# Appendices

For

Complex Climate and Network Effects on Internal Migration in South Africa Revealed by a Network Model

**Table of Contents**

## Appendix A. South Africa backgrounds

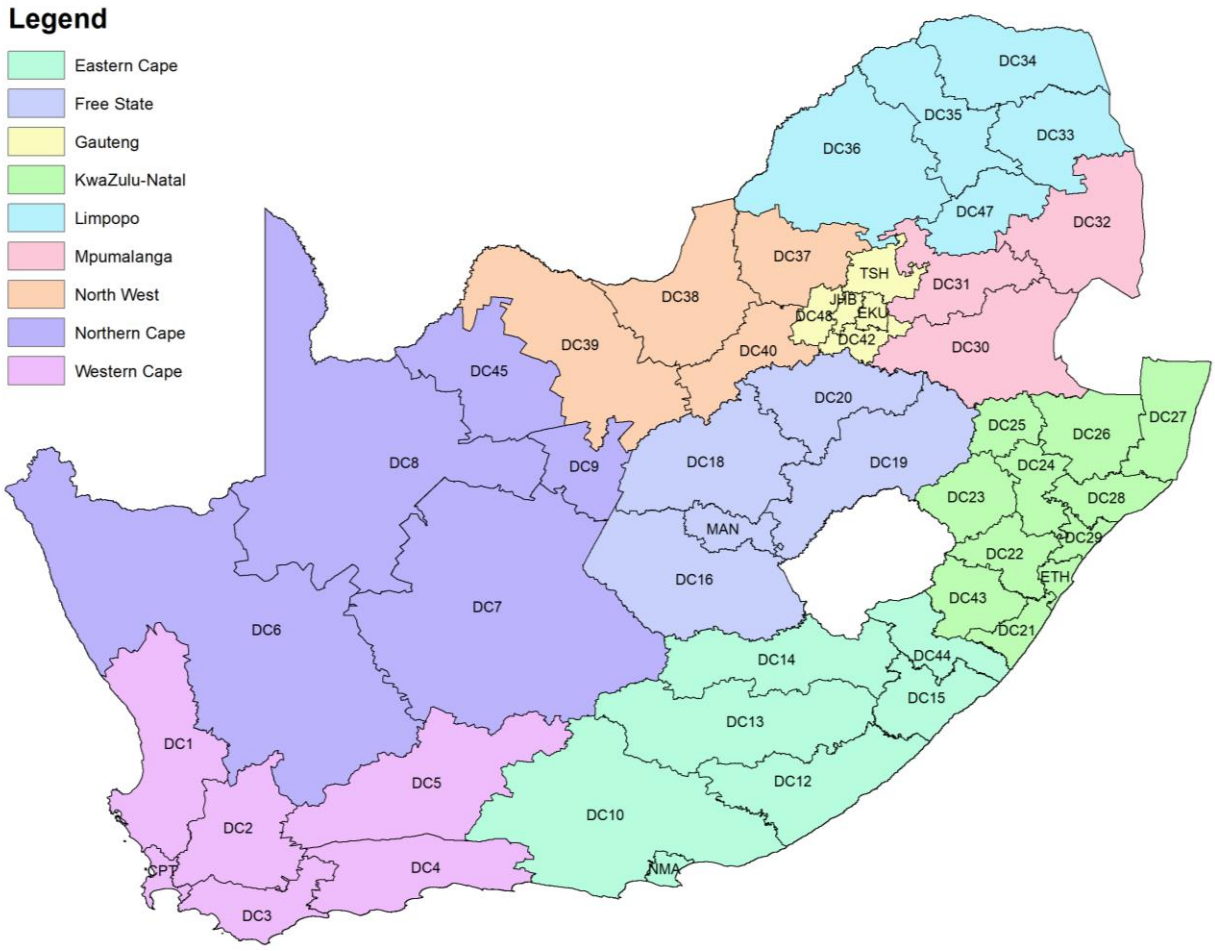


**Fig. A.1** Geographic map of South Africa districts with district code (2011 map). The district names corresponding to the district codes are listed in Table A.1. The colors denote the 9 provinces as in the legend. There are 52 districts in South Africa. In this study, district BUF is merged into DC12 to be consistent with other time steps, thus there are only 51 districts on the map

**Table A.1** South Africa provinces and districts code and name (2011). “District code” is used in Fig. 2 and the map in Fig. A.1, and the code for “Province.District” is used in Fig. 4 and Fig. G.1

| Province code | Province name | District code | District name | Province.District |
|---|---|---|---|---|
| EC | Eastern cape | DC10 | Cacadu | EC.DC10 |
| | | DC12 | Amathole | EC.DC12 |
| | | DC13 | Chris Hani | EC.DC13 |
| | | DC14 | Ukhahlamba | EC.DC14 |
| | | DC15 | O.R.Tambo | EC.DC15 |
| | | DC44 | Alfred Nzo | EC.DC44 |
| | | NMA | Nelson Mandela Bay | EC.NMA |
| FS | Free state | DC16 | Xhariep | FS.DC16 |
| | | DC18 | Lejweleputswa | FS.DC18 |
| | | DC19 | Thabo Mofutsanyane | FS.DC19 |
| | | DC20 | Fezile Dabi | FS.DC20 |
| | | MAN | Mangaung | FS.MAN |
| GT | Gauteng | DC42 | Sedibeng | GT.DC42 |
| | | DC48 | West Rand | GT.DC48 |
| | | EKU | Ekurhuleni | GT.EKU |
| | | JHB | City of Johannesburg | GT.JHB |
| | | TSH | City of Tshwane | GT.TSH |
| KZN | Kwazulu-Natal | DC21 | Ugu | KZN.DC21 |
| | | DC22 | UMgungundlovu | KZN.DC22 |
| | | DC23 | Uthukela | KZN.DC23 |
| | | DC24 | Umzinyathi | KZN.DC24 |
| | | DC25 | Amajuba | KZN.DC25 |
| | | DC26 | Zululand | KZN.DC26 |
| | | DC27 | Umkhanyakude | KZN.DC27 |
| | | DC28 | Uthungulu | KZN.DC28 |
| | | DC29 | iLembe | KZN.DC29 |
| | | DC43 | Sisonke | KZN.DC43 |
| | | ETH | eThekwini Metropolitan | KZN.ETH |

| Province code | Province name | District code | District name | Province.District |
|---|---|---|---|---|
| LIM | Limpopo | DC33 | Mopani | LIM.DC33 |
| | | DC34 | Vhembe | LIM.DC34 |
| | | DC35 | Capricorn | LIM.DC35 |
| | | DC36 | Waterberg | LIM.DC36 |
| | | DC47 | Greater Sekhukhune | LIM.DC47 |
| MP | Mpumalanga | DC30 | Gert Sibande | MP.DC30 |
| | | DC31 | Nkangala | MP.DC31 |
| | | DC32 | Ehlanzeni | MP.DC32 |
| NC | Northern cape | DC45 | John Taolo Gaetsewe | NC.DC45 |
| | | DC6 | Namakwa | NC.DC6 |
| | | DC7 | Pixley ka Seme | NC.DC7 |
| | | DC8 | Siyanda | NC.DC8 |
| | | DC9 | Frances Baard | NC.DC9 |
| NW | North west | DC37 | Bojanala | NW.DC37 |
| | | DC38 | Ngaka Modiri Molema | NW.DC38 |
| | | DC39 | Dr Ruth Segomotsi Mompati | NW.DC39 |
| | | DC40 | Dr Kenneth Kaunda | NW.DC40 |
| WC | Western cape | CPT | City of Cape Town | WC.CPT |
| | | DC1 | West Coast | WC.DC1 |
| | | DC2 | Cape Winelands | WC.DC2 |
| | | DC3 | Overberg | WC.DC3 |
| | | DC4 | Eden | WC.DC4 |
| | | DC5 | Central Karoo | WC.DC5 |

## Appendix B. Network description

### B.1. Network analysis terminology

In Fig. B.1:

- Node: the actor, which can be a person, a group of people, or an institution. In a geographic network, a node is a geographic location.

- Edge: the tie between a pair of nodes indicating the relationship or flow between them. For example, a friendship tie, or a communication flow.

- Direct vs. indirect edge:

Fig. B.1 shows an example of a directed network, in which an edge points from one node to another. The edge from A to B and the one from B to A are treated as two different edges. An edge can be indirect as well when the connection between two nodes does not have meaningful direction by the characteristics of the relationship. For example, in a school network, if a tie is defined between two students when they take the same class, then the tie does not have a direction.

