The Human Migration Database (HMigD)

Maciej J. Dańko[1,*], Arkadiusz Wiśniowski[2], Boris Barron[1], Domantas Jasilionis[1], Dmitri A. Jdanov[1], and Emilio Zagheni[1]

[1]Max Planck Institute for Demographic Research, Rostock, Germany
[2]Social Statistics Department, University of Manchester, Manchester, UK
[*]Presenting author, email:danko@demogr.mpg.de

EPC 2026

Abstract

International migration is a complex phenomenon with significant socioeconomic and political impacts on both origin and destination countries. However, the lack of reliable, comparable, and comprehensive migration data often leads to research gaps and reliance on unverified information. The Human Migration Database (HMigD) addresses these challenges by providing high-quality, harmonized data on international migration flows. Adhering to the principles of comparability, flexibility, accessibility, and reproducibility, HMigD employs a Bayesian modeling framework to integrate diverse data sources while accounting for variations in data quality and definitions. To incorporate high-quality net migration estimates as constraints, we are also developing information-theoretic methods to further refine the results. The database can be interactively explored through a Shiny app. By producing consistent and reliable migration flow estimates, HMigD aims to enhance the understanding of migration dynamics, initially focusing on EU countries, with plans to extend coverage globally.

1 Introduction

International migration is a complex process that affects both origin and destination countries across socioeconomic and political domains [1; 4]. Recent migration waves in Europe have intensified policy debates [6], yet many reports still rely on incomplete or unreliable evidence.

Reliable migration statistics remain difficult to obtain due to inconsistencies in data availability, definitions, and quality [26; 25]. Data come from censuses, surveys, population registers, and administrative sources [20; 16], each with limitations: censuses are infrequent [23; 2], surveys have large sampling errors [27], registers undercount [10], and administrative data are not designed for migration measurement. Consequently, reported figures often differ between sending and receiving countries.

The Human Migration Database (HMigD), developed at the Max Planck Institute for Demographic Research, provides harmonized estimates of international migration flows. Building on the principles of comparability, flexibility, accessibility, and reproducibility of the Human Mortality Database (HMD, [14]), HMigD offers full documentation and open data. Unlike the HMD, it is a synthetic database whose outputs are generated through statistical modeling.

Migration data quality depends on the ability of national agencies to trace migration flows and on consistent definitions across countries. Major issues include (i) data availability, (ii) accuracy, (iii) undercounting, (iv) coverage gaps, and (v) definitional inconsistencies with international standards.

We first evaluate and classify data quality across sources using metadata, expert judgment, and data-driven methods. We then integrate all data within a Bayesian framework, which harmonizes sources while accounting for data quality and definitional differences. Next we apply information projection (I-projection) to further tune up the estimates by applying net-migration constraints. The goal is to generate up-to-date and comparable bilateral migration flow estimates for EU countries, with future extensions beyond the EU planned.

A dedicated Shiny app [8] enables users to explore data issues, visualize migration flows, and download data, metadata, and estimates. It also allows comparisons with earlier projects such as IMEM [15] and QuantMig [21].

2 Data

2.1 Data Sources

We use administrative migration flow data and Labour Force Survey data provided by Eurostat. Administrative data originate from National Statistical Institutes, the UN, and Eurostat, complemented by historical datasets from earlier models. To capture completely missing corridors, we also incorporate online-based bilateral migration flow estimates for 2019–2022 derived from privacy-protected Facebook data [7]. These estimates provide high-frequency information on international migration patterns, complementing official statistics where data are delayed or incomplete. A detailed list of all data sources is available in the HMigD Shiny app [10].

2.2 Data Quality Assessment

The quality of migration statistics varies over time and depends on registration systems, legal incentives, and the methodologies applied by National Statistical Institutes. Unlike vital statistics, standardized and universally applied procedures for collecting migration data are lacking. Out-migration data are generally less reliable than in-migration data, and national estimates are often not comparable across countries or time due to differing data types, definitions, and adjustment methods [22; 25; 11; 10].

Five key issues affect data quality:

(1) Data availability. Migration data are often incomplete, outdated, or insufficiently disaggregated by age and sex. After the adoption of the 12-month stay definition in 2008 (Reg (EC) 862/2007), several countries–including the Czech Republic, Germany, Luxembourg, Poland, and Ireland–stopped providing bilateral flow data to Eurostat. To address structural gaps, particularly for missing corridors, we include online-based estimates 2019–2022 from Chi et al. [7].

(2) Accuracy. Random errors in registration and de-registration reduce accuracy. To assess data reliability, we analyzed metadata and collaborated with the QuantMig team [21], revealing substantial variation in data quality and frequent gaps in methodological documentation. Building a network of national experts is crucial for filling these gaps.

(3) Undercounting. Non-systematic biases arise when migration events are not properly recorded or when migrants fail to respond in surveys. We developed a score-based, partially data-driven method to assess undercounting [10]. The approach compares year-specific, duration-adjusted bilateral flow ratios across 29 European countries, incorporating metadata and expert input from IMEM and QuantMig. The resulting online tool, *UndercountMigScores* [9], enables users to generate and compare undercounting scores interactively. Figure 1 illustrates this assessment and its classification scale.

