural factors) to reduce the model’s error. This allows the model to omit data which may not be readily available and avoid possible confounding biases.
- Gravity models are highly flexible models which can be further developed to include stochastic processes and non-linearities and surpass current limitations.

Its main weaknesses are:

- Current gravity models may not be well suited to describing temporal patterns of human mobility[21]. This a crucial caveat which is overlooked in the literature.
- Structural factors captured by the fixed effect parameter are mainly unclear since one does not know what this quantity actually represents.
- Fixed effects in current gravity models do not allow for temporal changes, which is a strong modelling assumption since structural conditions can also change over time
- The utility maximization conceptual assumption may not be realistic in many contexts (e.g., non-economic reasons)
- Gravity models need to correct for multilateral resistance to migration, meaning that they do not consider alternatives of destination when migration decisions are in place, just the one that maximized the utility. This has a direct implication in applications targeting policy design[33].

- Gravity models are not able to capture complex patterns and are silent with regard to micro-level relationships and behaviours that give rise to mobility dynamics[12].

- Aggregation bias may be in place (including in the study of people who have not been subjected to the same environmental or socioeconomic conditions) if the regions of the study are large.

- The models' complexity in reviewed gravity models is linear, meaning it will not adjust for non-linear relationships between the accounted variables. Advances in the field that could modify the model to account for non-linearities still should not be discarded.

Radiation models

Radiation models are rooted in radiation and absorption processes in physics and based on the concept of "intervening opportunities". This means that the likelihood of migrating is influenced more by opportunities to settle in a destination and less by distance or population (i.e. people will migrate to the region that most closely fulfils their expectations)[34]. Radiation models have recently appeared as an approach and could perform better than gravity models in some situations[35]. Their field of application and purpose matches gravity models: they analyse spatial patterns of human mobility across regions, and for that reason share most of the gravity models' strengths and weaknesses. The recent criticisms of gravity models' ability to predict temporal patterns of migration also apply. to radiation models, which may not be well suited for describing temporal patterns. Only one of the reviewed articles (*Sibren* Isaacman et al., 2018) modelled migration patterns in Colombia employing a radiation model[34]. The study provided evidence of the influence of spatial patterns of rainfall variability on internal migration, using geolocated data registered by cell-phones.

Radiation models, however, represent a counterpart to gravity models in several aspects and overcome some of their limitations. Their strengths compared with gravity models include:

- Radiation models can be modelled without parameters. For that reason, they are less data hungry and can describe spatial patterns when specific data is not available for some regions.

- Radiation models make weaker assumptions about data distribution and thus may capture non-linear relationships. On the other hand, one key weakness compared with gravity models is that they are not good at estimating the effect of several factors on migration. Extensions to advance these limitations, however, have been proposed[36].

Fixed effects linear regression models

This modelling approach can be interpreted as a gravity model but includes only push factors. They are mainly used when bilateral flows of data are unavailable or when the case study is interested only in out-migration. Fixed effects coefficients are characterised following the same procedure as gravity models and their strengths and limitations are thus shared (see 5.2.1 Gravity models section). This is the case of a study conducted by the Joint Research Centre (European Commission) in 2021 in which model results returned a significant association between net migration and drought intensity, especially in rural areas[37].

Complex systems modelling: agent-based, system dynamics models, bayesian networks

Mobility responses and adaptions to environmental hazards are the result of complex interactions between drivers. In fact the most intricate and unknown processes of mobility dynamics match complex systems' behaviour: feedback effects, non-linear dynamics, tipping points, adaptive response to events, emergent behaviour, micro and macro-level interactions, etc.[38]. In this section we discuss the specific approaches found in the literature: Agent-based models, System dynamics and Bayesian networks.

Agent-based models (ABM)

Agent-based models were recently introduced in the field to simulate mobility dynamics emerging from the behaviour of many interacting agents (e.g. individuals or households) by explicitly programming the rules by which the agents interact[39]. They are useful when trying to model how individual behaviour gives rise to population-level patterns or possible adaptive strategies from socio-ecological feedbacks. In this sense, a successful ABM replicates the correct behavioural response of a system, such as the adaptation to drought of agricultural households or pastoralist communities.

Examples where this approach has been applied include a case by *Sakamoto, 2016* where a combination of satellite imagery and an agent-based model was used to examine how pastoralists access resources and adapt to unpredictable ecological changes in dryland regions of Nigeria[40]. Another example is a study in Northeast Thailand which models land use, social networks and household dynamics to inspect how in-migration and out-migration flows arise[41]. Still another is a modelling approach employed by *Nelson et al., 2020* in Somalia that tries to explore how conflict and environment interact and affect pastoralists' routes[16].

ABMs' strengths, include:

- Since ABMs simulate mobility using rules which generate artificial data and patterns the modeller wants to simulate, they can tackle problems which could demand unrealistic amounts of data concerning individual interactions and which could not be studied otherwise[40]. This includes having the potential to emulate the complex linkages between slow environmental changes, social network interactions and mobility outcomes of real-world scenarios.

- ABMs allow for interventional analysis to observe how changing a variable can affect mobility dynamics in different situations and in specific moments in time.

- ABMs allow for interactions, non-linear effects, feedback loops, autoregressive effects, and emergent and adaptive behaviour typical of mobility dynamics.

- ABMs are stochastic by nature. They can deal with associated uncertainties of mobility dynamics and return different scenarios for analysis.

Several weaknesses are present, however:

- Validation of ABMs requires real world data comparable with the artificially generated data which might not be available. For that reason, many of the reviewed models were not robustly validated. Models are limited to qualitative extrapolation at best, leaving quantitative projections outside of the models' capabilities.

- Another caveat is that ABMs require an extensive amount of parameter calibration and defining of the interactions driving the system. This is a critical part in model development: transitions within the system are determined by the causal mechanisms of socio-ecological networks that could be extracted from data or expert knowledge, but which are often assumed[39].

- ABMs also demand high computational power to return simulation results that are not appropriate for real-time policy discussion.

- Regarding interpretability, simulation outcomes must be interpreted by designing visualization methods that could bias model results in some cases.

- The conceptual design of the model that matches the real dynamics of human mobility may be challenging to implement. This is because the capturing of the fundamental rules that give rise to the mobility dynamics may require a robust and precise social theory.

System dynamics modelling (SD)

System dynamics models aim to understand complex phenomena using causal models that represent a system and its behaviour over time. SD models represent aggregated systems in the form of stocks, flows, feedback loops and time delays. In contrast to ABMs, instead of letting individual behaviour generate the dynamics of the system, SD modelling follows a top-down approach, in which a set of equations define how the network and interactions will evolve given an initial state (e.g., rainfall and pasture influence the livestock population, which will determine milk and livestock prices with an impact on the income of pastoralists who are finally displaced in extreme conditions)[42]. SD models must conceptualize and fit a framework able to represent the interrelationships between the components of the system. For this reason, an SD model can become very complex depending on the conceptualization of the system in situations where several variables need to be included[24]. Thus, SD modelling could provide a good support for testing the understanding of the causal interactions in drought displacement, its drivers, their interrelationships, and the data gaps needed to fit the model.

These are among the general strengths of an SD model:

- SDs are useful when trying to study how macro-level conditions (made up by designed rules) give rise to micro-level interactions.

- Forecasting is possible if the model is properly validated, but it is not the main objective of the use of SD as a methodology. Rather, these models could be suitable candidates for long-term strategic modelling and simulation (i.e., policy design tools), as they may show how intervention in the specific components (leverage intervention) of a system can change the overall behaviour towards a desired outcome.

- SD can model complex dynamics such as non-linear effects, feedback loops, autoregressive effects, and the emergent and adaptive behaviour typical of mobility dynamics.

- SD models can be computationally very efficient in comparison with ABMs (a matter of seconds vs. simulations potentially requiring days, depending on the complexity of the model). They also allow one to explore different scenario-based approaches.

There are also important limitations, however:

- Defining the system's components and interactions can be a time-consuming task and subject to the bias of the modeller, particularly when there is no clear understanding of the system (e.g., how does conflict affect water availability?) or when the underlying social theory is not precise.

- SDs equations must be designed to fit the real system's complex behaviour. The inherent complexity of mobility dynamics means that it is difficult to capture every relationship, making their mathematical representation complicated. This also requires a precise underlying social theory and expert knowledge of the context.

- SD models require an extensive amount of parameter adjustment and calibration: the causal links, the quantification of the relationships between the variables and the time-lags in which the different variables affect one another must be calibrated. Depending on the complexity of a given SD model, quantitative and qualitative data could be used to populate the models. The data, however, may not be readily available. This could become a major caveat for the model's calibration or validation and compromise the reliability of this type of model.

IDMC's system dynamics model was first developed in 2014 to explore the risk of drought displacement for Kenyan, Ethiopian and Somali pastoralists[24]. Another simplified version of the model was developed in 2021 by IDMC in collaboration with the Danish Refugee Council, focusing on pastoralists' livelihoods in Somalia and Ethiopia. Both versions of the model permitted an exploration of drought scenarios, the possible impacts on pastoralist livelihoods and the resulting internal displacement. It was also an effort to reveal the data gaps crucial for modelling the displacement of pastoralists. This approach uses an expert-based deterministic approach to define the different causal interactions that drive the system rather than a full characterization of its parameters by data-driven methodologies, such as regression analysis, because of data-availability limitations. For this reason, system dynamics simulations focus more on showing the broad outlines of possible scenarios, rather than offering specific and accurate predictions.

Bayesian networks (BN)

A Bayesian Network is a probabilistic tool which delivers a graph model made up by edges, arrows and probabilities, in which one can visualize the interactions between the variables of a system. For that reason, these models can be used for both qualitative and quantitative purposes[43]. The network is fully characterized by two parts: its structure and its parameters. The structure represents the causal connections between the variables of the network, while the parameters represent the conditional probabilities between the variables (i.e., a BN uses the fact that "X often causes Y"). Both parts need to be inferred from expert knowledge or estimated from data. Finally, the ultimate objective of a BN is to estimate the joint probability distribution (all events in the system happening together) of the system. This allows for an answer to questions such as "what is the probability drought displacement occurs (X) given below average rainfall (Y) when there are conflicts in the area (Z)?". For this reason, BNs are suitable to model and explore complex and multi-causal systems of many variables (e.g., drought displacement is caused by crop failure, rainfall, conflict, etc. combined), so applications directed to drought displacement modelling can be explored.

There are several strengths of modelling using a Bayesian approach.

- A key strength for its use in the field is that it can It can explicitly handle uncertainty and can work where data is missing or inconsistent[44]. This approach could allow modellers to quantify which aspects of human mobility are most unknown and uncertain. Secondly, and as stated above

- BNs construction is flexible in the sense that the causal graph can be constructed from expert knowledge or alternatively, derived from data, and then used to estimate the probabilities of the causal events. Estimating the structure of the network from observational data is difficult, but it can help discover unexpected relations between the variables.

- Once the network structure has been learned, a BN can be used not only for observational inference but also for interventional reasoning, meaning that we can ask the models questions like "can we stop X from happening if we decrease Y?"[45].

In contrast to ABMs and SD models, however:

- BNs are acyclic networks in the sense that they cannot include closed loops or feedback effects between variables and relationships are unidirectional.

- The estimation of many interactions between the variables can be computationally very costly, and the reliable inference of causal relationships from data is usually extremely challenging. For this reason, prior expert knowledge is often required.

- The quality of prior knowledge is crucial for the statistical model of the data: a BN just like an ABM or an SD, is only as useful as the prior knowledge is reliable.

Two articles in the literature used BN to assess drought influence on human mobility. *Groth et al., 2021* built a policy design tool to explore stakeholder's perception on the different processes driving displacement in rural Ethiopia[25]. Meanwhile the approach followed by *Drees and Liehr, 2015* in the Sahel consisted of building a BN model to understand the impact of climatic and environmental changes and socioeconomic drivers on human migration[46].

Subsection 5.3.4 compiles other types of models that have not been widely used.

Other models

They scarcely appear in the literature review, but the following model types should be considered because of their potential. Some machine-learning approaches, such as explainable AI (XAI) and causal discovery algorithms, are underrepresented in this sample although they deliver the most innovative solutions at present.

Tree-based ensembles

Tree-based ensembles are machine-learning algorithms that combine multiple decision trees to control for overfitting the data (i.e. the inability of a model to generalize). They have demonstrated remarkable performance among very different datasets and applications in recent decades and can account for non-linear effects and interactions between variables[47]. The main advantage of these algorithms is that they find relationships in the dataset while making no strong assumptions on the problem. In this way, they reduce the likelihood of bias introduced by the modeller and other conceptual assumptions such as linear relationships. One of their benefits is that they usually showcase superior performances in terms of sheer predictive power. Tree-based ensembles, however, suffer from interpretability issues and from low-sample datasets. They are very versatile algorithms which are also used in the literature to rank the most important factors influencing mobility, constituting a generalizable technique for targeted intervention[18, 48, 49]. Studies attempting to predict or discover relationships in very complicated datasets (even survey data) could greatly benefit from tree-based ensembles approaches and other non-parametric machine-learning approaches.

Exposure modelling using K-means clustering

This approach is proposed in the literature by Neumann et al, 2014 which tries to describe global patterns of environmental drivers of out-migration in drylands by using cluster analysis[17]. K-means clustering is what is called an unsupervised machine-learning method (i.e. algorithms that learn from unlabelled datasets). It aims to classify data into different groups based on how similar the patterns in the groups are. This allowed the authors to suggest that land degradation at a global scale is the most severe environmental constraint for out-migration when it comes to drought impacts and recommend monitoring these areas as potential hotspots of risk.

Main challenges and data gaps for modelling internal displacement triggered by drought

Because of the nature of the topic, and its scientific complexity, many challenges need to be faced in the collection of baseline data for models, modelling practices and the diversity of the cases in which the models are used. This subsection seeks to answer the question: what are the main challenges and data gaps that need to be addressed when modelling drought displacement?

Lack of monitoring of drought impacts

Drought-affected mobility is a multicausal and multidimensional phenomena where no driver can be determined as the single reason for displacement. Instead, its dynamics emerge from a combination of socioeconomic, political, and environmental drivers that interact at different spatio-temporal scales in non-linear ways.

Multisectoral information of pre-drought conditions is highly relevant to understanding the impacts of drought. Environmental data needed to monitor drought is ever more accessible (e.g., rainfall, Normalized Difference Vegetation Index (NDVI), or crop production), but disaggregated data on its economic, demographic and social impacts is not collected systematically or is not accessible.

Researchers should direct their efforts towards understanding the concrete pathways by which environmental, societal and climate stressors affect a specific region of interest. As the UN Office for Disaster Risk Reduction (UNDRR) states in its latest report, the damage and costs caused by drought are usually underestimated because of extended and cascading impacts, and often not attributed to the drought as most impacts are indirect[2]. Drought, for example, may trigger food and water security crises in regions where there is poor water and land management.