- Weight of an edge: the value on the edge that describes the strength of the tie. An unweighted network does not have meaningful values on the ties. For example, in Fig. B.1, we can define a tie from C to B if C knows B, which is unweighted. On the other hand, if the tie is defined as the number of emails that C has sent to B, the tie is weighted and the network is a weighted network.

- Dynamic vs. static: the static networks are not supposed to change, while a dynamic network can evolve over time.

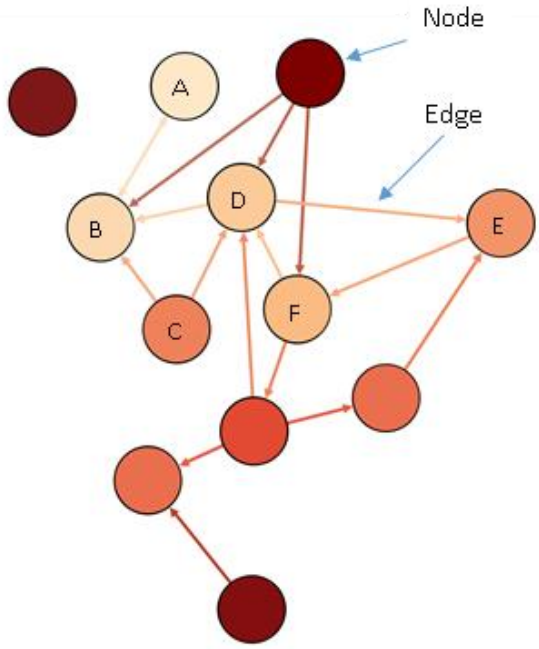


**Fig. B.1** Example of a directed network (drawn with Gephi)

## B.2. Network autocorrelation

### B.2.1. Common network autocorrelation structures in migration patterns

Network autocorrelation is defined as the dependencies between the flows for a migration network. The first to third-order dependencies correspond to the nodal, dyadic, and triadic properties of the network, which we describe in migration settings as follows:

• First-order:

- Sender (receiver) heterogeneity: Certain districts send (receive) more migrants than others on average due to their unobserved characteristics. It is also called out-degree (in-degree) heterogeneity in network analysis. For example, migration patterns in Fig. 2 display sender and receiver heterogeneities among the districts. During both time intervals, district JHB (City of Johannesburg) is a big sender and receiver of migrants, DC35 (Capricorn) does not receive many migrants, and DC48 (West Rand) does not send many of them. These are the second and third-order network autocorrelations among the flows.

• Second-order:

- Sender (receiver) correlation: Outbound (inbound) migration flows of a district may correlate with each other.

- Reciprocity: The tendency of a network to have mutual ties between nodes, such as the ties between A and B in Fig. B.1.

In the weighted migration network, generalized reciprocity describes the phenomenon that the number of migrants from place A to place B is proportional to the number flowing from B back to A. Corresponding to Fig. 1a in the main text, for established routes of bilateral migration, the aggregated migration flows should show the reciprocal patterns as in an equilibrium state, such like the migration flows between Mexico and the United States.

• Third order:

- Transitivity: In a social network, transitivity describes the triadic dependency that a friend of a friend is a friend, such as the triangle formed among nodes B, C, and D in Fig. B.1.

As illustrated in Fig. 1b in the main text, in a migration setting, generalized transitivity describes a situation where the number of migrants moving from place i to j is proportional to those moving from district i to k and from k to j.

Taking international transit migration corridors as an example, if people in country i desire to move to country j but not everyone has enough resources to do so, some of them may move to country k first and move into country j later when they are able, forming a transitive triangle.

- Cyclicality: In a social network, cyclicality describes the cyclic relationship, such as the triangle formed among nodes D, E, and F in Fig. B.1.

In a migration setting, there may be certain patterns of aggregated circular migration that form over years. For example, if some of the transit migrants return home later, the equilibrium state of aggregated migration flows forms cyclical triangles.

### B.2.2. The formulae for network statistics

Defining $\boldsymbol{Y} = [\boldsymbol{y_{ij}}]$ as the sociomatrix of migration flows with rows (i) denoting sending districts and columns (j) denoting receiving districts, we can calculate the above-mentioned network statistics (also follow the notation in Fig. 1 in the main text) as follows:

$$\boldsymbol{Reciprocity} = \frac{\boldsymbol{1}}{|\boldsymbol{S}|}\sum_{\boldsymbol{i,j\in S}} \boldsymbol{y'_{ij}y'_{ji}}\,, \boldsymbol{where\ S} = \{(\boldsymbol{i,j})|\boldsymbol{neither\ y_{ij}\ nor\ y_{ji}\ is\ null}\} \qquad (\boldsymbol{B.1})$$

$$\boldsymbol{Cyclicality} = \frac{\boldsymbol{1}}{|\boldsymbol{S}|}\sum_{\boldsymbol{i,j,k\in S}} \boldsymbol{y'_{ij}\,y'_{jk}y'_{ki}}\,, \boldsymbol{where\ S} = \{(\boldsymbol{i,j,k})|\boldsymbol{none\ of\ y_{ij}, y_{jk}, y_{ki}\ is\ null}\} \quad (\boldsymbol{B.2})$$

$$\boldsymbol{Transitivity} = \frac{\boldsymbol{1}}{|\boldsymbol{S}|}\sum_{\boldsymbol{i,j,k\in S}} \boldsymbol{y'_{ij}\,y'_{kj}y'_{ik}}\,, \boldsymbol{where\ S} = \{(\boldsymbol{i,j,k})|\boldsymbol{none\ of\ y_{ij}, y_{kj}, y_{ik}\ is\ null}\} \quad (\boldsymbol{B.3})$$

$$\boldsymbol{y'_{ij}} = \frac{\boldsymbol{y_{ij}} - \boldsymbol{\bar{y}_{ij}}}{\boldsymbol{std(Y)}} \qquad (\boldsymbol{B.4})$$

The two migration flow networks in Fig. 2 have reciprocity of 0.62 and 0.63, cyclicality of 0.10 and 0.14, and transitivity of 0.20 and 0.23, respectively, based on these equations.

### B.2.3. Model diagnostics

We can see in Fig. B.2 (top) that the trace plot for the β coefficients are all caterpillar-shaped, which indicates that the model is converged. By calculating the three statistics for each simulation iteration and comparing their distribution to the reference observations (Fig. B.2 (bottom)), we can see that the model simulations capture the transitivity and cyclicality well, but slightly overestimate the reciprocity.