(4) Coverage inconsistencies. Systematic biases occur when data collection rules exclude specific population groups, such as return migrants or foreigners omitted from official records.

(5) Definition inconsistencies. National migration definitions (e.g., minimum duration of stay) often diverge from international standards, and frequent revisions—such as the 2008 EU regulation—further complicate comparability [24].

3 The Hierarchical Bayesian Modeling Approach

We developed a hierarchical Bayesian model to estimate latent international migration flows among 31 European countries from 2002 to 2023, using data from administrative sources and the EU Labour Force Surveys (EU LFS). Our model builds upon previous Bayesian models that separately handled administrative and EU LFS data by integrating these approaches into a unified framework [22; 27]. The primary objective of combining these data sources is to estimate the true relocation rate [19], which is then used to predict the true latent migration flows, defining a long-term migration event as lasting at least 12 months.

The model comprises two main components: a measurement error model and a predictive model. The measurement error model harmonizes data from different sources and accounts for biases and inconsistencies. The predictive model addresses missing data through smooth functions of time and random effects, while incorporating a "shock" migration variable, such as the freedom of movement of workers, to capture rapid changes in migration flows that smooth models might not detect. See appendix for further details on the model.

3.1 Exemplary Results

In the examples shown below (Figures 3 and 4), we examine four cases of bilateral migration flows between selected countries. We also explore four model variations that differ by the inclusion or exclusion of EU LFS data (LFS(+) and LFS(-)) and the presence or absence of the freedom of movement variable (FM(+) and FM(-)). For additional results and interactive exploration, see our Shiny App [8].

The first example (Figure 3a) shows migration flows from the Netherlands to Italy. Predicted flows closely align with Dutch data points. Both countries use registers with the same 12-month definition of stay; however, Dutch data are of higher quality, with greater accuracy and lower undercounting. The fitted model excludes both the freedom of movement and EU LFS variables, as neither has a measurable effect for this corridor.

The second example (Figure 3b) presents flows from Greece to Cyprus. Only Cyprus provides data, collected through passenger surveys available until 2009, and additional LFS data. Two models are compared: both include the freedom of movement variable, but only one uses EU LFS data. The model incorporating EU LFS data performs better, capturing post-2012 changes and reducing uncertainty.

The third case (Figure 3c) shows flows from the United Kingdom to Norway. Norwegian data, based on high-quality registers, used a six-month definition of stay before 2008, causing overestimation of observed flows. Afterward, minor undercounting emerged. UK administrative data, collected via surveys, are less accurate. The model primarily aligns with Norwegian data, effectively correcting both over- and underestimation.

The fourth example (Figure 3d) presents flows from Poland to Germany, one of Europe's largest migration corridors. Polish data suffer from undercounting and a permanent-stay definition, while German data tend to overcount due to short stay definitions. Two models are compared—both exclude EU LFS data but differ in the inclusion of the freedom of movement variable. The FM(+) model performs slightly better, capturing the 2011 migration spike following Germany's opening of its labor market to Polish workers. The difference between FM(+) and FM(-) is even more pronounced for flows to Austria (see Figure 4).

In Figure 4, we show aggregated results illustrating the effect of the freedom of movement variable. Panel (a) displays flows from Bulgaria and Romania to the United Kingdom, where freedom of movement was granted in 2014; panel (b) shows flows from A8 countries to Austria, which granted it in 2011. These results highlight the importance of incorporating "shock" variables like freedom of movement, which capture sudden spikes in migration flows that would otherwise be overly smoothed.

3.2 Conclusions

Our approach introduces several innovations. We show that, despite the incompleteness and inconsistencies of data sources such as population registers and surveys (e.g., EU LFS), integrating these sources allows effective estimation of international migration flows among European countries. These inconsistencies stem from varying definitions of long-term migration, undercounting, coverage, and accuracy, as well as EU LFS sampling design issues. Our model accounts for these differences with a particular focus on time variation, advancing beyond earlier IMEM [22] and QuantMig [3] approaches.

The model also incorporates auxiliary data (metadata) to assess data quality, translating this information into prior distributions. A key feature is its dynamic adjustment of undercounting and accuracy classifications over time, using data-driven methodologies developed in [10] and metadata-based assessments for EU LFS data.

Our Bayesian framework is flexible and can incorporate new migration data sources, additional variables, or refined priors. The I-projection method currently being developed will help account for high-quality net migration estimates, ensuring consistency across all core databases (HMigD, HMD, and HFD). Enhanced information on data completeness and definitions could further refine our quality assessment. Collaboration with national statistical offices will also help clarify adaptations to EU Regulation No. 862/2007 and inform priors for accuracy and "shock" variables.

Finally, the model's results could be further processed using the approach of [28] to stratify migration by gender and age, providing deeper insights into population change and migrant heterogeneity.

