



Complex climate and network effects on internal migration in South Africa revealed by a network model

Tingyin Xiao¹ · Michael Oppenheimer^{1,2,6} · Xiaogang He^{3,4} · Marina Mastrorillo⁵

Accepted: 30 August 2021 / Published online: 6 January 2022
© The Author(s) 2021

Abstract

Climate variability and climate change influence human migration both directly and indirectly through a variety of channels that are controlled by individual and household socioeconomic, cultural, and psychological processes as well as public policies and network effects. Characterizing and predicting migration flows are thus extremely complex and challenging. Among the quantitative methods available for predicting such flows is the widely used gravity model that ignores the network autocorrelation among flows and thus may lead to biased estimation of the climate effects of interest. In this study, we use a network model, the additive and multiplicative effects model for network (AMEN), to investigate the effects of climate variability, migrant networks, and their interactions on South African internal migration. Our results indicate that prior migrant networks have a significant influence on migration and can modify the association between climate variability and migration flows. We also reveal an otherwise obscure difference in responses to these effects between migrants moving to urban and non-urban destinations. With different metrics, we discover diverse drought effects on these migrants; for example, the negative standardized precipitation index (SPI) with a timescale of 12 months affects the non-urban-oriented migrants' destination choices more than the rainy season rainfall deficit or soil moisture do. Moreover, we find that socioeconomic factors such as the unemployment rate are more significant to urban-oriented migrants, while some unobserved factors, possibly including the abolition of apartheid policies, appear to be more important to non-urban-oriented migrants.

Keywords Climate impact · Human migration · Network model · Bayesian · Drought · Migrant network

✉ Tingyin Xiao
tingyin.xiao@princeton.edu

Extended author information available on the last page of the article

Introduction

Studies using diverse methods have established that environmental factors, including climate variability and change, affect patterns of human migration both directly and indirectly (McLeman & Gemenne, 2018). Projected twenty-first-century climate change includes increased frequency and intensity of temperature extremes, increases in both wet and dry regional extremes of the hydrological cycle, increased tropical cyclone intensity, and sea level rise (IPCC, 2014, 2019a, b), each of which would trigger or enhance natural hazards that may lead to population displacement and migration (Bohra-Mishra et al., 2014, 2017; Cai et al., 2016; Feng et al., 2010; Gray & Mueller, 2012; Hauer, 2017; IPCC, 2019b; Mueller et al., 2014). In addition to episodic and often temporary responses, people migrate as a planned adaptation strategy against environmental risks (Black et al., 2011; Hunter et al., 2015). Human migration and other forms of mobility can influence the environment, economics, and epidemiology of sending and receiving locations. Understanding dynamic interactions of natural and societal systems that affect mobility is critical to estimating a range of measures of future migration.

Internal migration is the more likely choice compared to international migration in response to climate variation, according to one study, and total internal migration due to climate variability and change is projected to exceed 86 million by 2050 for sub-Saharan Africa (Rigaud et al., 2018). As previous studies found that the effects of climate variability on human migration are not easily generalized (Beine & Jeusette, 2018), even if only in sub-Saharan African countries (Gray & Wise, 2016; Mueller et al., 2020), in this study, we only focus on one country, South Africa, to investigate in detail the mechanisms through which the climate influence is exerted in a particular set of circumstances. South Africa is predicted to experience increases in temperature extremes and shifts in precipitation patterns, as well as the exacerbation of its already severe water scarcity (Liu et al., 2017; Mekonnen & Hoekstra, 2016). Coupled with population growth, its population exposure to drought is predicted to increase drastically under climate change (Supplementary Tables 9–12 in (Smirnov et al., 2016)). Besides, South Africa has a high rate of internal migration, with 12% (13.5%) of its population having moved internally during 1996–2001 (2006–2011), about 77.8% (83.5%) of which had an urban destination. The high migration rate and increasing inflow to urban areas greatly influence the population growth, resource distribution, and public health for individual districts and pose challenges for planning, especially under future climate change.

To analyze mechanisms controlling inter-district migration flows in South Africa and predict migration flows with higher accuracy, our study introduces a Bayesian model for networks that resolves a major limitation of the widely used gravity models—the assumption that network flows (e.g., human migration flows) are mutually independent (Tiefelsdorf, 2003). Assuming network independence may lead to estimation biases in both origin and destination effects. Common interdependent migration flow patterns include established bilateral migration

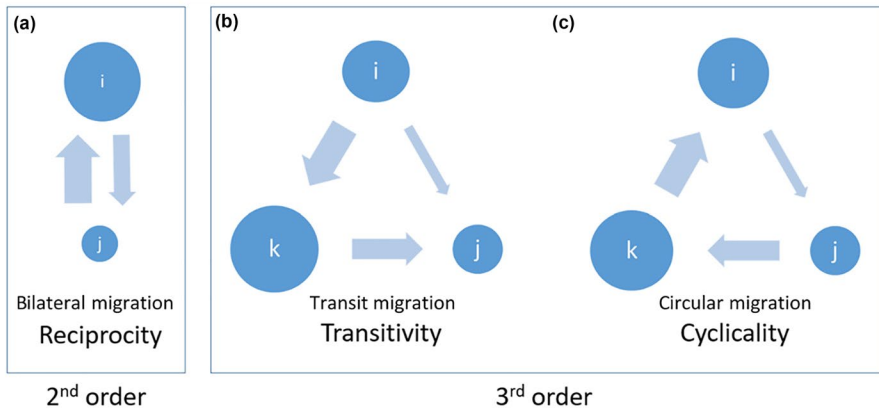


Fig. 1 Common migration patterns and corresponding network autocorrelations. **a** Second-order network autocorrelation (reciprocity)—bilateral migration; **b** third-order network autocorrelation (transitivity)—transit migration; **c** third-order network autocorrelation (cyclicity)—circular migration

corridors between two places (Fig. 1a), like between Mexico and the USA; transit migration corridors where some people make an intermediate stop before arriving at their desired destination (Fig. 1b), for instance, migrating from African countries to European countries and taking a third country (e.g., Libya) as an intermediate destination; and circular migration corridors where some transit migrants return home (Fig. 1c). These common migration patterns correspond to second-order network dependencies of reciprocity and third-order network dependencies of transitivity and cyclicity (see Appendix B for more details).

Moreover, migration, which is often modeled as a spatial problem, has both geographic and social network characteristics (Barbosa et al., 2018; Rozenblat & Melançon, 2013). For example, the social closeness between locations imbedded in their historical activities such as prior migration and trade, or in their similarities in social characteristics such as culture and ideology, may influence migration decisions like whether and where to move. Such closeness or distance is not defined in physical space but in an unobserved conceptual space, i.e., latent space (Barnett, 2012; Hoff et al., 2002). The gravity models and other spatial interaction models used in previous migration flow research only define the distance on physical space, though some included socioeconomic drivers (Garcia et al., 2015). Other studies included limited measurements, such as linguistic proximity, migration agreements, and previous colonial ties between nations (primarily using dummy variables) that can approximate social closeness (Beine et al., 2011; Coniglio & Pesce, 2015). Nevertheless, factors contributing to the latent distance between locations are difficult, if not impossible, to represent adequately for migration analysis.

However, conceptualizing migration as a network problem, we can apply social network analysis techniques, for which modeling of social structures and network dependencies are major objectives. As quoted from a comprehensive review of human mobility modeling (Barbosa et al., 2018), “The field of Social Network research is a huge enterprise in its own right and elucidating the connection between social networks and human movement is a nascent effort that should be explored

in more depth.” Nevertheless, social network analysis techniques are not often used in migration flow modeling, even less in those that involve any climate variables (Desmarais & Cranmer, 2012), though the influence of social networks on human migration decision-making is well recognized and documented (Beine et al., 2011; Carrico & Donato, 2019; Coniglio & Pesce, 2015; Curran & Rivero-Fuentes, 2003; Fussell & Massey, 2004; Fussell et al., 2014; Mahajan & Yang, 2020; Massey et al., 1993; Nawrotzki et al., 2015). Most of these previous studies consider a prior migrant network as an influence on subsequent migration. However, a preexisting migrant network is only one type of network that can impact migration, and omitted variable biases may arise if we cannot include a range of potentially influential networks (e.g., family ties and trade networks) in the model.

Numerous studies use agent-based models (ABMs) to investigate simultaneous environmental and social network impacts on, or interactions with, human migration (Bell et al., 2019, 2021; Entwisle et al., 2016; Klabunde & Willekens, 2016; McAlpine et al., 2021; Thober et al., 2018). ABMs use knowledge specific to agents in a particular society to reconstruct at the micro-level how these agents behave, interact with each other (e.g., by forming social networks), and interact with the environment. Such models then simulate an artificial society, subjecting it to social experiments that we cannot do in the real world, and dynamically derive collective behavioral outcomes at the macro-level. These are essentially mechanistic models where many of the mechanisms are not well known a priori. A major limitation to building useful ABMs is the lack of data on the actual behavior of the agents, which renders assigning rules of the behavior (i.e., mechanisms) for the agents in the models difficult.

Here we introduce an additive and multiplicative effects network model (AMEN; Hoff, 2021), an empirical model that can reveal the macro-level relationship between variables and network properties from the data. AMEN is suitable for both mechanisms analysis and prediction of migration. It illustrates the advantage of controlling for network effects in that we obtain unbiased estimations of the climate effect without the need to explicitly specify a potentially inexhaustible number of network properties (model details in the “[The network model](#)” section). The AMEN model can relax the strong assumption of flow independence by allowing up to third-order dependencies (e.g., transitivity) in a network using a set of latent factors. The latent factors allow us to control for unobserved variables and the effects that are difficult to quantify in the model, such as policy effects, like the impacts of the abolition of apartheid ([Migrant network effects and their interactions with climate variability](#)). As a result, a model trained with historical data should better simulate and predict future migration flows with limited information. Furthermore, we compare AMEN with a traditional gravity model as well as with a gravity model adding the Moran eigenvector spatial filter (MESF) (Chun, 2008; Chun & Griffith, 2011, 2013; Chun et al., 2016), which also addresses the network autocorrelation issue, and found AMEN outperforms both of them in migration flow prediction (MESF details in Appendix D; model comparison in the “[Model comparison](#)” section and Appendix E). In Appendix C, we also discuss other applicable network analysis approaches, such as a dynamic latent space model (Ward et al., 2013) and the weighted exponential random graph models (ERGMs) (Desmarais & Cranmer, 2012; Krivitsky, 2012;

Wilson et al., 2017), to model migration flows and provide our reasons for choosing AMEN.