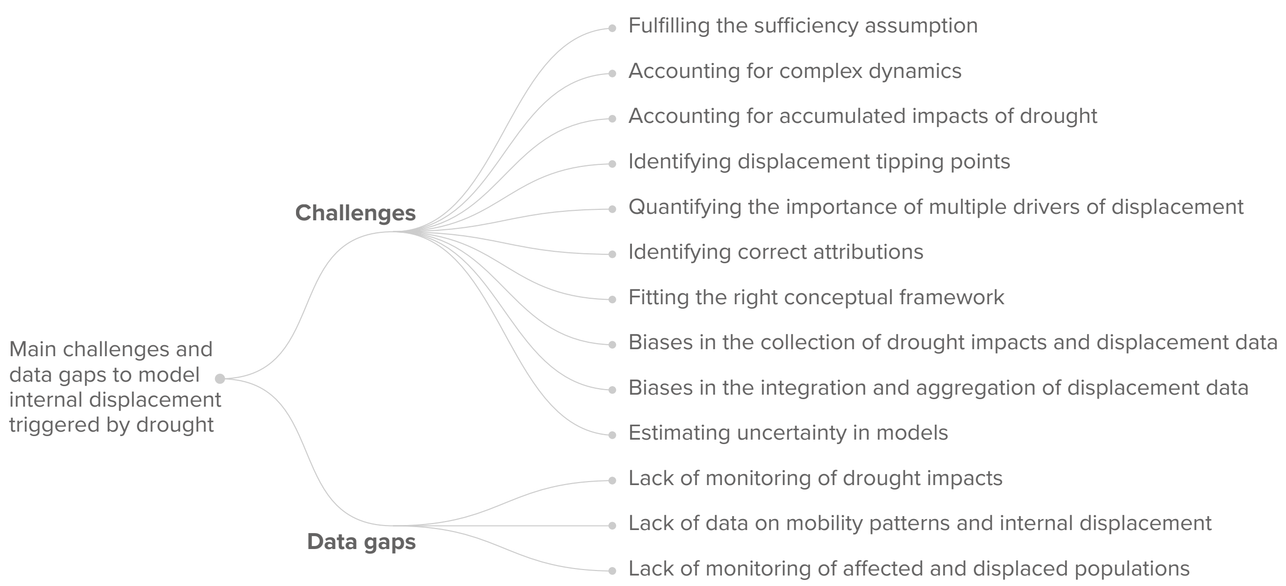


Figure 5: Main challenges and data gaps for modelling internal displacement triggered by drought

Conflict over scarce resources may then arise, generating demographic pressures and exacerbating drought's widespread impacts[50]. These processes are intertwined, affecting the social and ecological dimensions that may trigger displacement.

These factors are inherently challenging to quantify, and even more so when the **micro-level data of drought's impacts, such as household's capital losses (crop yields, livestock deaths or any other related loss), is not available.** The review of the literature also reveals how **different livelihoods require different approaches when characterizing drought impacts. Exclusively monitoring cropland regions, for example, may be useful in areas and communities dependent on agriculture but is not suitable for accounting for the degradation of pasture lands that affect pastoralist communities**.

At the same time, several studies point out how weather stations are sparse in rural areas of low-income countries and, for that reason, methodologies, such as gridded data interpolation or data assimilation methods producing climate data, must be chosen. When it comes to the meteorological characterization of droughts, a wide range of drought indicators have been used in the literature at different time scales. No unified approach has been established.

This calls for asking questions like: What is the right index for describing drought intensity? Which climate data products should be used? What are the environmental, meteorological, hydrological, and socioeconomic factors affecting drought and generated by drought? What is the drought onset and duration? All these questions, and perhaps others, must be addressed at a context-specific level to effectively assess the impact on human mobility.

Fulfilling the sufficiency assumption

Drought displacement is a multicausal and multidimensional problem in which multiple factors need to be accounted for. In any modelling exercise this is usually known as the "sufficiency assumption". This means that one assumes that all the relevant context-specific variables are included in the study. In the case of this report, that means that all potential relationships and interactions that trigger displacement associated with drought events are accounted for. Without fulfilling this assumption, model accuracy or conclusions derived from model results could be compromised depending on the relevance of the omitted variable in question.

Multiple factors are involved in drought displacement and mediating structural factors are key. The most immediate and extreme impacts of drought on displacement, however, are usually triggered by its effects on food and water security crises and their consequences[51]. The lack of contextual information on influences and mechanisms in many regions of the world is the main limitation in this regard. For more information on this point, the reader should revisit section 8, which presents covariates used to model drought mobility and provides examples of potentially key data.

That said, in modelling there is never a way to be totally certain that this sufficiency assumption is correct and that all relevant factors have been included in the model. Model performance and error metrics, however, allow us to measure how much of the displacement data the model was able to explain under its assumptions. This highlights the importance of implementing a model evaluation process as part of the modelling design one. Innovative approaches are under development to infer if there are unknown and unseen variables driving a system (latent factors) in the emerging causal discovery field[52]. Ultimately, however, quantitative and qualitative research both reveal and provide evidence on the specific factors driving displacement.

Lack of data on mobility patterns and internal displacement

Not all types of human mobility arise from the same causes or drivers. **Understanding how drought affects different patterns of human mobility is a research gap that needs to be filled.** Some of the key data necessary to solve the matter may be more readily available than others. For example, reliable data in rural areas is scarce, most literature employs non-displacement data, and the conditions of returning flows are, to a great extent, unexplored. Some articles also address migration in general terms only, without focusing on the different types of mobility involved. We recommend dedicating more attention to these points as they are relevant for understanding drought adaptation responses, for data acquisition and model implementation.

Some forms of human mobility may be part of an adaptation response to drought and can also be interpreted as a proxy for deeper crises resulting in forced displacement. For example, seasonal rural-urban migration, usually related to labour migration, is a common adaptation strategy when agricultural failure occurs[53, 54]. Understanding these pathways is closely related to understanding drought displacement.

According to data available on forced displacement (e.g., refugees, IDPs, asylum seekers, etc.), internally displaced people account for most of the forced displacement reported. There is scarce reporting and monitoring, however, of the number of people internally displaced by drought conditions. Of 13 countries where IDMC has reported internal displacements resulting from drought between 2008 and 2020, only one country has had long-term monitoring of the phenomenon (Somalia, where UNHCR and the Norwegian Refugee Council (NRC) collect the PRMN dataset. See section 7.4).

Lack of monitoring of affected and displaced populations

Affected populations include all those whose lives have been exposed to the drought hazard (e.g. EM-DAT dataset)[55]**. Establishing reliable figures has a direct effect when addressing exposure mapping and populations at risk of internal displacement. It also has a direct effect when estimating projections of internal displacement. For this reason, obtaining reliable figures of people exposed to a drought event is crucial to improving current methodologies' performance and advancing the understanding of the linkages of drought, climate-change and human mobility.**

Monitoring to ensure accurate displacement figures related to an event is also challenging. Because of the nature of drought as a slow-onset disaster, the magnitude of displacement is often invisible. Population movements can be scattered in time, in different destinations and in the way they occur. This is revealed by the fact that most of the reviewed studies did not address internal displacement but other types of human mobility, such as international or internal migration. They did so by indicating changes in household compositions or population distributions. Since this data is mainly collected from targeted surveys, the implication is that it is usually an underestimate of the magnitude of the displaced population.

Net migration estimates based on satellite measurements of changes in land and infrastructure may help in mapping urbanization or net migration that is missed by traditional approaches (See *Joint Research Centre, 2020*), but these are currently only available over long timespans[56]. They are unable to track emerging trends and sudden peaks of internal displacement. Other data acquisition approaches that could be implemented (e.g., cell-phone tracking) may leave the most vulnerable population in low middle-income countries out of the picture**. For the moment, only reliable in situ measurements such as IOM DTM, UNCHR PMT and PRMN data are available as reliable data sources of drought displacement for modelling purposes** (see sections 7.1 and 7.4)**.**

Accounting for complex dynamics

Mobility dynamics are intricate. Non-linear relationships, socio-ecological feedback effects, autoregressive effects, interactions between variables, non-stationarities and adaptive responses are to be expected in mobility dynamics. Here we discuss these cases from the modelling perspective and with model types that can account for such complexities. Modellers need to consider these key aspects and delve into these relationships to successfully explain mobility outcomes.

Non-linear dynamics are to be expected. Non-linear relationships are those which do not exhibit a direct linear relation between two variables. These are the overarching type of dynamic in any complex system. Some examples in the field are the existence of tipping points upon which the loss of livelihoods occurs or the "U-shaped" influence of climate and income on migration found in the literature: given the same climate impact, people with low incomes cannot migrate (trapped populations) while people with high incomes can.

Feedback effects or loops refer to the phenomena in which variables affect one another, critically resulting in a spiral of worsening livelihoods. Ecological degradation, for example, could result in the exploitation of agricultural land beyond its capacity, causing harsher degradation. Similarly, conflicts resulting from political violence can influence agricultural policies which, in turn, can generate conflicts over food availability. These effects are typical of intertwined systems such as drought displacement.

Autoregressive effects from the modelling standpoint are those in which past conditions or trends of human mobility affect future human mobility trends. Examples include the creation of family links, migrant networks and communities at the points of destination which drive migration or displacement towards a particular region or the constant erosion of livelihoods because of drought, also resulting in displacement. Such dynamics could also affect the decisions of the people that decide to stay, or the destination of displacements.

Non-stationarities are also to be expected. A non-stationary phenomenon is one in which its mean and variance over time is bound to change. Mobility drought dynamics are non-stationary in the sense that they present emerging mobility trends that are bound to change over time.

Describing temporal patterns of human mobility is inherently more challenging than characterizing spatial patterns because of **this complexity and the lack of consistent temporal data on displacement and its drivers**[21]**.**

Accounting for accumulated impacts of drought

Determining time-lags and windows in which the different variables affect drought displacement is one of the principal decisions in methodological implementations. The time when the deficit in rainfall starts affecting displacement is an example of an important time-lag to establish for crisis prevention applications. When it comes to windowing choice (accumulated sum or average of data), seasonal windows which can measure the accumulation of rainfall during the growing season is a common choice among reviewed studies to monitor crop failure and consequently, displacement risk. We know that in drought-prone regions, what happens in the rainy or crop-growing season is critical. The degree to which cascading effects disrupt livelihoods during subsequent growing seasons is mostly unknown, however.

Some models might be able to capture the accumulation of drought effects (i.e., autoregressive models). Including this information, however, is needed for other approaches such as the calibration of complex system models. The severity of accumulated effects is also defined by the intensity of the drought, which, in turn, depends on the frequency of exposure and duration of the hazard. It is these determinations which have the greatest impact on further model development and applications for crisis prevention. **Correct quantification and attribution of accumulated effects, however, is challenging because of issues around the availability of long-term time-series data.**

Identifying displacement tipping points

The question of what specific thresholds and tipping points trigger internal displacement remains largely unanswered. The question mainly refers to how and when environmental stressors caused and generated by drought surpass critical socio-ecological thresholds[57]**. Any small perturbation from this state may trigger displacement**. For example, drought could result in substantial reductions in crop yields or in the livestock of pastoralist communities. That would cause these communities to lose their livelihoods and push them to displacement. The critical baseline level is unknown for most cases. It is also highly context specific. If there are financial remittances from migrants or humanitarian actors, these tipping points may also never be reached. For that reason, the question is also about adaptation measures' influence on human mobility. Initiatives exist, such as the H2020 HABITABLE project, which try to address this question. In an operational context, modelling these thresholds precisely would allow one to design reliable early warning and crisis prevention systems that could alert populations at risk of being displaced[58]. It would permit answers to questions such as WHO will be at risk, WHAT is the magnitude of the situation, WHERE it will occur and WHEN humanitarian aid will be required[19].

Quantifying the importance of multiple drivers of displacement

The exact quantification of drivers is also critical. It **would allow simulation models to determine the links and influence between components with a certain level of confidence.** Such exact quantification is not possible at present and the importance of drivers is highly context specific. Evidence of qualitative importance would be highly beneficial, allowing regression models or data collectors to focus on the most important variables. Lastly, reliable, context specific rankings of drivers would allow for effective and targeted policy interventions addressing priority tasks that tackle the fundamental drivers of displacement.

Identifying correct attributions

Current hypothesis testing analysis relies on associations and correlations between variables from statistical tests or regression methods. Some of the relationships found in the literature might be causal, but correlation does not imply causation, and nowhere are causal discovery and causal inference methodologies explicitly explored in the literature**. Apart from a better knowledge of the factors driving displacement, a correct assessment of the relationship between the causation of drought-related factors and displacement gives modellers the advantage of focusing on just the variables which are generating the system, thereby simplifying, and not biasing the modelling approaches.** Dynamic models also need to assess the strength of causal links and its direction between variables, but most causal relationships are inferred in a qualitative manner because of data gaps or the lack of quantitative evidence. The correct attribution of the strength of each causal link is of tremendous importance for these methodologies.

Fitting the right conceptual framework

Modelling human mobility requires considering how society and the environment interact and conceptualizing these interactions. Depending on the model in question, the adopted conceptual framework can affect the results and the model's applicability and flexibility. For example, gravity models root their conceptual framework in economic theory, where a utility maximization criterion related to labour opportunities in urban regions is set and, in some cases, can explain a great part of migration flows[30]. This assumption, however, does not explain the influence of migration networks over large distances or other cultural reasons unless explicitly accounted for. This point is particularly delicate in the modelling of complex systems, since they fundamentally rely on replicating the plausible interactions of the processes involved in drought displacement. More concretely, **in the modelling of a complex system, a simple conceptual approach will be unable to fully characterize the system, while complex models could be difficult to calibrate, validate or interpret and could not do so without the required data.**

Biases in the collection of drought impacts and displacement data

Existing internal displacement data present different biases and uncertainties that need to be considered. After all, models extract their conclusions from the data they are trained on, and for that reason biased data can skew conclusions or produce inaccurate predictions. Here we cite the main issues.

Data acquisition biases happen when the data is not fully representative of the situation or the population that a survey aims to capture or represent. This could lead to data with **representation or selection bias**, which occurs when data is collected from unrepresentative populations. If a survey only captured feedback from a certain segment of the targeted population, the data would be skewed. This could happen when sampling methodologies are not randomized or when a decision is taken to not sample in regions where populations face similar drought exposure but are unequally affected. The inevitable solution here is to input the missing data (under certain assumptions) or to work with models that can incorporate missing data. Both approaches introduce further uncertainties into the models. Biases in drought displacement data are more concretely addressed in Section 11 on Improving internal displacement time series data.

Some drought impact and displacement data may focus on particular locations rather than the totality of areas affected. This could be the result of a lack of funds for the data collection, lack of access, or security restrictions for collecting data. It would also create a **location bias** in the datasets collected**.** Because of the lack of systemic collection of socio-economic data on the impacts of drought, some of the datasets available could present a **historical bias,** meaning that the data used in models no longer accurately reflects the present.