4 HMigD I App

The HMigD I App [8], a Shiny application, summarizes the assumptions and results of the Bayesian models for migration flow estimation. It offers multiple tools for data exploration, visualization, and download.

The app provides an overview of the input database, including administrative and EU LFS data, with detailed descriptions of sources, definitions, and data quality classifications. Users can explore data availability, investigate sources, and review covariates used in modeling.

For model comparison, the app enables users to contrast results from different model setups and assess data quality. Users can aggregate predicted flows under different scenarios (e.g., with or without EU LFS data or the freedom of movement variable) and visualize circular flow plots for selected countries and years.

Finally, the app allows users to select, combine, and download results for further analysis and provides a feedback feature for users to comment on specific flows.

5 Acknowledgements

This work was partially supported by the European Commission as part of the PREMIUM_EU project, which is funded under Horizon EUROPE Grant Agreement no. 101094345.

6 Figures

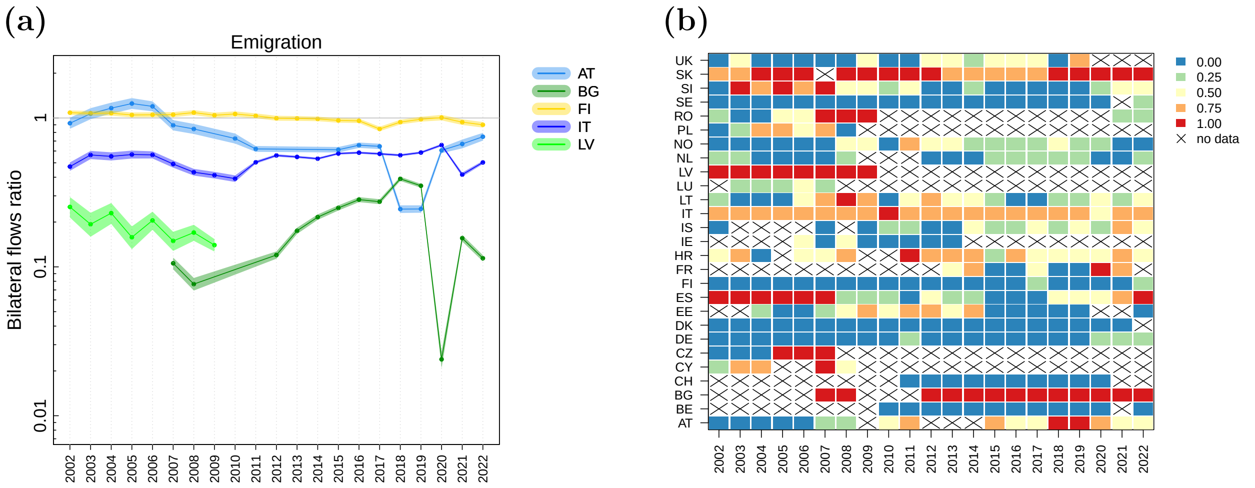

Figure 1. Assesment and classification of undercounting in migration flows. (a) Selected bilateral flow ratios for emigration data. The ratio is calculated by dividing flows from country X to a group of countries with good data quality (the reference group of countries) reported by country X by the flows in the same direction reported by the reference group of countries. Ratios higher than one indicate the overcounting of emigration flows, while ratios lower than one indicate the undercounting of emigration flows. The lower the ratio, the higher the level of undercounting. The 95% confidence intervals are calculated using the percentile bootstrap method.
(b) Undercounting scores predicted by the model for emigration data calculated by projecting bilateral flows ratios into discretized 0–1 scale (quantile method). It is not possible to calculate some results (denoted as "X") due to the lack of country-specific flows in the considered country or in the reference countries. To fill these gaps, the model offers PCA imputations (not shown). The both panels of the figure were generated using *UndercountMigScores* [9] at the default settings, but without PCA imputations.

