2nd Research Conference on Forced Displacement
World Bank-UNHCR Joint Data Center on Forced Displacement (JDC), School of Economics at Universidad de Los Andes, World Bank's Development Research Group

Drought-induced migration in Brazil

A quantitative analysis of internal displacement from the Northeast region due to water scarcity using survey microdata and gridded meteorological data

Guilherme Miranda Dutra

Survey and Data Consultant (Republic of Korea)

M.A. Legislative Drafting, evidence-based legislation
(University of London, Universidad Complutense de Madrid, LUISS Guido Carli Rome)
Postgraduate Specialization in Data Science
(PUC Minas Gerais)

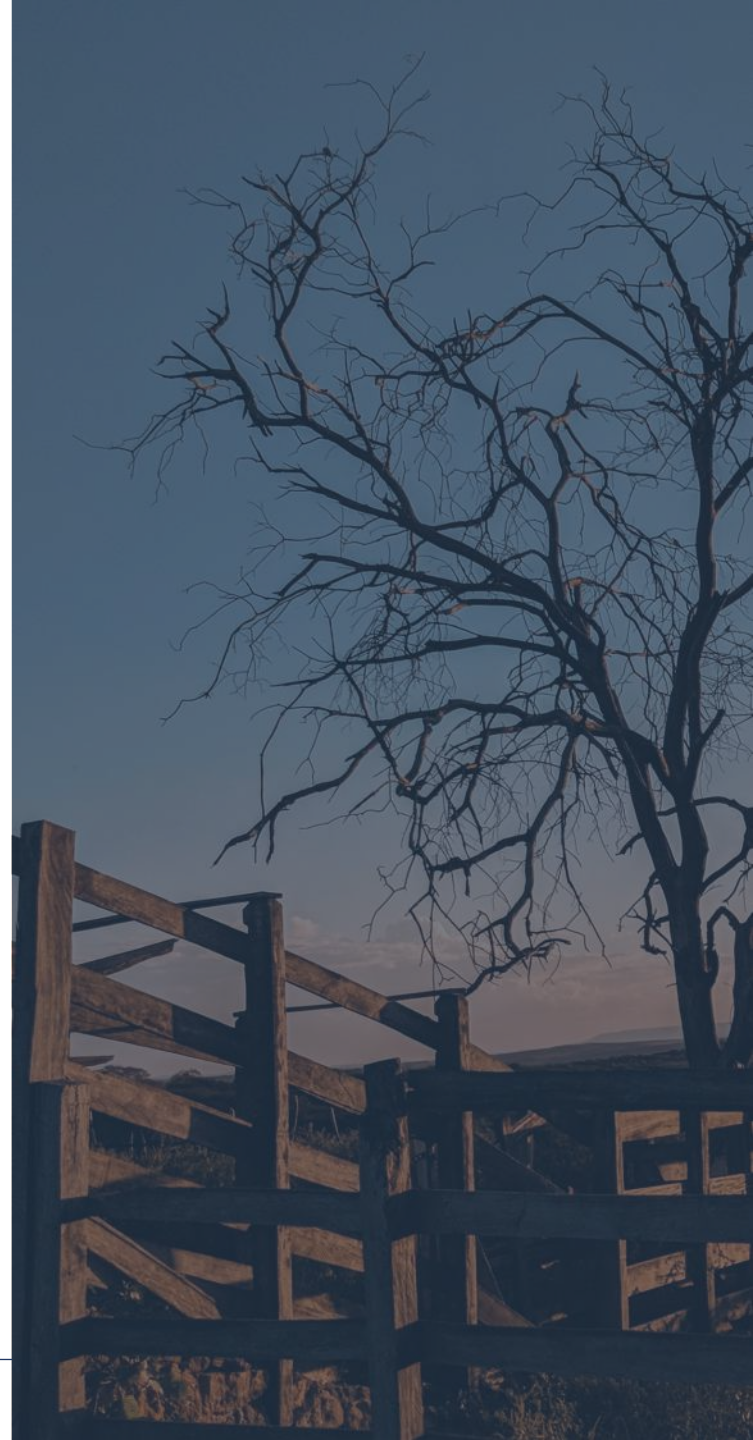
Marianne Bueno dos Passos Brum

Research Engineer Consultant (Germany)

