Probabilistic Population Projections: Theory

Hana Ševčíková

University of Washington, hanas@uw.edu

CSDE Workshop:
Part I (Theory): October 7th, 2025
Part II (Practice): October 9th, 2025 (https://github.com/PPgp/CSDE2025workshop)

Joint work with Adrian E. Raftery, Crystal Yu, Sara Curran, Patrick Gerland

Research supported by NICHD

Motivation

- Populations projections provide predictions of future population sizes for an area.
- Used by governments, private sector, researchers for variety of purposes (planning, decision making, health and social science research etc).
- Traditionally, they are produced deterministically using the cohort-component method (CCM) with no statements of uncertainty.
- Since 2015, the UN Population Division produces probabilistic population forecasts on national level using Bayesian models.
- More recently, methods extended to probabilistic subnational projections (states, counties).

Motivation (cont.)

- Approach – generate probabilistic projections of the individual components of the CCM:
  total fertility rate (TFR), life expectancy at birth (e_0), migration
  $\longrightarrow$ **probabilistic population projection**
- The WA Office of Financial Management (OFM) is actively exploring the method for the use in the next official forecast of population in WA counties.

Probabilistic Population Projection

- Demographic balancing equation:

$$Pop_t = Pop_{t-1} + Births_t - Deaths_t + Net\ Migration_t$$

  - solved using Cohort Component Method (CCM), while disaggregating the components into sex- and age-specific counts and rates

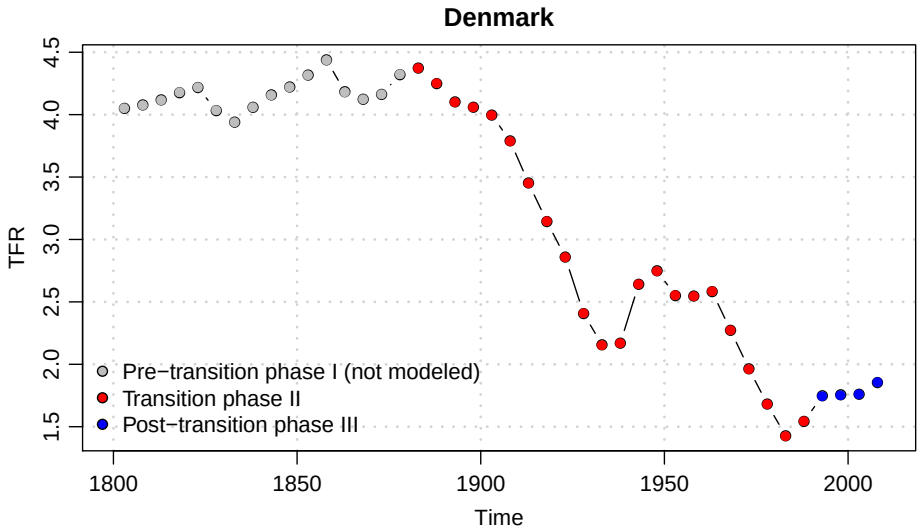
- **Probabilistic approach:**
  1. Using probabilistic models, generate sets of trajectories for TFR, e_0 and net migration.
  2. Convert each trajectory to age-specific rates and counts.
  3. Apply the CCM to each trajectory of age-specific fertility and mortality rates and migration counts.
  4. Result: Future trajectories of age- and sex-specific population quantities.

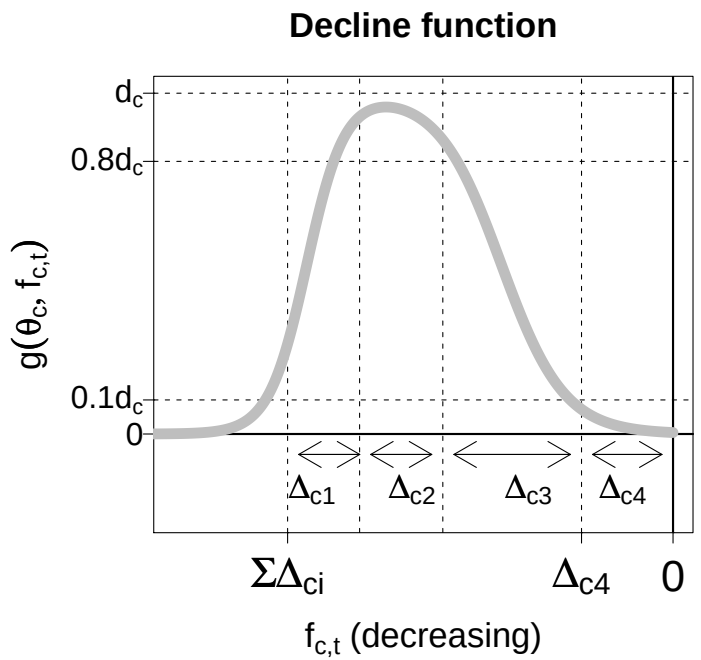
TFR and e_0: National vs. Subnational Model Objectives

- There are established Bayesian Hierarchical Models (BHM) for projecting TFR (Alkema et al. 2011) and e_0 (Raftery et al. 2013) that the UN uses for projections on national level.

- **Objectives for subnational models:**
  - Probabilistic approach built on established national BHM;
  - Works well for all regions of all countries;
  - Yields correlation between regions similar to observed data;
  - Is easy to use, understand and implement.

Modeling national TFR (Alkema et al. 2011)

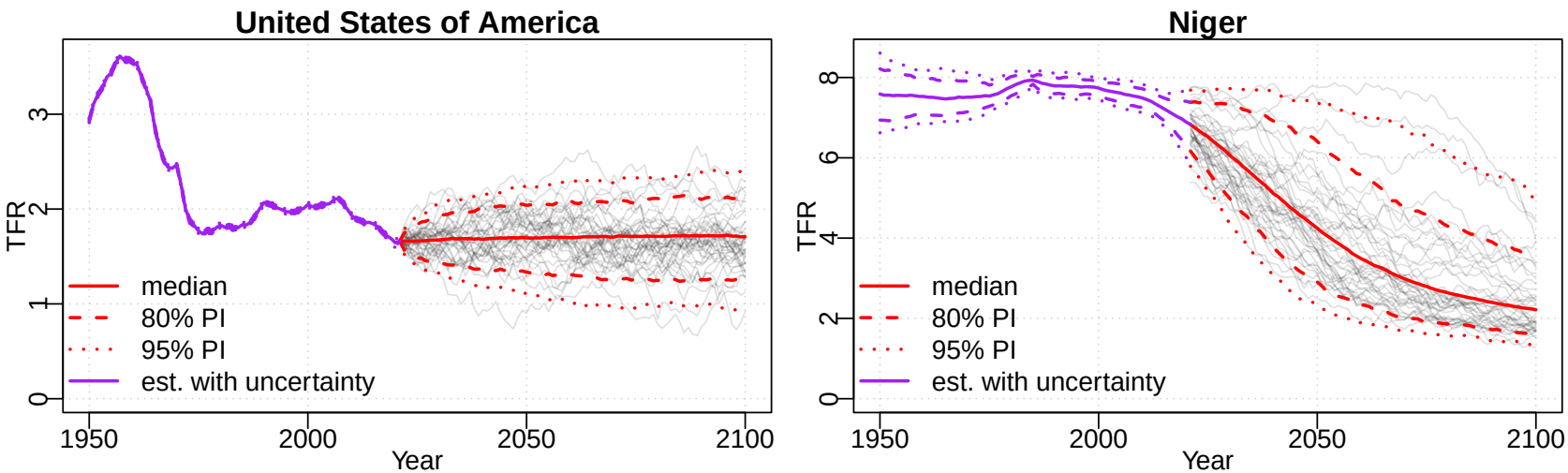




- Phase II: modelled via a BHM for double-logistic decline function.
- Phase III: BHM based on AR(1) with country-specific long-term means
- hierarchy: country's time points $\rightarrow$ country $\rightarrow$ world
- Result: Country-specific decline curves and AR1 curves
  $\longrightarrow$ Set of future TFR trajectories

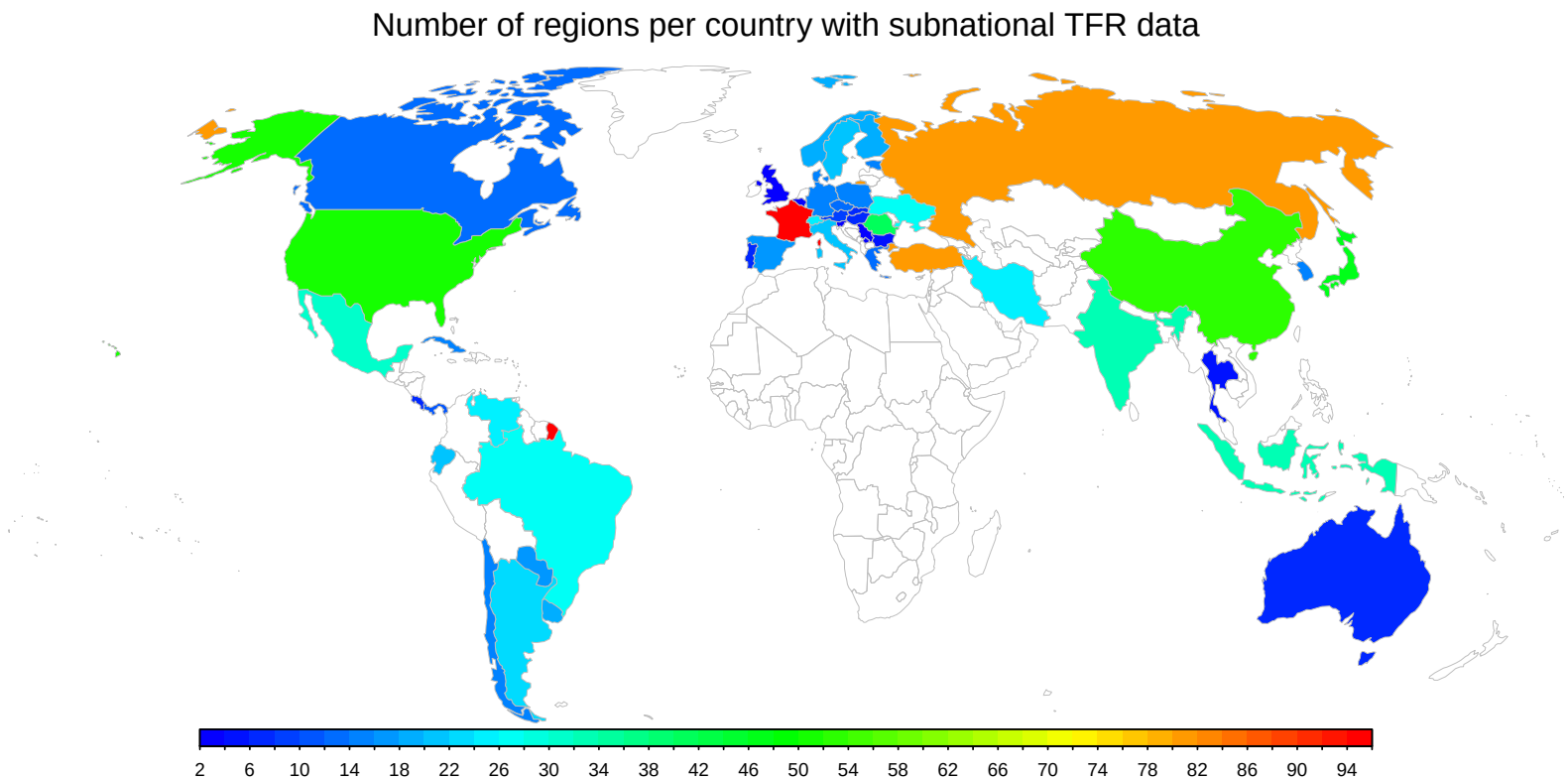
National TFR

- Method adopted by the UN for their official fertility projections since the *World Population Prospects* (WPP) 2012.
- Assessing uncertainty about the past data (Liu & Raftery 2020); adopted in WPP 2022
- Implemented in bayesTFR (Liu, Ševčíková, Raftery 2023, *JSS*)