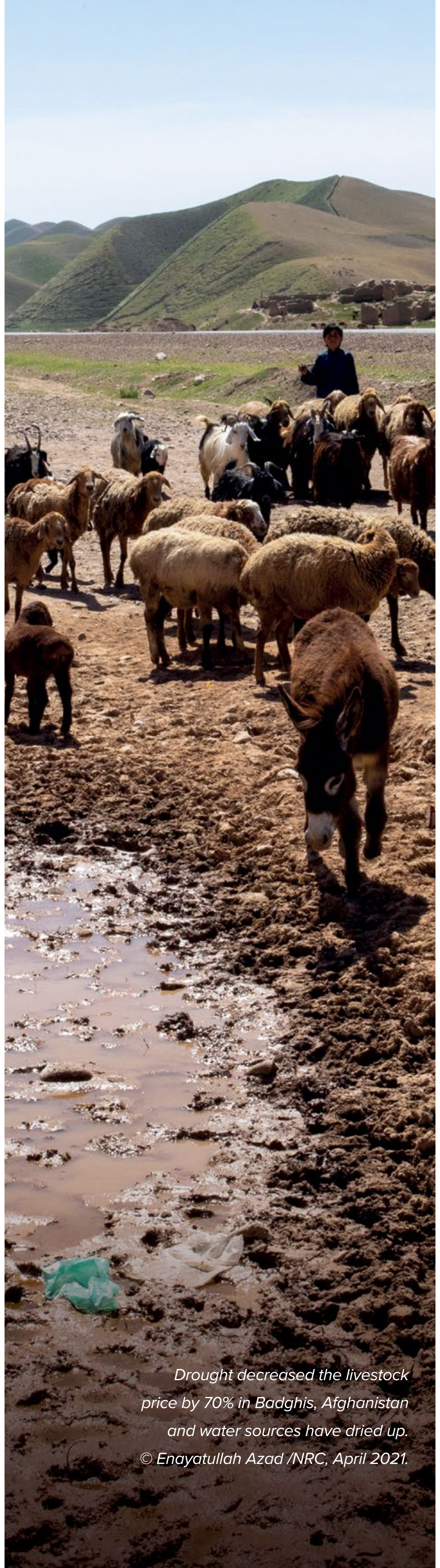
When an interviewer conducts a survey and influences its results, this could introduce **interviewer bias** (this could also be related to language bias). **Response bias** happens when the survey structure is constructed in a way that encourages or directs answers from the target population. We can identify biases arising from both the data acquisition process and from modelling methodologies.

Biases in the integration and aggregation of displacement data

Data processing for model ingestion and modelling methodologies could also introduce biases. A key example is possible aggregation bias. **It emerges from aggregating flows or stocks of displaced people over large time periods or over large regions. Both of these suffer from a similar problem: temporal or spatial aggregation may include people who were affected by different conditions. The result would be an inability to disentangle drought effects from other mobility patterns or baseline magnitudes of human mobility.** This is commonly found in the literature in spatial models of aggregated migration flows and must be addressed by developing context-specific, high-resolution models when possible. The way to integrate data into a model also has to be accounted for.

Estimating uncertainty in models

Because of the nature of displacement data (these are always estimates), assessing uncertainties may be the most difficult task. Uncertainty may arise from both the data and the modelling approach. **One must consider that models are as good as the data they are trained on, and their computed errors are usually based on how well the model replicates the observed data. For this reason, uncertainty estimations are highly related to the bias in the available data**. This is also a key limitation in displacement models, given displacement data availability. The other main source of uncertainty comes from how well the model assumptions capture the drought displacement phenomena to be modelled. The implication of these points is that error measurements and confidence intervals are bound to the modelling assumptions (used data and conceptual approach). As knowledge of the topic and data collection advances, these uncertainties might be harnessed by more robust and reliable models.



Drought decreased the livestock price by 70% in Badghis, Afghanistan and water sources have dried up. © Enayatullah Azad /NRC, April 2021.

Mobility datasets used in drought modelling

This section describes some common characteristics of human mobility datasets used in the literature. To better understand how models describe relationships between human mobility and drought, we need to keep in mind that the models' target variable is drought-related human mobility which is driven by explanatory variables (covariates). The covariates documented are described in Section 6. Figure 6 shows a summary of the main data sources of the target variables used to describe human mobility in the models reviewed. This figure illustrates that few attempts have been made to model drought displacement.

Our analysis showed that most of the data-driven modelling comes from surveys or census interviews (28 out of 42), mainly composed of retrospective information about the conditions and outcomes of human mobility. Net migration estimates follow, constituted by non-traditional or direct approaches of mobility data acquisition, and lastly, time-series data. One study by *Schutte et al., 2021* also used asylum seeking applications as its data source to study the climate effect on migration intentions[49].

As shown in Figure 6, five simulation models aimed to replicate mobility dynamics without having access to explicit mobility data. These modelling approaches could provide insight in situations where no data is available and explore plausible human mobility responses in the face of drought impacts. As an example, *Groth et al., 2021* seeks to replicate the movement patterns of pastoralists in Somalia across space and time and the way they might

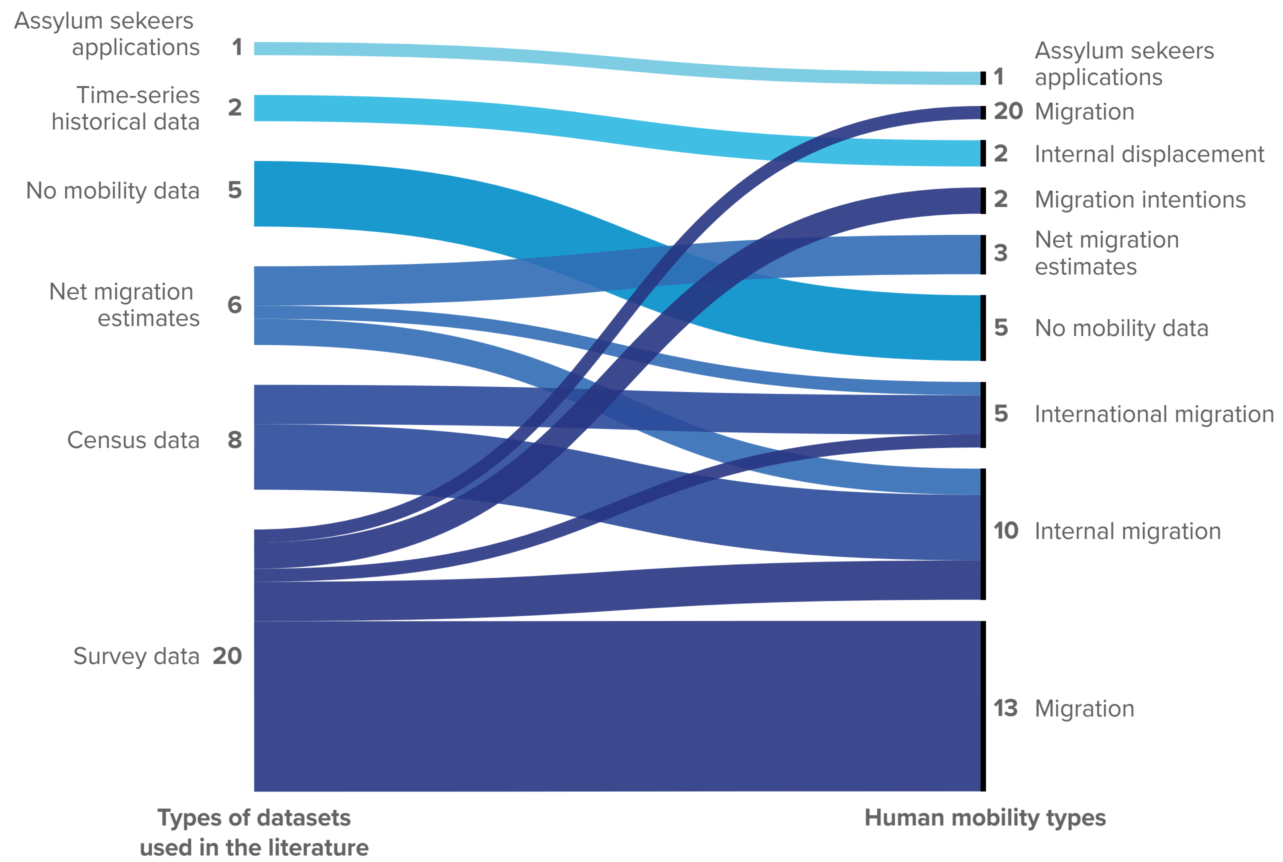


Figure 6: Frequency of data types used according to type of human mobility

continue in the future in the face of environmental stressors[25]. Following that example, modellers could include previous knowledge about pastoralist responses as rules in an agent-based model to explore adaptation responses. The reality is that geo-referenced data on pastoralist routes is extremely challenging to get. For that reason, trying to emulate this data is the only way to approach the problem. In these cases model validation is implausible, however, and results should be interpreted with caution. Their applicability under an operational context is infeasible

Datasets with the potential to characterize the slow and extended effects of droughts on displacement often do not exist or are not readily accessible. In most cases the data is also incomplete or presents high degrees of uncertainty. This implies constraints in model implementation and related policy making applications. It is especially problematic in medium- to low-income countries and particularly in rural areas, where there are data acquisition challenges and where droughts have tremendous impact on entire populations. Many of the reviewed articles draw attention to the lack of good quality longitudinal data (long-term systematic data) as the main data gap for building effective models[14, 46, 59, 60].

Regarding model results obtained from the mobility datasets used in the literature, we found that some approaches to modelling macro-aggregated flows between regions (e.g. cross-border displacements and migration) over long time periods could explain a significant part of internal and cross-border migration with simple assumptions and without demanding detailed micro-level data[61]. Mobility dynamics, however, are a product of much more complex interactions that require the exploration of displacement's driving forces by upscaling individual, household and social system interactions into community, village, district and even country level outcomes. In practice this means that although there are some situations where simple models would be helpful for defining policies and prevention strategies, these assumptions are too simplistic to effectively address populations at risk with the required time-frame and spatial disaggregation.

The implication is that there is a need for local and context-specific approaches when modelling drought displacement. Research findings indicate that conclusions of one study, are not necessarily pertinent to other areas, or to different population characteristics (i.e., income, gender, livelihood, age, etc.)or mobility patterns (i.e., rural vs. urban, internal vs. cross-border, permanent vs. temporary, voluntary vs. forced, etc.)[54, 62, 63, 61, 64]. Results vary depending on the case studies and hybrid/heterogeneous narratives of human mobility arise[60]. Several reviewed articles have also shown how demographic and socioeconomic factors have a bigger influence than environmental factors or specific climate shocks on displacement convoluting direct causal pathways and indicating that drought impact is primarily a social phenomenon, not just a natural hazard, where risk mitigation and resilience building is possible[46, 65, 48].

Survey data (microdata)

Individual or household surveys consist of targeted interviews for monitoring human mobility. They can describe flows of people or populations on the move (e.g., monitoring of population displacements and movements such as arrivals and returns at transit sites, established IDP settlements and ad hoc locations), retrospective information about household members who moved before the survey (e.g. total counting of people living in IDP camps, such as site assessments, also known as stock data), and future displacement intentions.

Some surveys strategically sampled and designed for displacement monitoring purposes were conducted in drought-prone areas[60]. So can more general surveys that are later curated for drought displacement modelling purposes[15]. These datasets often gather micro-level information about individual or household characteristics that are otherwise unavailable. Survey panel data are very valuable sources in the field which cannot be replaced by other methods. These datasets are collected on an ad hoc basis, however, hampering a long-term analysis of drought and displacement impacts.

The reviewed models use survey data applications to test whether climate variability or other confounding factors had any impact on mobility outcomes. For the same reason, survey data can also be exploited to extract micro-level information from affected populations that can be later included in simulation models. Some longitudinal surveys exist, but the time resolution is often not extensive enough to test the short- and long-term effects of drought, climate and other factors affecting forced displacement[66].

A unique dataset that deserves mention comes from the IOM DTM surveys in Ethiopia, collected since 2016[67]. This dataset consists of a series of surveys conducted through bi-monthly site assessments. They are focused on IDPs and the availability of services in their areas of displacement, with detailed information on household composition, income and other factors that are crucial to explaining mobility conditions with a level of detail and regularity not found in the literature. The DTM dataset is only used by the experimental IDMC System Dynamics model to extract micro-level information about pastoralists' livelihoods in Ethiopia[24]. We find this data collection effort to be incredibly useful for modelling purposes.

Census data

Mobility-related information in censuses can be exploited if it contains a tracking of changes in residence through successive census rounds or interviews. This should include retrospective information about past mobility, like in-survey data, with the difference being a non-targeted design for human mobility purposes. In this regard, there is no strategic sampling in drought-prone areas or specifically designed questionnaire that could accurately account for different human mobility conditions. Census panel data, on the other hand, covers a greater proportion of the population and often results in greater sample sizes. Such data collection is systematically conducted by governments or other agencies.

Thanks to the systematic and even collection across regions or countries, census data in the literature is mainly employed to train econometric spatial models (gravity or radiation models) which aim to predict net internal and cross-border migration flows between regions. For that reason, census data is usually aggregated at a regional or country level. For that reason, possible aggregation biases can be introduced (along with some of the same biases present in survey data). Among the reviewed literature, one can find models using national censuses to predict migration flows between provinces (Iran, 2018), or international censuses such as the one used in *Beine and Parsons, 2015* aiming to cover migration flows between countries at a global scale[31, 68]. For more information about the specifics of census data employed in drought modelling, see 3.1 Mobility Dataset Description in Appendix I.

Net migration estimates

Net migration estimates are compounded by data sources which are not explicitly on migration. Rather, they rely on exploiting demographic data or sources such as satellite imagery or cellphone-based data to build estimates or proxies for migration over regions. These estimates demonstrate strong potential as they are available on a timely basis, with the ability to monitor regions in which no other data acquisition method is possible and with little aggregated cost and deployment of resources. These approaches, however, include aggregated uncertainties because they target migration indirectly and may not track temporary movements because of time resolutions. The inclusion of micro-level information in an area also is challenging.

These datasets have been mainly used to build econometric models (gravity and radiation models) as they can successfully exploit spatial information given by gridded or country-scale estimates of migration in whole regions at an aggregated level. As their design varies greatly, a few examples will be cited.

A well-known approach that uses net internal migration estimates is presented in the *Groundswell: Preparing for Internal Climate Migration* report[12]. The authors use the Gridded Population of the World (GPW) census-adjusted product which models the population distribution over a continuous raster layer[69]. Migration flows are estimated from the difference of population across all grid cells in ten-year intervals under different development and climate scenarios. It focuses explicitly on slow-onset disasters such as droughts in rural and urban areas and aims to provide estimates on future migration outcomes and identify migration "hotspots" regions within countries. As stated before, these approaches provide mobility data over large regions that is otherwise unavailable. For that reason, they are very useful alternatives to in-situ data. Several low- and middle-income countries, however, present censuses at lower resolutions, are out-dated, or contain known inaccuracies, introducing uncertainties in the model's predictions which are challenging to quantify.

Another case is the Global Estimated Net Migration Grids By Decade database developed by the Center for International Earth Science Information Network (CIESIN) for 1970 to 2010[70]. This has the goal of filling the data gap for subnational migration estimates. It allows 1x1km grid global coverage and comparability of migration estimates across countries. This dataset is based on spatial population distribution data and natural population increases to estimate net migration. It is employed by Neumann et al. (2015) to extract land characteristics of areas with high magnitudes of drought displacement areas across the globe and by *Peri and Sasahara, 2019* to predict push factors of country-to-country migration worldwide [17, 63].