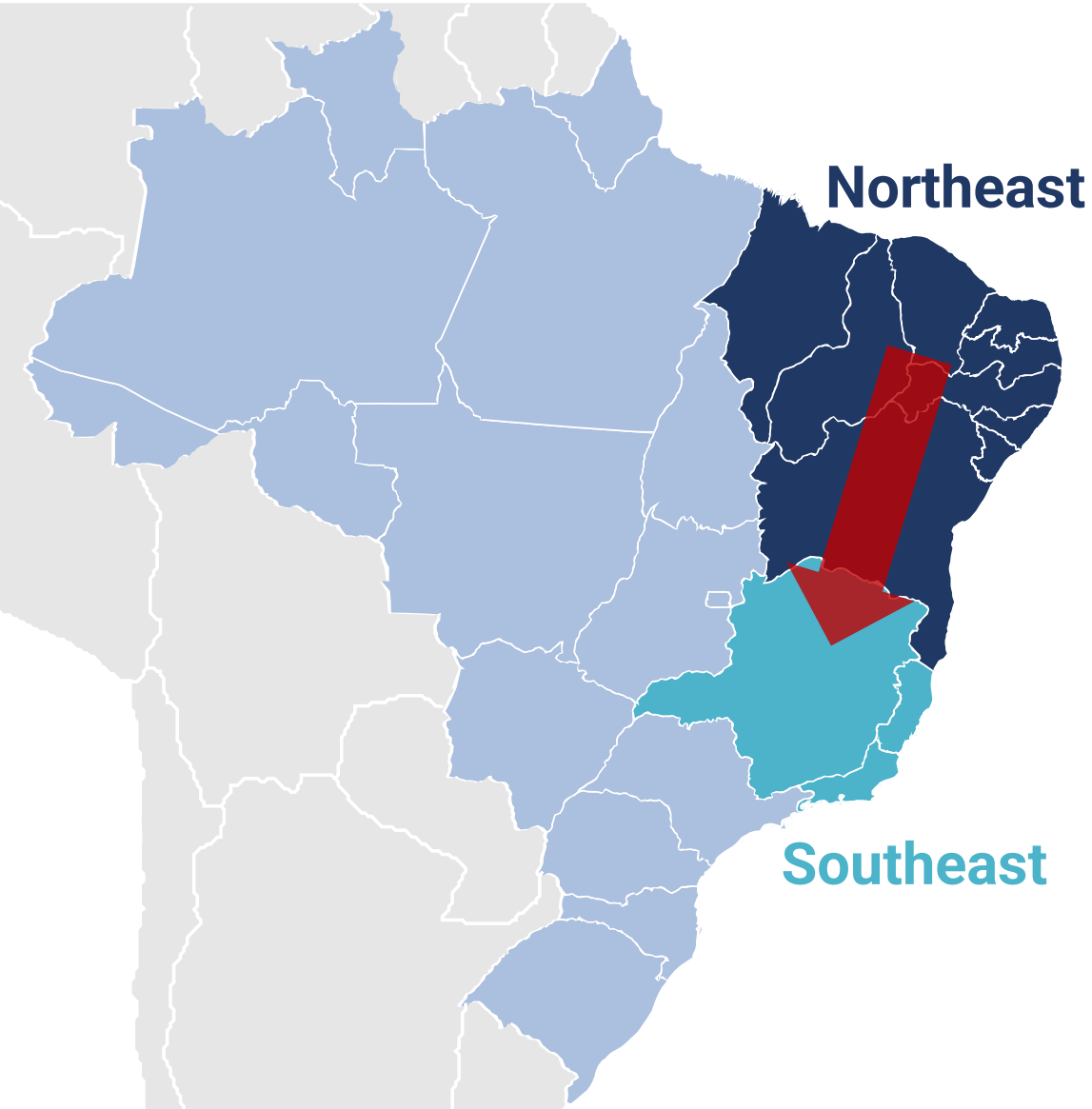
MSc Hydroinformatics and Water Management
(Polytechnical University of Nice-Sophia Antipolis, Brandenburg University of Technology at Cottbus, Warsaw University of Technology, Newcastle University, and Technical University of Catalonia)