Having established the modeling framework, we investigate how South African internal migration responds to climate variability, measured by temperature anomaly and water availability conditions. A subset of our results can be compared to a previous study using gravity modeling to examine climate variability impacts on South African internal migration (Mastrorillo et al., 2016). This previous analysis centered largely on “push effects,” factors that drive people to move away from a location, with “pull effects,” factors that attract potential migrants, receiving less attention. However, in our study, we include both the push and pull effects and interpret them as origin and destination effects, as the characteristics of sending nodes can drive away or retain residents, while those of receiving nodes can attract or discourage migrants. Furthermore, we also hypothesize that different metrics of a general physical factor, such as drought, may have distinct effects on human migration and thus include several drought measurements. We separately characterize migration destinations as urban or non-urban to answer whether people moving into urban areas respond to climate variables differently from migrants moving to non-urban locations.

We also examine migrant network effects and their interaction with climate variables to connect this study to the body of scholarship that investigates the effect of prior migrant networks on migration (Beine et al., 2011; Coniglio & Pesce, 2015; Mahajan & Yang, 2020). Prior migration can sustain future migration by helping lower moving costs and risks (Massey et al., 1993), but this cumulative causation effect is not universal (Fussell & Massey, 2004). Interacting with the environmental variables, Nawrotzki et al. (2015) use household- and community-level data in rural Mexico with multi-level discrete-time event-history models and reveal a “suppression” mechanism where social networks can enhance climate adaptation at origin locations by sharing information and financial resources and thus suppressing the climate migration association. Applying a similar type of analysis in Bangladesh, Carrico and Donato (2019) find little evidence of social network impact on migration after extreme weather events. On the other hand, Mahajan and Yang (2020) use country-level data and find that a larger preexisting United States immigrant stocks from a country amplify the push effect of hurricanes in these countries in sending immigrants to the USA. We test these effects in South Africa by investigating whether prior migrant networks affect the migration flows at the district-dyad level and whether they amplify or suppress the climate effects on the migration flows. At an aggregated population level, previous studies usually use the past migration flow (or migrant stocks) to proxy the migrant network. To demonstrate that prior migration flows from origin to destination in these studies are not complete representations of migrant networks, we also examine the corresponding effects of prior migration counterflows from destination to origin. Moreover, our results indicate that there are still other unobserved contributing variables besides migrant networks, which emphasizes the advantage of the AMEN model in controlling unknown variables even when we have data to represent some of the network factors.

Materials and methods

Data

Migration flows and socioeconomic variables

We generate the migration flows and socioeconomic variables from 1996, 2001, and 2011 census and 2007 community surveys of South Africa, provided by Statistics South Africa (<http://www.statssa.gov.za/>) and distributed by DataFirst (<https://www.datafirst.uct.ac.za/>). The original data are on the individual level. As the district boundaries change over time, to deal with the modifiable areal unit problem (MAUP), we assign each individual in each of the census and community surveys to a district based on the 2011 census map and then aggregate the socioeconomic variables at the district level and migration flows at the district-dyad level for each time interval. We also combine the Buffalo City Metropolitan Municipality (BUF) into the Amathole District Municipality (DC12) because BUF is only demarcated from DC12 in the 2011 map. Therefore, our data are unified on the 2011 census district boundaries for 51 districts in total (Fig. A.1).

Definitions of migration flows and socioeconomic variables are similar to those from the previous study (Mastrorillo et al., 2016), but we improve some of the variable definitions and add new variables in this study. Specifically, the definition of the migration flow, the dependent variable, is the number of adults (age 15–64) moving from one district to another during the past five years, including the census year. Here we use 2001 and 2011 census data to calculate two migration episodes: 1997–2001 and 2007–2011. In addition, we break down the migrants into two subgroups by land-use types of the receiving locations. The “urban” group includes the migrants whose current usual residence areas are classified as urban, and the “non-urban” group includes the rest of the migrants whose current residence areas are classified as rural, tribal, or other types. The descriptive statistics of the migration flow data and covariates are shown in Table 1.

All socioeconomic nodal control variables are measured at the previous census or community survey and thus are lagged covariates. In the following tables, the variable “population” is the natural logarithm of the total population of each district; “unemployed” is the unemployment rate of each district (population ratio of unemployed people among all working-age people, in 15–64 age range); “primary” is the population ratio of those who have finished primary school (7th grade based on South Africa education system) at most, among all people who are 25 or older; “White” is the population ratio of people whose self-identified race is White. The dyadic control variables used in the models include “distance” and “contiguity.” The “distance” is the natural logarithm of geographical distance between the sending and receiving districts of the migration flow dyads, and “contiguity” is a binary variable that indicates whether the origin and destination districts share a common border or not (equal to 1 if they share a common border, 0 if not). The two variables

Table 1 Descriptive statistics of the variables

Descriptive statistics	Migration flows (all)	Migration flows (urban)	Migration flows (non-urban)	Logarithm of population	Unemployed population ratio	At most finished primary education population ratio	White population ratio
Sample size (<i>n</i>)	5100	5100	5100	102	102	102	102
Sum	3,799,167.00	2,916,957.00	726,992.00	1361.17	37.31	48.76	9.18
Mean	744.93	571.95	142.55	13.34	0.37	0.48	0.09
Standard deviation	2276.63	2029.07	405.54	0.84	0.12	0.13	0.07
Minimum	0.00	0.00	0.00	10.86	0.10	0.20	0.00
Lower quartile	37.00	20.00	0.00	12.91	0.30	0.39	0.03
Median	131.00	79.00	25.00	13.45	0.36	0.49	0.08
Upper quartile	492.25	325.25	94.00	13.83	0.43	0.57	0.14
Maximum	53,902.00	48,572.00	10,492.00	15.17	0.70	0.78	0.28
Interquartile range	455.25	305.25	94.00	0.92	0.14	0.19	0.11
Descriptive statistics	Positive maximum temperature anomaly	Negative precipitation anomaly	Soil moisture percentile	Positive SPI	Negative SPI	Logarithm of prior migrant network (urban)	Logarithm of prior migrant network (non-urban)
Sample size (<i>n</i>)	102	102	102	102	102	2550	2550
Sum	14.60	-14.40	5164.40	35.94	-8.69	10,327.67	7224.68
Mean	0.14	-0.14	50.63	0.35	-0.09	4.05	2.83
Standard deviation	0.24	0.21	8.15	0.38	0.18	2.88	2.68
Minimum	0.00	-0.66	33.84	0.00	-0.98	0.00	0.00
Lower quartile	0.00	-0.28	44.57	0.00	-0.07	0.00	0.00
Median	0.00	0.00	50.43	0.24	0.00	4.47	3.09
Upper quartile	0.26	0.00	56.09	0.68	0.00	6.16	4.91
Maximum	1.08	0.00	65.79	1.25	0.00	11.55	10.65
Interquartile range	0.26	0.28	11.52	0.68	0.07	6.16	4.91

approximating preexisting migrant networks, past migration flow (*mij.past*) and counterflow (*mji.past*), are also dyadic. Using 1996 census data, “*mij.past*” is defined as the logarithm of the number of migrants who moved from district *i* to *j* during 1900–1996 and reside in district *j* at the time of the census (1996); similarly, “*mji.past*” is defined as the logarithm of the number of migrants who moved from district *j* to *i* during 1900–1996 and reside in district *i* in 1996.

Climate variables

As for the climate variables, we calculate a 12-month standardized precipitation index (SPI) (McKee et al., 1993), in addition to the positive maximum temperature anomaly, negative precipitation anomaly, and soil moisture percentile that were also used in (Mastrorillo et al., 2016). The negative SPI describes meteorological drought, which may influence the level of reservoirs and streamflow at the time-scale of 12 months. To calculate the 12-month SPI, we first fit the 12-month cumulative precipitation with a Gamma distribution at each grid cell (0.25 deg) based on the monthly precipitation data obtained from the most updated version of Princeton Global Forcing data (PGF) (He et al., 2020; Sheffield et al., 2006). Then, we transform the cumulative probability of the fitted gamma distribution to a standard normal distribution. With such transformation, positive/negative SPI reflects precipitation surplus/deficit compared to the long-term climatology (1950–2016). Thus, we input the positive SPI (“pos SPI”) and the negative SPI (“neg SPI”) into the models with dummy variables to evaluate the effect of rainfall surplus and deficit, respectively.

To obtain the positive maximum temperature anomalies (“pos Tmax anom”), we first calculate the mean maximum temperature over the 5-year periods before the time of the census (i.e., 1996–2000, 2006–2010), then subtract the long-run average and divide by long-run standard deviation of maximum temperature (long-run reference period 1950–2011), and last replace the negative values with zeros to enable the evaluation of the effect of heat excesses. The negative precipitation anomalies (“neg Precip anom”) are calculated similarly: the 5-year average of precipitation in the rainy season of the district is calculated first, then the anomalies are calculated by subtracting the long-run average and dividing by the long-run standard deviation, and, finally, positive values are replaced by zeros. Note that this is different from (Mastrorillo et al., 2016), in which the negative values are then replaced with their absolute values for the negative precipitation anomalies. The soil moisture percentile (“soil moisture”), which is measured in percentage on a 0–100 scale, is the 5-year average of relative soil moisture of the top layer (0–10 cm) computed dynamically in the land surface model (Sheffield et al., 2004).

The time intervals and temporal configuration of the migration flows, climate variables, and lagged demographic and socioeconomic variables are illustrated in Appendix Fig. F.1. The data are deposited in the Harvard Dataverse repository, along with the code used to generate the results in this study (Xiao et al., 2021) (link: <https://doi.org/10.7910/DVN/NSKZ45>).

Methods

Migration flow data as a network

The dependent variable we model is the 5-year aggregated migration flow of adults between South Africa districts for two time intervals: 1997–2001 and 2007–2011 (Fig. 2). In this study, we take the migration flow data as a dynamic, weighted, and directed network, in which the nodes are the districts in South Africa and the edges are the dyadic migration flows directed from origin to destination, weighted by the number of migrants, and changing over time (network analysis terminology in Appendix B.1). There are flow interdependencies in the migration data used in this study (details see Appendix B.2), which confirms the need to account for network autocorrelation in the model. The weights of the network (number of migrants) are count data (non-negative integers) that usually follow a Poisson distribution with overdispersion.