Following the previous dataset methodology, see a recent dataset built by the Joint Research Centre of the European Commission (JRC) to map rural-urban net migration in inaccessible regions at the country and subnational level[56]. It provides net migration by combining indirect demographic estimation techniques with satellite and census data from the JRC Human Settlement Layer (GHSL) at a 25x25 km grid level on a five-year basis. This dataset is later employed by the same authors to measure climate influence on migration in a drought-prone region in the Sahel[37]. For that reason, it is not restricted by national *boundaries* and is suitable to tracking environmental conditions using earth observation data. We did not find this dataset readily available to the public, however.

Another example worth-mentioning is one of the reviewed articles about La Guajira (Colombia) by *Sibren Isaacman et al., 2018*[34]. This used anonymized call-detail records to estimate the geographical position of people in a heavily drought-affected area, effectively displaying human displacement at a high spatiotemporal resolution across a whole region. A drawback is that there is sensitive data involved which has been anonymized, is company-owned. and not publicly available. At the same time, cell-phone data, and other possible approaches such as social media use is limited in marginalized rural-areas around the globe. Other similar datasets can be also found in the literature, such as the UN World Urbanization Prospects dataset, exploited by *Barrios et al., 2006* using urbanization data as a proxy for rural-urban migration[72, 71].

Time-series data

Time-series data consists of a sequence of data points taken at successive and (ideally) equally spaced-out points in time. Time-series data on drought displacement is the most versatile and desirable format from the modelling standpoint. and the most complete, and it is ideal for tracking changes across time and assessing climate impacts on displacement.

Some studies based on survey or census data try to replicate a time-series format by creating person-years or household-year time-series (Ecuador, 2013)[60] (Mexico 2019)[18] (Pakistan, 2014)[66]. Our findings indicate that authentic and systematically collected time-series data is only used in IDMC's experimental system dynamics model and Project Jetson[24, 13]. These models make use of the PMT (Population Movement tracking), PMN (Protection Monitoring Network), and PRMN (Protection & Return Monitoring Network) datasets collected since 2006 by a UNHCR-led project in Somalia[73]. We argue that these are the only datasets complying with timely resolution and the long historical record to effectively measure drought displacement dynamics. This data is also collected at a district level spatial resolution with an associated cause for displacement, which offers valuable ground truth labels for drought displacement monitoring.

UNHCR's PRMN dataset, however, focuses on displacement tracking and protection monitoring (e.g. displacement flows, such as internal and cross-border displacements, returns and relocations) and does not record additional information on the impacts or needs of the displaced population. Historical datasets (PMT and PMN) are also not openly accessible.

Openess of mobility data

Data availability issues are a principal concern in advancing existing modelling approaches. Here we display the findings on publicly available mobility data sources used to model drought displacement (Figure 7). Given the fact that none of the reviewed articles published the harmonized-ready data, only raw source data is considered (i.e. with no additional curation, harmonization or cleaning). This is yet another obstacle for advances in the field, since data curation is the most time-consuming task in model development. The link to these publicly available source datasets is provided on *Annex I*, in *3.4 Data Availability* column.

To provide a synthesis of existing mobility dataset types, we present the following table with their applications, strengths and weaknesses from the modelling perspective.

Figure 7: Raw data availability among reviewed models. Twenty-three of the models used publicly available raw data, 14 used data that was not publicly available.

Dataset type	Example of dataset	Applications	Strengths	Weaknesses
Survey data (micro data)	E.g. IOM DTM Site Assessments in Ethiopia. This dataset collects data on the total number of people displaced, the reasons for displacement, and some socio-economic impacts of the displacement.	Use survey data to quantify and calibrate simulation models' parameters	Surveys provide very valuable micro-level information about factors affecting human mobility that is otherwise unavailable. Surveys can be strategically sampled and designed for displacement monitoring purposes in drought-prone areas.	Mostly rely on self-reported information collected by household surveys and data is costly. This limitation is translated in some contexts into low sample sizes or big temporal resolutions. Could have the following bias: Interviewer bias (the way the interviewer conducts the survey, can impact its results, this could include language biases), representation or selection bias (when the survey results are skewed because the survey only captured feedback from a certain segment of the population, this could include location bias), Response bias (This happens when the survey structure is constructed in a way that encourages or directs some answers).
Census data (macro data)	E.g. Iran's national census data, covering the period 1996-2011.	Provide evidence on the factors affecting human mobility. Training of spatial models to predict internal and cross-border migration flows across regions.	Systematic and regular data collection that can be comparable across countries. This data could be employed to train spatial models.	Large timespans between census rounds. Census data is usually aggregated at a regional or country level. This type of data can also present aggregation bias and the same bias as the survey data.
Net migration estimates	E.g. Grid level net-migration rates. (CIESIN, 2011; Sherbinin et al., 2012).	Training of spatial models to predict internal and cross-border migration flows across regions	Available on a timely basis, with the ability to monitor regions in which no other data acquisition is in place with little aggregated cost. These datasets are available over large areas	Aggregated uncertainties as they target migration indirectly. Baseline migration data must be estimated. They may not track temporary movements because of time resolutions. Low- and middle-income regions may suffer from lower resolution.
Time-series data	E.g. Weekly, long historical record of drought displacement (E.g., UNHCR PMT and PRMN datasets)	Time-series forecasting	Most versatile format from the modelling standpoint, ideal for tracking changes through time and assess climate and other impacts on displacement	Time-series data is costly to obtain, and exhaustive monitoring needs to be in place to ensure an even sampling of displacement waves. The challenges of this type of dataset are similar to the biases presented in survey data.

Table 1: Applications, strengths and weaknesses of current mobility datasets used for modelling

Drought and internal displacement, why do we need this data?

Different international targets, goals, and processes (ex. internal recommendation on IDP statics) have highlighted the international community's need and commitment to define standards and develop tools for the collection, storage and long-term preservation and dissemination of displacement data[74, 5, 75]. This is intended to make forcibly displaced people visible for response, advocacy, and policy making, and to program actions aimed at preventing, responding, preparing and reducing forced displacement.

Forced displacement refers to situations in which people are forced or coerced to leave their homes or places of habitual residence to avoid the effects of armed conflict, generalized violence, violations of human rights or natural and human-made disasters[76]. Forced displacement includes people displaced across-borders (ex. refugees, returnees, and asylum seekers) and people displaced within internationally recognized state borders, defined as internally displaced people (IDPs)[77, 78]. According to the Global Trends Report published by UNHCR, the number of internally displaced people in 2020 represented more than 50 per cent of all forcibly displaced people reported[79]. This figure, however, refers just to forced displacements resulting from conflict and violence. Data on forced displacement resulting from disasters triggered by natural hazards is available, most of the time, only for internally displaced people. At the global level, only IDMC reports these figures.

The global statistics on new displacements published by IDMC between 2008 and 2020 shows that disaster-related displacements represented 77 per cent of all reported displacements (including new conflict displacements reported the same year). Of this percentage, just 0.7% were associated with internal displacements related to drought in 13 out of 65 countries.[vi] This percentage, however, is an underestimate resulting from the under reporting of this phenomenon. The Emergency Events Database (EM-DAT) reported that between 2008 and 2020 an average of 16 countries were affected by drought each year (or a total of 91 countries) with an annual average of about 62,7 million people affected[55].

The lack of reporting on internal displacement triggered by drought is related to multiple factors (for more information see section 8.1):

1. Not all drought events result in disasters or trigger displacement.
2. Because of the slow development and creeping nature of drought's impacts, determining that it has been a trigger of displacement is challenging, as drought can be a direct or indirect trigger of internal displacement.
3. There is often a reactive approach to monitoring displacement events resulting from drought. For this reason, several countries do not engage in regular monitoring of it.

According to the Sixth Assessment Report published by the Intergovernmental Panel on Climate Change (IPCC), human-induced climate change is already affecting many weather and climate extremes in every region of the globe, increasing the chance of compound extreme events and the frequency of concurrent heat waves and droughts. The report also emphasises that the societal impacts resulting from drought could increase in proportion with every degree of global temperature warming[80].

vi Afghanistan, Brazil, Burundi, Ethiopia, India, Iraq, Madagascar, Mongolia, Pakistan, Philippines, Senegal, Somalia, and South Sudan

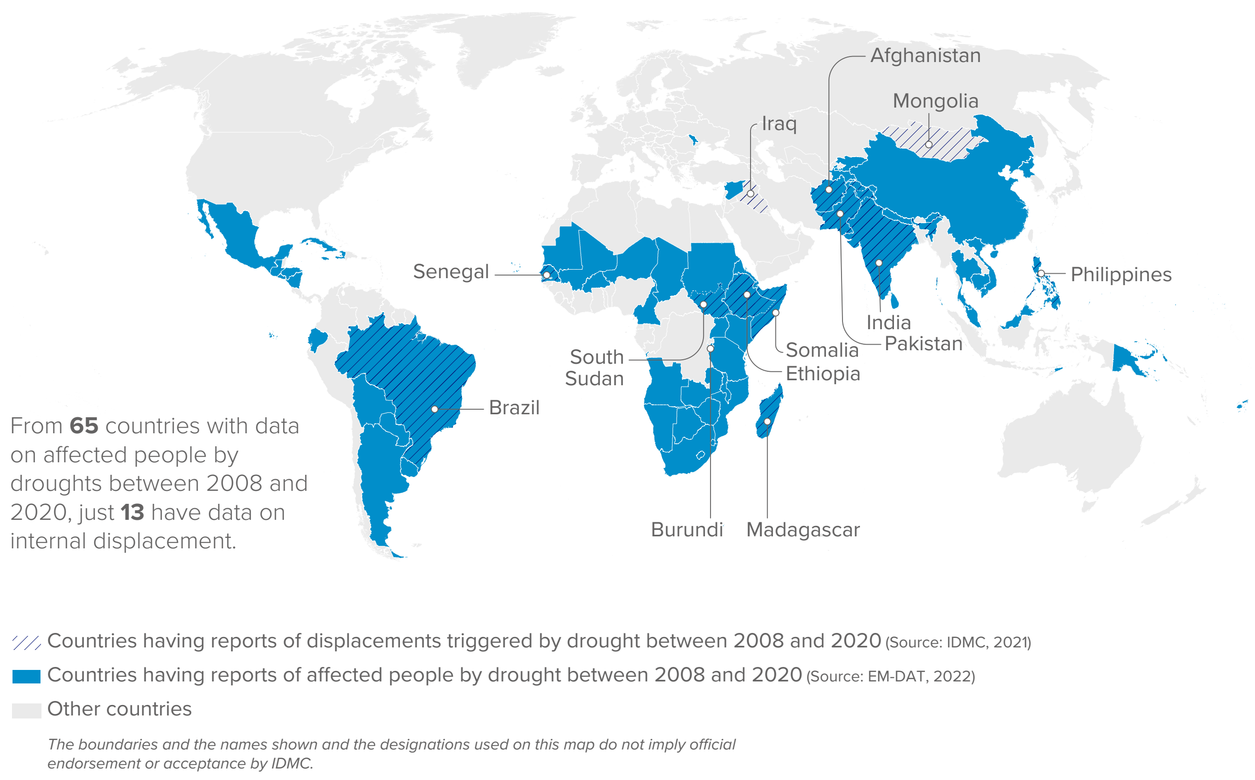


Figure 8: Countries reporting people affected and displacements triggered by drought between 2008 and 2020.

To fill the current data gap (under reporting of drought-displacements) and prepare responses to future displacements triggered by more frequent or intense droughts, we need to stress the importance of documenting and collecting systematic data on internal displacements and other forms of human mobility linked to drought. This could help us better understand the dynamics of slow-onset displacement, its drivers and its impacts on communities and livelihoods. Gathering displacement data is crucial to a better understanding of the magnitude of drought displacement and for the development of advocacy actions, informed policy formulation and programs aiming to prevent, reduce (ex. implementing forecast-based financing mechanisms), and respond to displacement situations and for the implementation of durable solutions for displaced populations already affected by drought.

Data on drought displacement could help identify trends and long-term impacts and develop robust drought displacement models that could support the assessment of short, medium and long-term displacement risk. It could also support the development or the improvement of displacement assistance, prevention and response systems. Statistics on the number of people displaced are also essential for measuring and evaluating the effectiveness of policies to reduce the risk of drought and related displacements, and to assess the effectiveness of the durable solutions implemented.

Two principal metrics collected during the monitoring of internal displacement are population stock and flows. Stocks refer to the total number of IDPs in a specified location at a defined moment in time. Flow refers to the dynamic measure of population movements over a period of time and represents the number of times that displacements are reported. Flows can be counted as inflows (ex. people entering the population stock) or outflows (ex. people leaving the population stock). An IDP stock could increase or decrease over time based on inflows (people who become displaced) and outflows[81].

“Confusion between stock and flow data is common and can lead to significant errors that result in an inaccurate assessment of the scale of displacement within a country,” said the International Recommendations on Internally Displaced Persons Statistics[81]. Understanding this metric is highly significant for the use of internal displacement data in the development of models and decision-support tools.

Covariates or driving factors used to model drought related to human mobility

Models use different explanatory variables (covariates) to understand or model human mobility resulting from drought. In our literature review, we found that data inputs used to model drought mobility are diverse, encompassing a wide range of macro-level, mezzo-level and micro-level factors which affect mobility at different scales and magnitudes[9]. Drought displacement is a context-specific phenomenon that could result from different environmental and socio-economic dynamics. At the same time, the data availability of different explanatory variables used in the reviewed models vary greatly from one context to another. The data used for one model may not have an equivalent in a different context[3, 82]. An ideal modelling practice calls for thoroughly considering every factor that could contribute to drought displacement by exhaustively and extensively studying the affected area and by incorporating local experts and knowledge into the models.

With the aim of illustrating different accounted dimensions from the literature review, we show the potential driving factors in different categories in Figure 9. All unique covariates derived from existing data sources are displayed in Table 2. with their associated category. For specific details about each model, the reader may look at *3.5 Covariates* and *3.8 Completeness of Covariates* sections from *Appendix I*.