Photo: Gabriel Tomaz, Ceará, Brazil

Table of Contents



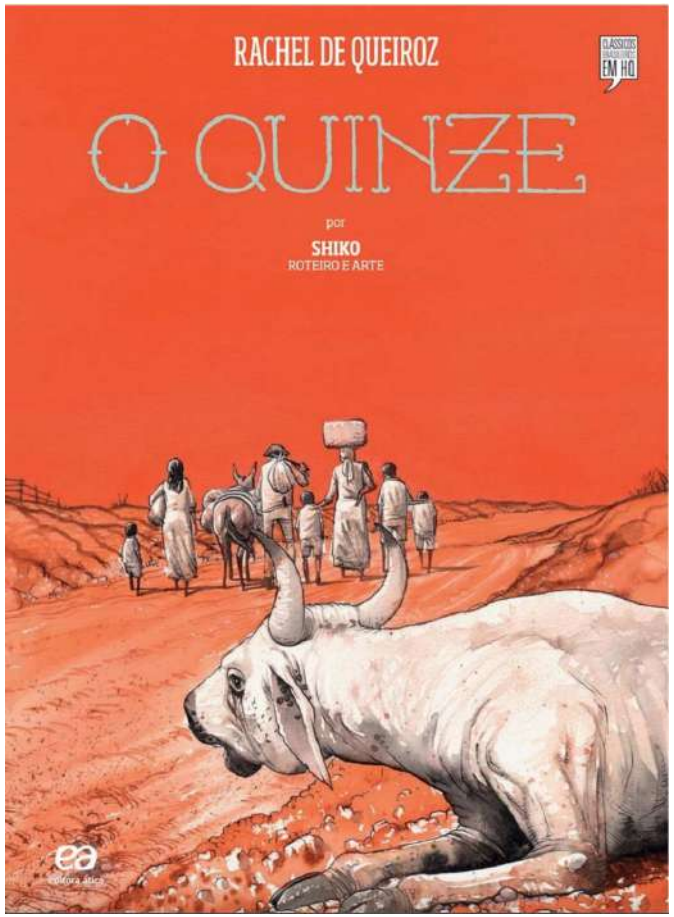
1 Introduction and starting points



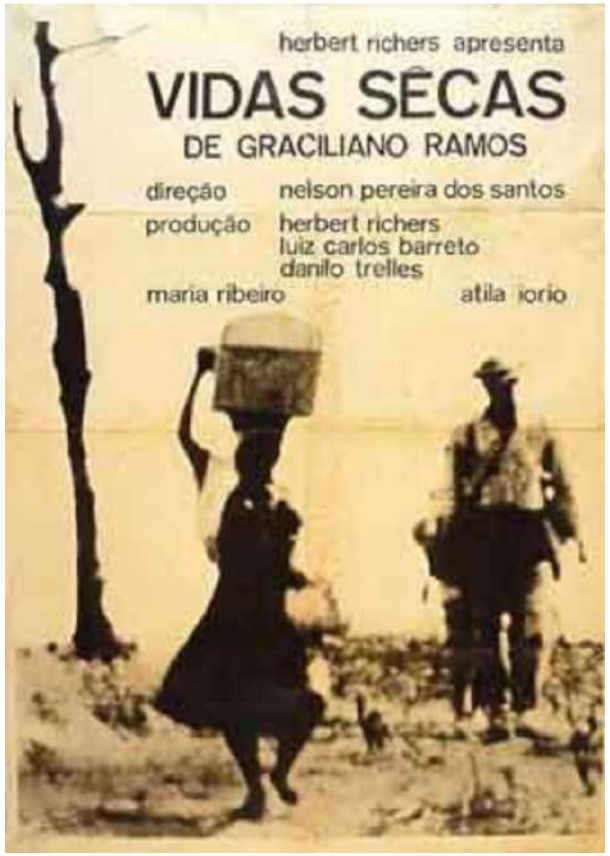
Extensive body of literature, anecdotal evidence and common association between droughts in the Northeast and migration to other regions of Brazil – especially the Southeast.

Limited research on the weather-related aspects of the migration, with a focus on socioeconomic factors.

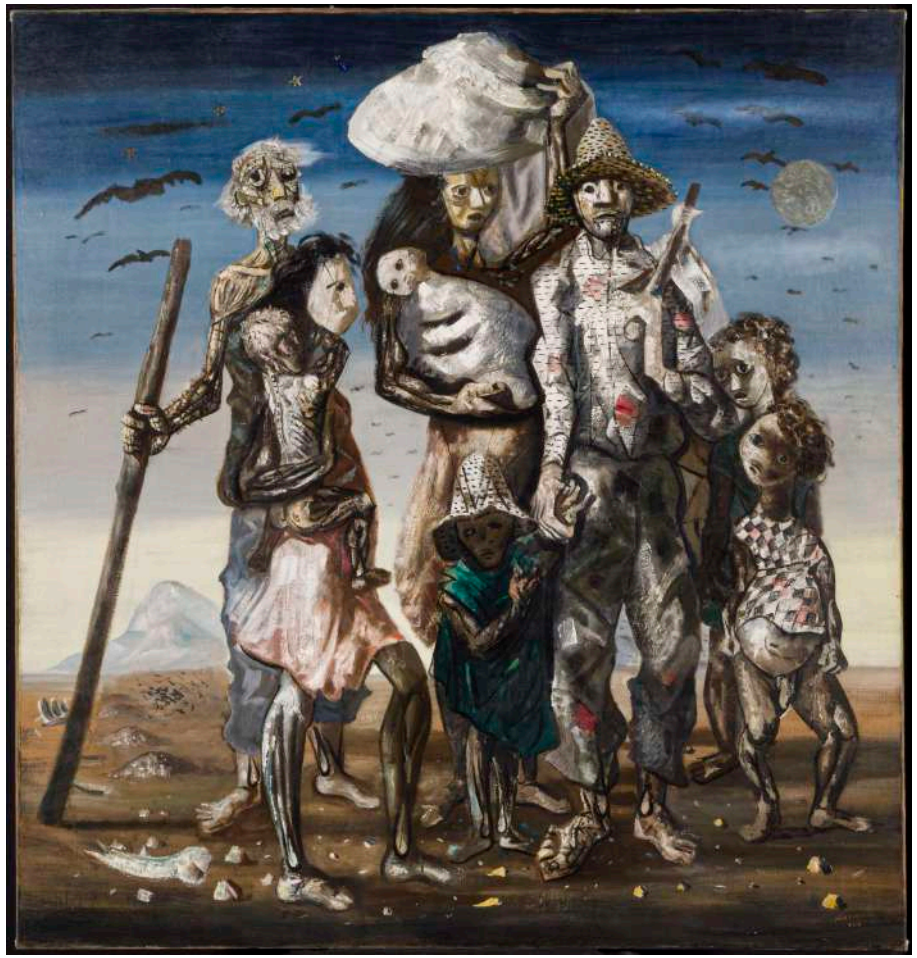
1 Introduction and starting points



O Quinze
Rachel de Queiroz, 1930



Vidas Secas
Book – Graciliano Ramos - 1938
Movie - Nelson Pereira dos Santos - 1963