The network model

We construct dynamic overdispersed Poisson AMEN models for migration flow networks considering the network autocorrelation and migration flow network properties mentioned above. To build and enable the estimation of the model, we consult the package website (<https://pdhoff.github.io/amen/>) and modify the source code of the AMEN R package (Hoff et al., 2018). The model is constructed as follows,

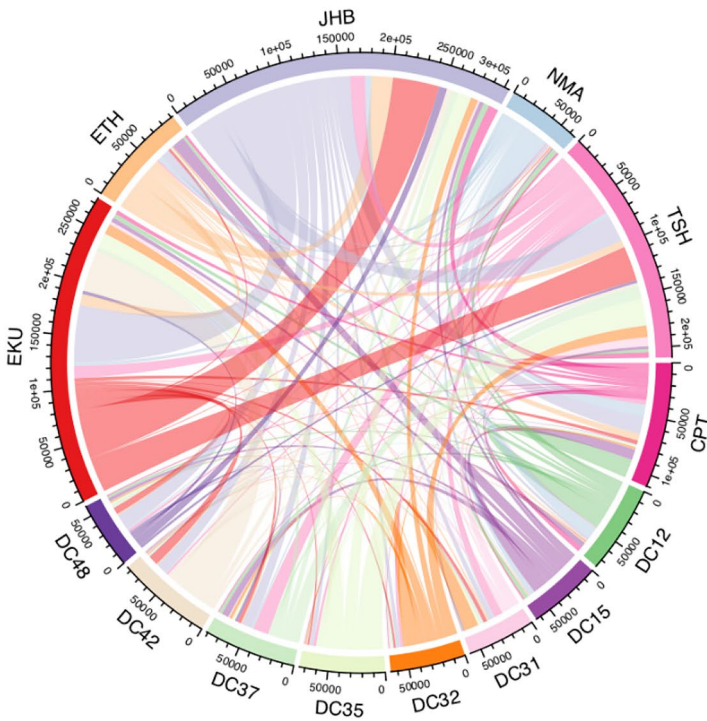
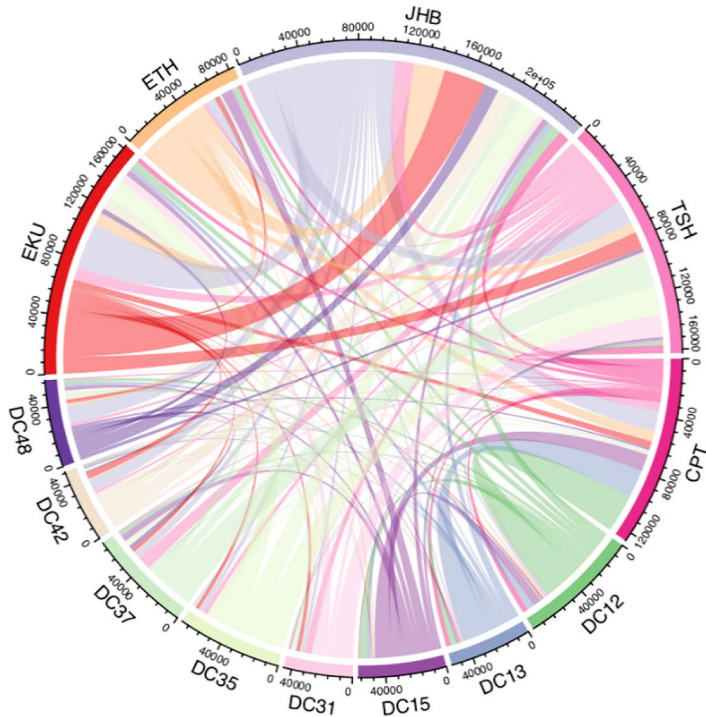
$$z_{i,j,t} = y_t + \beta_{dyad}^T x_{dyad,i,j,t} + \beta_o^T x_{o,i,t} + \beta_d^T x_{d,j,t} + a_i + b_j + u_i^T v_j + \epsilon_{i,j} \quad (1)$$

$$(m_{i,j,t} | z_{i,j,t}) \sim \text{Poisson}(e^{z_{i,j,t}}) \quad (2)$$

where $m_{i,j,t}$ is the migration flow from district i to j during time interval t . $z_{i,j,t}$ is the (i,j) th element of the unobserved latent sociomatrix (Z_t). y_t is the time fixed effect. $x_{dyad,i,j,t}$ is a vector of dyadic covariates representing the characteristics on the dyad $\{i,j\}$, including distance, contiguity, the proxies for preexisting migrant networks, and the interaction terms; $x_{o,i,t}$ ($x_{d,j,t}$) is a vector of origin (destination) nodal covariates representing the characteristics of origin (destination) district i (j). The origin and destination nodal covariates correspond to push and pull factors in the gravity model, respectively. They include lagged demographic and socioeconomic control variables, as well as measurements of climate variability.

The additive effect terms a_i, b_j in Eq. (1) are the random effects demonstrating the heterogeneity of the latent sociomatrix of the origin and destination node, respectively. In the multiplicative effect terms, $u_i^T v_j$, u_i , and v_j are vectors of unobserved latent factors describing district i and j 's characteristics as an origin and a destination, respectively; $\epsilon_{i,j}$ is the error term.

The covariance expressions are as follows:



◀**Fig. 2** South Africa internal migration between districts during the 5 years before (left) 2001 and (right) 2011. The abbreviations of district names outside the circles are also displayed geographically in Fig. A.1, and their corresponding full names are shown in Table A.1. We distinguish the districts by their distinct colors on the circle. Each hyperbolic line linking two districts represents a dyadic migration flow, colored the same as the origin district. The width of a line is weighted by the number of migrants. The values displayed on the circles show the sum of out-migrants and in-migrants of each district, and only the districts with the largest values are included for clarity and esthetics. The figure shows sender and receiver heterogeneity and temporal evolution of migration flows

$$\text{Var} \begin{pmatrix} a_i \\ b_j \end{pmatrix} = \Sigma_{ab}; \text{Var} \begin{pmatrix} u_i \\ v_j \end{pmatrix} = \Sigma_{uv}; \text{Var} \begin{pmatrix} \epsilon_{i,j} \\ \epsilon_{j,i} \end{pmatrix} = \Sigma_\epsilon = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad (3)$$

where ρ is the dyadic correlation and σ^2 is an overdispersion parameter. If $\sigma^2 = 0$, $m_{i,j,t}$ would follow a Poisson distribution with the expectation: $e^{y_i + \beta_{dyad}^T x_{dyad,i,j,t} + \beta_o^T x_{o,i,t} + \beta_d^T x_{d,j,t} + a_i + b_j + u_i^T v_j}$; if $\sigma^2 > 0$, the variance of $y_{i,j,t}$ will be larger than this expectation.

The AMEN models are Bayesian models and use a Gibbs sampler to approximate the joint posterior distribution of the unknown quantities, including the parameters $(\beta_d, \beta_r, \beta_c, \Sigma_{ab}, \Sigma_{uv}, \sigma^2, \rho)$, the latent nodal effects (a_i, b_j, u_i, v_j) , and the latent dyadic variables $(z_{i,j,t})$.

The AMEN models can capture the first-, second-, and third-order network autocorrelations. The latent effects a_i, b_j in Eq. (1) describe the sender and receiver heterogeneity, respectively, which corresponds to the first-order dependencies. The sender and receiver dyadic correlation ρ in Eq. (3) describes the second-order dependencies of reciprocity. The multiplicative effect term $u_i^T v_j$ in Eq. (1) constructed from latent factors captures the third-order dependencies like transitivity and cyclicity in the network by allowing non-zero values of them (Eq. B.2 and B.3; network statistics are formulated and calculated in Appendix B.2.2, model diagnostics of these network statistics are shown in Appendix B.2.3).

Results

Climate effects

The climate variability effects are the central concern discussed here. First, we examine the effects of excessive heat and rainfall deficit, both of which contribute to drought (Table 2, model 1). As seen in column “All,” the positive maximum temperature anomalies have significant origin effects; people tend to leave districts with excessive heat. Here we give an example that indicates the magnitude of these effects. As this is a Poisson model, the coefficient (0.29) means that when positive maximum temperature anomalies increase by one unit, the migration flows will increase by 1.336 (exponential of 0.29) times. The negative precipitation anomalies in the rainy season have significant origin and destination effects. The negative coefficient of the origin effect indicates that the increase of

Table 2 Additive and multiplicative effects network model (AMEN) estimation of climate and socioeconomic effects on migration flows between districts in South Africa

	Model 1			Model 2			Model 3		
	All	Urban	Non-urban	All	Urban	Non-urban	All	Urban	Non-urban
(O) population	0.83**	0.92**	1.26**	0.85**	0.94**	1.24**	0.75**	0.82**	1.2**
(O) unemployed	-0.94**	-1.48**	-0.22	-1.25**	-1.83**	-0.66	-1.57**	-2.32**	-0.68
(O) primary	-1.23	-1.56†	1.89	-1.58*	-1.84†	1.44	-2.53**	-3.19**	1.1
(O) White	1.06	0.66	4.59†	0.36	0.23	3.71	-1.44	-2.23	3.08
(O) pos Tmax anom	0.29**	0.18	0.69**						
(O) neg Precip anom	-0.44**	-0.56**	-0.46†						
(O) soil moisture				-0.03**	-0.03**	-0.03*			
(O) pos SPI							0.13†	0.19*	-0.04
(O) neg SPI							-0.13	-0.33†	0.21
(D) population	0.76**	0.81**	0.04	0.77**	0.78**	0.08	0.75**	0.84**	0.25
(D) unemployed	-1.07**	-1.64**	-1.39*	-0.54†	-1.23**	-0.73	-0.43	-0.92*	-0.23
(D) primary	-2.58**	-3.72**	-2.96†	-2.21**	-3.52**	-1.48	-2.14**	-2.78**	0.43
(D) White	5.83**	8.39**	-0.3	6.61**	8.79**	3.69	6.89**	10.17**	6.63*
(D) pos Tmax anom	0.08	0.19†	-0.85**						
(D) neg Precip anom	0.63**	0.64**	0.47†						
(D) soil moisture				-0.01	0	-0.01			
(D) pos SPI							-0.15*	-0.28**	-0.2
(D) neg SPI							0.01	0.04	0.79*
(Dyad) distance	-1.45**	-1.4**	-1.67**	-1.47**	-1.43**	-1.72**	-1.45**	-1.4**	-1.64**
(Dyad) contiguity	0.65**	0.62**	0.97**	0.64**	0.61**	0.95**	0.65**	0.62**	0.99**