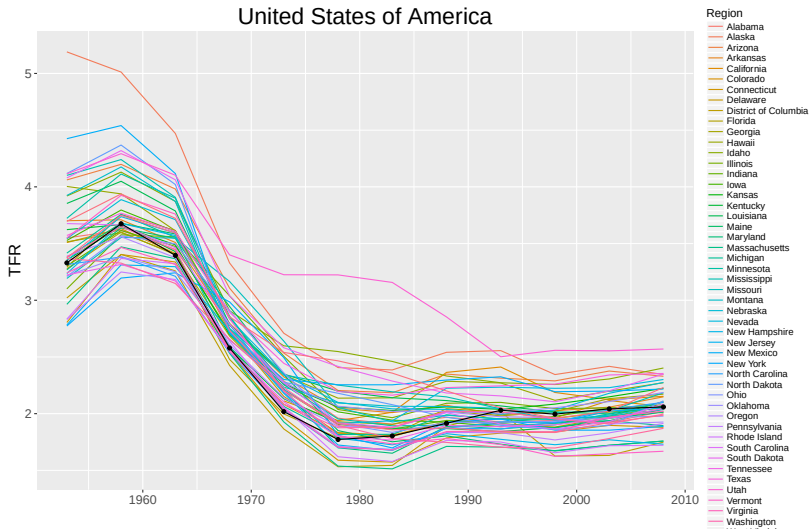


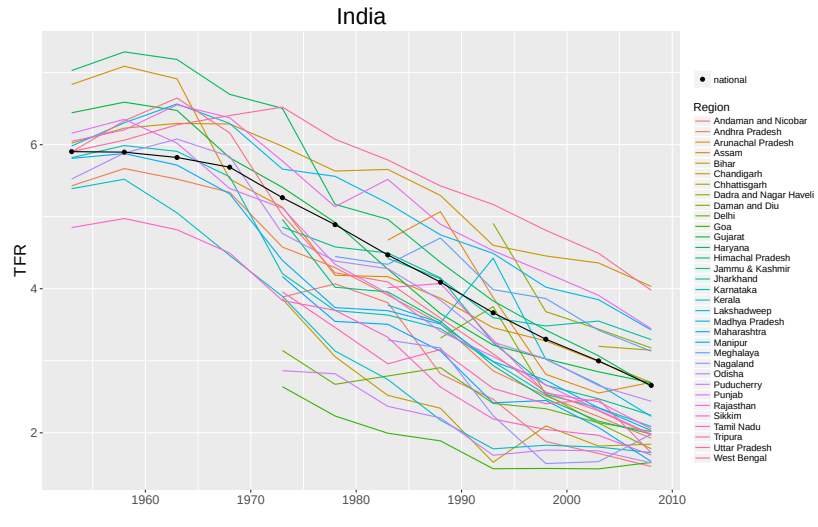
Subnational TFR: Data

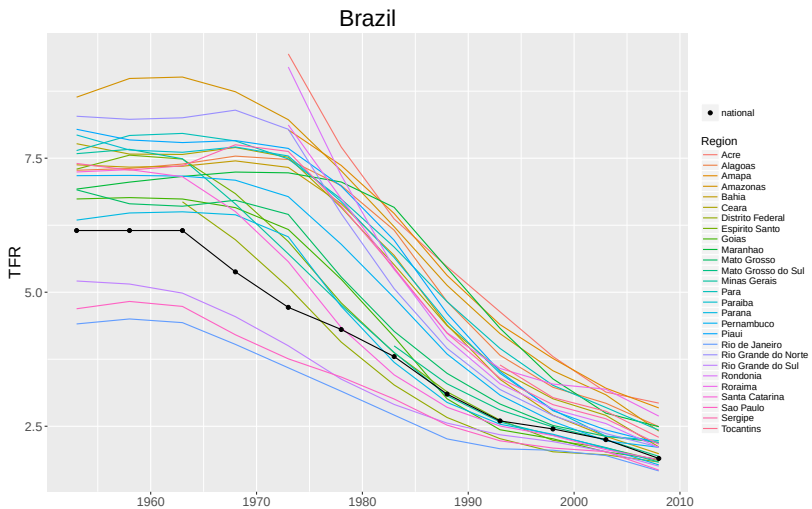
- Subnational TFR data collected by the UN, from 1950–2010: 47 countries, 1 092 regions.

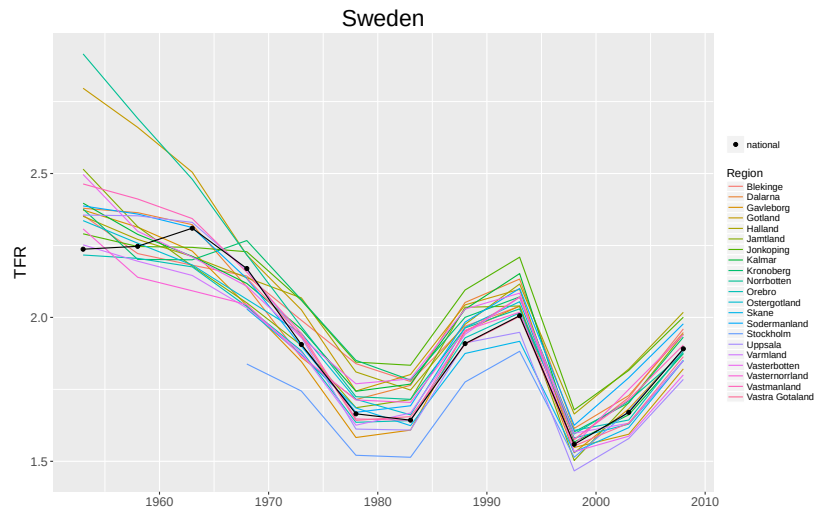


Data Examples









Method for Subnational TFR (Ševčíková et al., 2018)

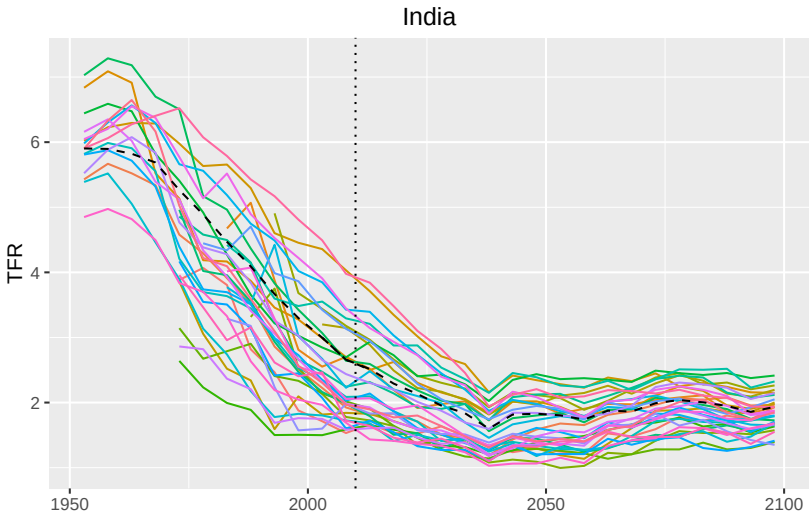
Scale-AR(1)

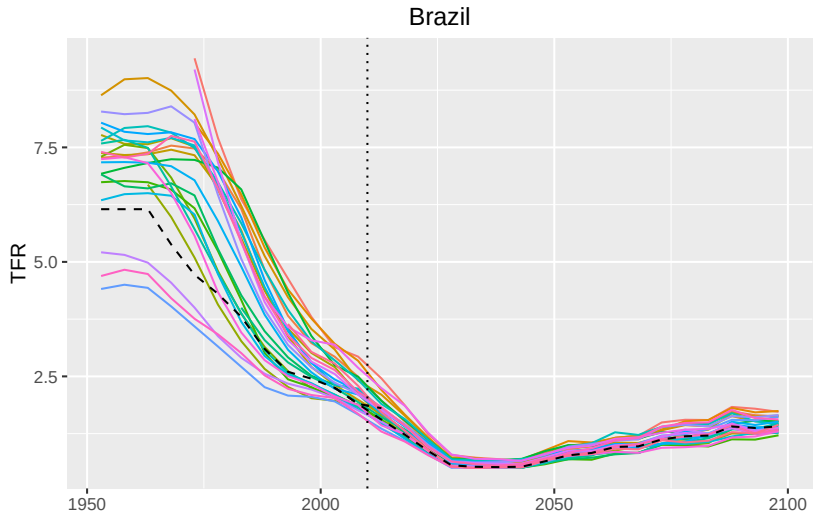
- Multiply each national trajectory by a region-specific scaling factor that changes slowly over time.

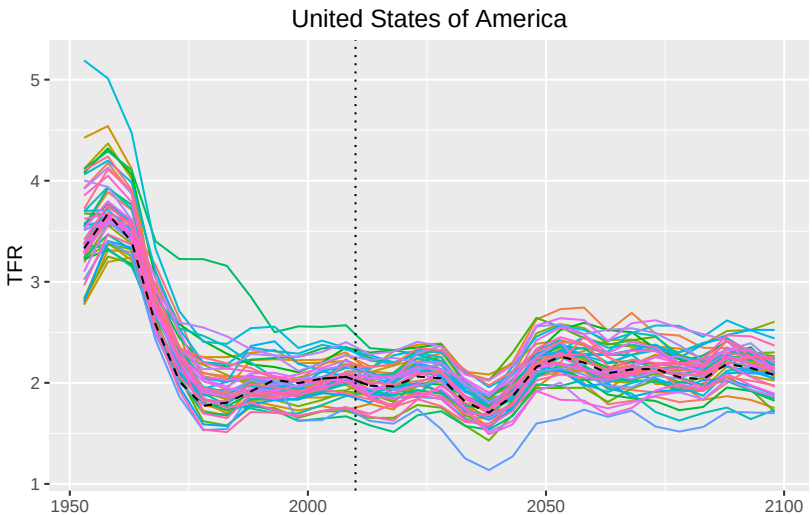
$$
\begin{aligned}
f^{(R)}_{r_c,t,i} &= \alpha_{r_c,t} f^{(C)}_{c,t,i} \\
\alpha_{r_c,t} - 1 &= \phi(\alpha_{r_c,t-1} - 1) + \varepsilon_{r_c,t}, \quad \text{with } \varepsilon_{r_c,t} \overset{iid}{\sim} N(0, \sigma^2_c)
\end{aligned}
$$

$f^{(R)}_{r_c,t,i}$	TFR for region r_c of country c at time t in the i-th trajectory
$f^{(C)}_{c,t,i}$	national TFR projection for country c at time t in the i-th trajectory (output of national BHM)
$\alpha_{r_c,t}$	regional scaling factor; initialized as $\alpha_{r_c,P} = f^{(R)}_{r_c,P} / f^{(C)}_{c,P}$ (P is the last observed time period) converges to a distribution centered around one
ϕ	determines the rate of convergence; $\hat{\phi} = 0.925$ (slow)