Retirantes (*Drought migrants)*
Candido Portinari, 1944

1 Introduction and starting points

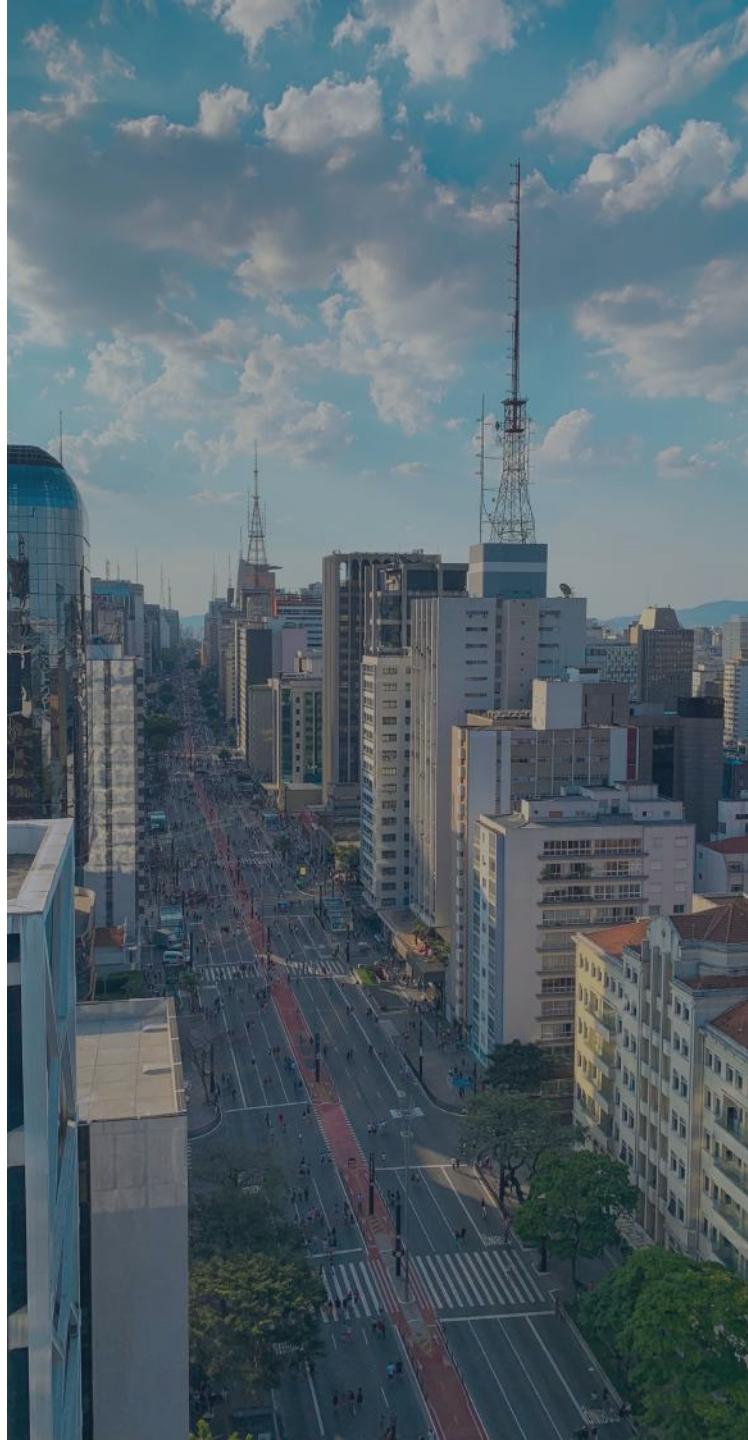
What are the methodological possibilities and challenges for integrating big data and survey microdata in internal migration research?

Using survey microdata and meteorological big data, is there evidence for drought-induced internal migration from the Northeast to the Southeast region of Brazil?

2 Literature Review

Interregional migration

- Economic crisis in the Northeast in the 1960s - emigration of 2.2 million people from the region (Ribeiro; Carvalho; Wong, 1996) and more than 3 million people in the 1970s.
- States of São Paulo and Rio de Janeiro in the Southeast region received the largest part of migration from the Northeast in the 1960s and 1970s.
- From the 1980s, limited economic opportunities in large cities of the Southeast and increased urban violence: change of the trend, reduction of migration from the Northeast and increased return migration to the Northeast (Oliveira, Costa, Ojima, 2019).
- In the 1990s, 2000s and 2010s, states of the Northeast started showing positive net migration rates, reversing the traditional trend – and drought as a push factor?