Table 2 (continued)

	Model 1		Model 2		Model 3	
	All	Urban	Non-urban	All	Urban	Non-urban
$R^2(\log)$	0.8337	0.8249	0.7023	0.8321	0.8236	0.6999
DIC	1.89e+10	5.7e+10	8.53e+09	2.17e+10	6.01e+10	8.73e+09
				2.24e+10	6.32e+10	7.66e+09

† $p < 0.10$
 * $p < 0.05$
 ** $p < 0.01$

Note: Estimated β coefficients (posterior mean) using AMEN models (Eqs. (1)–(2)). (O) stands for origin effects (β_o), (D) stands for destination effects (β_d), and (Dyad) stands for dyadic effects (β_{dyad}) of the covariates on all migrants (“All”) as well as on subgroups of migrants moving to urban areas (“Urban”) and non-urban areas (“Non-urban”) Variables:

- population: logarithm of population
- unemployed: unemployed population ratio among all working-age people (15–64 age range)
- primary: among people who are 25 or older, the population ratio of those who have finished primary school (7th grade) at most
- White: population ratio of people whose self-identified race is White
- pos Tmax anom: positive maximum temperature anomalies
- neg Precip anom: negative rainy season precipitation anomalies
- soil moisture: soil moisture percentile
- pos SPI: positive 12-month standardized precipitation index (SPI)
- neg SPI: negative 12-month SPI
- distance: logarithm of geographical distance between the sending and receiving districts of the migration flow dyads
- contiguity: binary variable that indicates whether the origin and destination districts share a common border or not (equal to 1 if they share a common border, 0 if not).

the covariate corresponds to migration flow reduction, as the exponential of negative values is smaller than 1. As this variable only has negative values, the coefficients mean that people are more likely to move out from the origins with a more severe rainfall deficit (decrease of this covariate) and less likely to move into the destinations with such a condition. The origin effects of the temperature and precipitation anomalies are qualitatively consistent with the push effects reported in (Mastrorillo et al., 2016) (their signs appear to be opposite ours for the negative precipitation anomalies because they take absolute values of this all-negative variable). In column “Urban,” we can see that the origin effect of the positive maximum temperature anomalies is not significant for migrants who moved to urban areas (we call them the “urban-oriented migrants”), and the destination effect of it is significant but weak for them. As for the migrants who moved to non-urban areas (“non-urban-oriented migrants”), shown in column “Non-urban,” the anomalously high temperature tends to push them away from their origins, and destinations with lower maximum temperature anomalies are more attractive to them. The results imply that the non-urban-oriented migrants are more sensitive to excessive heat than the urban-oriented ones. The origin and destination effects of negative precipitation anomalies are similar for both the subgroups as for all migrants, though less significant for the non-urban-oriented ones.

To investigate the impacts of different types of drought on migration, we include two measurements of drought. One is the soil moisture percentile (Sheffield et al., 2004) that can capture agricultural drought, which may directly influence the livelihood of agro-economic dependent populations (Table 2, model 2). Results in column “All” show that anomalously low soil moisture tends to drive people away from their origins, which is qualitatively consistent with its push effect in (Mastrorillo et al., 2016). These effects are similar in magnitude and significance for the two subgroups (columns “Urban” and “Non-urban”). The other measurement is the 12-month standardized precipitation index (SPI) (McKee et al., 1993), which can capture long-term meteorological droughts that may influence streamflow and reservoir levels (Table 2 model 3). Positive SPI has significant origin and destination effects on all migration (column “All”), indicating that people tend to leave an origin with a long-term rainfall surplus and avoid moving to such a destination. These effects are especially evident for urban-oriented migrants (column “Urban”), which is consistent with a tendency to leave urban areas with repeated urban flooding or agricultural areas experiencing destructive effects of flooding on crop yields, for other job opportunities in an urban location and aversion to destinations subject to urban flooding. Origin effects of negative SPI are also significant for urban-oriented migrants, which indicates that people tend to leave an origin with long-term meteorological drought and move to urban areas. Together with soil moisture percentile and rainy season rainfall anomaly effects for urban-oriented migrants, it suggests that this subgroup is sensitive to water deficit at the origin based on various measures. Also, neither the soil moisture anomalies nor negative SPI has a significant destination effect on urban-oriented migrants, implying that these types of drought are not affecting their destination choices either. On the other hand, non-urban-oriented migrants tend to avoid moving into locations with long-term meteorological drought (column “Non-urban”).

Model comparison

We compare the results from the network model, AMEN, to the gravity model with and without the Moran Eigenvector Spatial Filter (MESF, see Appendix D) to show the improvement of model performance in terms of goodness of fit, the robustness of the consistent effects across models, and to discuss the possible reasons for the discrepant ones. In Tables 3, 4, and 5, the columns named “AMEN” show the results for the AMEN models using Eq. (1), and the columns named “MESF” display the results for the gravity models with MESF components using Eq. D.12. The models used for the columns named “gravity” are the gravity models without MESF components using Eq. D.11. As the AMEN model is a Bayesian model and the other two are a frequentist, we only present one goodness of fit metric, R^2 -squared of the logarithm of the migration flows ($R^2(\log)$), for model performance comparison. The $R^2(\log)$ is used instead of R^2 considering the expectation-variance relationship of the Poisson distribution.

We find that the model performance measured by $R^2(\log)$ is improved using both models with network autocorrelation components (AMEN and gravity model with MESF) (Tables 3, 4, and 5). In Fig. 3, we visualize the comparison of the goodness of fit via plotting the observed migration flows against the fitted ones using the model for all migrants in Table 3 (first three columns). The AMEN model generally

Table 3 The same as model 1 of Table 2, except fitted by three models, denoted by “AMEN,” “MESF” (gravity model with MESF components), and “Gravity” (gravity model without MESF components)

	All			Urban			Non-urban		
	AMEN	MESF	Gravity	AMEN	MESF	Gravity	AMEN	MESF	Gravity
(O) population	0.83**	0.59**	0.62**	0.92**	0.63**	0.69**	1.26**	0.04	0.14
(O) unemployed	-0.94**	-0.06	0.03	-1.48**	-0.17	-0.04	-0.22	-0.11	-0.1
(O) primary	-1.23	2.8**	2.41**	-1.56†	2.66**	2.22*	1.89	1.37	1.79
(O) White	1.06	3.03**	3.5*	0.66	4.41**	5.15**	4.59†	-2.36	-1.99
(O) pos Tmax anom	0.29**	0.47**	0.48**	0.18	0.5**	0.52**	0.69**	0.41**	0.42**
(O) neg Precip anom	-0.44**	-0.42**	-0.35**	-0.56**	-0.33**	-0.26*	-0.46†	-0.64**	-0.56**
(D) population	0.76**	0.45**	0.39	0.81**	0.5**	0.44†	0.04	-0.22	-0.19
(D) unemployed	-1.07**	-0.91**	-0.93*	-1.64**	-1.81**	-1.85**	-1.39*	-0.63*	-0.61
(D) primary	-2.58**	-2.49**	-2.49*	-3.72**	-3.95**	-4.03**	-2.96†	-0.62	-0.69
(D) White	5.83**	3.03**	2.88	8.39**	4.66**	4.45**	-0.3	3.82*	3.73
(D) pos Tmax anom	0.08	0.08	0.06	0.19†	0.11*	0.1	-0.85**	-0.52**	-0.57**
(D) neg Precip anom	0.63**	0.67**	0.6**	0.64**	0.74**	0.68**	0.47†	0.41**	0.32†
(Dyad) distance	-1.45**	-1.04**	-0.94**	-1.4**	-1.1**	-0.95**	-1.67**	-0.85**	-1.18**
(Dyad) contiguity	0.65**	0.36**	0.54**	0.62**	0.32**	0.48**	0.97**	0.57**	0.62**
$R^2(\log)$	0.8337	0.7254	0.5203	0.8249	0.7073	0.5292	0.7023	0.5147	0.271

† $p < 0.10$

* $p < 0.05$

** $p < 0.01$

Table 4 The same as model 2 of Table 2, except fitted by three models, denoted by “AMEN”, “MESF” (gravity model with MESF components), and “Gravity” (gravity model without MESF components)

	All			Urban			Non-urban		
	AMEN	MESF	Gravity	AMEN	MESF	Gravity	AMEN	MESF	Gravity
(O) population	0.85**	0.73**	0.73**	0.94**	0.8**	0.83**	1.24**	0.43*	0.41
(O) unemployed	-1.25**	-0.48**	-0.48†	-1.83**	-0.53**	-0.51*	-0.66	-0.68**	-0.63†
(O) primary	-1.58*	2.83**	2.75**	-1.84†	3.16**	3.13**	1.44	0.36	0.62
(O) White	0.36	1.45	1.44	0.23	2.93**	3.02†	3.71	-2.91†	-2.71
(O) soil moisture	-0.03**	-0.13**	-0.13**	-0.03**	-0.15**	-0.15**	-0.03*	-0.1**	-0.08**
(D) population	0.77**	0.22	0.22	0.78**	0.22	0.24	0.08	-0.31	-0.33
(D) unemployed	-0.54†	-0.97**	-0.99*	-1.23**	-2.22**	-2.24**	-0.73	-0.17	-0.21
(D) primary	-2.21**	-1.87**	-1.73†	-3.52**	-3.03**	-2.99**	-1.48	-1.23	-1.19
(D) White	6.61**	-0.19	-0.07	8.79**	0.56	0.55	3.69	7.8**	7.89**
(D) soil moisture	-0.01	0.08**	0.06*	0	0.03	0.02	-0.01	0.18**	0.16**
(Dyad) distance	-1.47**	-1.07**	-0.94**	-1.43**	-1.11**	-0.95**	-1.72**	-0.85**	-1.18**
(Dyad) contiguity	0.64**	0.36**	0.54**	0.61**	0.33**	0.48**	0.95**	0.58**	0.62**
R ² (log)	0.8321	0.7268	0.5201	0.8236	0.7046	0.5304	0.6999	0.5157	0.2713

† $p < 0.10$
 * $p < 0.05$
 ** $p < 0.01$

yields a better fit, especially for zeros and smaller values (Fig. 3). We also compare their model performance in the out-of-sample predictions using a test dataset and find that the AMEN model also predicts the future migration flows better (Fig. E.2).