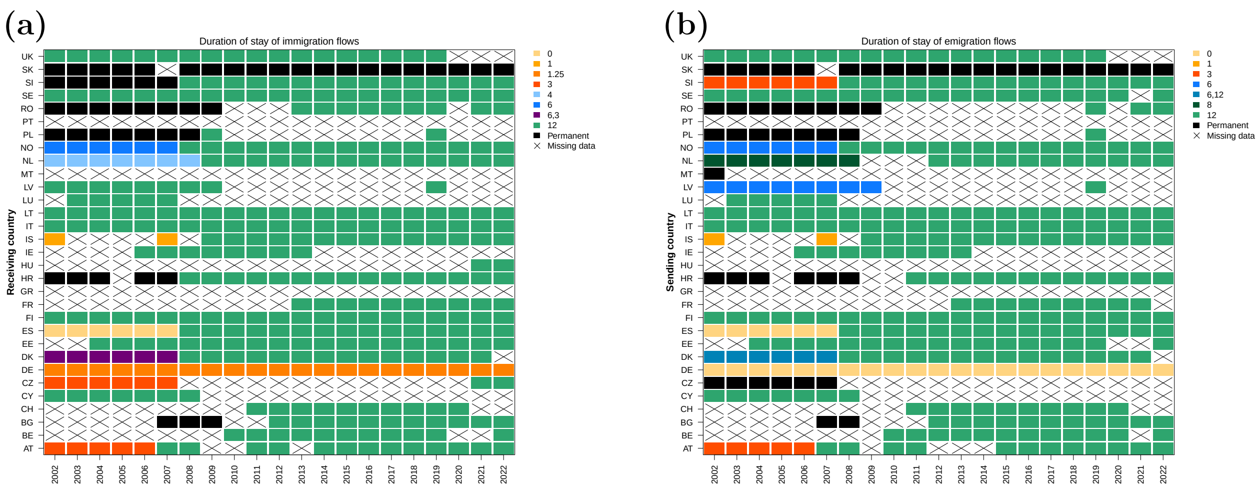

Figure 2. Duration of stay and availability of bilateral flows in the data used. (**a**) immigration (**b**) emigration. In the Danish (DK) data, "6, 3" means three months for immigration from Switzerland (CH) and six months for immigration from other countries, and "6, 12" means 12 months for emigration to Sweden (SE) or Finland (FI) and six months for emigration to other countries. In the case of Germany (DE), "1.25" is a mean duration of stay among different federal states.

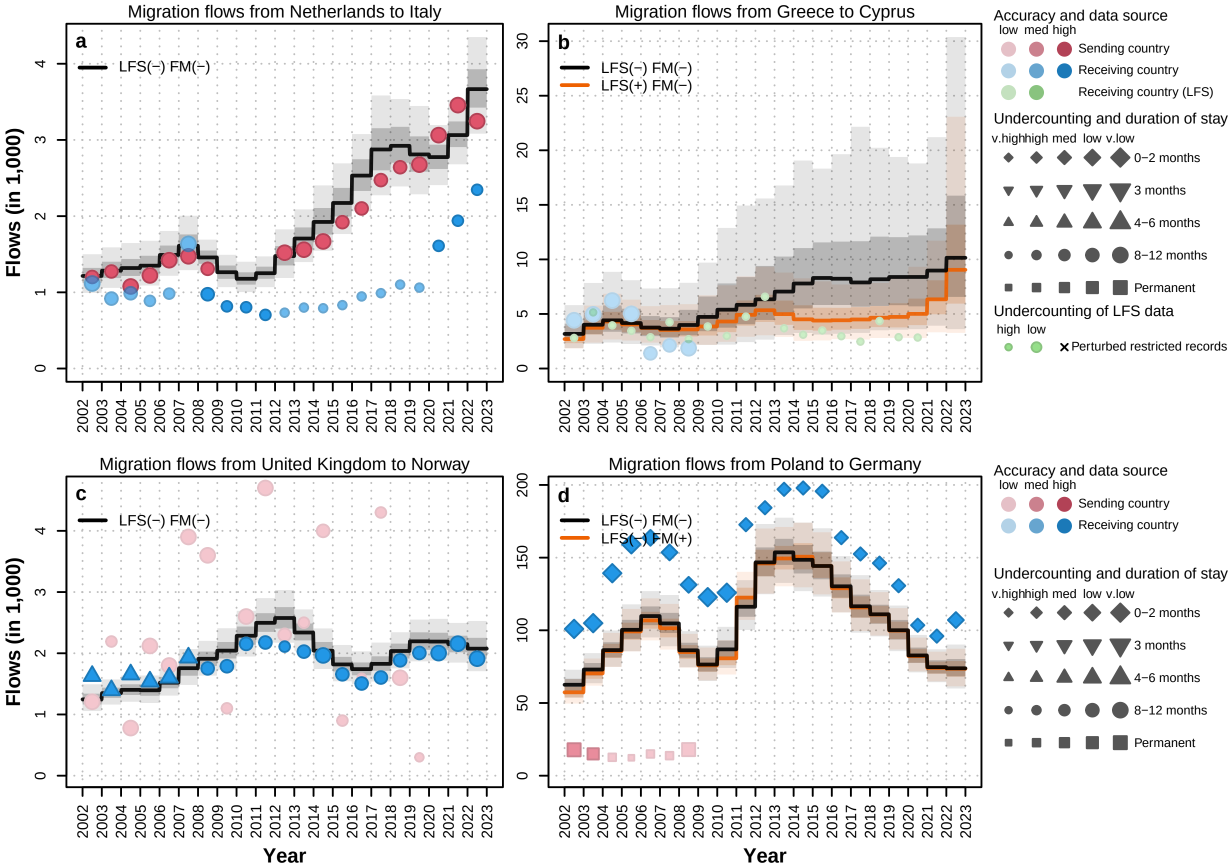

Figure 3. Four exemplary model-predicted flows and underlying data quality. The legend on the right side of the panel plot shows data quality codes using transparency (accuracy), size (undercounting), and shape (duration of stay). Orange lines indicate alternative models as described in the top-left legends.