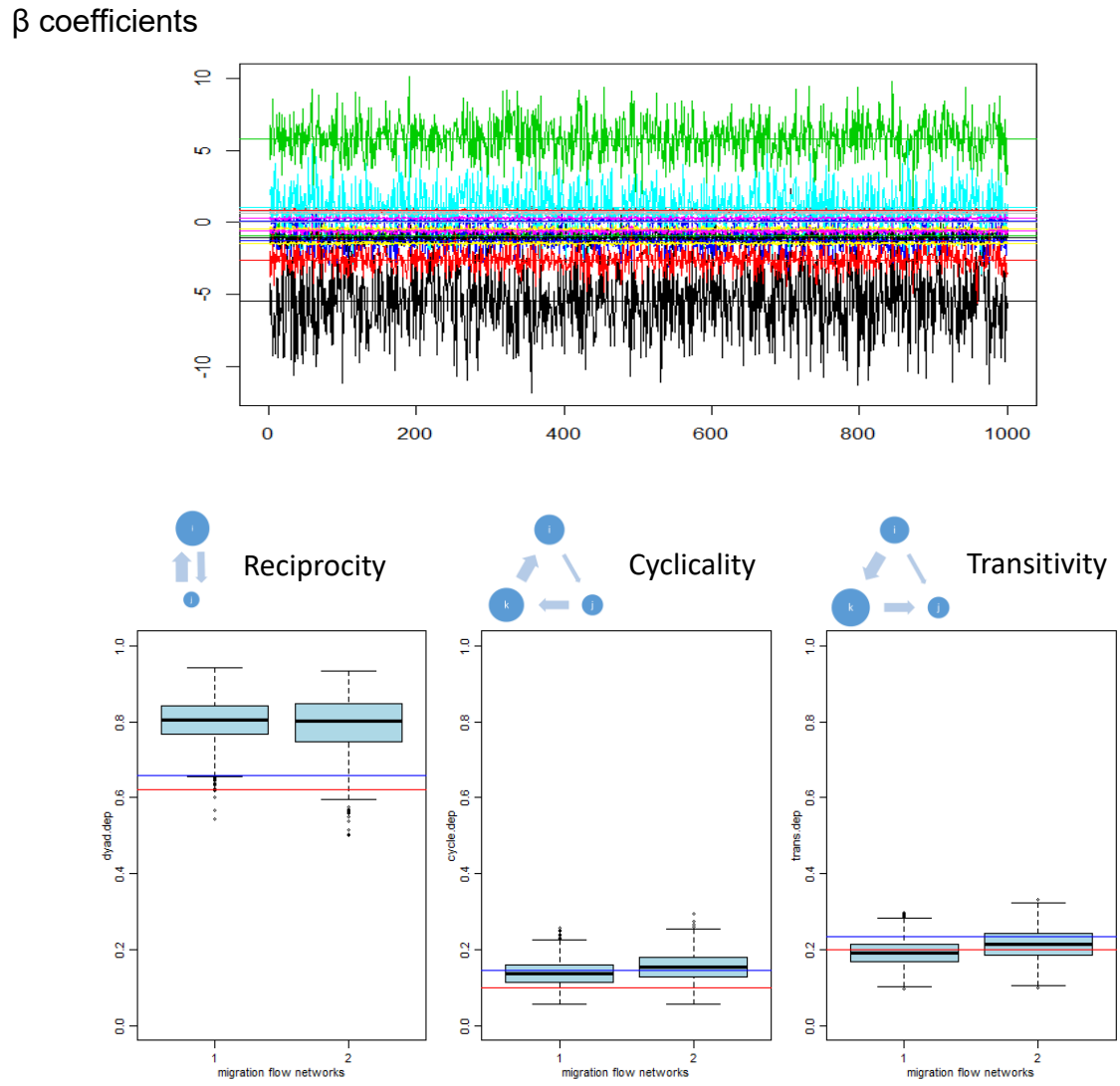


**Figure B.2.** Example diagnostics figures for the AMEN models: (Top) Values for all the β coefficients in equation (1) for every other 1000 iterations after the burn-in period (100,000 iterations for burn-in, 1,000,000 iterations in total); (Bottom) Distribution of network statistics calculated for each iteration for the two migration flow networks in Fig. 2. The red and blue horizontal lines are the reference lines for observed values for the two networks respectively (red: 1, blue: 2). Reciprocity, cyclicality, and transitivity are calculated using the formulae in Appendix B.2.2. The model is the same as used in Table 1 column 1.

## Appendix C. Other social network models for flow modeling and comparison to AMEN

We also considered applying previous studies that combine the log-linearized gravity model and Latent Space Model (LSM) (Hoff et al. 2002) to study trade flows (Ward et al. 2013). However, according to Jensen's inequality, the expectation of the log of y is not equal to the log of the expectation of y (Santos Silva and Tenreyro 2006). Thus, modeling our target, migration flow y, as the log of y with the log-linearized models may produce biased results. Instead, we simulate the migration flows in AMEN following a Poisson distribution, because the migration flows are count data that usually follows such a distribution.

Weighted exponential random graph models (ERGMs) provide another option for modeling migration flows (Desmarais and Cranmer 2012; Krivitsky 2012; Wilson et al. 2017). However, we decide not to use them in this study, mainly because of the model requirement of adding sufficient network structure statistics and our computational limitation. Directed and weighted ERGMs can model the whole migration network as a graph, taking the observed migration flows between all locations as one realization among all the possible graphs involving these locations. The generalized exponential random graph models (GERGM) (Desmarais and Cranmer 2012; Wilson et al. 2017) and the ERGM for count data (Krivitsky 2012) can be used to model migration flows. ERGMs are uniquely suited for studying the network structure effects. For example, if one's objective is to find out how the presence of mutual ties affects the outcome network (e.g., the migration flows), then one can add a term for reciprocity and estimate its coefficient. To our knowledge, this cannot be done with other network models. AMEN does not explicitly model these effects but allows for their existence.

However, there are limitations in applying the ERGMs in the current study. One major challenge is to deal with the model degeneracy, as we need to specify and feed in the network statistics, such as the observed number of transitive triangles, to account for higher-order flow dependencies. In addition, one also needs to add sufficient contributing network statistics terms to avoid estimation biases, which is difficult to accomplish (Minhas et al. 2019; Ward et al. 2013). Another challenge is related to ERGM's high demand for computational resources. The number of possible random graphs can be overwhelmingly large as the number of location points increases (for a network with n nodes, the possible random graphs will be on the order of $2^n$), thus this model is very computationally heavy for large networks. Besides, we need weighted ERGMs to model migration flows, but then much more computing resources and time is needed to achieve convergence compared to their binary counterparts.

The AMEN model can address network autocorrelation issues without specifying the network structures while running faster than ERGMs by estimating the unknown values with a Markov Chain Monte Carlo approach using the Gibbs sampler. For migration studies, which still lack a comprehensive theory of how network structures influence migration patterns, we prefer a model that can address the network autocorrelation problem with little or no assumption of specific network structures. Additionally, considering the scale of the study, we prefer a model with a lower computation burden, which enables efficient estimates for multiple experiments in a reasonable timeframe. Moreover, a recent study also demonstrates how AMEN notably outperforms ERGMs and LSMs in aspects including out-of-sample predictions (Minhas et al. 2019), which shows promise that this model could be a better choice for future projection.

## Appendix D. Methods – Moran eigenvector spatial filter (MESF)

Previous studies have shown that adding an Eigenvector Spatial Filter (ESF) to the spatial interaction model can account for the network autocorrelation (Fischer and Griffith 2008; Griffith 2003) and significantly improve the goodness of fit of the models (Chun 2008; Griffith 2007; LeSage and Pace 2008). This method has been applied to modeling several types of network flows, including migration (Chun 2008; Chun and Griffith 2011), trade (Krisztin and Fischer 2015; Metulini et al. 2016), commuting (Pace et al. 2013), and patent citation flows (Fischer and Griffith 2008). For comparison purposes in this study, we use the Moran Eigenvector Spatial Filter (MESF) to address network autocorrelation (Chun and Griffith 2011). The AMEN model has multiple advantages compared to this technique, including accounting for higher-order network autocorrelation. Nevertheless, the model comparison shows that taking autocorrelation into account using either method improves model performance (see section 3.2 and Appendix E for details). Thus the MESF technique is recommended for researchers preferring to preserve the gravity model structure and frequentist interpretations.

Here we review the MESF method (Chun 2008; Chun et al. 2016; Chun and Griffith 2011) and describe how to incorporate the MESF in the gravity model used in this study. The gravity model used in this study can include both the push and pull effects of nodal covariates, as well as the directional effects of dyadic covariates on migration.

Firstly, we describe a general format for the spatially filtered Poisson regression with a selected set of $k$ eigenvectors $\boldsymbol{E}'$ as follows:

$$\boldsymbol{Y} \sim Poisson(\exp(\boldsymbol{\mu})) \tag{D.5}$$
$$\boldsymbol{\mu} = \boldsymbol{X\beta} + \boldsymbol{E}'\boldsymbol{\beta_E} \tag{D.6}$$

where $\boldsymbol{Y} = (y_{ij})$ can denote any type of network flows.

There are various methods to select the eigenvectors, including forward stepwise selection, backward elimination, and LASSO-based approaches. The selection criterion can be based on p-values, F- or t-tests, (adjusted) R-squared, AIC, or BIC. In this study, set $\boldsymbol{E}'$ is selected via forward stepwise selection based on the significance level of the estimated coefficients (p-values).

To obtain the whole set of eigenvectors $\boldsymbol{E}$, MESF utilizes the spectral decomposition of a transformed spatial weights matrix, $\boldsymbol{C}$. The spectral decomposition of matrix $\boldsymbol{MCM}$ (where $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{11^T})/m$ and $\mathbf{1}$ is a vector of ones) produces a set of $m$ eigenvalues and their corresponding eigenvectors:

$$\boldsymbol{MCM} = \boldsymbol{E\Lambda E^{-1}} = \boldsymbol{E\Lambda E^T} \tag{D.7}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix, whose diagonal elements are the $m$ eigenvalues $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)$ ordered from the largest value to the smallest one; and $\boldsymbol{E} = (\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_m})$ represents the $m$ corresponding eigenvectors. In the network flows setting, $m = n^2$, where $n$ is the number of network nodes. Each eigenvector represents a distinct symmetric network pattern that has a certain level of spatial autocorrelation. The linear combination of the eigenvectors can account for the second-order network autocorrelations.