Table 5 The same as model 3 of Table 2, except fitted by three models, denoted by “AMEN,” “MESF” (gravity model with MESF components), and “Gravity” (gravity model without MESF components)

	All			Urban			Non-urban		
	AMEN	MESF	Gravity	AMEN	MESF	Gravity	AMEN	MESF	Gravity
(O) population	0.75**	0.35*	0.43†	0.82**	0.38**	0.46*	1.2**	0.13	0.26
(O) unemployed	-1.57**	-0.63**	-0.53*	-2.32**	-0.69**	-0.54*	-0.68	-0.8**	-0.77*
(O) primary	-2.53**	-0.22	-0.38	-3.19**	-0.08	-0.29	1.1	-2.26*	-1.82
(O) White	-1.44	-1.52	-0.89	-2.23	-0.22	0.65	3.08	-5.9**	-5.58*
(O) pos SPI	0.13†	0.18**	0.12†	0.19*	0.15**	0.08	-0.04	0.28**	0.29**
(O) neg SPI	-0.13	-0.42**	-0.4*	-0.33†	-0.42**	-0.43**	0.21	-0.37*	-0.37
(D) population	0.75**	0.14	0.08	0.84**	0.13	0.08	0.25	-0.5*	-0.6
(D) unemployed	-0.43	-0.59*	-0.67	-0.92*	-1.63**	-1.68**	-0.23	-0.24	-0.34
(D) primary	-2.14**	-0.81	-0.8	-2.78**	-2.19**	-2.21*	0.43	0.73	0.47
(D) White	6.89**	0.86	1.27	10.17**	1.99†	2.33	6.63*	6.11**	6.27*
(D) pos SPI	-0.15*	-0.38**	-0.33**	-0.28**	-0.45**	-0.4**	-0.2	-0.15*	-0.14
(D) neg SPI	0.01	-0.07	-0.09	0.04	0.02	0	0.79*	-0.26	-0.23
(Dyad) distance	-1.45**	-1.05**	-0.94**	-1.4**	-1.09**	-0.95**	-1.64**	-0.85**	-1.18**
(Dyad) contiguity	0.65**	0.36**	0.54**	0.62**	0.32**	0.47**	0.99**	0.57**	0.62**
R ² (log)	0.8326	0.7253	0.5197	0.8248	0.708	0.5299	0.6996	0.5102	0.2673

† $p < 0.10$
 * $p < 0.05$
 ** $p < 0.01$

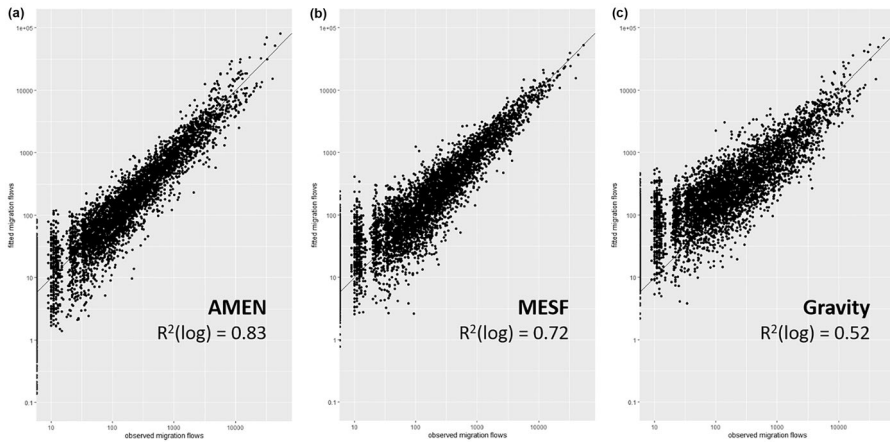


Fig. 3 Fitted (y-axis) using **a** AMEN, **b** gravity model with MESF components (“MESF”), and **c** gravity model without MESF (“gravity”) vs. observed (x-axis) migration flows (both in log₁₀ axis). Using models for all migrants in Table 3 (first 3 columns). Each dot in the plots represents one dyad during one time interval. The solid black line represents a perfect fit (1:1 line)

However, the migration observation data they are compared to has an underestimation problem. Therefore, we should treat this comparison with caution (see details in Appendix E). Moreover, most coefficients for the climate variables for the “urban” subgroup are robust in terms of being significant and obtaining the same signs using these two models (Tables 3, 4, and 5). The larger discrepancies between these two models for the “non-urban” subgroup could result from worse model performances dealing with data with excessive zero entries (zero-inflation, see Appendix E) because many fewer people moved to non-urban areas than to urban areas (Table 1). We choose to trust the AMEN model results for four reasons: better model performance, especially for the smaller values; accounting for a higher order of network autocorrelation; controlling for unknown influential variables; estimating the coefficients for the demographic and socioeconomic control variables more consistently across the three models using different climate variables for all migrants and each subgroup, respectively (see more details in Appendix E).

Socioeconomic and latent factor effects

In addition to climate variable impacts, here we discuss the effects of the demographic and socioeconomic control variables (“population,” “unemployed,” “primary,” and “White” in Table 2). Firstly, we find that these variables’ effects are generally more critical to urban-oriented migrants than non-urban-oriented ones. For the non-urban-oriented migrants, the only consistently significant effect across models 1–3 is the origin effect of population, suggesting their inclination to leave districts with a large population. Secondly, the destination effects of education level and race (“primary” and “White”) for all migrants are consistently significant across model settings. However, this pattern only shows up for the urban-oriented migrants,

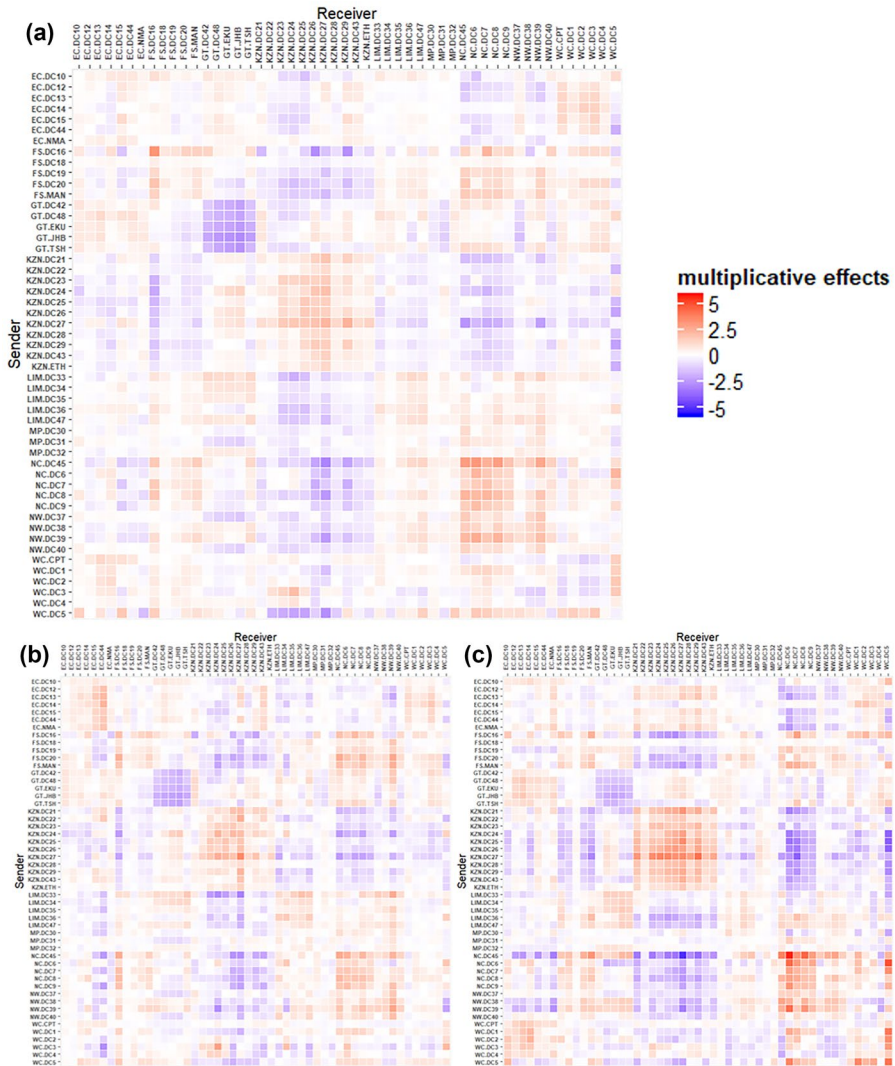


Fig. 4 Heat map of the total multiplicative effects ($u_i^T v_j$) in the AMEN model, Eq. (1) for **a** all migrants, **b** migrants moving to urban areas, and **c** migrants moving to non-urban areas. The values on the heat map show where the migration flows are underestimated (positive values) or overestimated (negative values) by other terms on the right-hand side of the equation and adjusted by the multiplicative effects term (model same as model 1 in Table 2, the sender and receiver codes are in the format province.district; see Table A.1 for the corresponding district and province names and Fig. A.1 for the geographic map)

implying that districts with higher education levels or higher White population ratio are attractive to them (for “primary” variable, larger values mean lower population ratio of those who have education beyond primary school). Thirdly, significant origin and destination effects of the unemployment rate (“unemployed”) have the same signs, just as in population effects. This phenomenon is analogous to the

gravitational attraction between populous locations; a pair of districts where each member is characterized by a relatively low unemployment rate may undergo higher reciprocal migration.

Moreover, by analyzing the latent factors in the AMEN model, we can gain insights into the unknown influential variables and seek plausible, relevant mechanisms. Based on the singular value decomposition theorem, the multiplicative effects term in AMEN ($u_i^T v_j$ in Eq. (1) can represent a (first order) matrix of residuals that are not explained by the observed regressors in the model (Hoff, 2015). Therefore, wherever the $u_i^T v_j$ values are positive, the observed regressors underestimated the migration flows there ($m_{i,j,t}$), and vice versa. Thus, we can see where the term adjusts the prediction biases in Fig. 4. For example, the cluster of positive values in the middle of Fig. 4a indicates that the migration flows among districts in province KwaZulu-Natal (KZN) were underestimated. In contrast, the negative values extending horizontally and vertically from this cluster suggest that migration flows between this province and other districts were generally overestimated. This means that after controlling for all the variables included in this model, we find that people are more likely to move within this province due to the latent factors and less likely to migrate to or from districts out of the province. By analyzing the latent positions of the first two pairs of latent factors (u_i, v_j in Eq. (1) that make the largest contribution to the multiplicative term, we find that the first pair contributes to the clustering of districts of KwaZulu-Natal (Appendix G). However, not every province exhibits such a clustering pattern. On the contrary, a cluster of negative values for Gauteng province (GT) indicates that people here are less likely to move among districts inside the province, holding all the variables included in this model constant. The latent position analysis in Appendix G shows that the second pair of latent factors contribute to this separation. We further deduce these latent factors in the following section. By comparing Fig. 4c to Fig. 4b, we can see that the abovementioned cluster of KwaZulu-Natal districts is much more evident for non-urban-oriented migrants than urban-oriented ones. The larger magnitude of both positive and negative values in Fig. 4c reflects the extra importance of these unknown variables for non-urban-oriented migration, while urban-oriented migrants are more sensitive to the demographic and socioeconomic origin and destination effects already included in the model as discussed above.