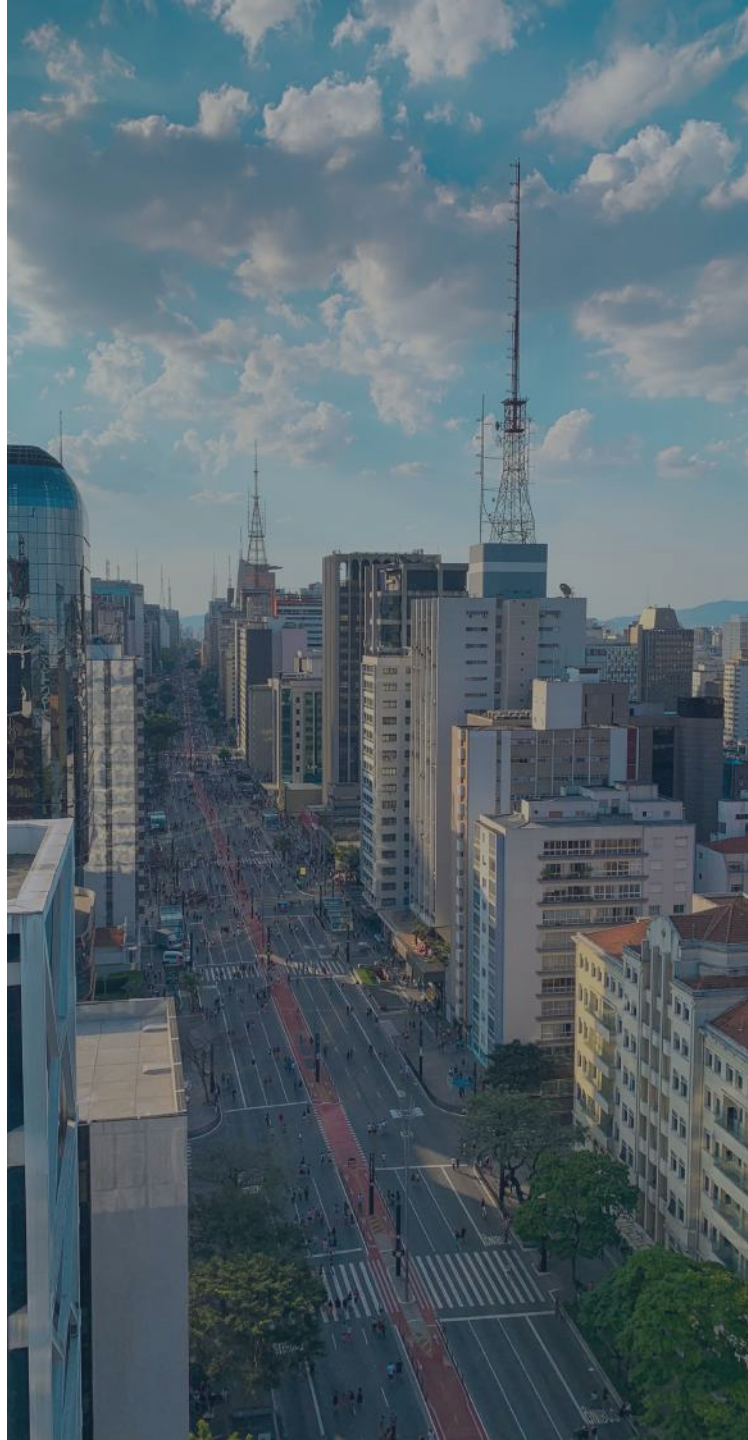


Bianca Monteiro
São Paulo, Brazil

2 Literature Review

Drought and interregional migration

- *Periods of drought* mentioned incidentally as reasons for interregional migration (Medeiros de Melo; Fusco, 2019)
- Studies mention that “migration due to drought do not happen to the extent they happened before” (Marengo; Cunha; Alves, 2015), but limited analysis on the association of these variables.
- In a study by Bastos; Busso; Miller (2013) using rainfall data and census from 1970-2010, evidence was found for increased drought frequency in the past decade inducing out-migration from the Northeast.
- Study from 2020 (Costa Olivieri) using census data found evidence of an increase of 6-7% in municipality migration rate following a drought in the Northeast.