The weighted matrix for the network flows in this study is defined as follows:

First, we let $B^S$ be the standard binary spatial link matrix among the nodes. Let $\boldsymbol{B^N}$ be the binary network link matrix given by

$$b^N_{ij,kl} = \begin{cases} 1 & \text{if } i = k \text{ and } b^S_{jl} = 1, \text{ or if } j = l \text{ and } b^S_{ik} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{D.8}$$

which is an indicator of a neighborhood relation among the network edges such that edge $ij$ is a neighbor of edge $kl$ if either the two share the same origin and have a spatially contiguous destination, or have spatially contiguous origins and shared destination.

Next, we define the network link matrix with a variance-stabilizing S-coding scheme (Chun 2008; M. Tiefelsdorf et al. 1999; Michael Tiefelsdorf and Griffith 2007) as

$$\boldsymbol{S^N} = \frac{n^2}{\sum_{i=1}^{n^2} a_i^{q+1}} \boldsymbol{AB^N} \qquad (D.9)$$

Where $a_i$ is the number of network neighbors for the $i$th network flow, $q = -0.5$, and $\boldsymbol{A}$ is a diagonal matrix with $\{a_1^q, \cdots, a_{n^2}^q\}$.

The weight matrix for network flows is then defined as

$$\boldsymbol{C} = \frac{1}{2}(\boldsymbol{S^N} + (\boldsymbol{S^N})^T). \qquad (D.10)$$

Other coding schemes can be used by changing the parameter q. For example, setting q = 0, we can obtain the globally standardized C-coding scheme, while setting q = -1, we can get the row-sum standardized W-coding scheme.

We use a forward stepwise selection algorithm to choose eigenvectors to include in the spatial interaction model (Chun et al. 2016; Chun and Griffith 2013).

Previous studies used both of the Linear Mixed Model (LMM) and the Generalized Linear Mixed Model (GLMM) to model the migration flows incorporating the MESF method (Chun 2008; Chun and Griffith 2011). In this work, we directly model the migration flows from Poisson distribution instead of using log-linearized models, considering Jensen's inequality (Santos Silva and Tenreyro 2006), and estimate the model using the Poisson Pseudo-maximum likelihood (PPML) method to account for potential dispersion (Burger et al. 2009).

Secondly, to be consistent with a previous study (Mastorillo et al. 2016), the gravity model without MESF is formulated as:

$$\boldsymbol{m_{ij,t}} = \kappa \cdot exp\{\psi_i + \phi_j + \tau_t + \boldsymbol{\beta Z_{ij}} + \boldsymbol{\theta_i X}_{i,\tau(t)} + \boldsymbol{\mu_i C}_{i,\omega(t)} + \boldsymbol{\theta_j X}_{j,\tau(t)} + \boldsymbol{\mu_j C}_{j,\omega(t)}\}\varepsilon_{ij,t} \qquad (D.11)$$

and the formula incorporating the MESF can be written as follows:

$$\boldsymbol{m_{ij,t}} = \kappa \cdot exp\{\psi_i + \phi_j + \tau_t + \boldsymbol{\beta Z_{ij}} + \boldsymbol{\theta_i X}_{i,\tau(t)} + \boldsymbol{\mu_i C}_{i,\omega(t)} + \boldsymbol{\theta_j X}_{j,\tau(t)} + \boldsymbol{\mu_j C}_{j,\omega(t)} + \sum_{k=1}^{K}(E_k)_{ij}\Phi_k\}\varepsilon_{ij,t} \qquad (D.12)$$

with the MESF term being

$$\exp\left\{\sum_{k=1}^{K}(E_k)_{ij}\Phi_k\right\} \qquad (D.13)$$

which accounts for network flow dependence with a linear combination of K selected eigenvectors $(E_k)$.

In the formula,

- $\boldsymbol{m_{ij,t}}$ are 5-year accumulated migration flows from i to j district before t = 2001, 2011. The time interval is defined as [t-4, t].

- $\boldsymbol{Z_{ij}}$ is a vector of dyadic variables. Here, we use the log of geographical distance between i and j and a contiguity dummy variable. We can also include other dyadic variables, such as the temperature difference between district i and district j.

- $\boldsymbol{X_{i,\tau(t)}}$ and $\boldsymbol{X_{j,\tau(t)}}$ are the vectors of lagged demographic and socio-economic control variables at year $\tau(t)$ = 1996, 2007 for the sender (i) and receiver (j) districts, respectively. $\boldsymbol{C_{i,\omega(t)}}$ and $\boldsymbol{C_{j,\omega(t)}}$ are the vectors of climatic variables computed over the 5-year time intervals $\omega(t)$ = [1996–2000], [2006–2010] for the sender (i) and receiver (j) districts, respectively.

- $\psi_i \; and \; \phi_j$ are the unobserved time-invariant individual effects (fixed effects) for the sender (i) and receiver (j) districts, respectively; and $\tau_t$ is the time fixed effect.

- $(E_k)_{ij}$ is the kth of the total K selected eigenvectors, while $\Phi_k$ is its coefficient.

## Appendix E. Model comparison

### Future prediction

We also tried to evaluate the model out-of-sample prediction performance by using the models trained on the training set of data to predict migration flows in the test set of data (Fig. E.1). In the test set, we generated climate variables during 2011-2015, and socioeconomic control variables based on 2011 census data, then used these new covariates to predict migration during 2012-2016 and compared them to the observed values based on the 2016 community survey.

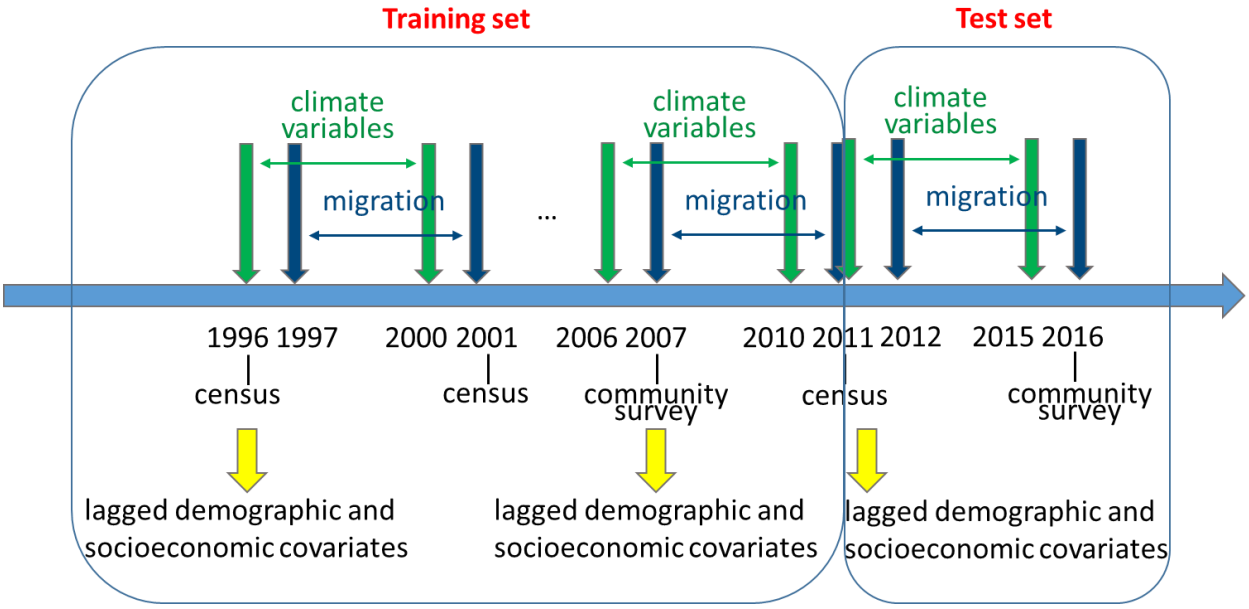


Fig. E.1 out-of-sample prediction test set data configuration chart

In the future prediction, we see that performance of the AMEN model and the gravity model with MESF, as measured by $R^2$ (log), are both better than the gravity model (Fig. E.2). The predicted values, in general, overestimate migration flows in all three models, but the overestimation may be attributable to migration underestimation problems in the 2016 community survey data. Migration flows based on 2016 community survey have a similar pattern to the 2011 census (Fig. E.3), but the values are too small (also see Table E.1) and are very likely to be underestimated. The 2016 community survey release report (Statistics South Africa 2016) explains, “This decrease may be due to underreporting from respondents or enumerator training deficiencies.” Therefore, we should use this data with caution. The gravity model also predicts larger values, especially for the smaller ones in the test set. We cannot conclude from this comparison that its performance is worse due to the observation data problem, but it tends to overestimate the small values even in the in-sample predictions (Fig. 3), which may be a systematical error. In addition, the spread of the dots for gravity model results are also much larger, which is also an indication of lower performance. Taken together, we conclude that the gravity model performance is the least reliable of these three. However, in comparing AMEN and the gravity model with MESF, we cannot distinguish one as superior to the other. Future census and community survey data will provide an improved basis for comparison.