Results: 1 trajectory out of 2000





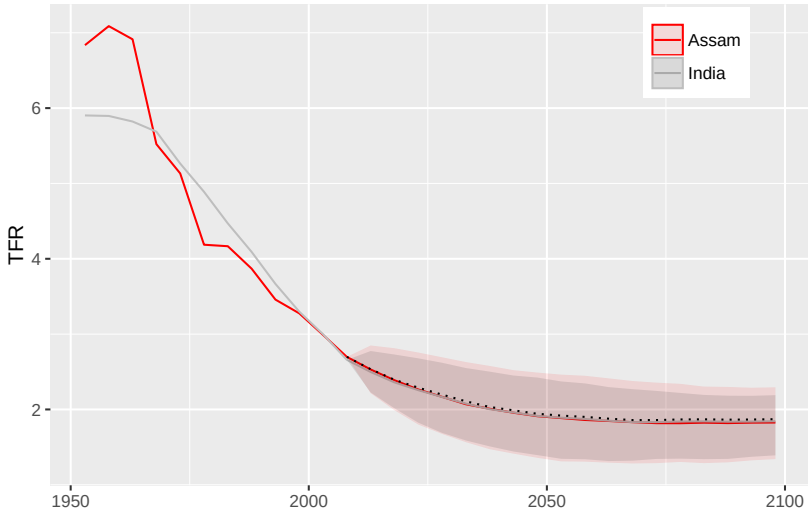


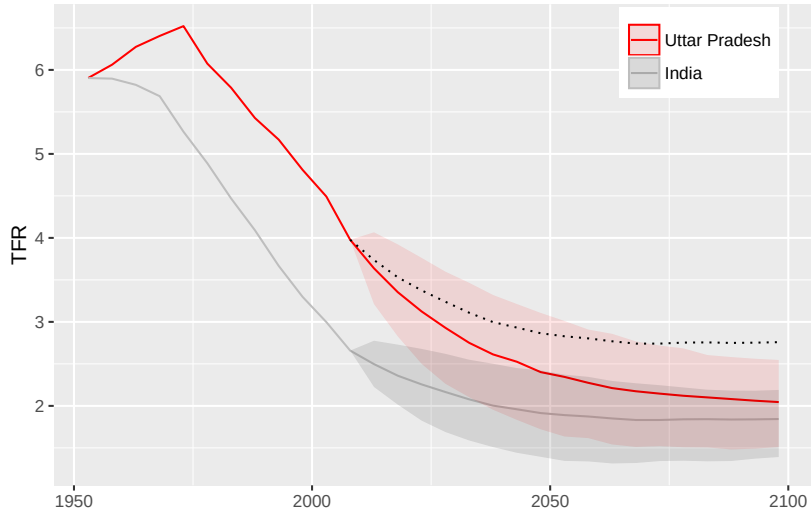
Frequency of TFR crossovers between regions within 30 years:

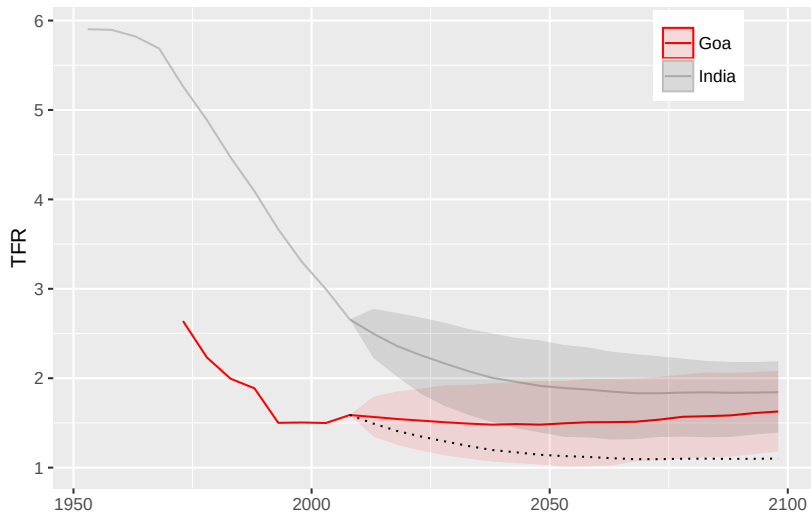
- Observed (1980–2010): 24%
- Predicted (2010–2040): 23%

Results: Projections for Scale-AR(1)

Three regions of India





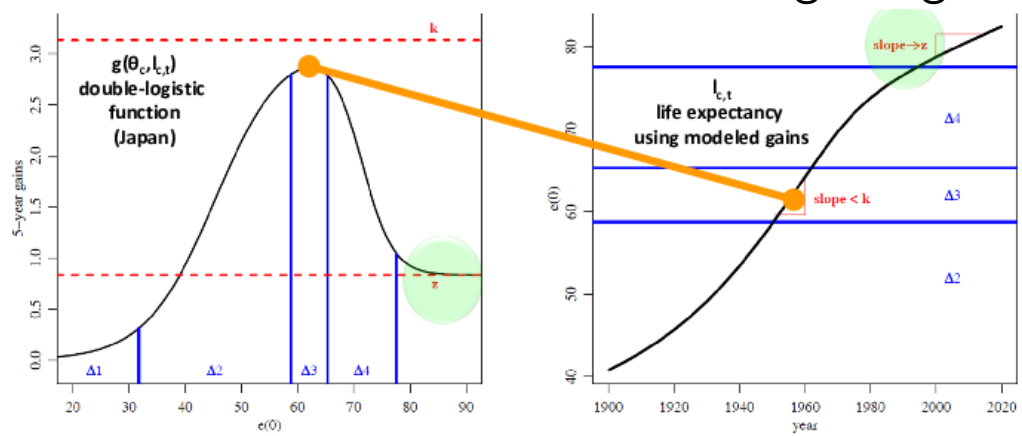


Summary of Probabilistic TFR

- **National model:**
  Use BHM by Alkema et al. (2011) that
  - estimates different phases of the fertility transition, while sharing information between countries, and
  - generates a set of future TFR trajectories for each country.
- **Subnational model:**
  Use Scale AR(1) by Ševčíková et al. (2018) to scale national TFR trajectories to subnational TFR trajectories.
  - inherits uncertainty from the national projection, while adding slightly more uncertainty contributed by the Scale AR(1).
- Implemented in **bayesTFR** R package.

Life expectancy at birth (e_0)

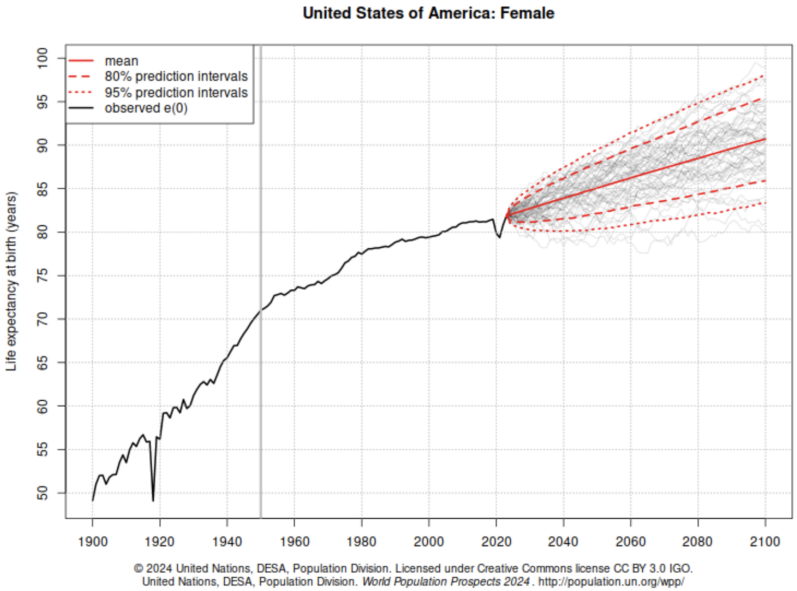
- e_0 reflects overall mortality level of a population.
- The UN Population Division projects probabilistic e_0 for all countries on the national level (Raftery et al. 2013, 2014)
  - Female: Modeled via a BHM for double-logistic gain function.

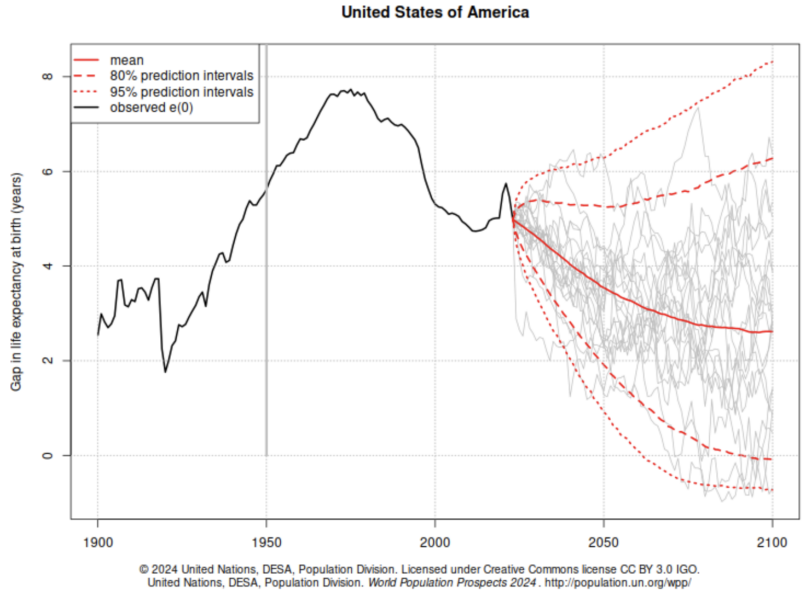


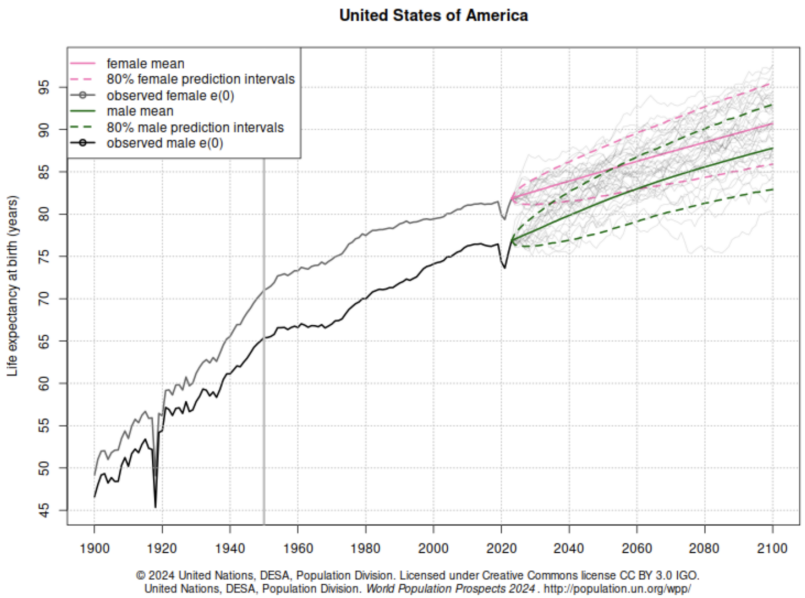
  - Male: Modeling gap between female and male via a lin. regression.
  - Result: Joint distribution of future female and male e_0

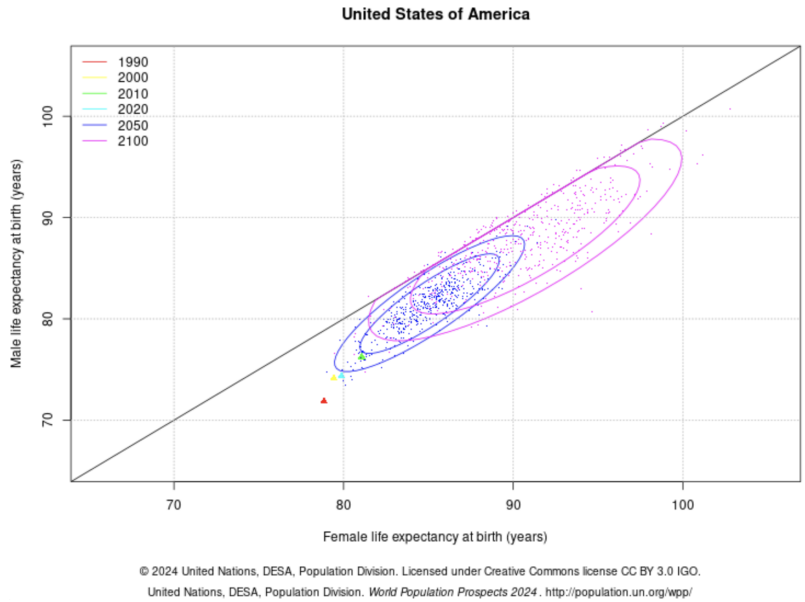
- Implemented in **bayesLife** R package.