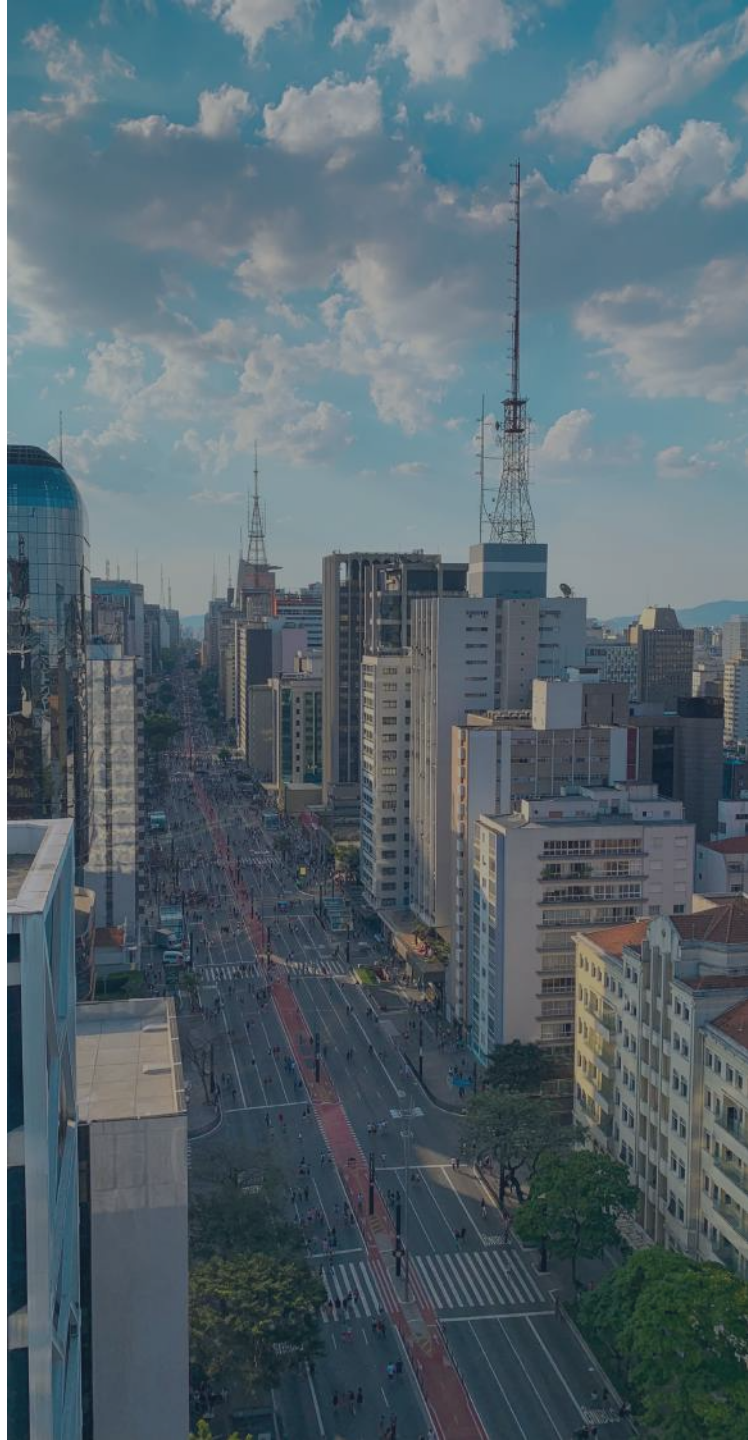
2 Literature Review

PNAD as a source of migration data

- Several references in the literature to the value of PNAD for migration studies due to its dedicated module to measuring migration, but limited empirical evidence/studies using the microdata/estimates (Pinto da Cunha; Eichman Jakob, 2002)

- Research using PNAD migration data usually adopts estimates of 5-year non-overlapping periods/fixed date (Dota; Queiroz, 2019; Hakkert; Martine, 2005)

- Main methodological issues flagged include the *non-inclusion of the rural zones of the Northern region until 2004* and issues with *estimates for small Northeastern states* due to the sample size. For migration, useful for understanding direction and trends, but more robust analysis using only PNAD in order to control for measurement error.