Migrant network effects and their interactions with climate variability

Although the AMEN model can control for the unknown variables, here, we explicitly include proxies of the migrant networks to evaluate their effects and demonstrate the connection and distinction between our work and previous studies involving migrant network effects. We assume that prior migrants may form networks connecting their origins and destinations, no matter why and when they moved. Therefore, we use the past migration flow and the counterflow during 1900–1996 as proxies of the preexisting migrant networks at the beginning of the period we are targeting (1996–2011) (data described in the “[Migration flows and socioeconomic variables](#)” section). The 1900–1996 migration flows are basically

1996 bilateral migration stocks between districts because earlier movers who were still alive in 1996 were very rare as the migration flows are mostly zeros before 1950 (Fig. 1.1). However, we still frame them as migration flows to introduce migration counterflows directed from destination to origin.

In Table 6, we compare the results of several models that estimate the migrant network effects. Results of the models that include these past migration terms (models 3–5) demonstrate that both prior migration flows ($m_{ij.past}$, from origin i to destination j) and counterflows ($m_{ji.past}$, from j to i) have significant positive effects on migration flows (from i to j), indicating that a more connected preexisting migrant network is associated with more future migration. Our results underscore not only that previous migrants flowing from origin to destination can facilitate future migration in the same direction but that prior migration counterflow from destination to origin has a similar effect on the same order, though in a slightly smaller magnitude.

Comparing model 3 to model 1 or model 4 to model 2, the $R^2(\log)$ does not change much, even though model 1 and model 2 do not include migrant network variables. This phenomenon demonstrates the bias adjustment ability of the multiplicative effects term. Adding the past migration flow variables reduces biases that need to be adjusted by multiplicative effects, especially for the ones involving the Gauteng province, implying that these migrant network proxies may be associated with the second pair of the latent factors that we discussed in the previous section (details see Appendix H). However, adding these variables does not weaken much of the clustering in the KwaZulu-Natal province, especially for the non-urban-oriented migration (Fig. H.1f). As apartheid-era (from 1948 to early 1990s) policies should have affected prior migration and thus be picked up to some degree by these variables, the unchanged clustering leads us to hypothesize that the cluster may be attributable to the 1994 abolition of the homeland (aka Bantustan) of KwaZulu and its merge with the Natal province, forming the current KwaZulu-Natal province. After the end of involuntary allocation of the black population to the homelands and the apartheid migration restrictions, migration may start to freely respond to family-based ties or the homophily among people who share characteristics like ethnicity and culture (e.g., Zulu, the majority population of KwaZulu-Natal), reflected in the social closeness of the districts within this province and thus the clustering. In any case, the essentially unchanged multiplicative effects indicate that other factors like policy effects still contribute to migration, which underscores the advantage of this model in controlling unknown variables.

In Table 6's model 4, we can see that the effects of the positive maximum temperature anomalies and the negative precipitation anomalies are consistent with the above results before adding the migrant network effects (also in Table 6, model 2). Using model 5, we examine the impacts of the migrant network on climate effects and find that a prior migration flow reduces the origin effects of the excessive heat, as the coefficient for the interaction term “(Dyad) $posTmax.i \times m_{ij.past}$ ” and “(O) $posTmax$ ” have opposite signs. This result is consistent with the “suppression” mechanism in (Nawrotzki et al., 2015). In analogy to their results, prior out-migrants may send remittances to their families remaining at the origin and share information about adaptation strategies to excessive heat, thus increasing their resilience and suppressing the push effect of this climate variable at the origin district. This

Table 6 AMEN estimation of the network effects and their interaction with climate variables

	Model 1	Model 2	Model 3	Model 4	Model 5
(O) population	0.819**	0.831**	0.652**	0.654**	0.69**
(O) unemployed	-1.345 **	-0.936 **	-1.434 **	-0.984**	-0.847**
(O) primary	-1.875 *	-1.229	-1.181	-0.682	-0.345
(O) White	-0.243	1.063	-2.232	-1.231	-1.272
(O) posTmax		0.286**		0.229*	0.209
(O) negPrec		-0.438**		-0.499**	-0.885**
(D) population	0.744**	0.764**	0.578**	0.586**	0.593**
(D) unemployed	-0.589 *	-1.068 **	-0.681*	-1.135**	-1.029**
(D) primary	-2.375**	-2.581**	-1.63*	-1.89**	-1.619*
(D) White	6.405**	5.83**	4.108**	3.425**	3.499**
(D) posTmax		0.078		0.009	-0.256†
(D) negPrec		0.634**		0.571**	0.08
(Dyad) mij.past			0.17**	0.167**	0.179**
(Dyad) mji.past			0.115**	0.114**	0.132**
(Dyad) posTmax.i × mij.past					-0.052†
(Dyad) posTmax.i × mji.past					0.05
(Dyad) posTmax.j × mij.past					0.08*
(Dyad) posTmax.j × mji.past					-0.029
(Dyad) negPrec.i × mij.past					0.054
(Dyad) negPrec.i × mji.past					0.03
(Dyad) negPrec.j × mij.past					0.029
(Dyad) negPrec.j × mji.past					0.075*
(Dyad) distance	-1.433**	-1.447**	-0.605**	-0.618**	-0.615**
(Dyad) contiguity	0.657**	0.649**	0.453**	0.452**	0.446**
R ² (log)	0.8321	0.8337	0.8331	0.8348	0.8368
DIC	2.19e+10	1.89e+10	1.6e+10	1.48e+10	1.67e+10

† $p < 0.10$

* $p < 0.05$

** $p < 0.01$

Note: Estimated β coefficients of the covariates on all migrants (“All”), adding dyadic effects of 1900–1996 migration flow (mij.past) and counterflow (mji.past) in models 3–5, and adding the interaction terms of the positive maximum temperature anomaly and the negative precipitation anomaly with these flows at origin (posTmax.i, negPrec.i) and destination (posTmax.j, negPrec.j), respectively, in model 5

– mij.past: logarithm of the number of migrants who moved from district i to j during 1900–1996 and reside in district j in 1996

– mji.past: logarithm of the number of migrants who moved from district j to i during 1900–1996 and reside in district i in 1996

hypothesis can be tested in a future study by including the remittance variable in the model.

From model 5 of Table 7, we can see that prior migration flows suppress the excessive heat impacts in the origin location for both urban-oriented and non-urban-oriented migrants, though the origin effect of this climate variable is not significant

Table 7 Results from models 2, 4, and 5 in Table 6 (“All”), compared with their counterparts for urban-oriented migration (“Urban”) and non-urban-oriented migration (“Non-urban”), respectively

	Model 2			Model 4			Model 5		
	All	Urban	Non-urban	All	Urban	Non-urban	All	Urban	Non-urban
	(O) population	0.831**	0.917**	1.26**	0.654**	0.701**	1.137**	0.69**	0.739**
(O) unemployed	-0.936**	-1.477**	-0.22	-0.984**	-1.558**	-0.154	-0.847**	-1.539**	0.095
(O) primary	-1.229	-1.563†	1.89	-0.682	-0.86	2.099	-0.345	-0.418	1.897
(O) White	1.063	0.659	4.59†	-1.231	-1.824	3.149	-1.272	-1.793	3.018
(O) posTmax	0.286**	0.183	0.69**	0.229*	0.126	0.62**	0.209	0.075	0.588*
(O) negPrec	-0.438**	-0.563**	-0.46†	-0.499**	-0.634**	-0.494*	-0.885**	-1.198**	-0.577†
(D) population	0.764**	0.807**	0.04	0.586**	0.595**	-0.014	0.593**	0.604**	-0.017
(D) unemployed	-1.068**	-1.644**	-1.39*	-1.135**	-1.757**	-1.428*	-1.029**	-1.41**	-1.427*
(D) primary	-2.581**	-3.718**	-2.96†	-1.89**	-2.777**	-2.373	-1.619*	-2.408**	-1.819
(D) White	5.83**	8.394**	-0.3	3.425**	5.458**	-1.517	3.499**	5.646**	-0.887
(D) posTmax	0.078	0.194†	-0.85**	0.009	0.12	-0.897**	-0.256†	-0.221	-1.236**
(D) negPrec	0.634**	0.64**	0.47†	0.571**	0.565**	0.441†	0.08	0.091	0.136
(Dyad) mij.past				0.167**	0.192**	0.177**	0.179**	0.235**	0.175**
(Dyad) mji.past				0.114**	0.105**	0.124**	0.132**	0.093**	0.125**
(Dyad) posTmax.i X mij.past							-0.052†	-0.096*	-0.111†
(Dyad) posTmax.i X mji.past							0.05	0.098*	0.129*
(Dyad) posTmax.j X mij.past							0.08*	0.032	0.09
(Dyad) posTmax.j X mji.past							-0.029	0.044	0.04

Table 7 (continued)

	Model 2			Model 4			Model 5		
	All	Urban	Non-urban	All	Urban	Non-urban	All	Urban	Non-urban
(Dyad) negPrec.i × mij.past							0.054	0.156**	-0.143*
(Dyad) negPrec.i × mji.past							0.03	-0.021	0.158*
(Dyad) negPrec.j × mij.past							0.029	0.057	0.083
(Dyad) negPrec.j × mji.past							0.075*	0.06	-0.007
(Dyad) distance	-1.447**	-1.401**	-1.67**	-0.618**	-0.774**	-0.925**	-0.615**	-0.751**	-0.942**
(Dyad) contiguity	0.649**	0.619**	0.97**	0.452**	0.384**	0.583**	0.446**	0.386**	0.589**
R ² (log)	0.8337	0.8249	0.7023	0.8348	0.8246	0.7015	0.8368	0.8282	0.7025
DIC	1.89e+10	5.7e+10	8.53e+09	1.48e+10	4.14e+10	4.13e+09	1.67e+10	3.47e+10	6.75e+09

† $p < 0.10$

* $p < 0.05$

** $p < 0.01$

for urban-oriented migration (see model 4). On the other hand, prior counterflows of migration amplify this origin effect of maximum temperature for both the urban-oriented and non-urban-oriented migration. The mechanism for amplification still needs theoretical support. Moreover, preexisting migrant networks have different interaction effects for urban-oriented and non-urban-oriented migration associated with negative precipitation anomalies at the origin location, and the migrant network interaction effects are also different for the two climate variables we investigate. Due to the limitation of data availability, we cannot reach conclusions about the mechanism causing the differences. However, this study underscores that social network modification of climate effects on migration is context-sensitive and should not be generalized.