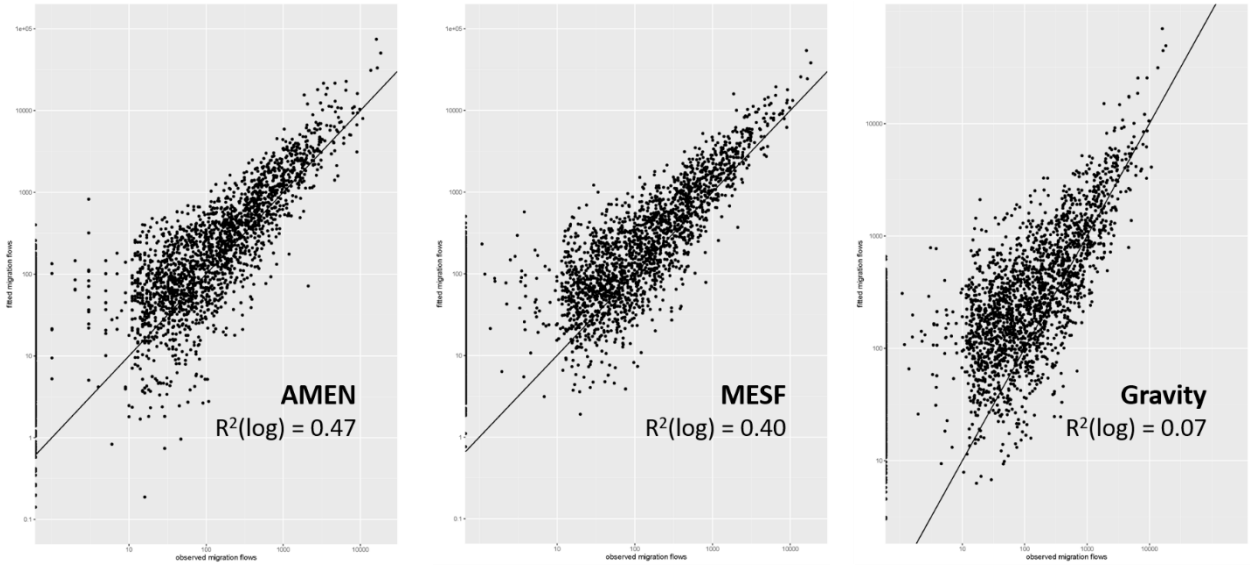


Fig. E.2 Fitted (y-axis) using (a) AMEN, (b) gravity model with MESF components ("MESF"), and (c) gravity model without MESF ("Gravity"), vs. observed (x-axis) migration flows (both in log10 axis). Using models for all migrants in Table 5. Each dot in the plots represents one dyad during this time interval (2012-2016). The solid black line represents a perfect fit (1:1 line). Note: y-axis scale slight different

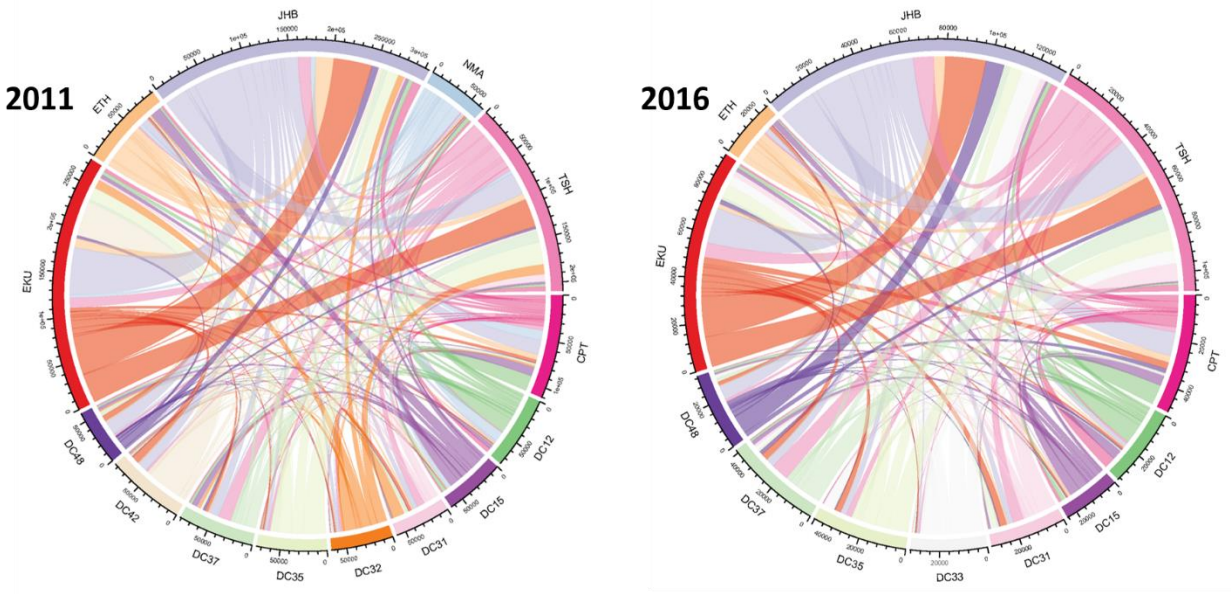


Fig. E.3 South Africa internal migration between districts during the 5 years before (left) 2011 and (right) 2016

Table E.1 Descriptive

| descriptive statistics | migration flows training set | migration flows test set |
|---|---|---|
| sample size (n) | 5100 | 2550 |
| sum | 3799167.00 | 991915.00 |
| mean | 744.93 | 388.99 |
| standard deviation | 2276.63 | 1106.60 |
| minimum | 0.00 | 0.00 |
| lower quartile | 37.00 | 13.00 |
| median | 131.00 | 62.00 |
| upper quartile | 492.25 | 291.75 |
| maximum | 53902.00 | 18349.00 |
| interquartile range | 455.25 | 278.75 |

## Zero-inflation

As we can see in Fig. 3, the two gravity models tend to overestimate the smaller migration flows and do not capture zero flows very well, while the AMEN model seems to perform better than them (the fitted values < 1 can be considered 0, as migrants are integers). The effect is minor for the "all migrants" runs because only about 10% of those observations are zeros. However, the issue appears to be influential when the zeros in the data are abundant. In Table 3-5, the columns named "urban" and "non-urban" displaying the results for the climate effects on the subgroups of migrants moving to urban and non-urban areas, respectively. Comparing them, we can see that the model performance in terms of $\boldsymbol{R^2}$(log) for "non-urban" subgroup are much worse, especially for the gravity models. Part of the reason is that there are much fewer people moving towards non-urban areas than urban locations and thus more zero entries in the migration matrix.

The zero-inflation issue arises when the migration flow is frequently zero between the locations, like when we divide the migrants into two subgroups. In such cases, it would be more appropriate to model the non-randomly missing migration flows as well to avoid biased estimation. In other words, we need a model that allows for both types of zero migration flows, structural (type I zeros) and incidental (type II zeros). Previous efforts have dealt with this issue in other related model settings (Krisztin and Fischer 2015; Metulini et al. 2016; Ward et al. 2013), but we have not done so in this study.