3 Data and Methodology

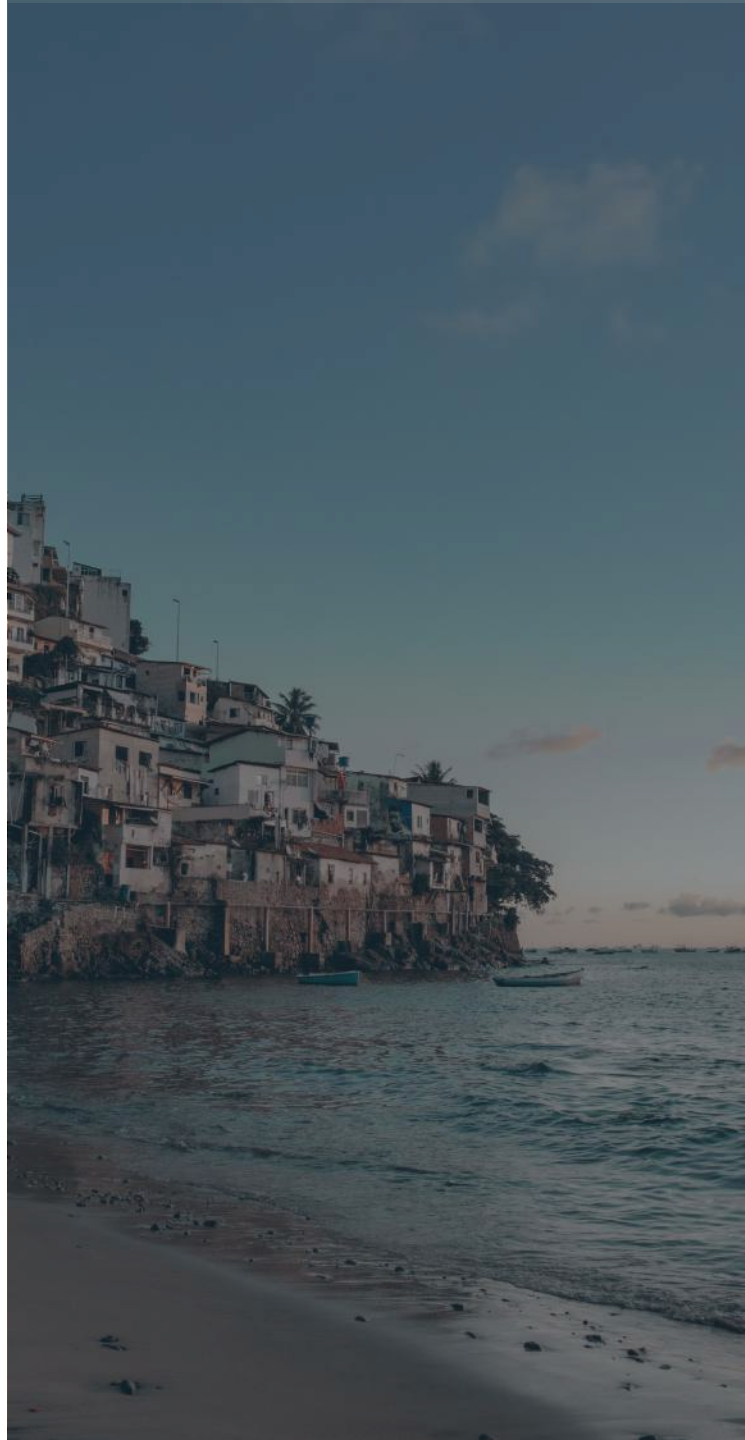
Microdata

IBGE
Instituto Brasileiro de Geografia e Estatística

PNAD

Pesquisa Nacional por
Amostra de Domicílios

2001 – 2015
Yearly



Marcus Santos
Bahia, Brazil

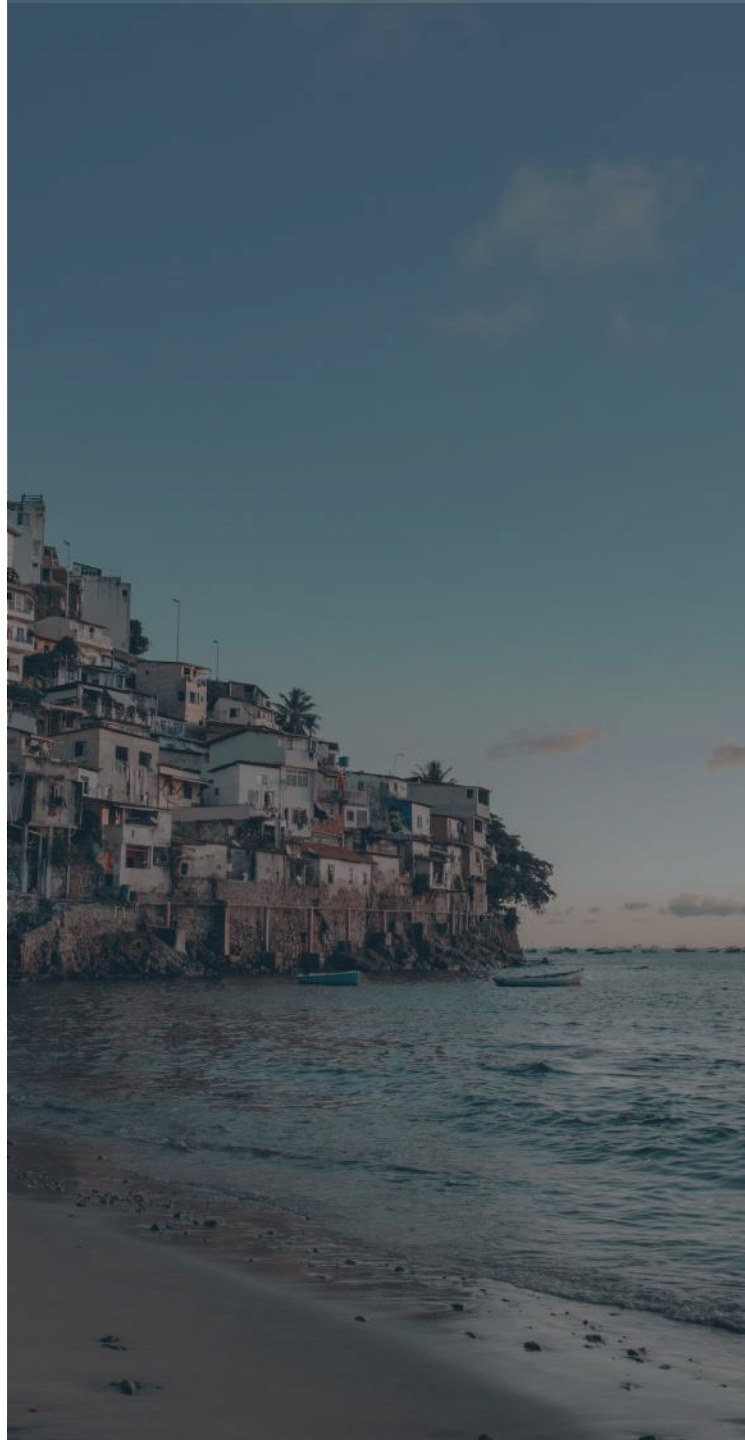
3

Data and Methodology

Design
PNAD
Pesquisa Nacional por Amostra de Domicílios

- PNAD - National Household Sample Survey – measured general characteristics of the population, education, labor, income and housing – included a migration module
- No PNAD in census years (2010)
- Complex sample design, estimation using R *survey* package
- 36 strata – 18 Federal Units (UFs/states) independent strata + 9 UFs with two strata, one for the metropolitan region of the capital and other with the rest of the UF
  - Metropolitan region strata – two-stage sampling (PSU census sectors; SSU households)
  - Other strata – three-stage sampling (PSU municipalities, SSU census sectors; TSU households)
  - All residents in the HH are included

- Approximately 150 000 HHs
- Research utilizes yearly PNAD data from 2001 – 2015 (methodological comparability)



3 Data and Methodology

PNAD

Criteria for estimating migration

Fixed date – 5 years before reference period

- Migrant defined as the individual older than 5 years old who was living in a UF/state different from the current state of residence **five years before the reference period**
- Decision to utilize **UF/state** estimates (not municipal)