Previous studies using the prior migration flows/stocks as predictors often discuss potential endogeneity, meaning that some unobserved omitted variables may affect both prior migration flows/stocks (the independent variables) and migration flows (the dependent variable) (Beine et al., 2011; Coniglio & Pesce, 2015; Mahajan & Yang, 2020). In our model, however, we are less concerned about omitted variables as the multiplicative effect $u_i^T v_j$ can be interpreted as representing omitted variables and the error terms should be random (Hoff, 2021). Nevertheless, we test the sensitivity of our modeling by substituting migration flows during different intervals within the 1900–1996 period in place of our original proxy for prior migrant networks. The comparison confirms the robustness of the results as it shows that the coefficients are qualitatively consistent using different intervals (comparison see Appendix I).

Conclusion and discussion

This study uses a network model, which can allow network autocorrelation, control unobserved covariates, and improve prediction performance (Fig. 3), to investigate the impacts of climate variability, migrant networks, and their interaction on South African internal migration. By including both the origin and destination factors in the network model, we find that origin effects of the same climate variables as included in (Mastrorillo et al., 2016) are qualitatively consistent with the push effects reported therein, while our results show that some climate variables also have significant destination effects on migration. In addition, we find that migrants moving to an urban versus a non-urban destination do not always respond to climate variabilities in the same manner. Nuance in climate variables, such as different types of drought, also may evoke different migration responses. Moreover, both prior migration flows and counterflows, which could be taken as proxies of preexisting migrant networks, have significant positive effects on subsequent migration. Finally, preexisting migrant networks could modify the association of climate variability and migration in distinctive ways for different climate variables.

Specifically, we find that a higher maximum temperature anomaly tends to push people away from origins, and this origin effect is only significant for non-urban-oriented migrants. Non-urban-oriented migrants also avoid moving into destinations with excessive heat. Both effects may reflect the higher dependencies on the temperature-sensitive livelihood of these migrants. Differently, all migrants, urban or non-urban-oriented, are

inclined to leave districts with rainy season rainfall deficits and avoid moving into districts with such conditions. The soil moisture percentile that can capture agricultural drought conditions has significant origin effects on all migrants and both subgroups, reflecting that people are more likely to leave a district with agricultural drought.

Using the 12-month SPI, we investigate the effect of long-term water surplus and deficit at a much longer timescale. The long-term water surplus measured by positive 12-month SPI has significant origin and destination effects. People tend to leave a district with long, excessively wet episodes and avoid moving into such a district. These effects are only significant for urban-oriented migrants, which would be consistent with the behavior of fleeing from places with repeated flooding and avoiding destinations prone to urban flooding. Meanwhile, the long-term meteorological drought measured by negative 12-month SPI has significant deterrent effects on the urban-oriented subgroup at origins and the non-urban-oriented subgroup at destinations. As rainfall in South Africa is highly variable, agriculture there may be largely irrigation dependent. Thus, long-term meteorological drought, perhaps indicative of streamflow and reservoir levels, could be a superior variable for representing the type and timescale of drought that influences agriculturally dependent migrants' destination choices.

Not surprisingly, socioeconomic variables like the unemployment rate have significant origin and destination effects on migration. However, we find that socioeconomic effects are especially significant to urban-oriented migrants and less critical to non-urban-oriented ones. As for the non-urban-oriented migrants, some unknown factors may be more important. For example, after the abolition of apartheid-era black homelands (Bantustans) in 1994, migrants within the KwaZulu-Natal province, especially the non-urban-oriented ones, may have begun to move more freely to locations where their families were or where people ethnically and culturally similar to them were located. To explore the nature of these factors, the model could be expanded to include additional variables, such as family-based ties and ones describing similarity in ethnicity. Modeling with data collected during the apartheid could reveal whether these effects differed before the abolition of apartheid to see if this policy change played a role.

To situate this study in the literature concerning migrant network effects in environmental migration, we also include prior migration flows and counterflows as proxies of preexisting migrant networks, which are one type of previously unobserved network variables. The results show that both prior migration flows and counterflows have significant positive effects on internal migration in South Africa. Prior migrant networks can also modify relationships between climate variability and migration flows. However, the modification effects are different for the two climate variables we investigate here and, in some instances, different for urban-oriented and non-urban-oriented migration. Thus, we find that network effects on environmental migration are context-sensitive and should not be generalized.

In summary, we extend the understanding of climate variability origin and destination effects on South African internal migration using a network model. By comparing results using various measurements of water surplus and deficits, we discover that migrants respond to different types of drought differently, indicating the importance of deploying multiple climate variables. Moreover, we find that

climate variations have different origin and destination effects on urban-oriented and non-urban-oriented migration flows. We also demonstrate that prior migrant networks can influence migration and impact the association between climate variability and migration. Due to the intrinsic network character of migration data and the model advantages demonstrated in this study, we consider AMEN a promising method for conceptualizing and analyzing migration flow problems. It can be applied in other migration case studies, modeling other network flows, and analyzing various problems, including transportation, trade, and conflicts. With the AMEN model, we establish multiple relationships that are not significant or not consistent when using gravity models (with or without MESF) (Tables 3, 4, and 5), for example, the pull effect of a large population. By visioning migration as a network problem, we can apply various network analysis techniques that expand the toolbox of migration studies and may reveal previously obscure mechanisms and relationships.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s11111-021-00392-8>.

Acknowledgements We acknowledge the High Meadows Foundation and Princeton University’s Center for Policy Research on Energy and the Environment for funding and supporting this research. Tingyin Xiao would like to thank Peter Hoff of Duke University for discussions of the AMEN model and Christopher Schwarz of New York University who first pointed out the potential utility of the model to her. Tingyin Xiao is also grateful to her ICPSR summer program lecturers for their courses and related discussions, especially to Bruce Desmarais and John Poe. We also thank the reviewers and our audiences at many research group meetings and conferences for valuable feedback.

Author contributions (CRediT) Tingyin Xiao: conceptualization, methodology, resources—provision of data, data curation, software, formal analysis, visualization, writing—original draft, writing—review & editing. Michael Oppenheimer: conceptualization, methodology, resources—provision of computing resources, writing—review & editing, supervision, funding acquisition. Xiaogang He: resources—provision of data, writing—review & editing. Marina Mastrorillo: resources—provision of data, writing—review & editing. All the authors participated in the discussion of the content.

Funding This research is funded and supported by High Meadows Foundation and Princeton University’s Center for Policy Research on Energy and the Environment.

Availability of data and material The data that support the findings of this study are deposited and published in the Harvard Dataverse repository (<https://doi.org/10.7910/DVN/NSKZ45>).

Code availability The code of this study is deposited and published in the Harvard Dataverse repository (<https://doi.org/10.7910/DVN/NSKZ45>).

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended

use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


References

- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., et al. (2018). Human mobility: Models and applications. *Physics Reports*, 734, 1–74. <https://doi.org/10.1016/j.physrep.2018.01.001>
- Barnett, G. (2012). Encyclopedia of social networks. *Encyclopedia of Social Networks*, 722–723. <https://doi.org/10.4135/9781412994170>
- Beine, M., Docquier, F., & Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, 95(1), 30–41. <https://doi.org/10.1016/j.jdeveco.2009.11.004>
- Beine, M., & Jeusette, L. (2018). A meta-analysis of the literature on climate change and migration. *CREA Discussion Paper*, (05), 46. http://www.wfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers
- Bell, A., Calvo-Hernandez, C., & Oppenheimer, M. (2019). Migration, intensification, and diversification as adaptive strategies. *Socio-Environmental Systems Modeling*, 1, 1–18. <https://doi.org/10.18174/sesmo.2019a16102>
- Bell, A. R., Wrathall, D. J., Mueller, V., Chen, J., Oppenheimer, M., Hauer, M., et al. (2021). Migration towards Bangladesh coastlines projected to increase with sea-level rise through 2100. *Environmental Research Letters*, 16(2). <https://doi.org/10.1088/1748-9326/abdc5b>
- Black, R., Bennett, S. R. G., Thomas, S. M., & Beddington, J. R. (2011). Climate change: Migration as adaptation. *Nature*, 478(7370), 447–449. <https://doi.org/10.1038/478477a>
- Bohra-Mishra, P., Oppenheimer, M., Cai, R., Feng, S., & Licker, R. (2017). Climate variability and migration in the Philippines. *Population and Environment*, 38(3), 286–308. <https://doi.org/10.1007/s11111-016-0263-x>
- Bohra-Mishra, P., Oppenheimer, M., & Hsiang, S. M. (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences of the United States of America*, 111(27), 9780–9785. <https://doi.org/10.1073/pnas.1317166111>
- Cai, R., Feng, S., Oppenheimer, M., & Pytlikova, M. (2016). Climate variability and international migration: The importance of the agricultural linkage. *Journal of Environmental Economics and Management*, 79, 135–151. <https://doi.org/10.1016/j.jeem.2016.06.005>
- Carrico, A. R., & Donato, K. (2019). Extreme weather and migration: Evidence from Bangladesh. *Population and Environment*, 41(1), 1–31. <https://doi.org/10.1007/s11111-019-00322-9>
- Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, 10(4), 317–344. <https://doi.org/10.1007/s10109-008-0068-2>
- Chun, Y., & Griffith, D. A. (2011). Modeling network autocorrelation in space-time migration flow data: An eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, 101(3), 523–536. <https://doi.org/10.1080/00045608.2011.561070>
- Chun, Y., & Griffith, D. A. (2013). *Spatial statistics and geostatistics: Theory and applications for geographic information science and technology*. SAGE Publications Ltd.
- Chun, Y., Griffith, D. A., Lee, M., & Sinha, P. (2016). Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters. *Journal of Geographical Systems*, 18(1), 67–85. <https://doi.org/10.1007/s10109-015-0225-3>
- Coniglio, N. D., & Pesce, G. (2015). Climate variability and international migration: An empirical analysis. *Environment and Development Economics*, 20(4), 434–468. <https://doi.org/10.1017/S1355770X14000722>
- Curran, S. R., & Rivero-Fuentes, E. (2003). Engendering migrant networks: The case of Mexican migration. *Demography*, 40(2), 289–307. <https://doi.org/10.2307/3180802>
- Desmarais, B. A., & Cranmer, S. J. (2012). Statistical inference for valued-edge networks: The generalized exponential random graph model. *PLoS One*, 7(1). <https://doi.org/10.1371/journal.pone.0030136>
- Entwisle, B., Williams, N. E., Verdery, A. M., Rindfuss, R. R., Walsh, S. J., Malanson, G. P., et al. (2016). Climate shocks and migration: An agent-based modeling approach. *Population and Environment*, 38(1), 47–71. <https://doi.org/10.1007/s11111-016-0254-y>