The results of our analysis show that most of the reviewed models use climate, demographic, and socioeconomic data as the principal drivers for modelling drought displacement. A smaller number of models account for an explicit agricultural or livestock pathway to drought, *(*See for example *Entwisle et al., 2020 or Cattaneo and Massetti, 2015*) or environmental stressors such as vegetation indexes (See *Justin Ginnetti, 2014, Project Jetson, 2018 or Sakamoto*

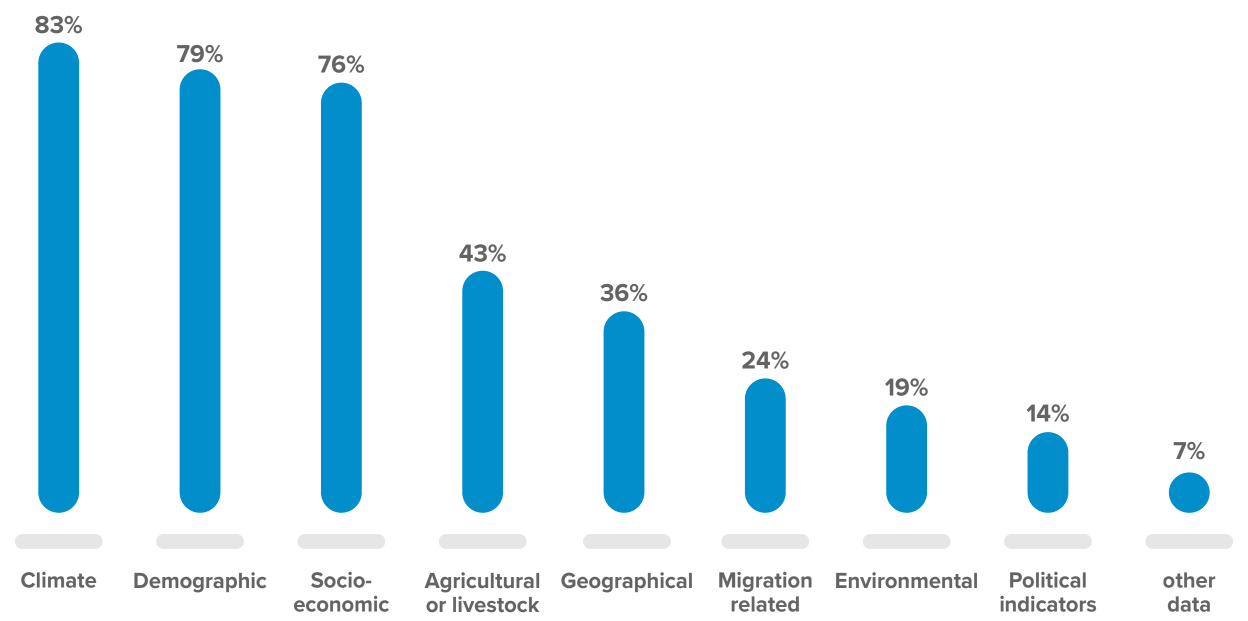


Figure 9: Counts of covariate types used across models (e.g., 83% of the models used climate variables).

2016), surface water indexes (See *Nelson et. al., 2020)*, or land degradation variables (See *Henry et al., 2003*)[13, 24, 40, 16, 84, 41, 83]. These are key factors for describing drought-related phenomena and their impacts on populations. Geographical data also plays a huge role, particularly in spatial models. It mostly reveals the distance to the nearest available destination as the most important factor in model performance, and indicates a prevalence of short-distance movements in coping with drought (See *Shiva and Molana, 2018,* or *Garcia et al., 2015*)[31, 85]. Fewer studies use political indicators and data on violent conflict (See *Barrios et. al., 2006*, *Schutte et al., 2021* or *Project Jetson, 2018*), or migration- related data, such as migration networks (See *Gray and Bilsborrow, 2013, Entwisle et al., 2020,* or *Beine and Parsons, 2015*) (e.g. transhumant movements in the case of pastoralists)[72, 49, 13, 60, 41, 68]. All of these are found to have a significant impact in describing climate-related mobility across the literature.

To avoid missing crucial relationships and falling into simplistic assumptions, our recommendation is to consider all relevant covariates in the analysis and later discard them if data analysis or expert discussion finds them to be unimportant or unusable because of uncertainties or data quality.

Bellow we present some examples of displacement covariates (or drivers of displacement) found in the literature:

IDMC's Experimental System Dynamics Model and Project Jetson identified livestock prices as crucial to model displacement magnitudes in Somalia and Ethiopia because of pastoralists' drought adaptation strategies[24, 13]. An agent-based model approach in Nigeria by *Sakamoto, 2016* includes tsetse fly distribution, as pastoralists' routes are intended to avoid encounters with the insect[40]. Another study in India claims that data on drought-resistant crops would help predict migration flows (*Dallmann and Millock, 2017*)[82]. A study by *Barrios et al., 2006* found that a decolonization factor had the most impact on model performance to explain country-to-country migration across sub-Saharan Africa[72].

These examples and others point to the fact that more effort should be made to collect qualitative and data-ready factors to improve drought displacement modelling approaches. A more detailed review of data gaps in modelling is presented in section 6.

Category	Covariates used in the literature
Climate	Precipitation anomalies, Decadal precipitation, Yearly precipitation, Seasonal precipitation, Monthly precipitation, Standardized Precipitation Index (SPI), Standardized Precipitation-Evapotranspiration Index (SPEI), Drought hazard, Drought frequency, Drought magnitude, Drought duration, Mean temperature, Temperature anomalies, Number of Droughts, Climate projection scenarios, Soil Adjusted Vegetation Index (SAVI)
Demographic	Age, Gender, Marital status, Population, Ethnicity, Household size, Religion, Birthplace, Urban population, Language, Urbanization rate, Percentage of white population, Household number of children, Female is the head of household, Ethnic minority
Socioeconomic	Household income, Household assets, Household expenditure, Land ownership, Ties to wealthy households, Household's economically active members, Educational level, Employment, Food prices, Normal food expenditure, Milk produced per head of livestock, Cash expenditure, GDP per capita, Remittances from relatives living abroad, Value of land owned, Market prices data, GINI index, Inflation rate, Wage differential, Marginalization index, Household vulnerability
Agricultural or livestock related	Crop yields, Land parcel size, Soil suitability/fertility, Income earned from crops, Crop market prices, Household members participating in farming, Livestock ownership, Livestock market prices, Number of livestock, Pasture area, Livestock death rate, Share of men in agriculture, Cropland area, Ratio of net irrigated land, Total cultivated area, Local main crop calendar, Means of ploughing plots
Geographical	Slope of terrain, Distance to nearest village, Elevation, Settlements, Area, Border Contiguity, Distance between regions, Distance to nearest trading centre, Household location, Rural or urban location, Distance to paved road
Migration related	Migrant Network, Migration experience, Personal attitude towards migration, Return migration rate, Peers opinion on migration
Environmental	Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Soil Adjusted Vegetation Index (SAVI), Flooding potential, Land use, Soil degradation, Pastureland area, River levels, Forest cover
Political indicators	Severity of armed conflict, Number of fatalities, Number of conflict incidents, Perception of policies, Number of international violence episodes, Decolonization, Civil wars
Other data	Indicated migration reasons, Malnutrition data, Disease data, Tsetse fly distribution

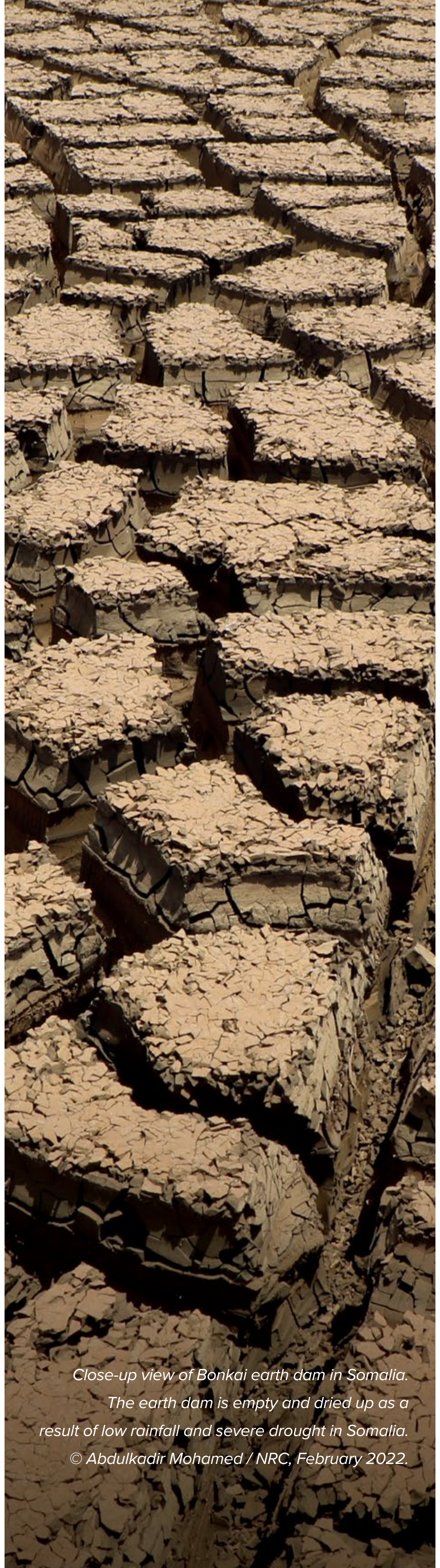
Table 2: Different covariates used for drought mobility monitoring

For these reasons, to fully describe mobility dynamics, the modeller should account for all possible relations between the context-specific factors and spatiotemporal interaction scales, while at the same time avoiding overcomplicating the model or including redundant information (i.e. using an analysis of intercorrelated variables) that could in some cases compromise the model's results[28].

There is always a trade-off between simplicity and complexity in any modelling effort, where too simple models will be unable to characterize the system and too complex models could be difficult to calibrate, validate or interpret. The final choice for modelling depends on the specific goal, the availably and quality of data, and on the modeller's expertise.

Availability of covariates

In the reviewed literature we found that most of the model covariates used are publicly available, except for variables extracted from restricted survey data. Covariates are generally much more available for model ingestion (especially those variables derived from Earth observation sources that periodically cover the planet).There are still micro-level data gaps, however, and, as some authors remark, high spatiotemporal resolution of migration conditions is often lacking (See *Justin Ginnetti and Travis Franck, 2014*, *Mastrorillo* et al., 2016 or *Sakamoto, 2016*)[24, 61, 40]. Most socioeconomic and demographic micro-data covariates are often present in survey or census interviews while macro-level indicators are often extracted from global indicators databases or government agencies. Geographic data is mostly extracted from geographic information system (GIS)-based sources and environmental and climate data from in-situ weather stations or from satellite imagery that provides a worldwide, systematic record of environmental stressors (precipitation, vegetation indices, soil moisture, etc.). Sources and references for all these datasets can be found *in 3.6 Covariates Source & Availability* section of the model catalogue in *Appendix I.*



Close-up view of Bonkai earth dam in Somalia. The earth dam is empty and dried up as a result of low rainfall and severe drought in Somalia. © Abdulkadir Mohamed / NRC, February 2022.

Main data gaps identified

One of the objectives of the literature review was to identify data gaps that were hampering model development and affecting model results. More than half of the works (25 out of 42 studies) specified concrete data gaps to be solved in their respective field of application. We found that internal displacement is greatly underrepresented by current modelling approaches. Only two of 42 studies used internal displacement as mobility data.

Readily available internal displacement data is a large gap. Model implementation, however, must meet certain data requirements. **To fully characterize slow-onset displacement dynamics long-term and high spatiotemporal resolution data is required. Longitudinal (long-term) geospatial data is needed to follow the slow accumulation of environmental and societal changes leading to displacement.** This is also relevant to the number of points and the number of drought events that have been recorded. Low-sample sizes greatly increase the margin of error in model fitting, while records of single drought events imply challenging generalizations (i.e., extrapolating results to other drought events). At the same time, **the temporal resolution must be sufficient to map and monitor the exact conditions and thresholds triggering displacement flows. The subject requires the inclusion of diverse sources with different granularity and level of detail (micro, mezzo, and macro-level interactions) that add layers of complexity to the model design, as described in the covariates' section.**

Table 3 compiles all the data gaps cited by the authors in the reviewed articles. Our principal focus is on internal displacement. Other types of factors affecting drought mobility dynamics, are also included, however, because they could act as proxies for more threatening socio-ecological impacts (food and water security) that trigger forced displacement.

The modellers and end users should also acknowledge that as the understanding of mobility dynamics increases, other key aspects for monitoring will arise.

Category	Cited data gaps
Mobility dataset	Reliable data (particularly in drought prone areas)[60, 85]
	Lack of long-term studies on areas that suffer drought impacts with human mobility trends[14, 66, 84]
	Bilateral flow data , rather than just unilateral flow data (i.e., data for both arrival and departure in a region, not only departure or arrival)[86]
	Survey data at higher temporal resolution, and geo-referenced survey data[46, 87]
	Field observations of pastoralists' mobility patterns and baseline data on seasonal transhumance movements[24, 16, 40]
	Missing values in displacement time-series[13]
	Lack of data on the composition of migrants[68]
Socioeconomic	Lack of long-term data with the micro-level information needed to model displacement dynamics (e.g., temporal tracking of household assets such as cash holdings, grain stocks, market purchases per household, water purchases, etc.)[59, 65, 68, 86]
	Monthly estimates of household income, household expenditure, returning flows for each livelihood zone (of IDPs to their previous livelihoods prior to displacement), price of fodder per market, access to water, etc.[24]
	Disaggregated data on the economic and social impacts of drought hazards[24]
	Remittance and humanitarian cash-assistance data as a pull/push-factor[13]
	Disaggregated employment or unemployment data[54]
Environmental	Long-term micro-level environmental data at the household level (e.g.., in situ soil moisture content or soil composition data)[88]
	Ground truth land cover data[46]
	Reliable land degradation indicators at higher resolution[14]
Demographic	Disaggregated data on the demographic impacts of drought hazards[24]
Agricultural or livestock related	Quantitative data on livestock movement patterns[24, 40]
	Monthly estimates on livestock birth rates, mortality, milk production and livestock sales per market, among others[24]
	Disaggregated data on the agricultural impact of drought, such as agricultural productivity or household agricultural gains[24, 83]
	Drought-resistance crop data in agricultural areas to control for the drought impact on agriculture[34]
	Rain-fed vs. irrigation land mappings[14]
Migration related	Information about migration networks (if migrant communities are at destiny or if temporal economic migration is a result of transhumant or pastoralist movements)[83]
	In situ questionnaires on the drivers of affected and internally displaced populations[61]

Table 3: Specific data gaps as cited by the authors

Recommendations to improve drought displacement modelling

Our findings show how internal displacement modelling is underrepresented in climate literature even though displacement is the most extreme response to drought conditions. Displacement modelling is still at an early stage with only two experiments conducted: IDMC's system dynamics model that seeks to understand pastoralists' responses to drought and the time series forecasting approach conducted by Project Jetson[24, 13].

As we previously noted, there are significant challenges in modelling drought displacement. The availability of drought displacement and contextual data is the main bottleneck hampering advances. Filling any of these cited research gaps would greatly improve modelling approaches. Assumptions in model development are crucial to building operational models: the stronger the assumptions, the more targeted and controlled operational models can be. The problem lies in the degree to which assumptions are correct or verifiable, and how well they fit the data.

Improving drought displacement modelling can be very useful for building tools that have a large impact on decision-making in the humanitarian sector. Models can provide a deeper understanding of the mechanisms and processes of drought displacement and insights into the bigger picture of displacement, depending on the scale of aggregation. Validated and reliable models can be used to perform interventions or turn from "what is" scenarios to "what if" scenarios once causal relationships are established. Such models could permit the carrying out of experiments that are otherwise infeasible on the ground by intervening in the inputs of the model. For example, by providing Forecast-based Financing (FbF) models, the humanitarian community could support the evaluation of different preparedness, early warning and response actions targeting the prevention and reduction of displacement or aimed at providing better assistance to drought-affected and displaced populations. Having such models is linked to the potential of automated detection systems of critical drought conditions and their respective trigger warnings for internal displacement risk. In order to reliably implement any of these technologies, however, overcoming existing challenges from the modelling perspective is crucial.

As stated before, current limitations and challenges can be broadly summarized in three principal categories:

1. Lack of access to relevant data,
2. Improving existing methodologies and modelling of drought-related mobility,
3. Lack of context-specific understanding of the problem.