National e_0: Results Example



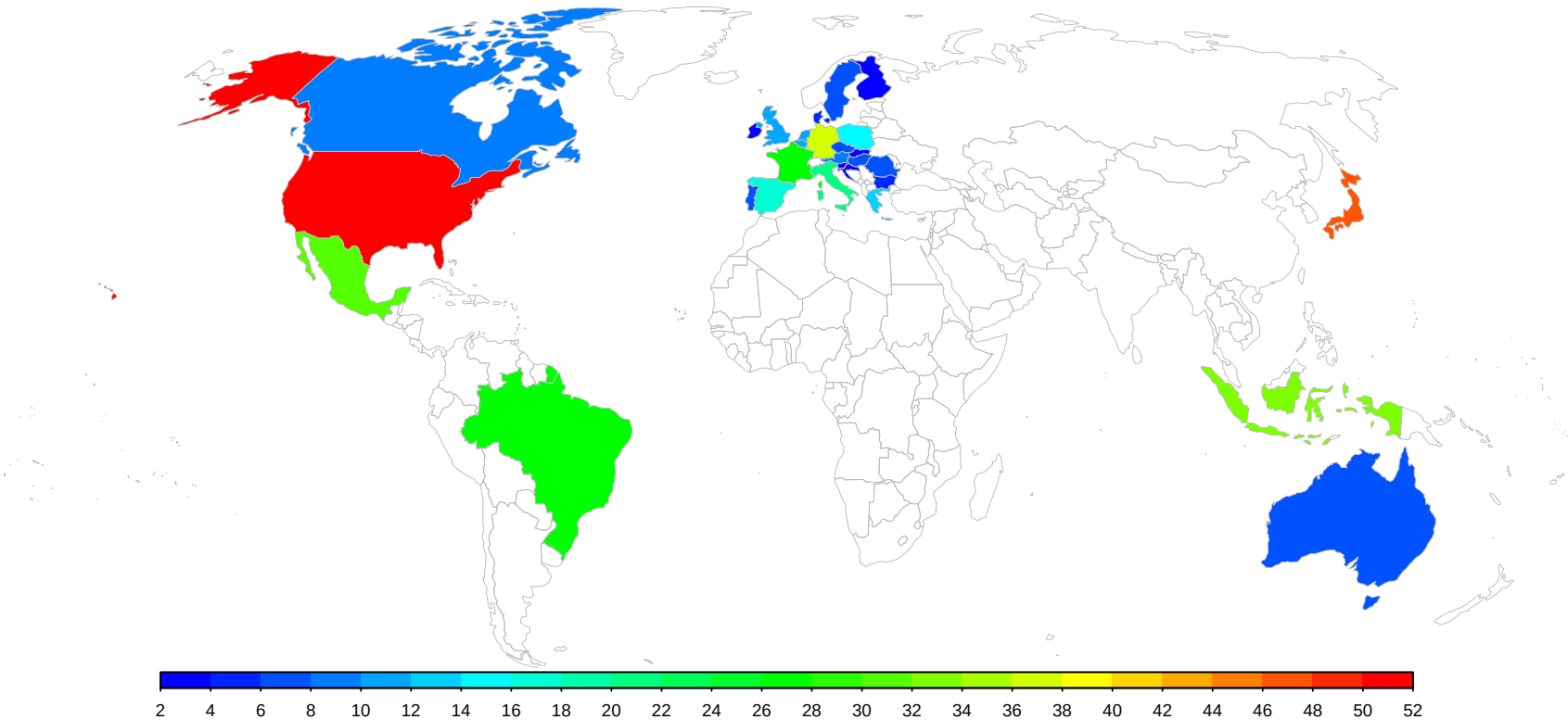




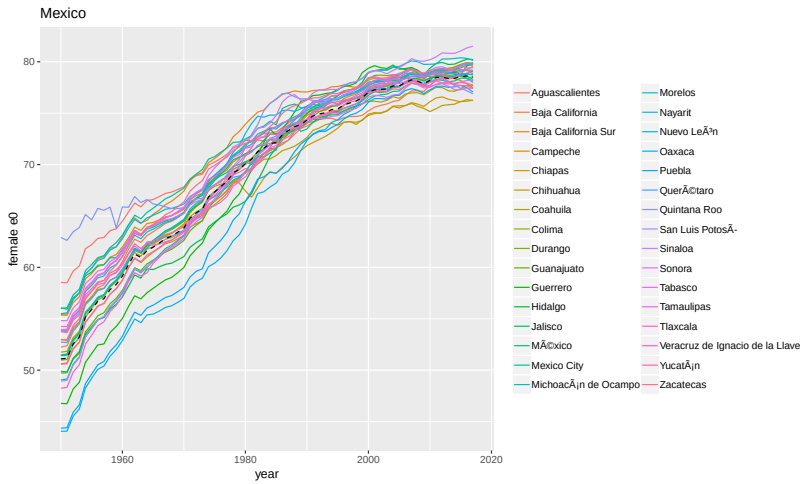


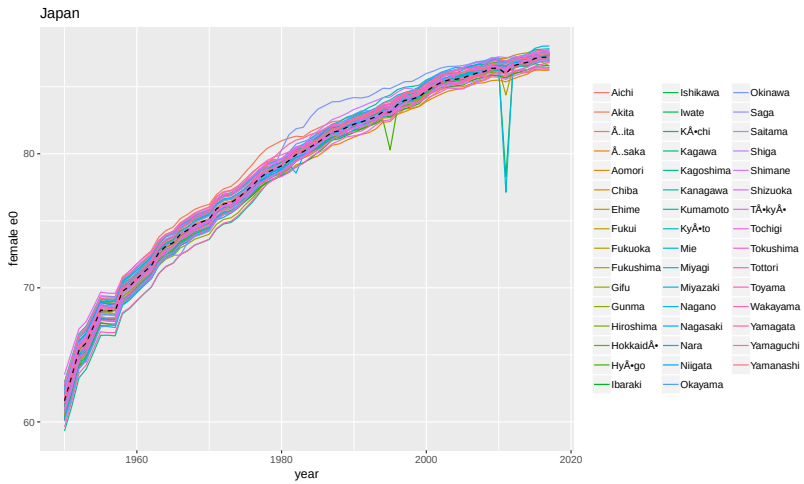
Subnational e_0: Data

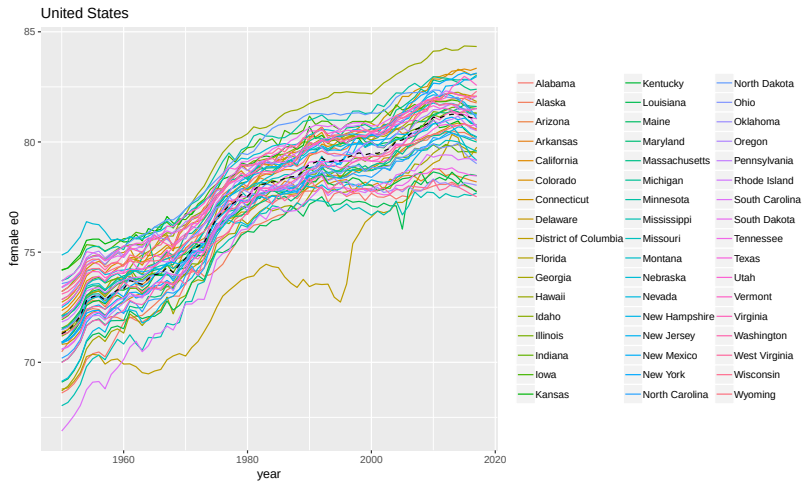
Data from 28 countries, 447 regions (EuroStat, IHME, StatCan, ABS, UN)