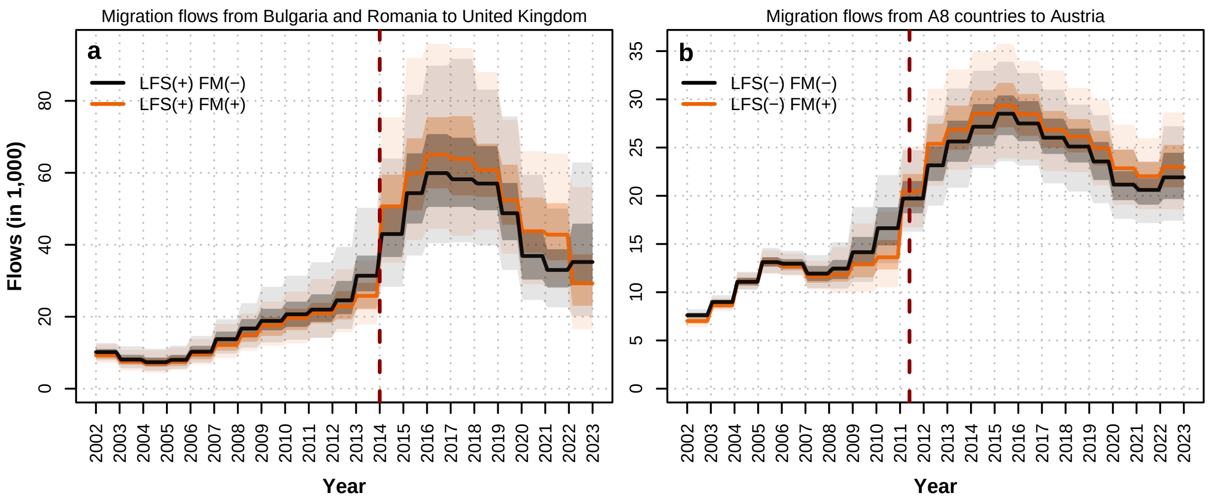

Figure 4. Aggregated flows for selected groups of countries and models. The vertical dashed line indicates the year when the freedom of movement of workers was granted.