The following section addresses these caveats.

Improving internal displacement time series data

We found that internal displacement data with the level of required spatiotemporal detail is almost non-existent. To our knowledge, there are few detailed studies on the spatiotemporal effects of drought on displacement. Quantitative evidence that can be translated into the models of tipping points, accumulated and cascading effects, and time-lags are mostly unknown. For that reason, only qualitative indications about the needed resolution can be suggested until the problem is addressed. To close these gaps, displacement data, through which the effects of droughts can be measured, is needed. This implies both long-term and high spatiotemporal resolution. Our findings reveal that only the UNHCR PMT and PRMN time series dataset complies with timely requirements.

In section 9, we presented data gaps in the literature that are crucial to modelling drought displacement. Multiple articles pointed out the lack of micro-level mobility and socioeconomic data needed to characterize displacement conditions (See section 7.1). The IOM DTM site assessments in Ethiopia deliver extensive survey information of these micro-level factors for displaced households. This kind of data collection represents an outstanding effort that could allow modellers to account for the relationships in drought displacement.

For that reason, we will present UNHCR PMT and PRMN time series dataset in Somalia (flow data) and IOM DMT site assessments in Ethiopia (total population counts data), as reference points. The best practise would come from adapting both methodologies to capture population flows, stocks and the impacts of drought displacement.

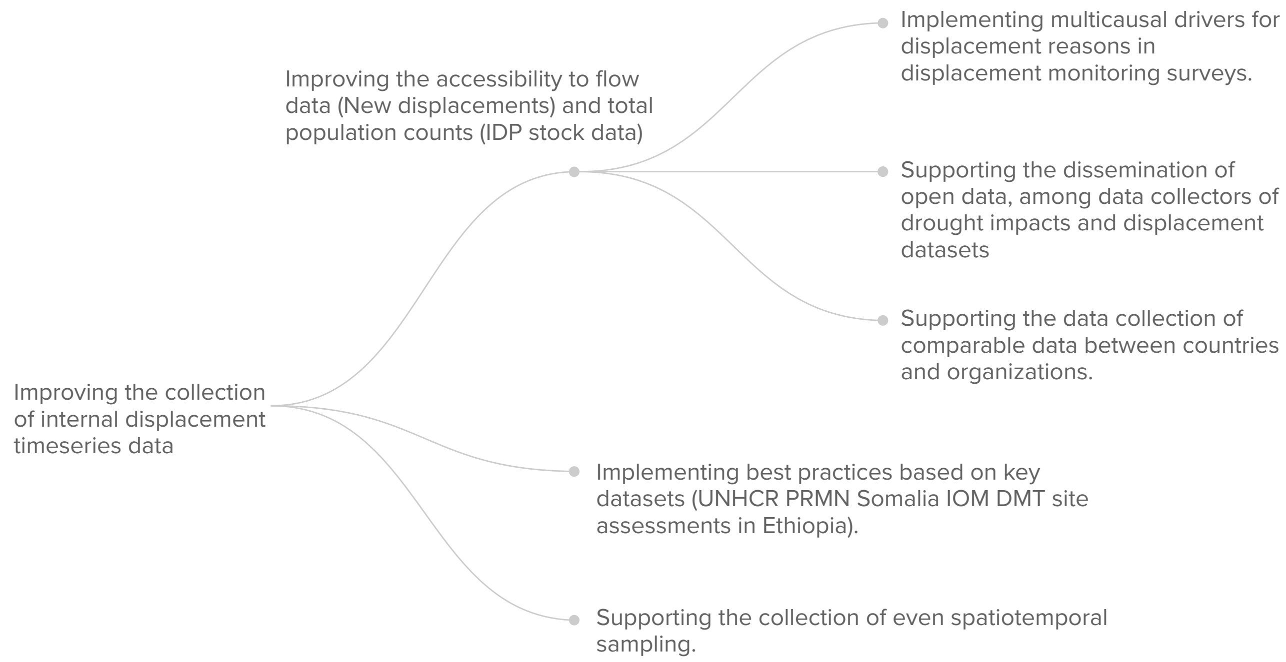


Figure 10: Main recommendations to improve internal displacement time series data

Implementing best practices based on key datasets (UNHCR PRMN Somalia IOM DMT site assessments in Ethiopia)

The PMT and PRMN datasets have been compiled since 2006. These datasets collect data on population flows, displacements and movements, such as returns, by targeting strategic points, including transit sites, established IDP settlements, border crossings and other ad hoc locations. The network works with trained enumerators whose geographical coverage or access depends on local security situations. The PRMN dataset structures the information by region of origin and destination, type of population flow (e.g., arrival, return), date of arrival, reason for the displacement and additional comments[89]. We encourage data collection initiatives to follow similar procedures so that reliable models and quantitative proof of needed dynamics and causal relationships can be correctly addressed.

In accordance with the cited requirements, these datasets together cover (1) long-term periods (2006-ongoing) that allow for comparisons of different drought events across time (2) and weekly time resolution that is enough to monitor both short-term effects and detect drought displacement waves. (3) They also are collected with information about settlement-level departures that allow for the monitoring of drought impacts at specific locations within a district.

Some caveats and the need for improvements must also be mentioned, however. As stated in the PMT and PRMN methodology report "reports of displacement figures can be seen as indicators of potentially larger movements and their underlying causes[90]. Some types of movement such as short-term displacement of individuals or groups and subsequent returns, may not always be easily identified by the network." The main caveat from a modelling standpoint is that data is not always sampled across all districts in Somalia. For that reason, the time series has many missing data points, creating unavoidable uncertainties.

Data collected in Ethiopia by IOM assesses the number of IDPs and their multisectoral needs across more than 300 sites. The coverage of this dataset, compiled since 2016 and available since December 2017, depends on access and security constraints. The structure and number of fields of information collected by IOM site assessments changed over time. The data collected, however, provides a rich variety of information related to the location of displacement, type of sites, site starting dates, displacement reasons, type of management in the displacement sites, sex- and age-disaggregated data, P-codes, accessibility, ownership of displacement sites, new arrivals to the sites, reasons preventing the return of IDPs, preferred durable solutions, shelter types, access to services, needs of IDPs, access to food, assets damaged by the trigger of displacement and livelihood data (e.g. pastoralism and agro-pastoralism), among other metrics.

The limitations of the dataset are the generalization of the displacement reason at site level, the lack of displacement flow data, and the limited number of rounds of data collection. We highlight this dataset as one of the most comprehensive in terms of the characteristics of IDPs displaced by drought and other drivers.

Supporting the collection of even spatiotemporal sampling

To better understand a specific drought displacement crisis, we need to address the question of what happened to the people (if any) that, faced with the same circumstances, were not forcibly displaced, and we need to do so with the same level of spatiotemporal resolution. Current data collection efforts focus on heavily affected areas of drought displacement. Providing reliable measures of the magnitude of displacement waves has an immense importance. This is where the UNCHR dataset excels. **As counterintuitive as it may seem, however, monitoring areas in which drought impacts are less severe is equally important from the modelling standpoint. Data collection in less affected areas allows for the intercomparison of affected regions and the identification of the specific factors (climatic and structural) that are driving displacement and with what magnitude**. We recommend that data collecting organizations compile information in less sampled districts, when the political climate allows it, to reduce uncertainties and allow for model intercomparison.

Improving the accessibility to flow data (new displacements) and total population counts (IDP stock data)

UNHCR datasets collect flow and the IOM dataset collects stock data on IDPs. From the modelling standpoint, flow data is always desirable. It is critical to have reliable information about arrival and departure dates for a forcibly displaced household and to know which households have returned. Flow data allows one to map the conditions at the time of displacement at the place of origin and contributes to understanding displacement routes. The exact monitoring resulting from flow data addresses possible information errors from interviewees' memory and recall that could come from stock surveys. DMT stock data, however, has micro-level information available which is not available in the PRMN dataset. We will address this point in the following section. The **departure date must be estimated from the arrival date by calculating the average time it takes to travel the shortest distance or the road distance between locations. In order to improve data collection, we suggest that this type of survey include a question dedicated to this matter (ex. what was the departure date of displacement?). This type of data will be very helpful for data scientists with no aggregated cost involved in the data collection.**

As described previously, IDP counts collected by IOM Ethiopia could be considered good practice in the sector. This type of survey data could be complemented with additional questions that could provide information on the livelihood, income, food security and access to water in IDP's areas of origin. This type of information could help in understanding the duration of displacement, triggers of displacement, and the characteristics of the population most affected by droughts. It could be also used to understand the needs of IDPs and propose long-term solutions to their displacement.

Implementing multicausal drivers for displacement reasons

Another caveat is that a multicausal explanation for displacement is not accounted for during PRMN data collection, as only one causal label is identified. Drought displacement does not have a single cause and entangled relationships exist. We propose that this limitation could be solved by simply asking for and collecting primary and secondary reasons for displacement, such as in the proceedings of the site assessment conducted for the IOM DTM dataset in Ethiopia[67]. This increased effort would shed light on the multicausal nature of drought displacement.

Supporting the dissemination of open data among data collectors of drought impacts and displacement datasets

PRMN datasets have been compiled since 2006, but an aggregated dataset of these two PMT and PRMN systems is not openly available. For that reason, we recommend that all data not containing sensitive information compromising the wellbeing of affected populations follow FAIR (findable, accessible, interoperable, reusable) principles[91]. This would accelerate innovation. It would also maximize the research impact for developing tools to better assist affected populations and prevent and reduce future displacements.

This recommendation also applies to certain ad hoc data collection exercises on the socio-economic impacts of drought (e.g. livelihood data, livestock market data and food security). We have found from our review of the literature that this type of data is collected in some countries in sub-Saharan Africa. The surveys' results are often published as reports, however, without readily accessible raw data.

Supporting the data collection of comparable data between countries and organizations

The greatest incremental efforts should be made to collect the best quality data possible. The ideal scenario is to merge the strengths of two data collection methodologies: the high frequency monitoring of displacement flows from UNCHR, and the dedicated questionnaires gathered by DMT about the conditions of internally displaced households. We cannot emphasize enough how valuable it would be to include in-situ, context-specific vulnerabilities affecting households at the time of displacement. At the same time, **we strongly recommend that data collection methodologies present comparable and harmonizable data between countries. Model and empirical evidence intercomparisons could allow for the establishment of common frameworks of data and model evaluation.** The need for such data is probably the main findings of this report. It would permit the achievement of robust modelling methodologies and the accumulation of scientific evidence on displacement responses to drought.

Improving modelling methodologies: ways forward

This report provided an overview of the state-of-the-art of drought displacement modelling, and challenges and limitations to overcome. Drought displacement modelling is a young field. It is heavily underrepresented in drought-related human mobility modelling in which advanced methodologies have been applied only in the recent decade (e.g., ABMs, Bayesian networks and other machine-learning techniques). *Appendix I* presents a summarized overview of key considerations in different data-driven drought mobility research, including how these studies were developed (data and methods) and their intended applications and results.

This report targets donors, policymakers, researchers and operational actors from the humanitarian and development sector that have an interest in developing drought displacement models. The catalogue developed as part of this work could allow for an understanding of the inherent complexities of drought-induced mobility and how different approaches tried to conceptualize drought impacts. We also suggest paying special attention to sections *4. Model Development* and *5. Model Evaluation* of the catalogue of reviewed models so as to be able to evaluate the variety of employed strategies and some of their key weaknesses and strengths.

In general terms, we suggest, that in achieving advances in drought displacement modelling, efforts should be to continue developing models which seek to address the context-specific situation in the field. We provide general suggestions on where efforts could be directed to improve existing methodologies and develop new models in the following section.

Using innovative modelling techniques

Work in machine-learning (ML) could provide a way to advance current modelling methodologies and cope with some of the previously mentioned modelling challenges. For example, tree-based ensembles, or neural networks, are data-driven methods which make no strong assumptions on the relationships between the variables. This can overcome the possible bias introduced by the modeller and strong conceptual assumptions that may not represent the actual dynamics of a system. They are much more flexible than static equations and can combine many input features. As a result, improved performances are expected. Current articles employing state-of-the-art modelling techniques claim that these methodologies can outperform traditional physics-based approaches or include more flexible descriptive analysis of the data[92].

A big disadvantage is that these models can be very complex and impossible to interpret (i.e., they are black boxes), but this limitation is being tackled by a new emerging field called *Explainable AI (XAI)*. Another caveat is that some data-driven methodologies cannot incorporate domain knowledge to constrain the plausible solutions, and model fitting can return results that do not match domain expert expectations. This is a well-known challenge, and it is handled by mitigating the sufficiency assumption and by collecting a wide diversity of potentially involved factors. Modelling efforts should also be concerned with data quality, filtering and harmonization.

We are not advocating for traditional methods to be discarded in favour of these techniques, but for the techniques to complement them. Traditional mobility modelling has proven successful in a variety of contexts. This might not change in the near future. In an operational context in which decision making could have great impact on livelihoods, the implications are crucial. For a successful implementation of a model to predict people's movements between regions, it may be more useful to have models that can be controlled, interpreted, and changed at will than models that, in spite of offering higher predictive power, learned relationships difficult to understand and thereby are not interpretable. **The combination of different approaches could help deliver quality operational products.** These innovative data-driven approaches could be used to learn about the current problem (this is what machine-learning really meant when the term emerged); to compare results between different methodologies to ensure robust results; to see if more complex relationships in the data exist than what was initially considered; and to then try to emulate the more complex functional forms (i.e., representation of the relationships in the data) in one's operational ready models, among other possible applications (including direct applications) in which these techniques could be applied.

It is important to also comment on the recent advances in complex system modelling and their applicability in social sciences that are emerging in the field[93]. These modelling approaches can greatly contribute to understanding mobility responses emerging from the intricate web of relationships. They can do so by explicitly incorporating rules or prior information on how drought conditions lead to forced displacement. All the reviewed methodologies (agent-based, system dynamics and Bayesian networks) have different strengths and weaknesses depending on the application. These approaches can also benefit from advanced machine-learning algorithms. In agent-based and system dynamic models, for example, the strength of the influence between variables (parameters) must be calibrated based on empirical findings. Data shortages are what mainly hamper parameter calibration, but if data exists, machine-learning techniques could be used to calibrate such parameters.

Better assessing of models using explainable AI (XAI) approaches

Data-driven algorithms that can learn very complex relationships can usually deliver better performance. This comes, however, at the expense of interpretability. This is probably the principal reason why opaque models cannot be used for critical decision making. In the final instance, practitioners cannot know how the model arrived at its conclusion. This may make modellers reticent to employ such approaches. XAI is a new field of AI, however, that tries to trace accountability and ensure the trustworthiness of ML models[94]. It allows us to visualize the models' decisions, and to understand (and eventually correct) the why/when/how of model predictions. This is of tremendous importance in terms of trusting models that will have an impact on policy decisions and where ethical issues must be considered. For these reasons, utilizing XAI techniques could contribute in operational settings, such as the humanitarian field and in modelling advances. We recommend keeping an eye on research advances that strive towards reliable implementation based on these techniques[95]. The business sector is already investing in such solutions to target consumers. There are also dedicated projects, such as the EU-funded H2020 XMANAI, that seek to advance XAI techniques[96]. Some of these techniques are still being developed, however, as there is no solid consensus about the meaning and conclusions of some of the employed approaches.