## Robustness of the results

Considering their better performance, we compare the β coefficients estimated by the two models that accounting for network autocorrelation but under different frameworks ("AMEN" and "MESF" columns in Table 3-5). Although the values of the estimates are not directly comparable, as AMEN is a Bayesian model and its coefficients have a different interpretation from the ones for the gravity model with MESF using frequentist statistics, we can consider the results to be robust if they are qualitatively consistent in both models (both significant or both not significant; if significant, have the same sign). For "all" migrants, almost all results are robust except for the destination effect of soil moisture and the origin effect of the negative SPI. They are significant using the gravity model with MESF but not significant using AMEN. We can see these discrepancies also show up for the non-urban-oriented subgroup, but not for the urban-oriented one. For urban-oriented migrants, all the effects are consistent, except for the origin effect of positive maximum temperature anomalies. It is not significant using the AMEN model but significant using the gravity model with MESF. As for the results for migrants moving to non-urban areas, most of them are not consistent between the two models. Besides the above-mentioned destination effect of soil moisture and origin effect

of the negative SPI, the origin and destination effects of positive SPI are also not significant using the AMEN model while being significant using the gravity model with MESF. Additionally, the destination effect of the negative SPI is significant using the AMEN model but not significant using the gravity model with MESF.

The large discrepancies for the “non-urban” subgroup may be due to the zero-inflation issue because there are much more zero entries in the data of the “non-urban” subgroup. Considering the gravity models (with or without MESF) also overestimate the smaller values, we suspect the estimates from them are less reliable, as much fewer people moved to non-urban areas than to urban areas. Moreover, in Tables 3-5, we can see that AMEN’s estimates of the demographic and socio-economic effects are more consistent across the three models using different climate variables for all migrants and each subgroup, respectively. Together with the fact that it can account for higher-order of network autocorrelation (AMEN: up to third order, MESF: second-order) and its latent factors can control for unknown influential variables, we consider the results from AMEN models to be more reliable.

## Appendix F. Data

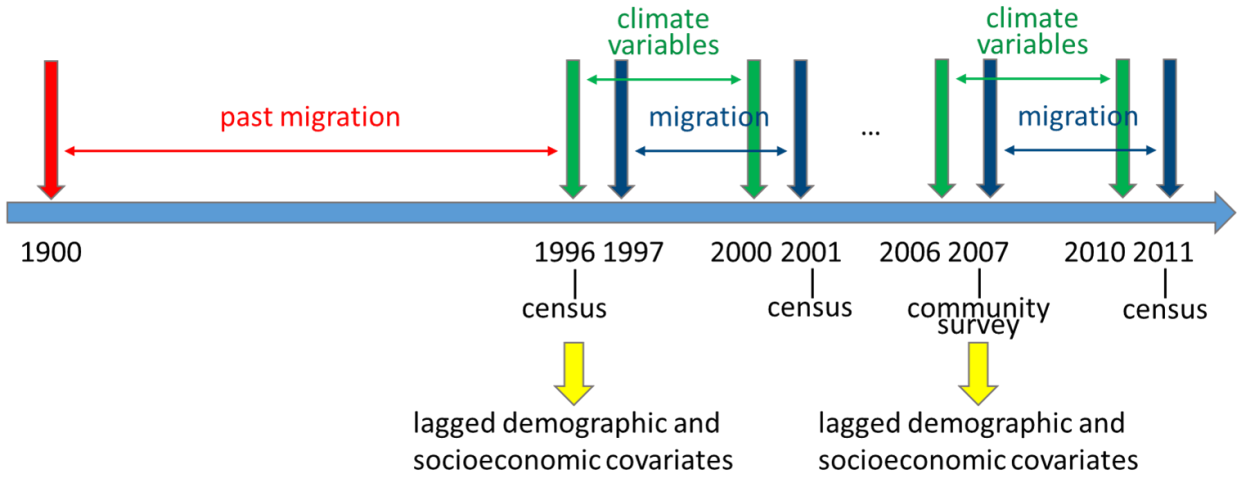


**Fig. F.1** Variable temporal configuration for the model inputs. The demographic and socio-economic covariates for the second time interval for migration (2007-2011) are observed in the first year of the interval, but for the first time interval (1997-2001) are observed one year ahead. This is because there is no census data in 2006 (usually every 5 years) and we use 2007 community survey data instead

## Appendix G. Latent factors of the multiplicative terms

In Fig. G.1, which shows 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 note the clustering of districts of KwaZulu-Natal (in the green circle). In this figure, the clustering of the sending and receiving districts indicates the closeness among them in the latent space, and this horizontal clustering indicates that the first pair of the latent factors should be the contributing latent factors.

We can also see from Fig. G.1 that the Gauteng districts acting as senders (in the red circle) are distributed at the bottom and lie far from the Gauteng districts acting as receivers (in the blue circle) at the top. The vertical separation in Fig. G.1 suggests that the second pair of latent factors are contributing to the division of Gauteng districts as senders and receivers.

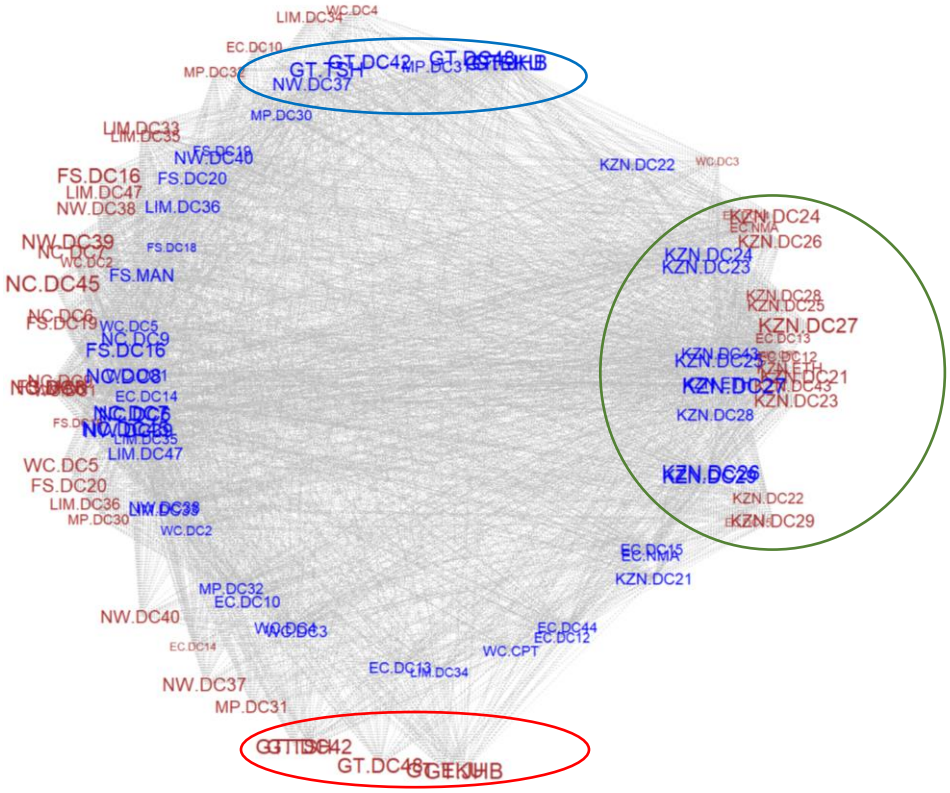


**Fig. G.1** Latent position of the first two pairs of latent factors ($\boldsymbol{u_i}$, $\boldsymbol{v_j}$) in Eq. 1, using the first model for all migrants in Table 2. The red letters denote the sender districts, blue letters denote the receiver districts, and grey lines link district dyads, which have migration flows. As in Fig. 4, 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). The clustering in the green circle indicates the closeness among the sending and receiving districts in the latent space. It appears that these sending and receiving districts are mostly in the province of KwaZulu-Natal (KZN). The red circle highlights the latent positions of the Gauteng (GT) districts as senders and the blue circle the latent positions of them as receivers

## Appendix H. Multiplicative effects before and after adding the prior migration terms

We compare the results from model 2 and 4 in Table 7, and display the changes in their multiplicative effects before and after adding the migrant network variables in the models in Fig. H.1. Comparing the figures to the right (added migrant network variables) to the ones to the left (same models without these variables), we can see that adding the migrant network variables somewhat reduces the magnitude of the values of the multiplicative effects and almost wipes out the cluster involving GT province, which means these variables are especially important predictors of migration moving between GT districts. For the urban-oriented migrants, the values in the rows and columns extending from the GT cluster are also largely reduced, implying these variables are also important in predicting migration between GT districts and districts in other provinces in South Africa (Fig. H.1d). As for KZN province, values still show a clear cluster, especially for migration to non-urban areas (Fig. H.1f). Based on this pattern, we suspect that the prior migrant network is associated with the second pair of latent factors. By calculating the correlations of the proxies (1900-1996 migration flow and counterflow), with each pair of the latent factors, we found that their correlations with the second pair of latent factors are indeed the largest (with negative values) among all the latent factors. The remaining multiplicative effects shown in the subplots in the right column of Fig. H.1 indicate that there are still other factors we have not accounted for, which leaves space for future theory development. Nevertheless, we reiterate that specifying the network variables or other unobserved factors is not required in our model setting, which is a major advantage of this model framework.