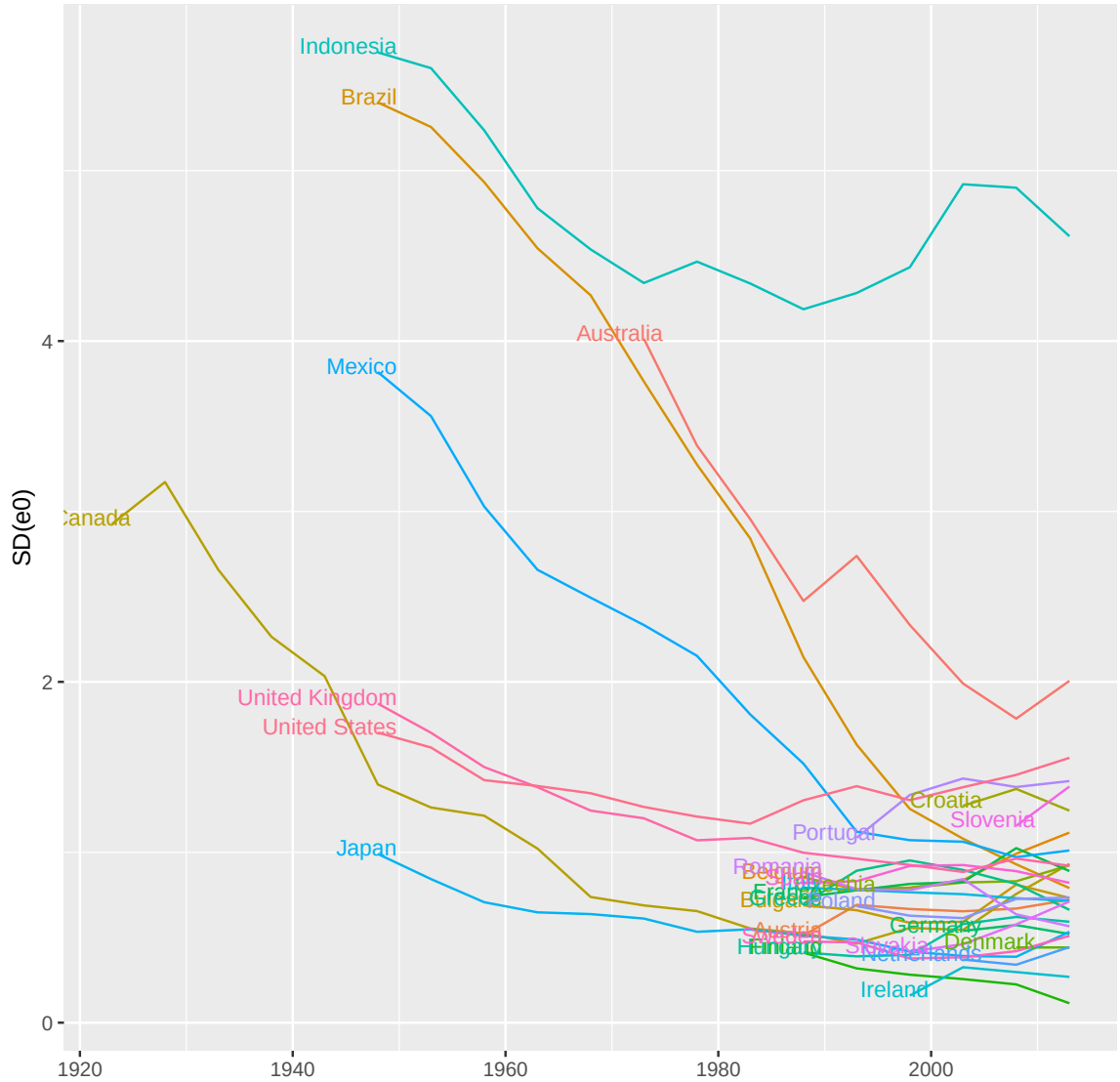
Data Examples







Within-country between-region standard deviation



Method for subnational e_0 (Ševčíková & Raftery, 2021)

Shift-AR(1) method

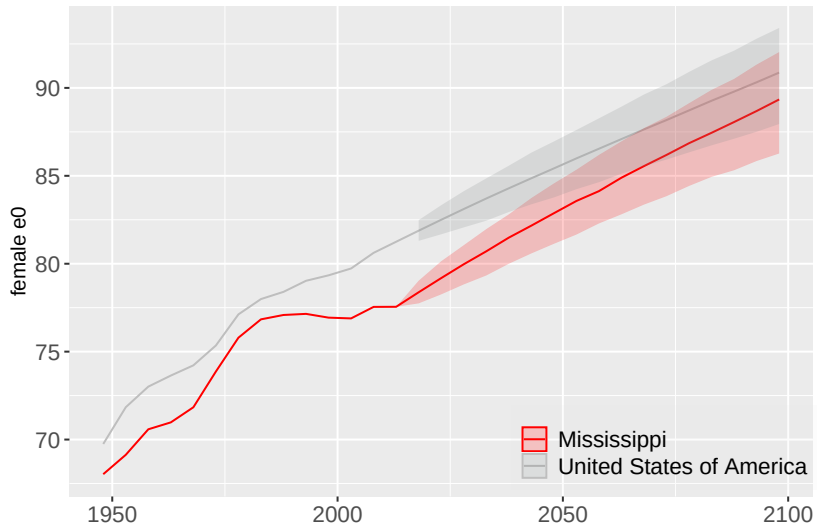
$$
\begin{aligned}
e^{(R)}_{r,c,t} &= e^{(C)}_{c,t} + \alpha_{r,c,t} \\
\alpha_{r,c,t} &= \rho\alpha_{r,c,t-1} + \varepsilon_{r,c,t}, \quad \text{with } \varepsilon_{r,c,t} \overset{\text{ind}}{\sim} N(0, \sigma^2_{c,t})
\end{aligned}
$$

- $\alpha_{r,c,t}$ is time dependent and additive, modeled via an AR(1)
- ρ is constant across countries and regions
- Residual variance $\sigma^2_{c,t}$ varies between countries and is defined as

$$
\sigma^2_{c,t} = \begin{cases} a + b(e^{(C)}_{c,t} - U) & e^{(C)}_{c,t} < U \\ a & e^{(C)}_{c,t} \geq U \end{cases}
$$

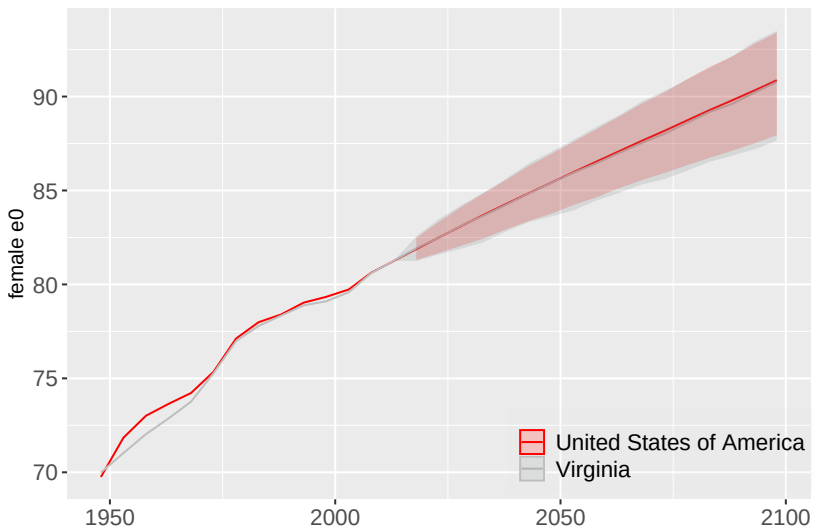
- Using observed data, estimation yields
  $\hat{\rho} = 0.95$, $\hat{U} = 82.5$, $\hat{a} = 0.0482$, and $\hat{b} = -0.0154$

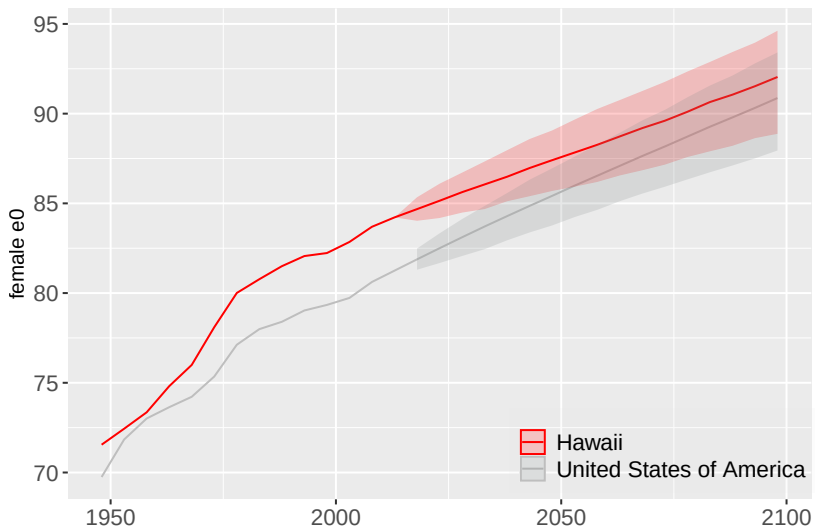
Subnational predictive distribution of female e_0 (USA)