Exploring causal discovery and causal inference techniques

Reports, including those using scientific approaches, have sometimes assumed a link between climate change, conflict and forced migration. Such a chain of cause-effect, however, should be gauged empirically from reliable data employing rigorous and advanced statistical tools and uncertainty assessment. Causal discovery is an emerging field of statistics, and, with the proper assumptions, it allows us to discover the causal relationships driving a system without strong prior assumptions on the data. These techniques could be extremely useful in discovering unknown causal pathways present in drought displacement. They are being employed, for example, in earth sciences to help identify unknown causes and effects of climate processes. They could also be applicable to drought displacement as they present similar challenges from the modelling standpoint (non-linear dynamics, feedback loops, autoregressive effects, etc.)[97].

Causal inference techniques are statistical methods aiming to capture the strengths of the causal relationships driving a system. One of their main applications could be quantifying the causal strengths in order to calibrate complex system models, which are often qualitatively assumed, with rigorous data-driven methodologies. Another application is the ability to perform rigorous counterfactual analysis, which could have greater impact in policy design and crisis prevention[98]. Once a causal graph is correctly verified, it permits the exploration of what-if scenarios and policy intervention impacts without performing the experiments with rigorous uncertainty assessments. The H2020 DeepCube project case 2, *Climate Induced Migration in Africa*, is focused on uncovering the underlying mechanisms of how changing climate conditions cause drought displacement through modern observational causal inference models[99]. We believe these technologies could have a positive impact in operational contexts.

Exploiting available data sources and state-of-the-art drought indexes

Earth observation data guarantees a globally consistent and continuous record of climatic and environmental variables (e.g., soil moisture, vegetation indices, drought indicators) and human processes (night lights, human infrastructure, etc.). It does so at reasonable spatiotemporal resolutions at grid level and in perpetual evolution, with better resolution, coverage and accessibility of information day after day. Advances in the exploitation of these datasets are at the frontier of a better understanding of drought displacement drivers, mainly because spatiotemporal data can be consistently included in the models. Spatially explicit climate models of drought displacement can be subsequently built. Following this line, a wide range of drought indicators were documented in the literature review (e.g., drought indexes, rainfall mean, temperature anomalies, severity of soil degradation, vegetation indexes, etc.), meaning that no unified approach has been established.

Some studies used the Standardized Precipitation-Evapotranspiration Index (SPEI), a well-validated drought index[61]. SPEI makes use of precipitation, soil moisture and temperature and can account more accurately for agricultural failure. Recent advances in vegetation index measurements such as the KNDVI can also be used to more effectively measure the impact of drought on vegetation in areas such as drylands (where there is a need to correct for soil reflectance)[100].

Implementing model evaluation or peer review systems

Models that could support decision-making processes (operational models) are constructed with different goals than research-oriented, experimental ones. An operational model must be submitted to additional tests because of the ethical implications inherent in model failure. This implies that practitioners should validate developed models based on predictive capabilities with out of sample data (unseen data), goodness of fit metrics, sensitivity analysis and robustness analysis. The limitations of such models must be stated clearly to ensure operational readiness. Following this line, **initiatives such as the OCHA model cards of the Peer Review Framework for Predictive Models[1] can be useful for introducing model standards. We recommend that the humanitarian and scientific community dedicate efforts to track, improve and refine these standards depending on specific modelling applications.**

A view of Shahrak-e Sabz informal IDP settlement located in Herat city, Afghanistan. © Maisam Shafiey / NRC, May 2022.

Model access and reusability

The findings in the literature indicate that only two of the reviewed models were openly accessible. This is a major caveat in knowledge transfer and a huge breach in actionable science in the domain. Implementing model evaluation or peer review mechanisms, however, depend on the model's openness and ease of access. Because of the intensive time and costs potentially associated with the data processing and modelling of drought impacts and associated internal displacement, we strongly recommend developing and implementing reusable models. This should not be done, of course, without assessing potential ethical risks, data protection, and other unintended harms to vulnerable populations resulting from these models.

Uncertainties and bias assessment

This, as has been previously stated, may be one of the greatest challenges in the field with regard to data inherent biases and unexpected outcomes. That is because of the nature and lack of knowledge of drought displacement dynamics. Properly validating the models and measuring their robustness using "stress tests" can indicate how reliable these models are. If possible, employing different modelling approaches (based on totally different assumptions) and then comparing the results could reveal how different models are in their predictions and provide an uncertainty measure. Such is the approach employed by Project Jetson. In the end, the model uncertainties are subject to model assumptions. Ultimately there is no other way than advancing in the research to improve current uncertainty estimates. Bias in modelling, on the other hand, can be addressed, by carefully inspecting stated model assumptions and biases in the data and, if possible, correcting for them.

Interdisciplinary and multisectoral collaboration

Understanding and modelling the links between drivers of drought, forced displacement, human mobility and climate change, require interdisciplinary collaborative approaches to understand how different components and dynamics that trigger displacement are interconnected, and to validate the scenarios and results of projections and forecast models. We believe that improvements in displacement modelling are likely to occur if interdisciplinary collaboration between technical experts (ex. data scientists), domain knowledge experts and community experts is fostered.

Collaboration between national and international organizations and funding mechanisms or entities that support long-term data collection, preservation, and dissemination of relevant data, and the development of modelling initiatives, is also required. It is needed to encourage the communication and implementation of good practices, knowledge exchange, the implementation of peer review processes for models and models outcomes. and support efficient allocation and use of funding resources. This is particularly relevant in contexts where models are intended to be used as decision-support tools in operational settings (ex. the humanitarian and development sector or by national disaster risk reduction agencies). Encouraging and fostering multisectoral collaboration (ex. between the development and humanitarian sector) is also needed to fill data and information gaps, to increase the awareness and communication of data collection and good practices in modelling, and to foster innovative data-driven approaches. This could support actions and policymaking processes to better prepare, respond, recover, manage and prevent forced displacement triggered by drought.

We also encourage donors, humanitarian and development organizations, and those in academia to improve their collaboration and synergies in promoting the cataloguing and transparency of models. We encourage the use of initiatives such as OCHA's catalogue of predictive models in the humanitarian sector (to keep a record of models used in operational settings) and the use of the peer review framework of predictive analytics in humanitarian response (for the peer review of models and model outcomes).

Water Distribution for drought-affected IDPs in Baidoa, Somalia. © Will Seal/NRC, March 2022.

Endnotes

1 “Peer Review Framework 2020,” p. 14.

2 UNDRR, “GAR Special Report on Drought 2021,” 2021. Accessed: Aug. 04, 2021. [Online]. Available: https://reliefweb.int/sites/reliefweb.int/files/resources/GAR%20Special%20Report%20on%20Drought%202021.pdf

3 J. Ginnetti and T. Franck, “Assessing drought displacement risk for Kenyan, Ethiopian and Somali pastoralists,” Apr. 2014. Accessed: Aug. 05, 2021. [Online]. Available: https://www.internal-displacement.org/publications/assessing-drought-displacement-risk-for-kenyan-ethiopian-and-somali-pastoralists

4 E. L. Owain and M. A. Maslin, “Assessing the relative contribution of economic, political and environmental factors on past conflict and the displacement of people in East Africa,” *Palgrave Commun.*, vol. 4, no. 1, Art. no. 1, Apr. 2018, doi: 10.1057/s41599-018-0096-6.

5 IDMC, “Global Internal Displacement Database,” *Internal Displacement Monitoring Centre*, 2022. https://www.internal-displacement.org/database/displacement-data (accessed Mar. 09, 2022).

6 “Somalia: Humanitarian Response Plan 2022 (December 2021) - Somalia,” *ReliefWeb*. https://reliefweb.int/report/somalia/somalia-humanitarian-response-plan-2022-december-2021 (accessed Mar. 12, 2022).

7 M. D. Svoboda *et al.*, *Handbook of drought indicators and indices.* 2016.

8 NCDC, “DROUGHT: Monitoring Economic, Environmental, and Social Impacts | National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC),” *National Centers For Environmental Information (NCEI)*, 2022. https://www.ncdc.noaa.gov/news/drought-monitoring-economic-environmental-and-social-impacts (accessed Feb. 19, 2022).

9 “Migration and global environmental change: future challenges and opportunities | United Nations Network on Migration.” https://migrationnetwork.un.org/resources/migration-and-global-environmental-change-future-challenges-and-opportunities (accessed Dec. 25, 2021).

10 Y. Yihdego, B. Vaheddoost, and R. A. Al-Weshah, “Drought indices and indicators revisited,” *Arab. J. Geosci.*, vol. 12, no. 3, p. 69, Jan. 2019, doi: 10.1007/s12517-019-4237-z.

11 A. Mehran, O. Mazdiyasni, and A. AghaKouchak, “A hybrid framework for assessing socioeconomic drought: Linking climate variability, local resilience, and demand,” *J. Geophys. Res. Atmospheres*, vol. 120, no. 15, pp. 7520–7533, 2015, doi: 10.1002/2015JD023147.

12 K. K. Rigaud *et al.*, “Groundswell: Preparing for Internal Climate Migration,” World Bank, Washington, DC, Mar. 2018. doi: 10.1596/29461.

13 “Project Jetson.” https://jetson.unhcr.org/ (accessed Dec. 21, 2021).

14 R. J. Nawrotzki, F. Riosmena, and L. M. Hunter, “Do Rainfall Deficits Predict U.S.-Bound Migration from Rural Mexico? Evidence from the Mexican Census,” *Popul. Res. Policy Rev.*, vol. 32, no. 1, pp. 129–158, Feb. 2013, doi: 10.1007/s11113-012-9251-8.

15 M. Ezra, “Ecological degradation, rural poverty, and migration in Ethiopia: A contextual analysis,” *Poverty Gend. Youth*, Jan. 2001, doi: 10.31899/pgy6.1052.

16 E. L. Nelson, S. A. Khan, S. Thorve, and P. G. Greenough, “Modeling pastoralist movement in response to environmental variables and conflict in Somaliland: Combining agent-based modeling and geospatial data,” *PLOS ONE*, vol. 15, no. 12, p. e0244185, Dec. 2020, doi: 10.1371/journal.pone.0244185.

17 K. Neumann, D. Sietz, H. Hilderink, P. Janssen, M. Kok, and H. van Dijk, “Environmental drivers of human migration in drylands – A spatial picture,” *Appl. Geogr.*, vol. 56, pp. 116–126, Jan. 2015, doi: 10.1016/j.apgeog.2014.11.021.

18 Filiz Garip and Mario Molina, “Predictive Machine Learning Models for Exploring Environment and Migration Linkages in Historical Mexico-U.S. Flows,” *Popul. Assoc. Am. Conf.*, 2019, [Online]. Available: http://paa2019.populationassociation.org/uploads/192918

19 K. Hernandez and T. Roberts, “Predictive analytics in humanitarian action: a preliminary mapping and analysis,” Jun. 2020, Accessed: Mar. 12, 2022. [Online]. Available: https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/15455

20 “Forecasting International Migration in Europe: A Bayesian View | SpringerLink.” https://link.springer.com/book/10.1007/978-90-481-8897-0 (accessed Mar. 12, 2022).

21 R. M. Beyer, J. Schewe, and H. Lotze-Campen, “Gravity models do not explain, and cannot predict, international migration dynamics,” *Humanit. Soc. Sci. Commun.*, vol. 9, no. 1, Art. no. 1, Feb. 2022, doi: 10.1057/s41599-022-01067-x.

22 M. D. Ward, B. D. Greenhill, and K. M. Bakke, “The perils of policy by p-value: Predicting civil conflicts,” *J. Peace Res.*, vol. 47, no. 4, pp. 363–375, Jul. 2010, doi: 10.1177/0022343309356491.

23 F. Gemenne, “Why the numbers don’t add up: A review of estimates and predictions of people displaced by environmental changes,” *Glob. Environ. Change*, vol. 21, pp. S41–S49, Dec. 2011, doi: 10.1016/j.gloenvcha.2011.09.005.

24 Justin Ginnetti and Travis Franck, “Assessing drought displacement risk for Kenyan, Ethiopian and Somali pastoralists.” Koninklijke Brill NV. doi: 10.1163/2210-7975_HRD-9806-2014039.

25 J. Groth, K. Hermans, C. Wiederkehr, E. Kassa, and J. Thober, “Investigating environment-related migration processes in Ethiopia – A participatory Bayesian network,” *Ecosyst. People*, vol. 17, no. 1, pp. 128–147, Jan. 2021, doi: 10.1080/26395916.2021.1895888.

26 C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, “An Introduction to Logistic Regression Analysis and Reporting,” *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, Sep. 2002, doi: 10.1080/00220670209598786.

27 Alessandrini Alfredo, Ghio Daniela, and Migali Silvia, “Heterogeneity in Migration Responses to Climate Shocks: Evidence from Madagascar.” https://www.iza.org/publications/dp/14052/heterogeneity-in-migration-responses-to-climate-shocks-evidence-from-madagascar (accessed Dec. 21, 2021).

28 J. I. Daoud, “Multicollinearity and Regression Analysis,” *J. Phys. Conf. Ser.*, vol. 949, p. 012009, Dec. 2017, doi: 10.1088/1742-6596/949/1/012009.

29 A. DeMaris, “A Tutorial in Logistic Regression,” *J. Marriage Fam.*, vol. 57, no. 4, pp. 956–968, 1995, doi: 10.2307/353415.

30 J. Poot, O. Alimi, M. P. Cameron, and D. C. Maré, “The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2864830, Nov. 2016. doi: 10.2139/ssrn.2864830.

31 M. Shiva and H. Molana, “Climate Change Induced Inter-Province Migration in Iran,” Apr. 2018, Accessed: Dec. 21, 2021. [Online]. Available: https://abdn.pure.elsevier.com/en/publications/climate-change-induced-inter-province-migration-in-iran

32 M. Beine, S. Bertoli, and J. Fernández-Huertas Moraga, “A Practitioners’ Guide to Gravity Models of International Migration,” *World Econ.*, vol. 39, no. 4, pp. 496–512, 2016, doi: 10.1111/twec.12265.

33 S. Bertoli and J. Fernández-Huertas Moraga, “Multilateral resistance to migration,” *J. Dev. Econ.*, vol. 102, pp. 79–100, May 2013, doi: 10.1016/j.jdeveco.2012.12.001.

34 Sibren Isaacman, Vanessa Frias-Martinez, and Enrique Frias-Martinez, “Modeling human migration patterns during drought conditions in La Guajira, Colombia | Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies.” https://dl.acm.org/doi/10.1145/3209811.3209861 (accessed Dec. 21, 2021).

35 F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, Apr. 2012, doi: 10.1038/nature10856.

36 C. Alis, E. F. Legara, and C. Monterola, "Generalized radiation model for human migration," *Sci. Rep.*, vol. 11, no. 1, p. 22707, Nov. 2021, doi: 10.1038/s41598-021-02109-1.

37 Joint Research Centre (European Commission), A. Alessandrini, D. Ghio, and S. Migali, *Population dynamics, climate change and variability in Western Africa: the case of Sahel regions*. LU: Publications Office of the European Union, 2021. Accessed: Dec. 21, 2021. [Online]. Available: https://data.europa.eu/doi/10.2760/797541

38 "Complexity Explorer." https://www.complexityexplorer.org/home (accessed Dec. 26, 2021).