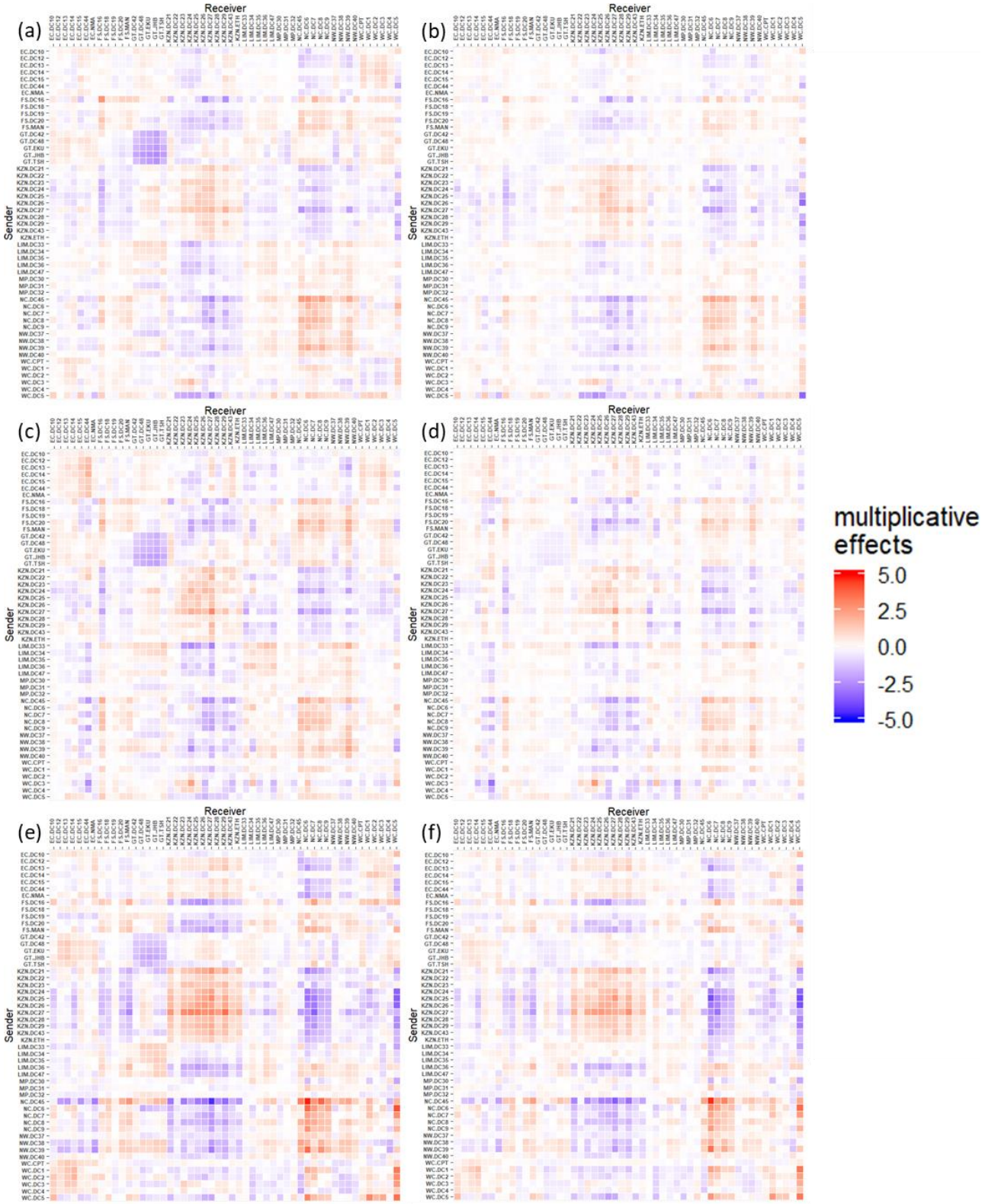


**Fig. H.1** Same as Fig. 4 except (a), (c), and (e) use models in Table 7 model 2: (a) all, (c) to urban, (e) to non-urban; (b), (d), and (f) use models in Table 7 model 4, which include migrant networks variables: (b) all, (d) to urban, (f) to non-urban

## Appendix I. Robustness test of the migrant network effects

We cut the 1900-1996 periods to different intervals to test the sensitivity of our modeling to the prior migrant network proxies using migration flows during different periods. Here we show the results based on the cutting that divides the period into two periods with the different political environments, 1948-1990 (apartheid) and 1990-1996 (abolishment of apartheid). The migration flows are mostly zeros before 1950, as they are measured in the 1996 census (Fig. I.1). We excluded 1948 and 1990 to generate the period 1949-1989, to make sure the interval is totally within apartheid. The comparison results show that the coefficients for the variables are basically consistent even using migration proxies during such different periods (Tables I.1-I.3).

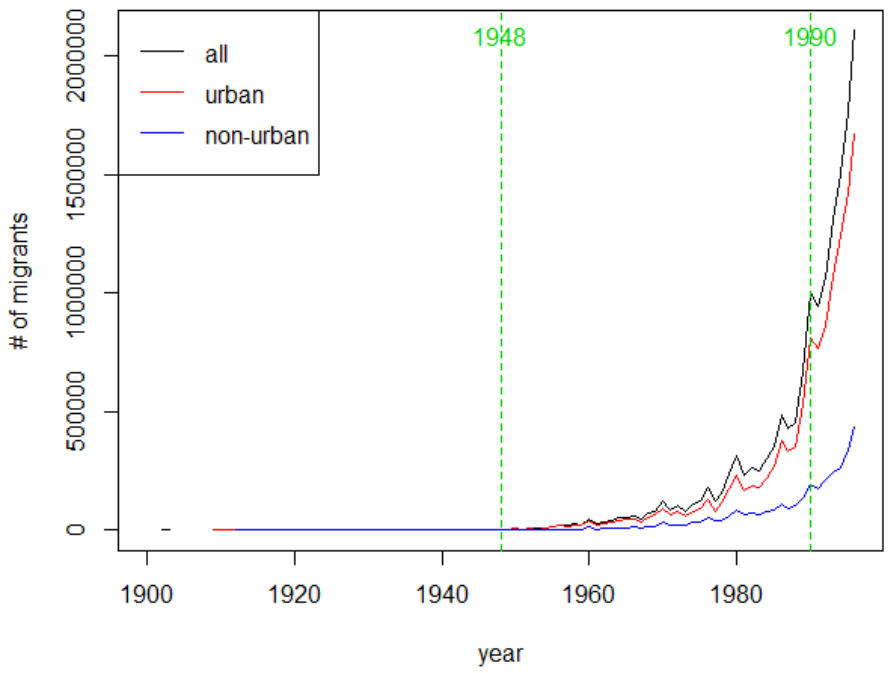


**Fig. I.1** Total migrants moved during 1900-1996, provided in 1996 census. The green dashed lines denotes the two cut-off years for the sensitivity test, based on the apartheid of South Africa starting from 1948

**Table I.1.** Comparison of the results in Table 6 Model 3, which uses past migration flow during 1900-1996, with the results using past migration flow during other periods (1990-1996, 1949-1989, and 1948-1996).