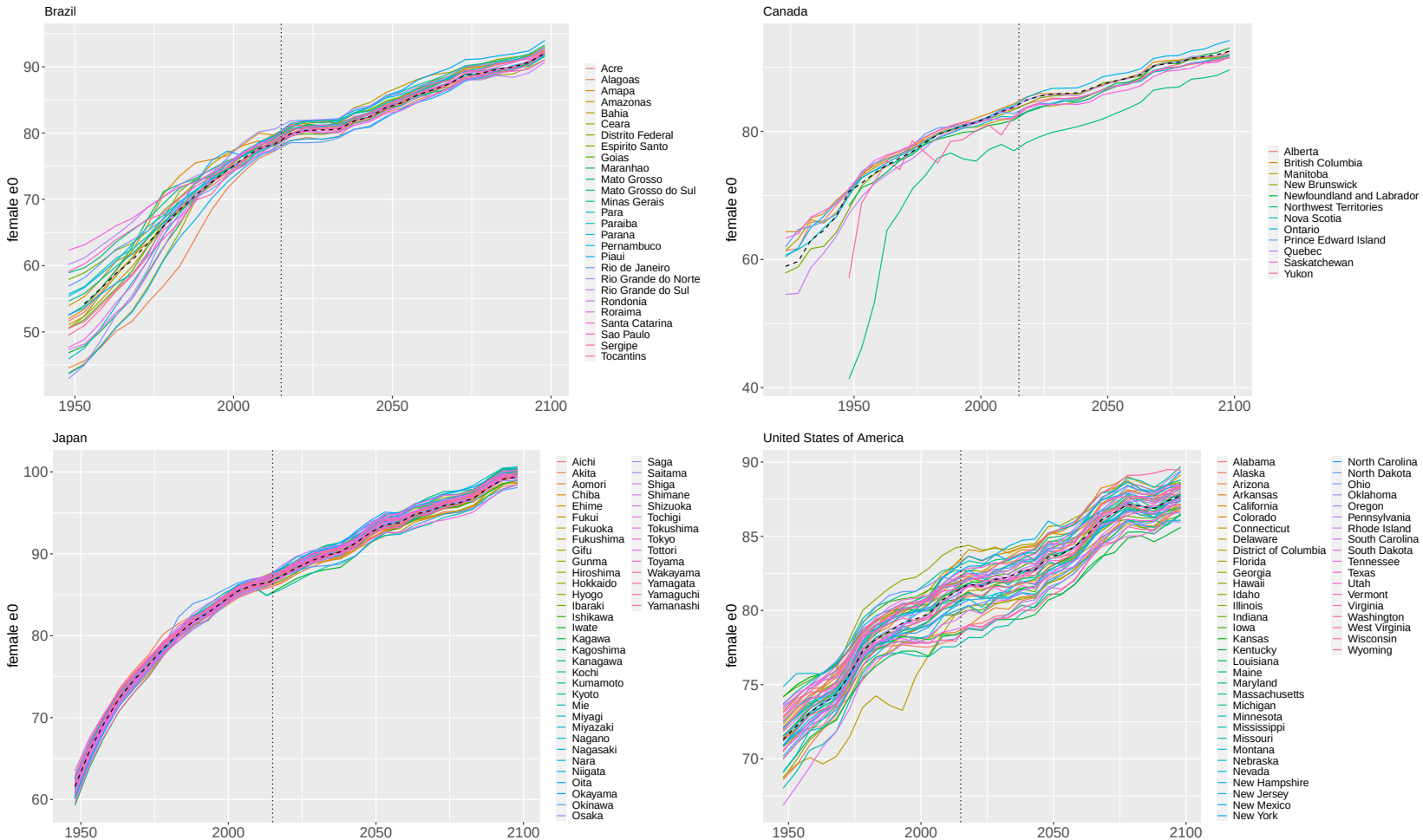
Red: Shift-AR(1) method

Shaded: 80% probability intervals



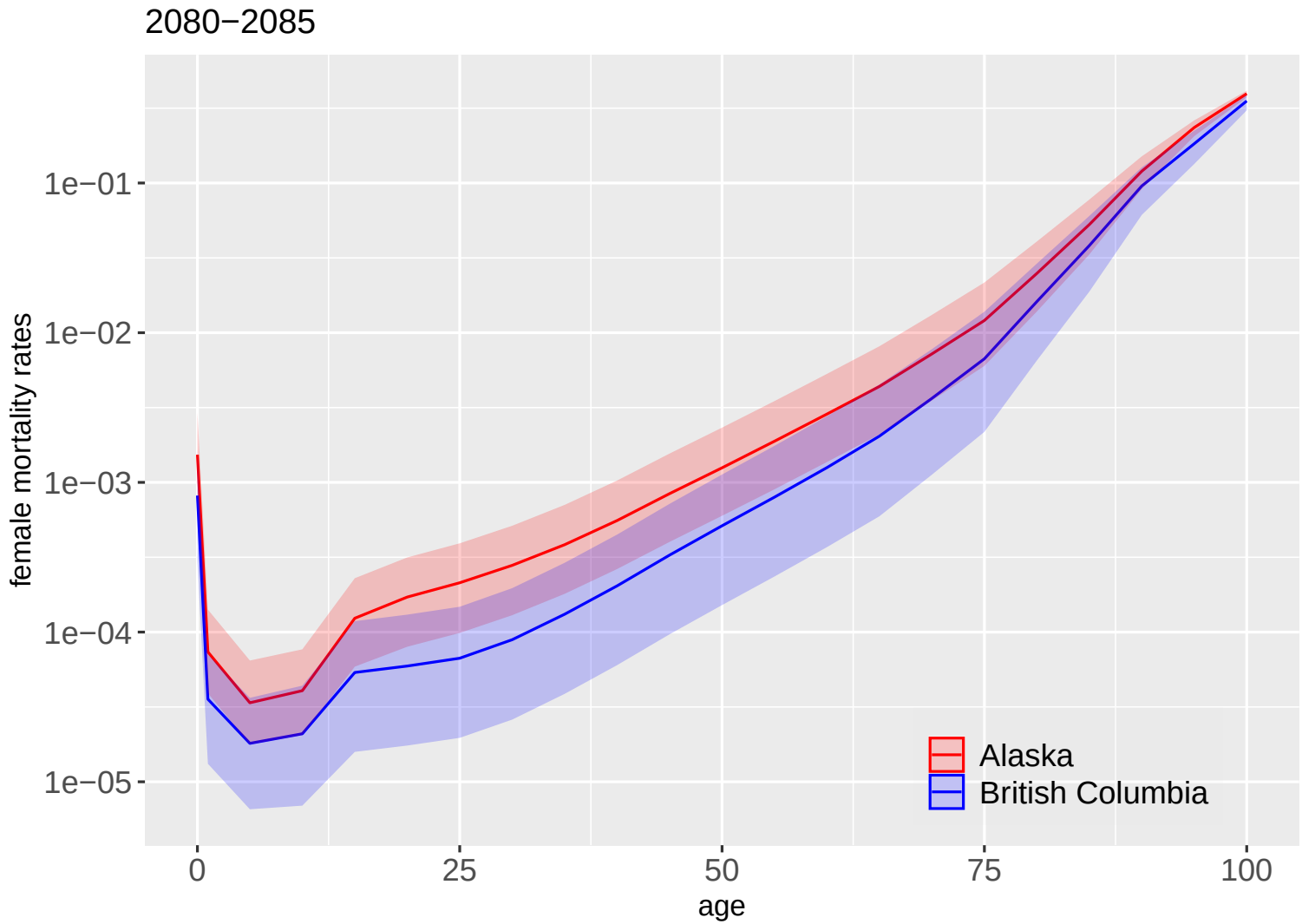


Crossover patterns for Shift-AR(1): 1 trajectory out of 1000



Probabilistic Subnational Mortality Rates

Each trajectory of e_0 converted into one trajectory of m_x (Ševčíková et al 2016; MortCast R package) $\longrightarrow$ probabilistic distribution of m_x



Summary of Probabilistic Mortality

- **National model:**
  - Generate a set of future **female** e_0 trajectories for each country via an BHM by Raftery et al. (2013).
  - Generate a set of future **male** e_0 trajectories via the gap model by Raftery et al. (2014).
  - This results in a joint distribution between the future female and male e_0.
- **Subnational model:**
  - Use Shift AR(1) by Ševčíková & Raftery (2021) to shift national **female** e_0 trajectories to subnational e_0 trajectories.
  - Apply the gap model to derive subnational **male** e_0 trajectories.
  - This results in a joint distribution between the future female and male e_0 for each subnational unit.
- Implemented in **bayesLife** R package.
- The **MortCast** package can be used to disaggregate e_0 into age-specific mortality rates.

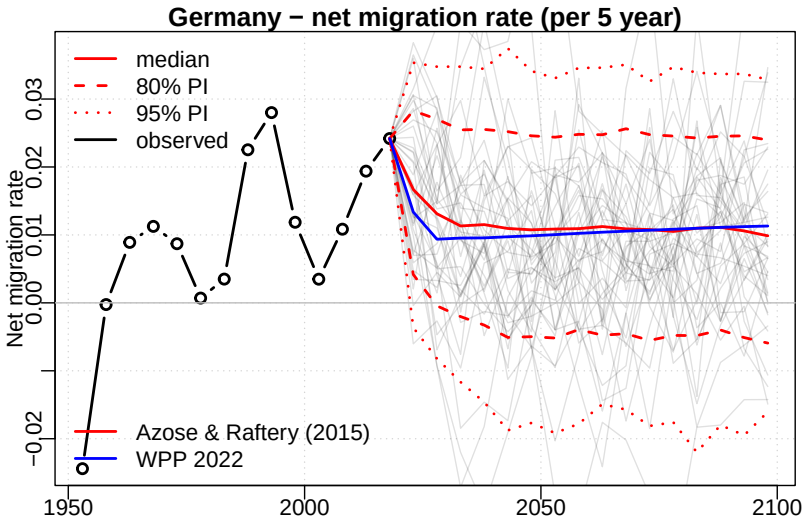
Probabilistic Projection of Net Migration

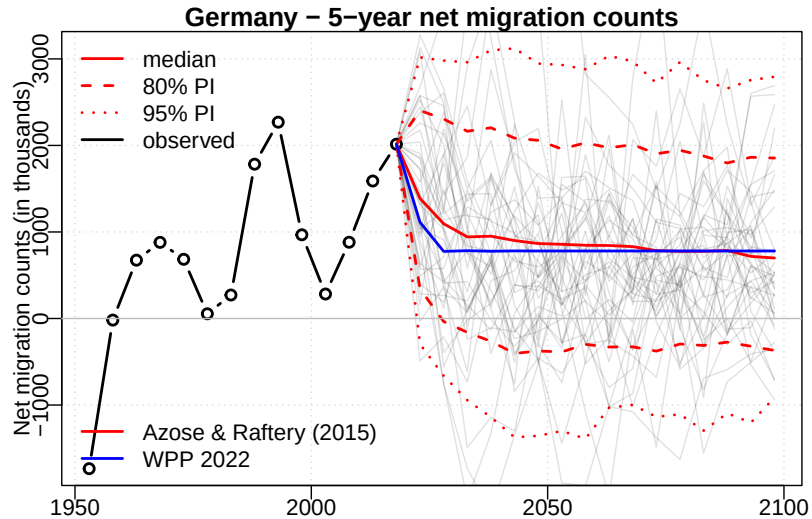
- Model of Azose & Raftery (2015) for projecting international net migration rates.
- Bayesian hierarchical AR(1) model

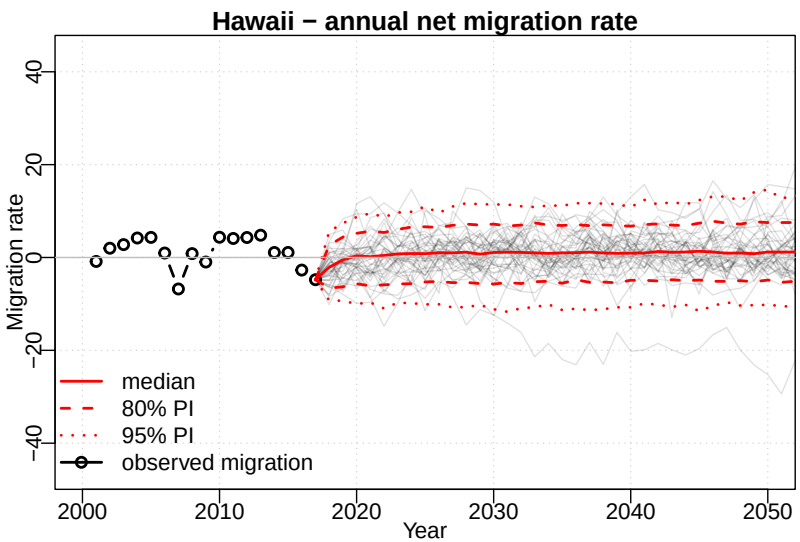
$$(r_{c,t} - \mu_c) = \phi_c(r_{c,t-1} - \mu_c) + \epsilon_c$$

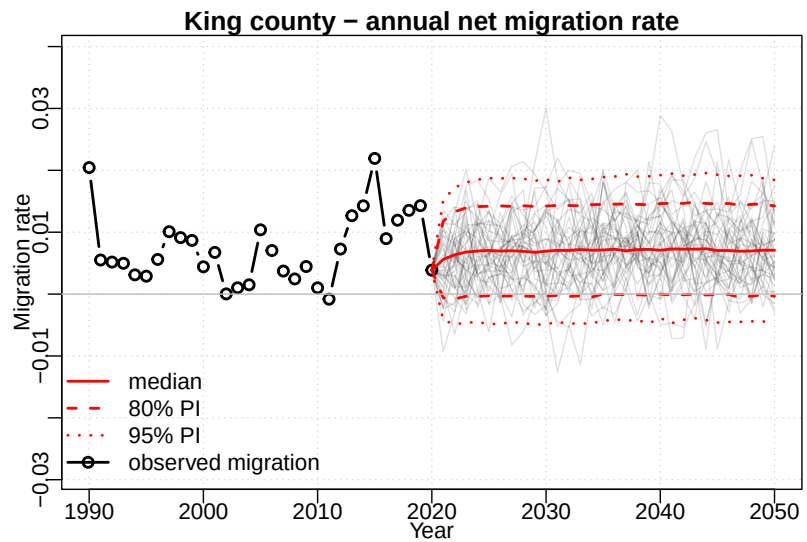
- Yield trajectories of future net migration rates for each country.
- Migration rates are converted to net migration counts within CCM.
- Optional constraints on population density and/or zero sum.
- Works well for projecting subnational migration.
- In the BHM setting, migration experience shared among regions.
- Results in a set of future trajectories of net migration rates (converted to counts) for each region.
- Implemented in the bayesMig R package