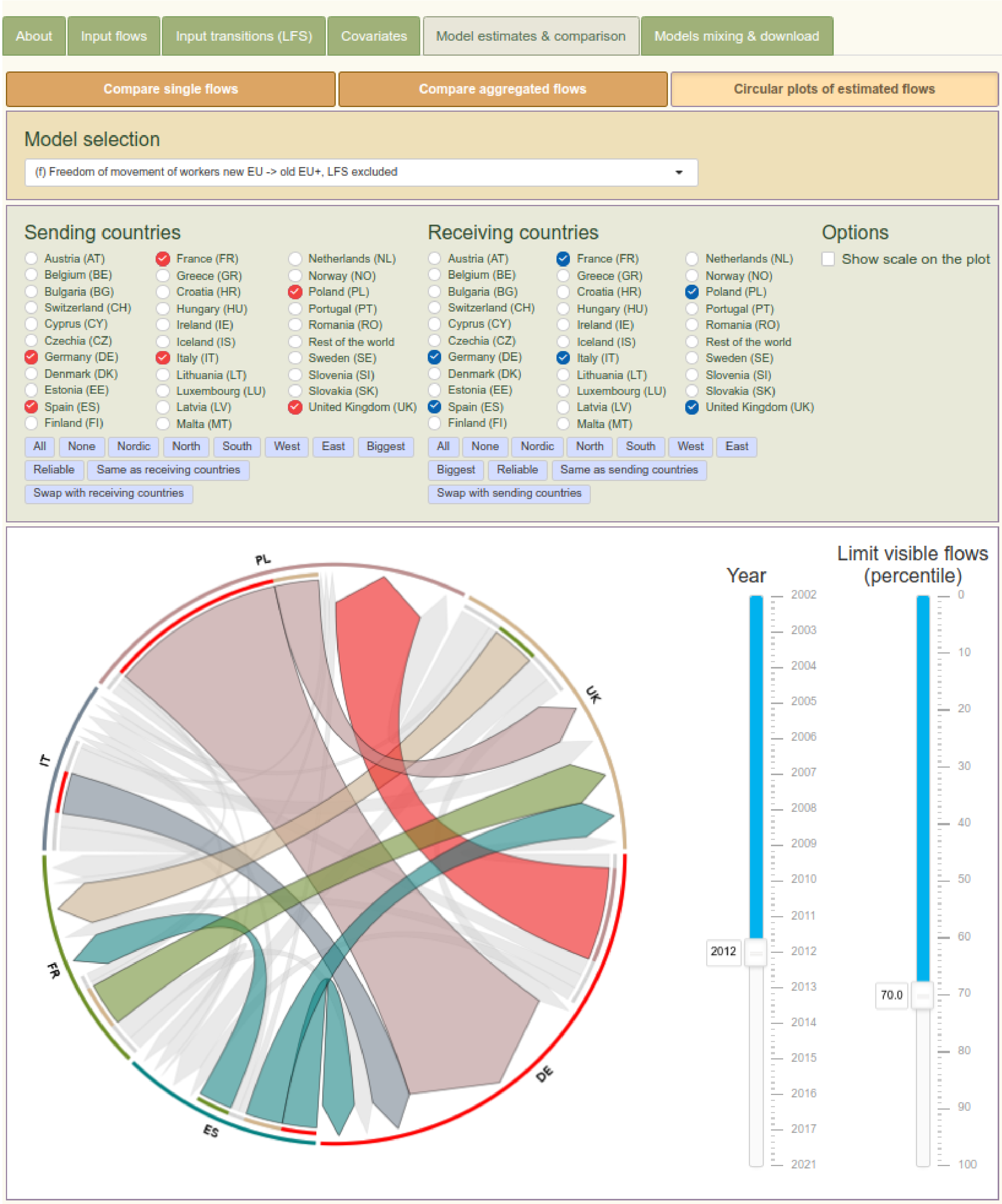

Figure 5. Exemplary screenshot of the HMigD I App showing a panel to plot circular migration plots.

References

[1] Abel, G. J. and Cohen, J. E. [2019], 'Bilateral international migration flow estimates for 200 countries', *Scientific Data* **6**(1), 1–13.

[2] Aparicio Castro, A., Wiśniowski, A. and Rowe, F. [2024], 'A bayesian approach to estimate annual bilateral migration flows for south america using census data', *Journal of the Royal Statistical Society Series A: Statistics in Society* **187**(2), 410–435.

[3] Aristotelous, G., Smith, P. W. F. and Bijak, J. [2022], *Technical report: Estimation methodology. QuantMig Project Deliverable 6.3.*, The Hague: Netherlands Interdisciplinary, Demographic Institute (NIDI-KNAW)/University of Groningen.
URL: http://quantmig.eu/res/files/QuantMig%20deliverable_6_3.pdf#page=11

[4] Azose, J. J. and Raftery, A. E. [2019], 'Estimation of emigration, return migration, and transit migration between all pairs of countries', *Proceedings of the National Academy of Sciences* **116**(1), 116–122.
URL: https://www.pnas.org/doi/abs/10.1073/pnas.1722334116

[5] Bürkner, P.-C. [2017], 'brms: An R Package for Bayesian Multilevel Models Using Stan', *Journal of Statistical Software* **80**, 1–28.

[6] Castles, S. [2017], 'Migration policies are problematic–because they are about migration', *Ethnic and racial studies* **40**(9), 1538–1543.

[7] Chi, G., Abel, G. J., Johnston, D., Giraudy, E. and Bailey, M. [2025], 'Measuring global migration flows using online data', *Proceedings of the National Academy of Sciences of the United States of America* .

[8] Dańko, M. J. [2024*a*], 'HMigD I App v4.1.1', *GitHub repository* .
URL: https://github.com/MaciejDanko/HMigD_Shiny_App_I

[9] Dańko, M. J. [2024*b*], 'Undercountmigscores v1.3.0. assessing the level of undercounting in the internationalmigration flows reported by Eurostat', *GitHub repository* . **URL:** https://github.com/MaciejDanko/UndercountMigScores.

[10] Dańko, M. J., Wiśniowski, A., Jasilionis, D., Jdanov, D. A. and Zagheni, E. [2024], 'Assessing the quality of data on international migration flows in Europe: The case of undercounting', *Migration Studies* **12**(2), mnae014.
URL: https://doi.org/10.1093/migration/mnae014

[11] Del Fava, E., Wiśniowski, A. and Zagheni, E. [2019], 'Modelling international migration flows by integrating multiple data sources', *SocArXiv* .
URL: osf.io/preprints/socarxiv/cma5h

[12] Gelman, A. [2006], 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian Analysis* **1**(3), 515–534.

[13] Gelman, A., Jakulin, A., Pittau, M. G. and Su, Y.-S. [2008], 'A weakly informative default prior distribution for logistic and other regression models', *The Annals of Applied Statistics* **2**(4), 1360 – 1383.
URL: https://doi.org/10.1214/08-AOAS191

[14] Human Mortality Database [2024], 'Human mortality database'. https://www.mortality.org.
URL: https://www.mortality.org

[15] IMEM [2012], 'Integrated modeling of european migration (IMEM)'. IMEM project funded by New Opportunities for Research Funding Agency Cooperation in Europe (NORFACE Migration).
URL: https://www.imem.cpc.ac.uk/

[16] Kupiszewska, D. and Nowok, B. [2008], Comparability of Statistics on International Migration Flows in the European Union, *in* 'International Migration in Europe', John Wiley & Sons, Ltd, pp. 41–71.

[17] Nowok, B. [2007], Evolution of International Migration Statistics in Selected Central European Countries, *in* J. Raymer and F. Willekens, eds, 'International Migration in Europe', John Wiley & Sons, Ltd, Chichester, UK, pp. 73–87.

[18] Nowok, B. [2010], Harmonization by simulation: a contribution to comparable international migration statistics in Europe, PhD thesis, University of Groningen, Groningen, Netherlands. OCLC: 955134815.