39 J. Thober, N. Schwarz, and K. Hermans, "Agent-based modeling of environment-migration linkages: a review," *Ecol. Soc.*, vol. 23, no. 2, Jun. 2018, doi: 10.5751/ES-10200-230241.

40 T. Sakamoto, "Computational Research on Mobile Pastoralism Using Agent-Based Modeling and Satellite Imagery," *PLoS ONE*, vol. 11, no. 3, p. e0151157, Mar. 2016, doi: 10.1371/journal.pone.0151157.

41 B. Entwisle, A. Verdery, and N. Williams, "Climate Change and Migration: New Insights from a Dynamic Model of Out-Migration and Return Migration," *Am. J. Sociol.*, vol. 125, no. 6, pp. 1469–1512, May 2020, doi: 10.1086/709463.

42 H. Van Dyke Parunak, R. Savit, and R. L. Riolo, "Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and Users' Guide," in *Multi-Agent Systems and Agent-Based Simulation*, Berlin, Heidelberg, 1998, pp. 10–25. doi: 10.1007/10692956_2.

43 "Bayesian Network - an overview | ScienceDirect Topics." https://www.sciencedirect.com/topics/mathematics/bayesian-network (accessed Dec. 27, 2021).

44 J. L. Puga, M. Krzywinski, and N. Altman, "Bayesian networks," *Nat. Methods*, vol. 12, no. 9, Art. no. 9, Sep. 2015, doi: 10.1038/nmeth.3550.

45 L. G. Neuberg, "CAUSALITY: MODELS, REASONING, AND INFERENCE, by Judea Pearl, Cambridge University Press, 2000," *Econom. Theory*, vol. 19, no. 4, pp. 675–685, Aug. 2003, doi: 10.1017/S0266466603004109.

46 L. Drees and S. Liehr, "Using Bayesian belief networks to analyse social-ecological conditions for migration in the Sahel," *Glob. Environ. Change*, vol. 35, pp. 323–339, Nov. 2015, doi: 10.1016/j.gloenvcha.2015.09.003.

47 R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A Comparison of Decision Tree Ensemble Creation Techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 173–180, Jan. 2007, doi: 10.1109/TPAMI.2007.250609.

48 J. Aoga, J. Bae, S. Veljanoska, S. Nijssen, and P. Schaus, "Impact of weather factors on migration intention using machine learning algorithms," *ArXiv201202794 Cs Econ Q-Fin*, Dec. 2020, Accessed: Dec. 21, 2021. [Online]. Available: http://arxiv.org/abs/2012.02794

49 S. Schutte, J. Vestby, J. Carling, and H. Buhaug, "Climatic conditions are weak predictors of asylum migration," *Nat. Commun.*, vol. 12, no. 1, p. 2067, Apr. 2021, doi: 10.1038/s41467-021-22255-4.

50 G. J. Abel, M. Brottrager, J. Crespo Cuaresma, and R. Muttarak, "Climate, conflict and forced migration," *Glob. Environ. Change*, vol. 54, pp. 239–249, Jan. 2019, doi: 10.1016/j.gloenvcha.2018.12.003.

51 S. Adaawen, C. Rademacher-Schulz, B. Schraven, and N. Segadlo, "Chapter 2 - Drought, migration, and conflict in sub-Saharan Africa: what are the links and policy options?," in *Current Directions in Water Scarcity Research*, vol. 2, E. Mapedza, D. Tsegai, M. Bruntrup, and R. Mcleman, Eds. Elsevier, 2019, pp. 15–31. doi: 10.1016/B978-0-12-814820-4.00002-X.

52 A. Gerhardus and J. Runge, "LPCMCI: Causal Discovery in Time Series with Latent Confounders," pp. EGU21-8259, Apr. 2021, doi: 10.5194/egusphere-egu21-8259.

53 H. Lyu, Z. Dong, M. Roobavannan, J. Kandasamy, and S. Pande, "Rural unemployment pushes migrants to urban areas in Jiangsu Province, China," *Palgrave Commun.*, vol. 5, no. 1, Art. no. 1, Aug. 2019, doi: 10.1057/s41599-019-0302-1.

54 R. J. Nawrotzki, J. DeWaard, M. Bakhtsiyarava, and J. T. Ha, "Climate shocks and rural-urban migration in Mexico: exploring nonlinearities and thresholds," *Clim. Change*, vol. 140, no. 2, pp. 243–258, Jan. 2017, doi: 10.1007/s10584-016-1849-0.

55 "EM-DAT | The international disasters database." https://www.emdat.be/ (accessed Mar. 14, 2022).

56 Joint Research Centre (European Commission), A. Alessandrini, D. Ghio, and S. Migali, *Estimating net migration at high spatial resolution*. LU: Publications Office of the European Union, 2020. Accessed: Dec. 25, 2021. [Online]. Available: https://data.europa.eu/doi/10.2760/383386

57 D. J. Kaczan and J. Orgill-Meyer, "The impact of climate change on migration: a synthesis of recent empirical insights," *Clim. Change*, vol. 158, no. 3, pp. 281–300, Feb. 2020, doi: 10.1007/s10584-019-02560-0.

58 "Linking Climate Change, Habitability and Social Tipping Points: Scenarios for Climate Migration - HABITABLE." https://habitableproject.org/ (accessed Dec. 27, 2021).

59 C. D. Smith, "Modelling migration futures: development and testing of the Rainfalls Agent-Based Migration Model – Tanzania," *Clim. Dev.*, vol. 6, no. 1, pp. 77–91, Jan. 2014, doi: 10.1080/17565529.2013.872593.

60 C. Gray and R. Bilsborrow, "Environmental Influences on Human Migration in Rural Ecuador," *Demography*, vol. 50, no. 4, pp. 1217–1241, Aug. 2013, doi: 10.1007/s13524-012-0192-y.

61 M. Mastrorillo, R. Licker, P. Bohra-Mishra, G. Fagiolo, L. D. Estes, and M. Oppenheimer, "The influence of climate variability on internal migration flows in South Africa," *Glob. Environ. Change*, vol. 39, pp. 155–169, Jul. 2016, doi: 10.1016/j.gloenvcha.2016.04.014.

62 C. Gray and E. Wise, "Country-specific effects of climate variability on human migration," *Clim. Change*, vol. 135, no. 3, pp. 555–568, Apr. 2016, doi: 10.1007/s10584-015-1592-y.

63 G. Peri and A. Sasahara, "The Impact of Global Warming on Rural-Urban Migrations: Evidence from Global Big Data," National Bureau of Economic Research, Working Paper 25728, Apr. 2019. doi: 10.3386/w25728.

64 C. Gray and V. Mueller, "DROUGHT AND POPULATION MOBILITY IN RURAL ETHIOPIA," *World Dev.*, vol. 40, no. 1, pp. 134–145, Jan. 2012, doi: 10.1016/j.worlddev.2011.05.023.

65 V. Koubi, G. Spilker, L. Schaffer, and T. Bernauer, "Environmental Stressors and Migration: Evidence from Vietnam," *World Dev.*, vol. 79, pp. 197–210, Mar. 2016, doi: 10.1016/j.worlddev.2015.11.016.

66 V. Mueller, C. Gray, and K. Kosec, "Heat stress increases long-term human migration in rural Pakistan," *Nat. Clim. Change*, vol. 4, no. 3, pp. 182–185, Mar. 2014, doi: 10.1038/nclimate2103.

67 "Ethiopia | DTM." https://dtm.iom.int/ethiopia (accessed Dec. 23, 2021).

68 M. Beine and C. Parsons, "Climatic Factors as Determinants of International Migration," *Scand. J. Econ.*, vol. 117, no. 2, pp. 723–767, 2015, doi: 10.1111/sjoe.12098.

69 "Gridded Population of the World (GPW), v4 | SEDAC." https://sedac.ciesin.columbia.edu/data/collection/gpw-v4 (accessed Jan. 17, 2022).

70 A. de Sherbinin *et al.*, "Global Estimated Net Migration Grids By Decade: 1970-2000." Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2015. doi: 10.7927/H4319SVC.

71 "World Urbanization Prospects – Population Division – United Nations." https://population.un.org/wup/ (accessed Dec. 25, 2021).

72 S. Barrios, L. Bertinelli, and E. Strobl, "Climatic change and rural–urban migration: The case of sub-Saharan Africa," *J. Urban Econ.*, vol. 60, no. 3, pp. 357–371, Nov. 2006, doi: 10.1016/j.jue.2006.04.005.

73 "UNHCR Somalia – Interactive Internal Displacements Visualisation," *UNHCR Somalia Internal Displacement*. https://unhcr.github.io/dataviz-somalia-prmn/index.html (accessed Dec. 23, 2021).

74 European Union and United Nations, *Technical report on statistics of internally displaced persons: current practice and recommendations for improvement*, March 2018. Luxembourg: Publications Office of the European Union, 2018. doi: 10.2785/414331.

75 U. N. H. C. for Refugees, "UNHCR – Refugee Statistics," *UNHCR*, 2022. https://www.unhcr.org/refugee-statistics/ (accessed Mar. 09, 2022).

76 "Key Migration Terms, Migration Glossary | IOM, UN Migration." https://www.iom.int/key-migration-terms (accessed Mar. 14, 2022).

77 "Data disaggregation of SDG indicators by forced displacement," *Migration data portal*. https://www.migrationdataportal.org/resource/data-disaggregation-sdg-indicators-forced-displacement (accessed Mar. 14, 2022).

78 U. N. H. C. for Refugees, "Guiding Principles on Internal Displacement," *UNHCR*. https://www.unhcr.org/protection/idps/43ce1cff2/guiding-principles-internal-displacement.html (accessed Mar. 14, 2022).

79 "UNHCR Global Trends – Forced displacement in 2020." https://www.unhcr.org/flagship-reports/globaltrends/ (accessed Mar. 14, 2022).

80 "AR6 Climate Change 2021: The Physical Science Basis — IPCC." https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/ (accessed Mar. 14, 2022).

81 "International Recommendations on Internally Displaced Persons Statistics (IRIS)." https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-20-005 (accessed Mar. 14, 2022).

82 I. Dallmann and K. Millock, "Climate Variability and Inter-State Migration in India," *CESifo Econ. Stud.*, vol. 63, no. 4, pp. 560–594, Dec. 2017, doi: 10.1093/cesifo/ifx014.

83 C. Cattaneo and E. Massetti, "Migration and Climate Change in Rural Africa," Fondazione Eni Enrico Mattei (FEEM), 2015. Accessed: Dec. 21, 2021. [Online]. Available: https://www.jstor.org/stable/resrep01147

84 S. Henry, P. Boyle, and E. F. Lambin, "Modelling inter-provincial migration in Burkina Faso, West Africa: the role of socio-demographic and environmental factors," *Appl. Geogr.*, vol. 23, no. 2, pp. 115–136, Apr. 2003, doi: 10.1016/j.apgeog.2002.08.001.

85 A. J. Garcia, D. K. Pindolia, K. K. Lopiano, and A. J. Tatem, "Modeling internal migration flows in sub-Saharan Africa using census microdata," *Migr. Stud.*, vol. 3, no. 1, pp. 89–110, Mar. 2015, doi: 10.1093/migration/mnu036.

86 M. Berlemann and T. X. Tran, "Climate-Related Hazards and Internal Migration Empirical Evidence for Rural Vietnam," *Econ. Disasters Clim. Change*, vol. 4, no. 2, pp. 385–409, Jul. 2020, doi: 10.1007/s41885-020-00062-3.

87 B. Hassani-Mahmooei and B. W. Parris, "Climate change and internal migration patterns in Bangladesh: an agent-based model," *Environ. Dev. Econ.*, vol. 17, no. 6, pp. 763–780, Dec. 2012, doi: 10.1017/S1355770X12000290.

88 M. Call and C. Gray, "Climate anomalies, land degradation, and rural out-migration in Uganda," *Popul. Environ.*, vol. 41, no. 4, pp. 507–528, Jun. 2020, doi: 10.1007/s11111-020-00349-3.

89 UNHCR, "Protection & Return Monitoring Network (PRMN) - Notes on methodology - UNHCR Somalia." 2017. Accessed: Aug. 08, 2021. [Online]. Available: https://data2.unhcr.org/en/documents/details/53888

90 "Document - Protection & Return Monitoring Network (PRMN) - Notes on methodology - UNHCR Somalia." https://data2.unhcr.org/en/documents/details/53888 (accessed Dec. 27, 2021).

91 S. Hagstrom, "The FAIR Data Principles," *FORCE11*, Sep. 03, 2014. https://www.force11.org/group/fairgroup/fairprinciples (accessed Dec. 25, 2021).

92 C. Robinson and B. Dilkina, "A Machine Learning Approach to Modeling Human Migration," in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, New York, NY, USA, Jun. 2018, pp. 1–8. doi: 10.1145/3209811.3209868.

93 "Complex systems modeling for humanitarian action – The Centre for Humanitarian Data." https://centre.humdata.org/complex-systems-modeling-for-humanitarian-action/ (accessed Dec. 25, 2021).

94 A. Holzinger, "From Machine Learning to Explainable AI," in *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, Aug. 2018, pp. 55–66. doi: 10.1109/DISA.2018.8490530.

95 N. Savage, "Breaking into the black box of artificial intelligence," *Nature*, Mar. 2022, doi: 10.1038/d41586-022-00858-1.

96 "Explainable Manufacturing Artificial Intelligence | XMANAI Project | Fact Sheet | H2020," *CORDIS | European Commission*. https://cordis.europa.eu/project/id/957362 (accessed Mar. 31, 2022).

97 P. Nowack, J. Runge, V. Eyring, and J. D. Haigh, "Causal networks for climate model evaluation and constrained projections," *Nat. Commun.*, vol. 11, no. 1, p. 1415, Mar. 2020, doi: 10.1038/s41467-020-15195-y.

98 I. Boas *et al.*, "Climate migration myths," *Nat. Clim. Change*, vol. 9, no. 12, pp. 901–903, Dec. 2019, doi: 10.1038/s41558-019-0633-3.

99 "Climate induced migration in Africa," *DeepCube H2020*. https://deepcube-h2020.eu/use-cases/climate-induced-migration-in-africa/ (accessed Dec. 28, 2021).

100 G. Camps-Valls *et al.*, "A unified vegetation index for quantifying the terrestrial biosphere," *Sci. Adv.*, Feb. 2021, doi: 10.1126/sciadv.abc7447.

Annexes

Click here to download all annexes tables

Every day, people flee conflict and disasters and become displaced inside their own countries. IDMC provides data and analysis and supports partners to identify and implement solutions to internal displacement.

Join us as we work to make real and lasting change for internally displaced people in the decade ahead.

iDMC internal displacement monitoring centre

The Internal Displacement Monitoring Centre
La Voie-Creuse 16, 1202 Geneva, Switzerland
+41 22 552 3600 | info@idmc.ch

internal-displacement.org
twitter.com/IDMC_Geneva
facebook.com/IDMC.Geneva
youtube.com/c/InternalDisplacementMonitoringCentreIDMC
linkedin.com/company/idmc-geneva