Probabilistic projection of migration - Examples

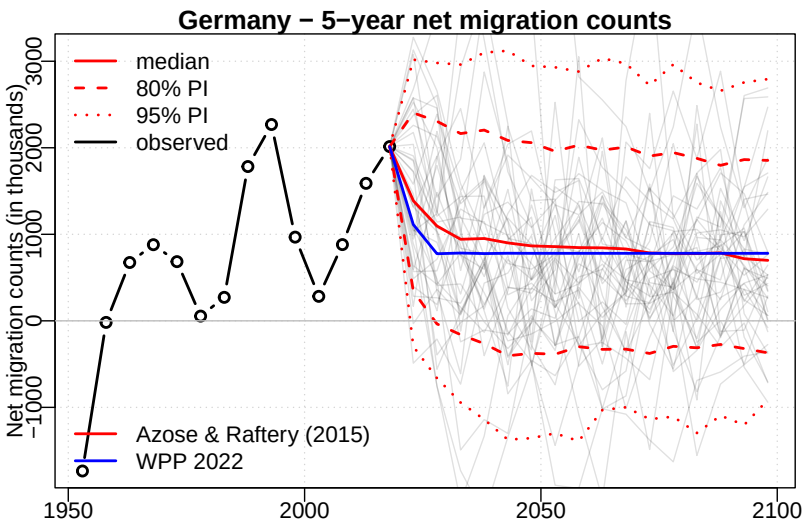


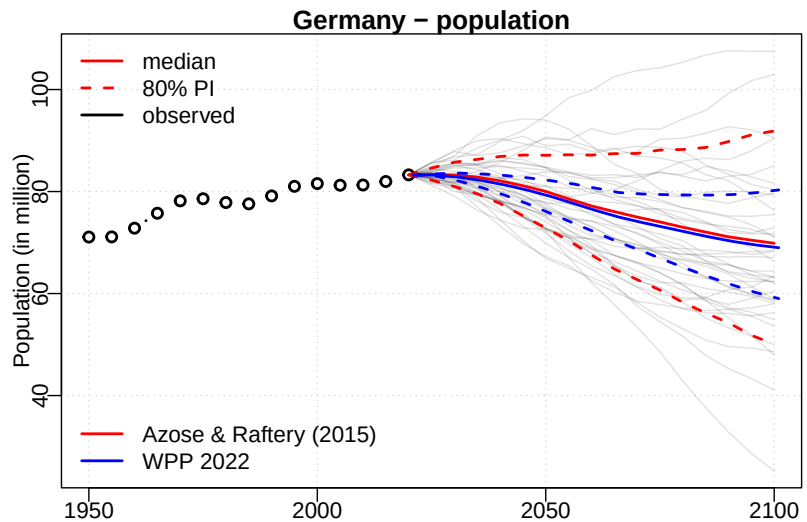






Deterministic vs. probabilistic migration





Probabilistic Population Projection

- Demographic balancing equation:

$$Pop_t = Pop_{t-1} + Births_t - Deaths_t + Net\ Migration_t$$

  - solved using Cohort Component Method (CCM), while disaggregating the components into sex- and age-specific counts and rates

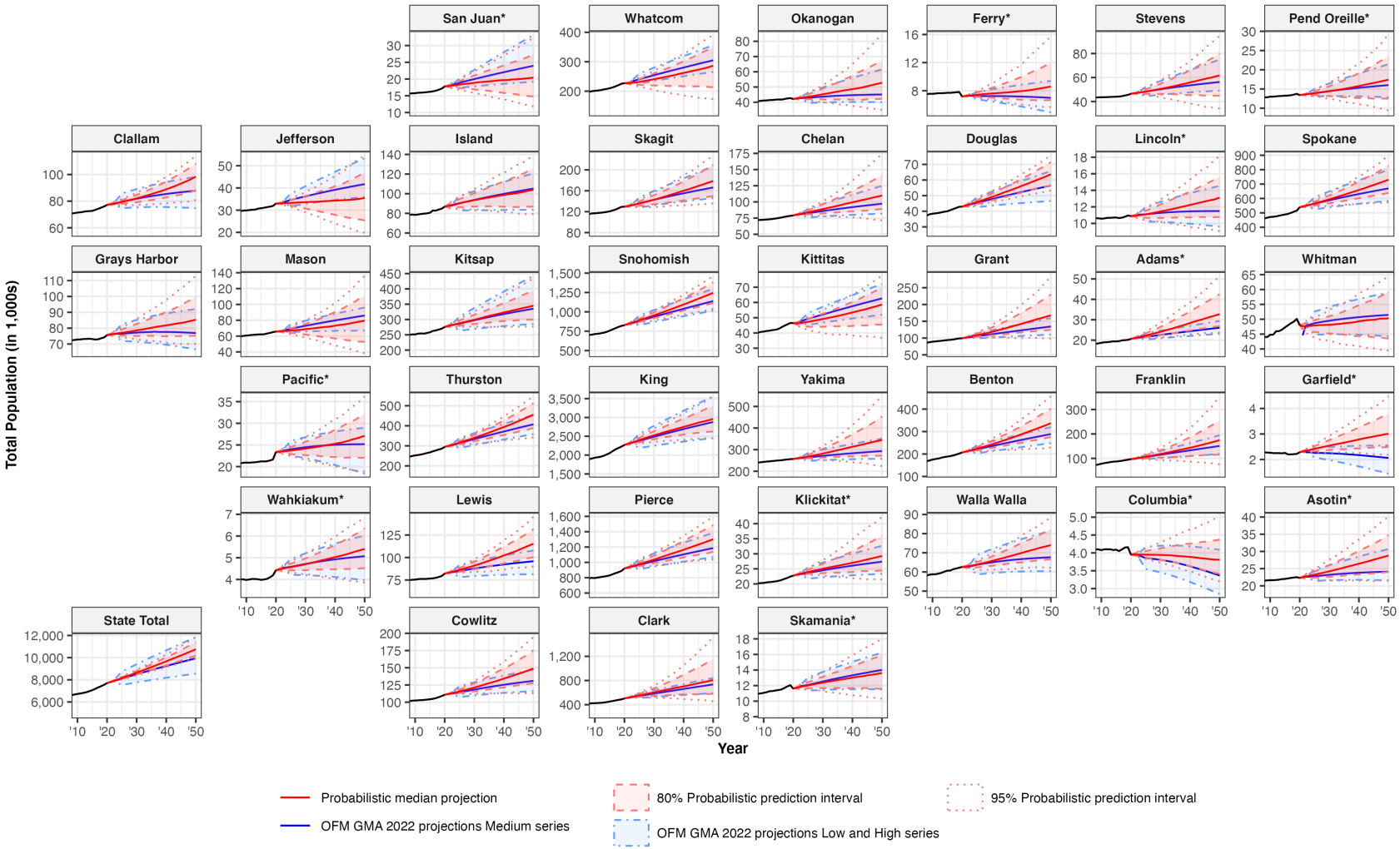
- **Probabilistic approach:**
  1. Using probabilistic models, generate sets of trajectories for TFR, e_0 and net migration.
  2. Convert each trajectory to age-specific rates and counts.
  3. Apply the CCM to each trajectory of age-specific fertility and mortality rates and migration counts.
  4. Result: Future trajectories of age- and sex-specific population quantities.

- Implemented in the bayesPop R package

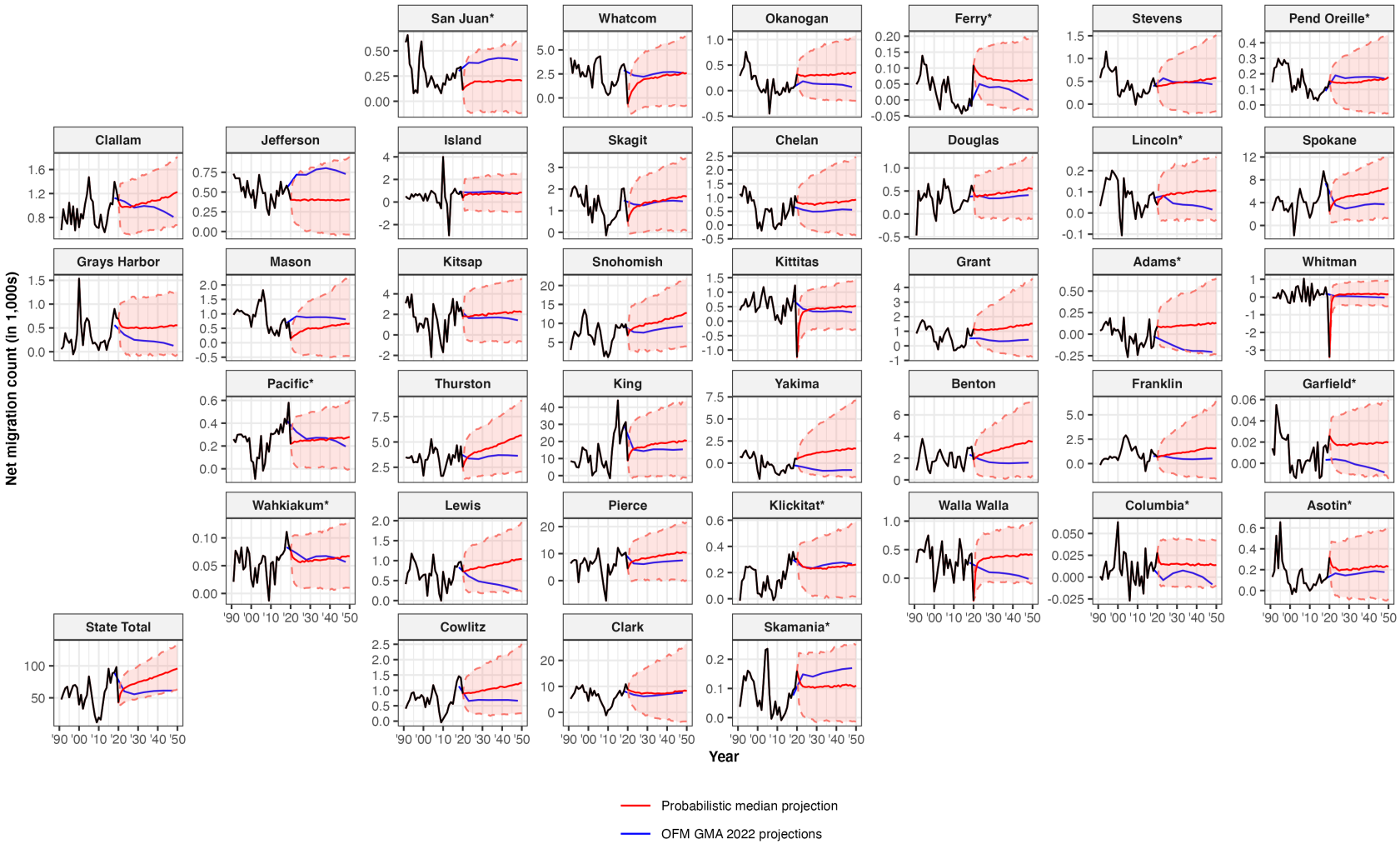
Subnational Probabilistic Population Projection

- Methodology applied to 39 counties of Washington state (12 counties smaller than 25 000).
- For each county, TFR and e_0 projections generated to 2050.
- Special treatment of TFR in counties with large college population $\rightarrow$ projected noncollege TFR
- Net migration BHM excluded 12 small counties from influencing the global experience.
- Cohort-component method applied to each trajectory of the three components of population change.
- College population was excluded from the CCM and added back after advancing the remaining population.
- Yu et al. (2023), *Demography*

Probabilistic Population Projection for WA counties

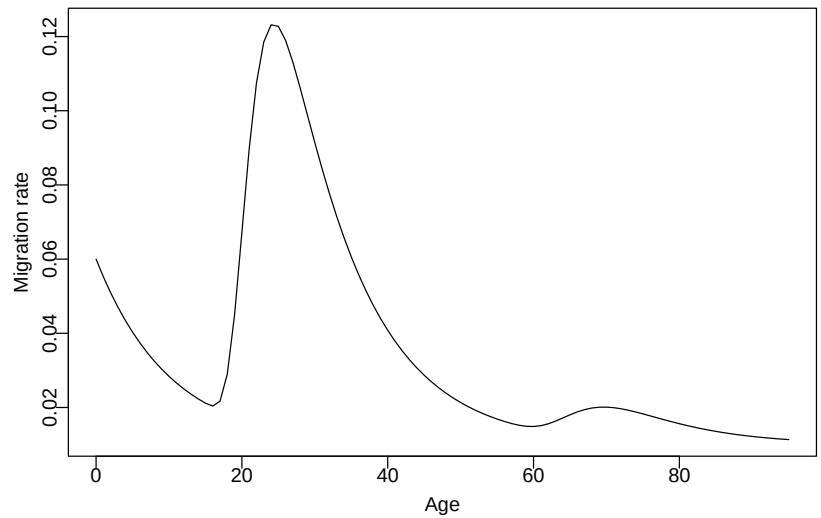


Probabilistic Migration Projection for WA counties



Age-specific Net Migration

- Practitioners often distribute total net migration (= difference between in- and out-migration) into ages via the Rogers-Castro model



- The Rogers-Castro model was developed for out-migration.
- For subnational population projections, applying Rogers-Castro to net migration can yield unrealistic future population age patterns.