- Feng, S., Krueger, A. B., & Oppenheimer, M. (2010). Linkages among climate change, crop yields and Mexico-US cross-border migration. *Proceedings of the National Academy of Sciences*, 107(32), 14257–14262. <https://doi.org/10.1073/pnas.1002632107>
- Fussell, E., Curtis, K. J., & DeWaard, J. (2014). Recovery migration to the City of New Orleans after Hurricane Katrina: A migration systems approach. *Population and Environment*, 35(3), 305–322. <https://doi.org/10.1007/s11111-014-0204-5>
- Fussell, E., & Massey, D. S. (2004). The limits to cumulative causation: International migration from Mexican urban areas. *Demography*, 41(1), 151–171. <https://doi.org/10.1353/dem.2004.0003>
- Garcia, A. J., Pindolia, D. K., Lopiano, K. K., & Tatem, A. J. (2015). Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies*, 3(1), 89–110. <https://doi.org/10.1093/migration/mnu036>
- Gray, C., & Mueller, V. (2012). Drought and population mobility in rural Ethiopia. *World Development*, 40(1), 134–145. <https://doi.org/10.1016/j.worlddev.2011.05.023>
- Gray, C., & Wise, E. (2016). Country-specific effects of climate variability on human migration. *Climatic Change*, 135(3–4), 555–568. <https://doi.org/10.1007/s10584-015-1592-y>
- Hauer, M. E. (2017). Migration induced by sea-level rise could reshape the US population landscape. *Nature Climate Change*, 7(5), 321–325. <https://doi.org/10.1038/nclimate3271>
- He, X., Pan, M., Wei, Z., Wood, E. F., & Sheffield, J. (2020). A global drought and flood catalogue from 1950 to 2016. *Bulletin of the American Meteorological Society*, 101(5), E508–E535. <https://doi.org/10.1175/BAMS-D-18-0269.1>
- Hoff, P. D. (2015). Dyadic data analysis with amen. <http://arxiv.org/abs/1506.08237>
- Hoff, P. D. (2021). Additive and Multiplicative Effects Network Models. *Statistical Science*, 36(1), 34–50. <https://doi.org/10.1214/19-sts757>
- Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098. <https://doi.org/10.1198/016214502388618906>
- Hoff, P., Fosdick, B., & Volfovsky, A. (2018). amen: Additive and multiplicative effects models for networks and relational data. <http://github.com/pdhoff/amen>
- Hunter, L. M., Luna, J. K., & Norton, R. M. (2015). Environmental dimensions of migration. *Annual Review of Sociology*, 41, 377–397. <https://doi.org/10.1146/annurev-soc-073014-112223>
- IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. (Core Writing Team, R.K. Pachauri, & L. A. Meyer, Eds.) Intergovernmental Panel on Climate Change (IPCC) [Core Writing Team, Pachauri, R.K. and Reisinger, A. (eds.)]. http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full.pdf
- IPCC. (2019a). Summary for policymakers. In H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, M. T. P. Zhai, E. Poloczanska, K. Mintenbeck, et al. (Eds.), *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. https://www.ipcc.ch/site/assets/uploads/sites/3/2019/11/03_SROCC_SPM_FINAL.pdf
- IPCC. (2019b). Summary for policymakers. In P. R. Shukla, J. Skea, E. C. Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, et al. (Eds.), *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems* (Vol. In press). https://www.ipcc.ch/site/assets/uploads/sites/4/2019/12/02_Summary-for-Policymakers_SPM.pdf
- Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration : State of the art and challenges. *European Journal of Population*, 32(1), 73–97. <https://doi.org/10.1007/s10680-015-9362-0>
- Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electron. J. Statist.*, 6, 1100–1128. <https://doi.org/10.1214/12-EJS696>
- Liu, J., Yang, H., Gosling, S. N., Kummu, M., Flörke, M., Hanasaki, N., et al. (2017). Water scarcity assessments in the past, present, and future Earth' s Future. *Earth's Future*, 5(6), 545–559. <https://doi.org/10.1002/ef2.200>
- Mahajan, P., & Yang, D. (2020). Taken by storm: Hurricanes, migrant networks, and US immigration. *American Economic Journal: Applied Economics*, 12(2), 250–277. <https://doi.org/10.1257/app.20180438>
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration : A review and appraisal. *Population and Development Review*, 19(3), 431–466.

- Mastrorillo, M., Licker, R., Bohra-Mishra, P., Fagiolo, G., Estes, D., & L., & Oppenheimer, M. (2016). The influence of climate variability on internal migration flows in South Africa. *Global Environmental Change*, 39, 155–169. <https://doi.org/10.1016/j.gloenvcha.2016.04.014>
- McAlpine, A., Kiss, L., Zimmerman, C., & Chalabi, Z. (2021). *Agent-based modeling for migration and modern slavery research: A systematic review*. *Journal of Computational Social Science* (Vol. 4). Springer Singapore. <https://doi.org/10.1007/s42001-020-00076-7>
- Mckee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Eighth Conference on Applied Climatology*.
- McLeman, R., & Gemenne, F. (Eds.). (2018). *Routledge handbook of environmental migration and displacement*. Routledge.
- Mekonnen, M. M., & Hoekstra, Y. A. (2016). Four billion people experience water scarcity. *Science Advances*, 2(2), 1–7. <https://doi.org/10.1126/sciadv.1500323>
- Mueller, V., Gray, C., & Hopping, D. (2020). Climate-Induced migration and unemployment in middle-income Africa. *Global Environmental Change*, 65(June), 102183. <https://doi.org/10.1016/j.gloenvcha.2020.102183>
- Mueller, V., Gray, C., & Kosec, K. (2014). Heat stress increases long-term human migration in rural Pakistan. *Nature Climate Change*, 4(January), 182–185. <https://doi.org/10.1038/NCLIMATE2103>
- Nawrotzki, R. J., Riosmena, F., Hunter, L. M., & Runfola, D. M. (2015). Amplification or suppression: Social networks and the climate change-migration association in rural Mexico. *Global Environmental Change*, 35, 463–474. <https://doi.org/10.1016/j.gloenvcha.2015.09.002>
- Rigaud, K. K., de Sherbinin, A., Jones, Bryan Bergmann, J., Clement, V., Ober, K., Schewe, J., et al. (2018). *Groundswell: Preparing for Internal Climate Migration*. Washington, DC. <https://openknowledge.worldbank.org/handle/10986/29461>
- Rozenblat, C., & Melançon, G. (2013). Introduction. In C. Rozenblat & G. Melançon (Eds.), *Methods for multilevel analysis and visualisation of geographical networks* (pp. 1–15). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6677-8_1
- Sheffield, J., Goteti, G., Wen, F., & Wood, E. F. (2004). A simulated soil moisture based drought analysis for the United States. *Journal of Geophysical Research d: Atmospheres*, 109(24), 1–19. <https://doi.org/10.1029/2004JD005182>
- Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, 19(13), 3088–3111. <https://doi.org/10.1175/JCLI3790.1>
- Smirnov, O., Zhang, M., Xiao, T., Orbell, J., Lobben, A., & Gordon, J. (2016). The relative importance of climate change and population growth for exposure to future extreme droughts. *Climatic Change*, 138(1–2), 41–53. <https://doi.org/10.1007/s10584-016-1716-z>
- Thober, J., Schwarz, N., & Hermans, K. (2018). Agent-based modeling of environment-migration linkages: a review. *Ecology and Society*, 23(2):41. <https://doi.org/10.5751/ES-10200-230241>
- Tiefelsdorf, M. (2003). Misspecifications in interaction model distance decay relations: A spatial structure effect. *Journal of Geographical Systems*, 5(1), 25–50. <https://doi.org/10.1007/s101090300102>
- Ward, M. D., Ahlquist, J. S., & Rozenas, A. (2013). Gravity's rainbow: A dynamic latent space model for the world trade network. *Network Science*, 1(1), 95–118. <https://doi.org/10.1017/nws.2013.1>
- Wilson, J. D., Denny, M. J., Bhamidi, S., Cranmer, S. J., & Desmarais, B. A. (2017). Stochastic weighted graphs: Flexible model specification and simulation. *Social Networks*, 49, 37–47. <https://doi.org/10.1016/j.socnet.2016.11.002>
- Xiao, T., Oppenheimer, M., He, X., & Mastrorillo, M. (2021). Replication data for: Complex climate and network effects on internal migration in South Africa revealed by a network model. *Harvard DataVerse*. <https://doi.org/10.7910/DVN/NSKZ45>

Authors and Affiliations

Tingyin Xiao¹  · **Michael Oppenheimer**^{1,2,6} · **Xiaogang He**^{3,4} · **Marina Mastrorillo**⁵

¹ School of Public and International Affairs, Princeton University, Robertson Hall, 20 Prospect Ave, Princeton, NJ 08544, USA

² Department of Geosciences, Princeton University, Guyot Hall, Princeton, NJ 08544, USA

³ Department of Civil and Environmental Engineering, Princeton University, 59 Olden St, Princeton, NJ 08544, USA

⁴ Department of Civil and Environmental Engineering, National University of Singapore, 1 Engineering Drive 2, Singapore 117576, Singapore

⁵ Agrifood Economics Division (ESA), Food and Agriculture Organization of the United Nations (FAO), Viale Delle Terme Di Caracalla, 00153 Rome, Italy

⁶ High Meadows Environmental Institute, Princeton University, Guyot Hall, Princeton, NJ 08544, USA