| Model 3 | 1900-1996 | 1990-1996 | 1949-1989 | 1949-1996 |
|---|---|---|---|---|
| (O) population | 0.652 ** | 0.655 ** | 0.719 ** | 0.653 ** |
| (O) unemployed | -1.434 ** | -1.429 ** | -1.408 ** | -1.435 ** |
| (O) Primary | -1.181 | -1.226 | -1.311 † | -1.183 |
| (O) white | -2.232 | -2.386 † | -1.578 | -2.246 |
| (D) population | 0.578 ** | 0.586 ** | 0.651 ** | 0.579 ** |
| (D) unemployed | -0.681 * | -0.663 * | -0.641 * | -0.68 * |
| (D) Primary | -1.63 * | -1.678 ** | -1.777 ** | -1.626 * |
| (D) white | 4.108 ** | 3.988 ** | 4.886 ** | 4.111 ** |
| (Dyad) mij.past | 0.17 ** | 0.161 ** | 0.125 ** | 0.17 ** |
| (Dyad) mji.past | 0.115 ** | 0.121 ** | 0.101 ** | 0.116 ** |
| (Dyad) distance | -0.605 ** | -0.625 ** | -0.688 ** | -0.603 ** |
| (Dyad) contiguity | 0.453 ** | 0.455 ** | 0.483 ** | 0.452 ** |
| R2(log) | 0.8331 | 0.833 | 0.8332 | 0.8331 |
| DIC | 1.6e+10 | 1.68e+10 | 1.86e+10 | 1.51e+10 |

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

**Table I.2.** Comparison of the results in Table 6 Model 4, which uses past migration flow during 1900-1996, with the results using past migration flow during other periods (1990-1996, 1949-1989, and 1948-1996).

| Model 4 | 1900-1996 | 1990-1996 | 1949-1989 | 1949-1996 |
|---|---|---|---|---|
| (O) population | 0.654 ** | 0.655 ** | 0.717 ** | 0.653 ** |
| (O) unemployed | -0.984 ** | -0.987 ** | -0.981 ** | -0.986 ** |
| (O) Primary | -0.682 | -0.73 | -0.783 | -0.68 |
| (O) white | -1.231 | -1.4 | -0.538 | -1.249 |
| (O) posTmax | 0.229 * | 0.22 * | 0.239 ** | 0.227 * |
| (O) negPrec | -0.499 ** | -0.495 ** | -0.481 ** | -0.499 ** |
| (D) population | 0.586 ** | 0.592 ** | 0.655 ** | 0.585 ** |
| (D) unemployed | -1.135 ** | -1.125 ** | -1.113 ** | -1.134 ** |
| (D) Primary | -1.89 ** | -1.941 ** | -2.033 ** | -1.891 ** |
| (D) white | 3.425 ** | 3.303 ** | 4.219 ** | 3.434 ** |
| (D) posTmax | 0.009 | 0.002 | 0.022 | 0.009 |
| (D) negPrec | 0.571 ** | 0.574 ** | 0.592 ** | 0.571 ** |
| (Dyad) mij.past | 0.167 ** | 0.158 ** | 0.123 ** | 0.168 ** |
| (Dyad) mji.past | 0.114 ** | 0.12 ** | 0.1 ** | 0.115 ** |
| (Dyad) distance | -0.618 ** | -0.642 ** | -0.707 ** | -0.617 ** |
| (Dyad) contiguity | 0.452 ** | 0.452 ** | 0.479 ** | 0.449 ** |
| R2(log) | 0.8348 | 0.8346 | 0.8349 | 0.8347 |
| DIC | 1.48e+10 | 1.57e+10 | 1.9e+10 | 1.64e+10 |

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

**Table I.3.** Comparison of the results in Table 6 Model 5, which uses past migration flow during 1900-1996, with the results using past migration flow during other periods (1990-1996, 1949-1989, and 1948-1996).

| Model 5 | 1900-1996 | 1990-1996 | 1949-1989 | 1949-1996 |
|---|---|---|---|---|
| (O) population | 0.69 ** | 0.694 ** | 0.754 ** | 0.689 ** |
| (O) unemployed | -0.847 ** | -0.858 ** | -0.96 ** | -0.845 ** |
| (O) Primary | -0.345 | -0.43 | -0.507 | -0.345 |
| (O) white | -1.272 | -1.463 | -0.47 | -1.28 |
| (O) posTmax | 0.209 | 0.243 | 0.221 † | 0.209 |
| (O) negPrec | -0.885 ** | -0.811 ** | -0.699 ** | -0.883 ** |
| (D) population | 0.593 ** | 0.602 ** | 0.674 ** | 0.592 ** |
| (D) unemployed | -1.029 ** | -1.031 ** | -1.179 ** | -1.033 ** |
| (D) Primary | -1.619 * | -1.693 * | -1.818 ** | -1.613 * |
| (D) white | 3.499 ** | 3.406 ** | 4.262 ** | 3.497 ** |
| (D) posTmax | -0.256 † | -0.234 | -0.064 | -0.255 † |
| (D) negPrec | 0.08 | 0.153 | 0.323 * | 0.08 |
| (Dyad) mij.past | 0.179 ** | 0.168 ** | 0.128 ** | 0.18 ** |
| (Dyad) mji.past | 0.132 ** | 0.135 ** | 0.11 ** | 0.132 ** |
| (Dyad) posTmax.i x mij.past | -0.052 † | -0.046 | -0.035 | -0.052 † |
| (Dyad) posTmax.i x mji.past | 0.05 | 0.035 | 0.034 | 0.049 |
| (Dyad) posTmax.j x mij.past | 0.08 * | 0.065 * | 0.04 | 0.079 * |
| (Dyad) posTmax.j x mji.past | -0.029 | -0.014 | -0.019 | -0.029 |
| (Dyad) negPrec.i x mij.past | 0.054 | 0.05 | 0.067 † | 0.053 |
| (Dyad) negPrec.i x mji.past | 0.03 | 0.028 | 0.002 | 0.031 |
| (Dyad) negPrec.j x mij.past | 0.029 | 0.024 | -0.012 | 0.028 |
| (Dyad) negPrec.j x mji.past | 0.075 * | 0.075 * | 0.093 * | 0.076 * |
| (Dyad) distance | -0.615 ** | -0.648 ** | -0.74 ** | -0.614 ** |
| (Dyad) contiguity | 0.446 ** | 0.447 ** | 0.478 ** | 0.445 ** |
| R2(log) | 0.8368 | 0.8366 | 0.8365 | 0.837 |
| DIC | 1.67e+10 | 1.57e+10 | 1.75e+10 | 1.6e+10 |

† $p < 0.10$

* $p < 0.05$

** $p < 0.01$

## Reference


Burger, M., van Oort, F., & Linders, G.-J. (2009). On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. *Spatial Economic Analysis*, *4*(2), 167–190. https://doi.org/10.1080/17421770902834327

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

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

Fischer, M. M., & Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the european union. *Journal of Regional Science*, *48*(5), 969–989. https://doi.org/10.1111/j.1467-9787.2008.00572.x

Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering*. Springer.

Griffith, D. A. (2007). Spatial structure and spatial interaction: 25 years later. *The Review of Regional Studies*, *37*(1), 28–38.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis, *97*(460), 1090–1098. https://doi.org/10.1198/016214502388618906

Krisztin, T., & Fischer, M. M. (2015). The Gravity Model for International Trade: Specification and Estimation Issues. *Spatial Economic Analysis*, *10*(4), 451–470. https://doi.org/10.1080/17421772.2015.1076575

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

LeSage, J. P., & Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, *48*(5), 941–967. https://doi.org/10.1111/j.1467-9787.2008.00573.x

Mastrorillo, M., Licker, R., Bohra-Mishra, P., Fagiolo, G., D. Estes, 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

Metulini, R., Patuelli, R., & Griffith, D. A. (2016). *A Spatial-Filtering Zero-Inflated Approach to the Estimation of the Gravity Model of Trade*. *SSRN*. https://doi.org/10.2139/ssrn.2854075

Minhas, S., Hoff, P. D., & Ward, M. D. (2019). Inferential Approaches for Network Analysis: AMEN for Latent Factor Models. *Political Analysis*, *27*(2), 208–222. https://doi.org/10.1017/pan.2018.50

Pace, R. K., Lesage, J. P., & Zhu, S. (2013). Interpretation and Computation of Estimates from Regression Models using Spatial Filtering. *Spatial Economic Analysis*, *8*(3), 352–369. https://doi.org/10.1080/17421772.2013.807355

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4), 641–658. https://doi.org/10.1162/rest.88.4.641

Statistics South Africa. (2016). *Community Survey 2016, Statistical release P0301*. *Statistics South Africa*. http://www.statssa.gov.za

Tiefelsdorf, M., Griffith, D. A., & Boots, B. (1999). A variance-stabilizing coding scheme for spatial link matrices. *Environment and Planning A*, *31*(1), 165–180. https://doi.org/10.1068/a310165

Tiefelsdorf, Michael, & Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environment and Planning A*, *39*(5), 1193–1221. https://doi.org/10.1068/a37378

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, *49*, 37–47.