Age-specific Net Migration: Flow Difference Method (FDM)

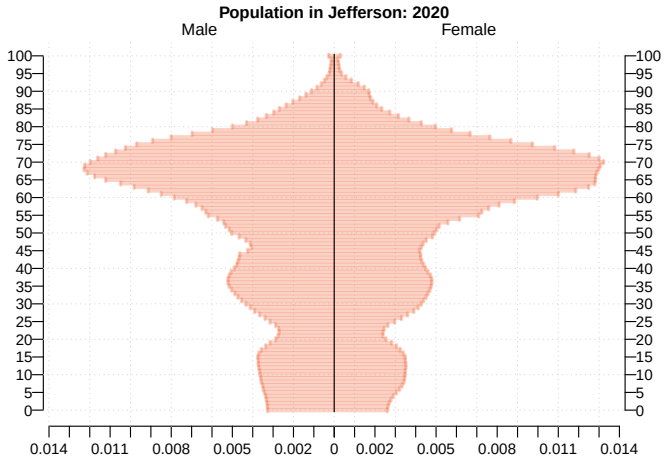
(Ševčíková, Raymer & Raftery, 2024, arXiv:2411.09878)

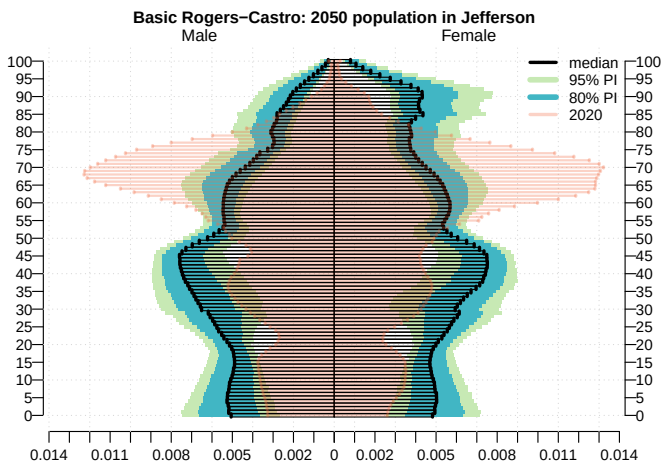
Idea:

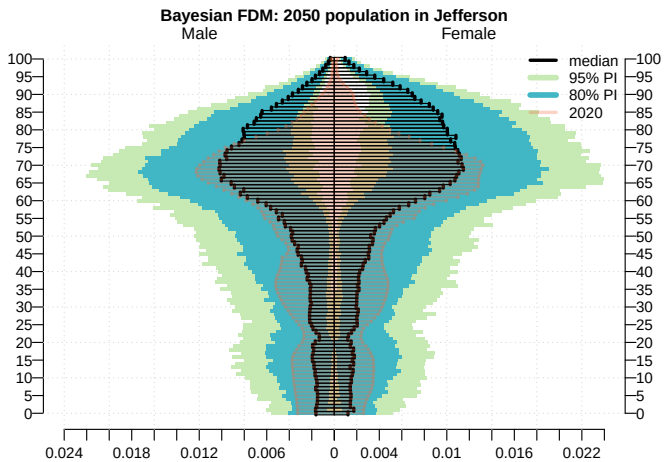
1. Split total net migration into total in-migration and total out-migration.
2. Estimate Rogers-Castro parameters for each flow and use it to distribute each total flow into age-specific in- and out-migration.
3. Take their difference.

FDM: Use in Probabilistic Population Projections

Jefferson county:

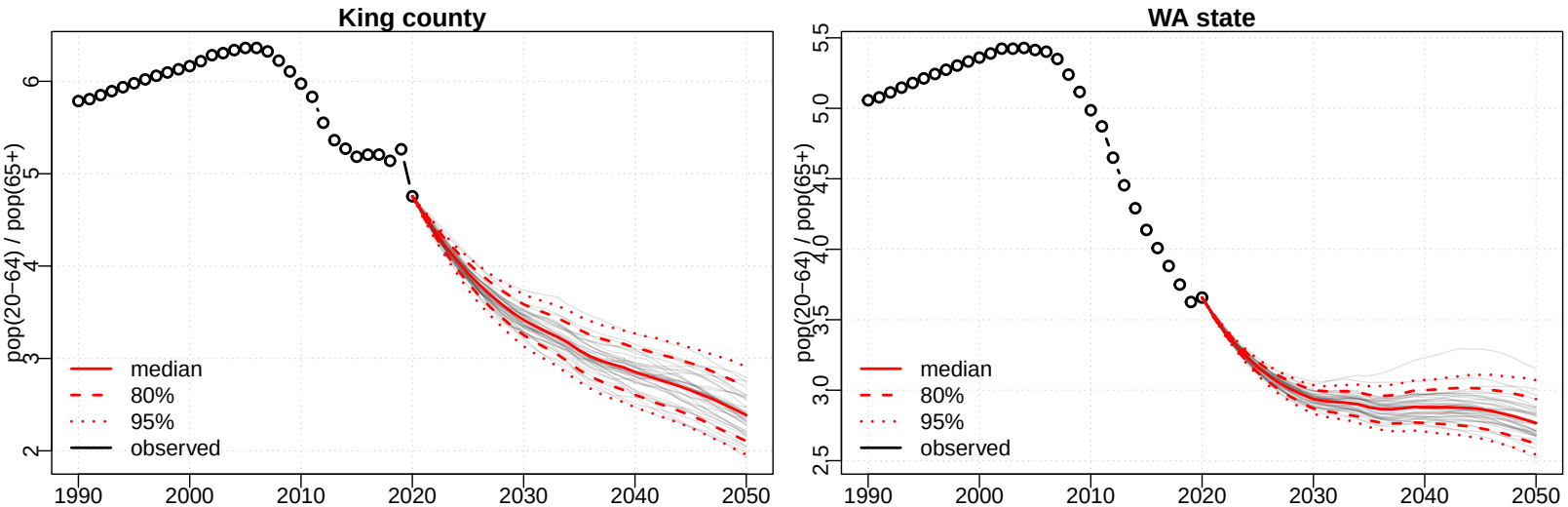






Probabilistic Indicators: Potential Support Ratio

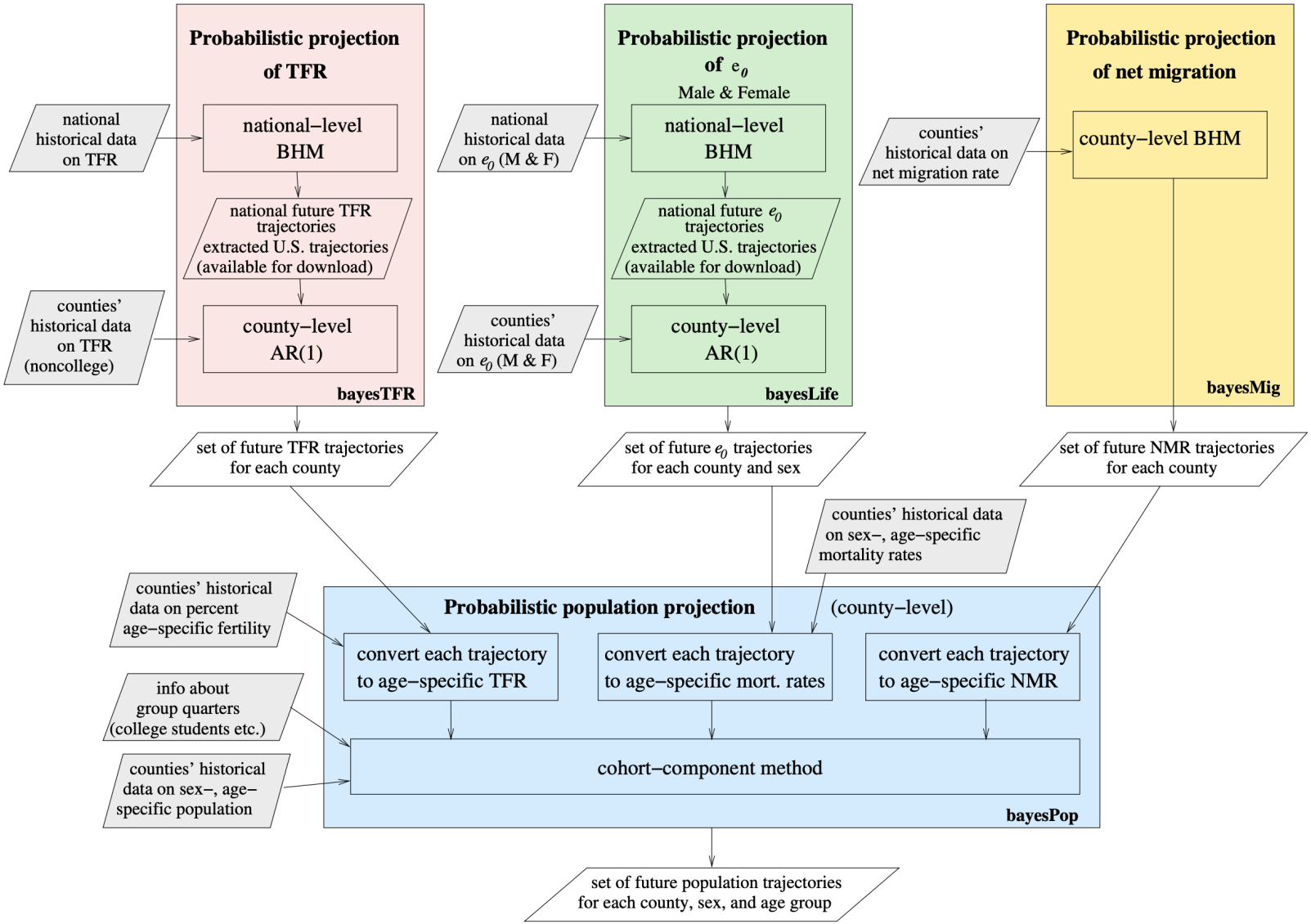
PSR: Population 20-64 / Population 65+



probability(PSR $<$ 3 by 2050) = 98% (King)

= 79% (WA state)

Probabilistic Population Projections: Workflow



Summary

- Developed framework for probabilistic projections of subnational population.
- It includes:
  - New methods for probabilistic projections of subnational TFR and e_0.
  - Based on established national models used by the UN.
  - Method for probabilistic projection of net migration applied to subnational units.
  - Special treatment of large college population.
- Methods implemented in publicly available software: R packages bayesTFR, bayesLife, MortCast, bayesMig, bayesPop

hanas@uw.edu
https://bayespop.csss.washington.edu