[19] Nowok, B. and Willekens, F. [2011], 'A probabilistic framework for harmonisation of migration statistics: Harmonisation of Migration Statistics', *Popul. Space Place* **17**, 521–533.

[20] Poulain, M., Perrin, N. and Singleton, A. [2006], *THESIM: Towards harmonised European statistics on international migration*, Presses univ. de Louvain.

[21] QuantMig [2020], *QuantMig: Quantifying Migration Scenarios for Better Policy*, The Hague: Netherlands Interdisciplinary, Demographic Institute (NIDI-KNAW)/University of Groningen. QuntMig project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 870299.
URL: https://www.quantmig.eu

[22] Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F. and Bijak, J. [2013], 'Integrated Modeling of European Migration', *J. Am. Stat. Assoc.* **108**, 801–819.

[23] Rees, P. H. [1977], 'The measurement of migration, from census data and other sources', *Environment and Planning A* **9**(3), 247–272.

[24] The European Parliament and the European Council [2007], 'Regulation (EC) No 862/2007 of the European Parliament and of the Council', https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32007R0862.

[25] Willekens, F. [2019], 'Evidence-Based Monitoring of International Migration Flows in Europe', *J. Off. Stat.* **35**, 231–277.

[26] Willekens, F., Massey, D., Raymer, J. and Beauchemin, C. [2016], 'International migration under the microscope', *Science* **352**(6288), 897–899.

[27] Wiśniowski, A. [2017], 'Combining Labour Force Survey data to estimate migration flows: the case of migration from Poland to the UK', *J. R. Stat. Soc. Ser. A. Stat. Soc.* **180**, 185–202.

[28] Wiśniowski, A., Forster, J. J., Smith, P. W. F., Bijak, J. and Raymer, J. [2016], 'Integrated modelling of age and sex patterns of European migration', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **179**(4), 1007–1024.

APPENDIX

A The Hierarchical Bayesian Modeling Approach

A.1 Measurement Error Model for Administrative Data

The measurement error model estimates the true number of *relocations* (i.e., all changes of usual residence, regardless of duration) while addressing inconsistencies in data sources related to undercounting, coverage, and accuracy. The model is tailored to the characteristics and limitations of each data source.

Migration flow data are collected from administrative sources, including population and foreigner registers and sample surveys, with primary sources from Eurostat, the UN, and national statistical offices (see supplementary HMigD I App [8]). The data represent migration events, x^k_{ijt}, showing flows from country i to j in year t, from source $k \in \{IR, ER\}$, where IR refers to immigration by previous residence and ER to emigration by next residence.

We assume that the count data follow a Poisson distribution:

$$x^k_{ijt} \sim \text{Poisson}\left(\lambda^k_{ijt}\right), \tag{1}$$

where λ^k_{ijt} is the expected number of migration events, which follows a log-normal distribution:

$$\log \lambda^{IR}_{ijt} \sim \text{normal}\left(\log Y^R_{ijt} + \delta_{ijt} + \log \upsilon^{IR}_{jt} - \log(1 + \exp(-\kappa^{IR}_j)), \tau^{IR}_{jt}\right), \tag{2}$$

$$\log \lambda^{ER}_{ijt} \sim \text{normal}\left(\log Y^R_{ijt} + \delta_{ijt} + \log \upsilon^{ER}_{it} - \log(1 + \exp(-\kappa^{ER}_i)), \tau^{ER}_{it}\right). \tag{3}$$

The true number of relocations, Y^R_{ijt}, is derived from:

$$Y^R_{ijt} = R_{ijt} \exp(-\mu_{,j,t} d^m_j), \tag{4}$$

where $R_{ijt} = \mu_{ijt} N_{it}$, with μ_{ijt} as the relocation rate, N_{it} as the origin-country population, and $\exp(-\mu_{,j,t} d^m_j)$ adjusting for the minimum duration of stay d^m_j [18].

We rewrite Equations (2–4) as:

$$\log \lambda^{IR}_{ijt} \sim \text{normal}\left(\log R_{ijt} - \mu_{,j,t} d^m_j + \delta_{ijt} + \log \upsilon^{IR}_{jt} - \log(1 + \exp(-\kappa^{IR}_j)), \tau^{IR}_{jt}\right), \tag{5}$$

$$\log \lambda^{ER}_{ijt} \sim \text{normal}\left(\log R_{ijt} - \mu_{,j,t} d^m_j + \delta_{ijt} + \log \upsilon^{ER}_{it} - \log(1 + \exp(-\kappa^{ER}_i)), \tau^{ER}_{it}\right). \tag{6}$$

We use Bayesian methods to estimate parameters for duration of stay, undercounting, coverage, and accuracy, specifying prior distributions to incorporate prior knowledge. Posterior distributions are summarized using medians and credible intervals.

The duration of stay, d^m_j, varies from 0 (no limit) to 5 years [17]. The parameter δ_{ijt} adjusts for the minimum duration of stay affecting migration counts. Parameters υ^{IR}_{jt} and υ^{ER}_{it} address undercounting bias (0 = high undercounting, 1 = low undercounting), using non-informative priors based on bilateral flow ratios [9; 10]. Parameters κ^{IR}_j and κ^{ER}_i measure coverage bias (0 = low coverage, 1 = high coverage), modeled with weakly informative priors [5]. Parameters τ^{IR}_{jt} and τ^{ER}_{it} reflect the accuracy of data collection systems, with higher precision indicating lower variability in $\log \lambda^k_{ijt}$ [12].

A.2 Measurement Error Model for Survey Data

The EU LFS tracks transitions over a year, assuming individuals made at most one relocation in that period. Observed transitions should therefore match the actual number of migration events, consistent with administrative data.

We model the transition counts, x_{ijt}^{IS} (survey data in a country of destination), using a Poisson distribution with parameter λ_{ijt}^{IS}, which follows a log-normal distribution:

$$\log \lambda_{ijt}^{IS} \sim \text{normal}\left(\log R_{ijt} + \log\left(1 - \exp\left(-2\sum_{i\neq j}\mu_{jit}\right)\right) - \log\left(2\sum_{i\neq j}\mu_{jit}\right) + \log \upsilon_{jt}^{IS} - \log\left(1+\exp(-\kappa_j^{IS})\right),\right.$$

The parameter λ_{ijt}^{IS} is related to the true number of transitions Y_{ijt}^{S} by:

$$\frac{Y_{ijt}^S}{Y_{ijt}^R} = \frac{\exp(\mu_{,j,t}d_j^m)(1-\exp(-2\mu_{,j,t}))}{2\mu_{,j,t}}, \tag{8}$$

where $Y_{ijt}^R = R_{ijt}\exp(-\mu_{,j,t}d_j^m)$.

Parameter υ_{jt}^{IS} accounts for undercounting in surveys, calculated as $1-(1-p^{nr})(1-p^{mi})$, where p^{nr} is the non-response rate and p^{mi} the missing migration-data rate. We use Beta priors for low and high undercounting scenarios: υ_1^S (Beta$(47.39, 6.28)$) and υ_2^S (Beta$(11.42, 5.18)$).

The parameter κ_j^{IS} measures population coverage bias. Countries are categorized as having "standard" or "excellent" coverage, and we apply the same priors as for administrative data. Precision τ_{jt}^{IS} reflects survey accuracy, classified as "high" or "low" based on the average coefficient of variation (CV). We use Gamma priors: τ_1^S (Gamma$(3.2674, 0.0185)$) and τ_2^S (Gamma$(3.2217, 0.1305)$). Accuracy is adjusted for undercounting, ensuring that high undercounting corresponds to low accuracy.

A.3 Predictive Model

The predictive model forecasts missing migration flows, smooths data, and captures time trends. Unlike measurement error models, which are tailored to each data source, this model estimates true migration flows consistently across sources. The true latent number of relocations, R_{ijt}, follows a log-normal distribution:

$$\log R_{ijt} \sim \text{normal}\left(\beta_1 + \beta_2 A_{ijt} I_{i,j}^{E\to W} + \sum_{k=1}^{12} b_{k,i,j} Z_{t,k} + \gamma_{ij}, \omega_R\right), \tag{9}$$

where β_1 and β_2 are regression coefficients with $t_3(0, 2.5)$ priors [13]. A_{ijt} indicates freedom of movement, $I_{i,j}^{E\to W}$ selects flows from new to old EU/EFTA countries, γ_{ij} represents corridor-specific random intercepts ($\Gamma(0.1, 0.1)$ prior), $b_{kij}Z_{t,k}$ models time effects using cubic B-splines with weakly informative priors, and ω_R is a precision parameter ($\Gamma(0.01, 0.01)$ prior).

The smoothing component captures temporal flow patterns between country pairs but may miss sudden migration shocks due to policy or economic changes. To address this, we include a migration shock variable—freedom of movement of workers. No socioeconomic (gravity) variables are added, allowing subsequent analyses to incorporate such factors independently.

We are establishing a network of national experts to better understand the impact of legislation and other factors on migration flows. Their insights will guide model refinements and the inclusion of additional shock variables relevant to changing socio-political contexts.

A.4 Predicting Migration Flows

Accurately predicting migration flows is essential for understanding demographic changes and informing policy decisions. Our model predicts migration flows based on a minimum 12-month duration of stay, addressing data quality issues such as accuracy and undercounting, and providing consistent estimates even for incomplete data. Consequently, predictions are comparable across migration corridors.

The calculation of predicted flows is realized using:

1. The true latent relocation rate:
$$\mu_{ijt} = \frac{R_{ijt}}{N_{it}}, \tag{10}$$
2. The estimated number of migration flows with a minimum 12-month stay:
$$Y_{ijt}^{12} = \mu_{ijt} N_{it} \exp\left(-d_{jt}^m \mu_{*,j,t}\right), \tag{11}$$
where $d_{jt}^m = 1$ for all countries and years, and $\mu_{*,j,t} = \sum_{i\neq j}\mu_{jit}$.