1 UF Respondent + **V5080**

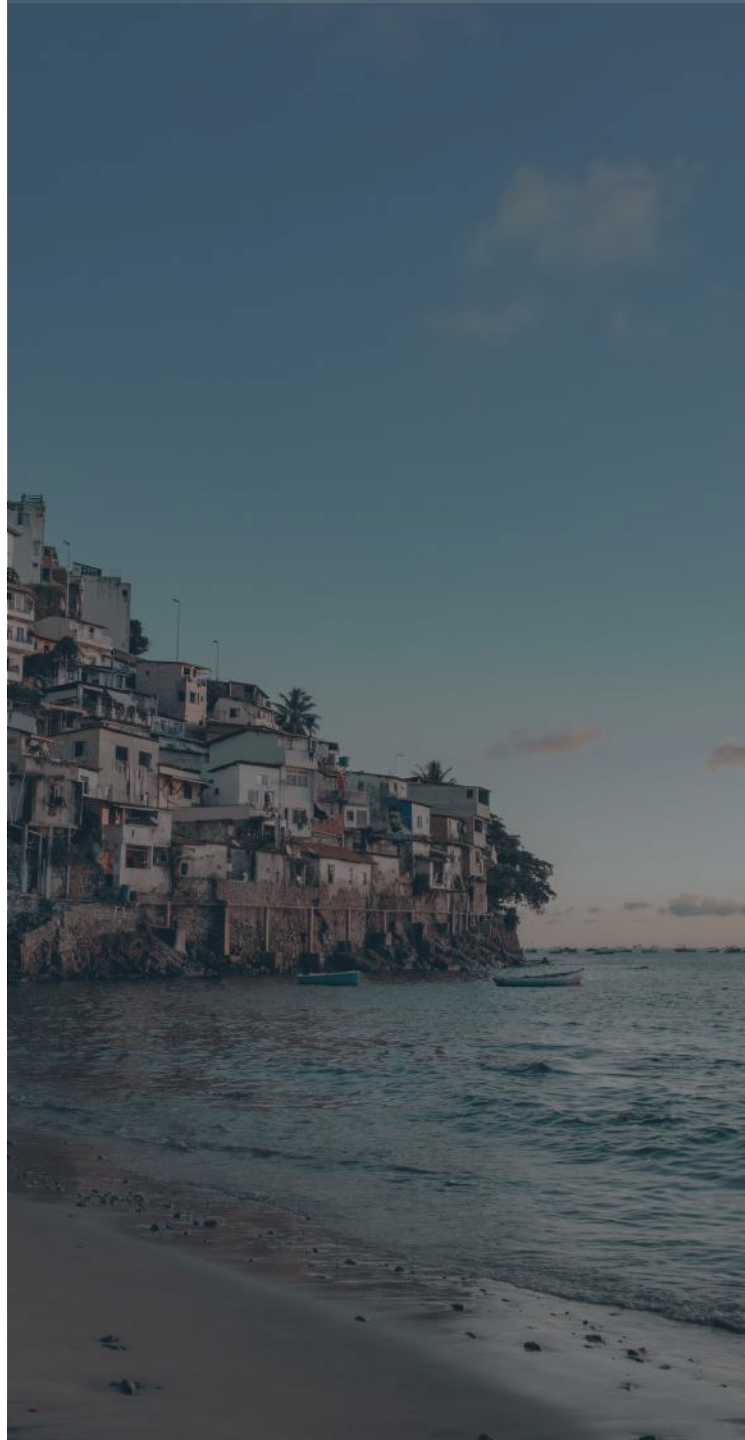
"[Date 5 years before reference period], in which UF/state [person] lived?

- Migration flows during 5-year periods – **overlapping and non-overlapping periods**
- Common in the literature.

2 UF Respondent + **V5061 / V5062 + V5080**

"[Reference period], for how long has [person] lived, uninterruptedly, in this state/UF?"

V5061 – Up to 4 years
V5062 – Number of years



3 Data and Methodology

João Dourado
Bahia, Brazil

GPCC-DI
Criteria for estimating drought

- Global Precipitation Climatology Centre Drought Index (GPCC-DI)
- Open data
- **Gridded** product
- Derived from the SPI-DWD and SPEI indexes
- Monthly time resolution
- Global spatial resolution
- 1, 3, 6, 9, 12, 24 and 48 month-aggregation periods
- Data is provided in the netCDF4 format
- Coverage: 1959 to 2021
- 1º spatial resolution

Grid properties

Property	Value
Gridtype	Lonlat
Gridsize	64 800
Xname	Lon
Xlongname	Longitude
Xunits	Degrees east
Yname	Lat
Ylongname	Latitude
Yunits	Degrees north
Xsize	360
Ysize	180
Xfirst	−179.50
Xinc	1.0
Yfirst	−89.50
Yinc	1.0

3 Data and Methodology

What is drought?

Time and space-wise aggregations

Meteorological

Precipitation anomaly

Hydrological

Low discharge, groundwater

time

extension

Socio-Economic

Impacts on agriculture, jobs

3 Data and Methodology

Drought index

Calculated indicators

- We use the 3-month aggregation product
- GPCC-DI is averaged over all micro-basins
- For each time period and each micro-basin, we define whether there is a drought event or not
- The drought-events are aggregated over each state
- Normalized and non-normalized results (over the number of micro-basins) are produced
- Aggregation periods of 1 year and of 5 years (overlapping and non-overlapping)

GPCC-DI Interpretation

GPCC-DI value	Category
2.00 or more	Extremely wet
1.50 to 1.99	Severely wet
1.00 to 1.49	Moderately wet
0.00 to 0.99	Mildly wet
−0.99 to 0.00	Mild drought
−1.00 to −1.49	Moderate drought
−1.50 to −1.99	Severe drought
−2.00 or less	Extreme drought

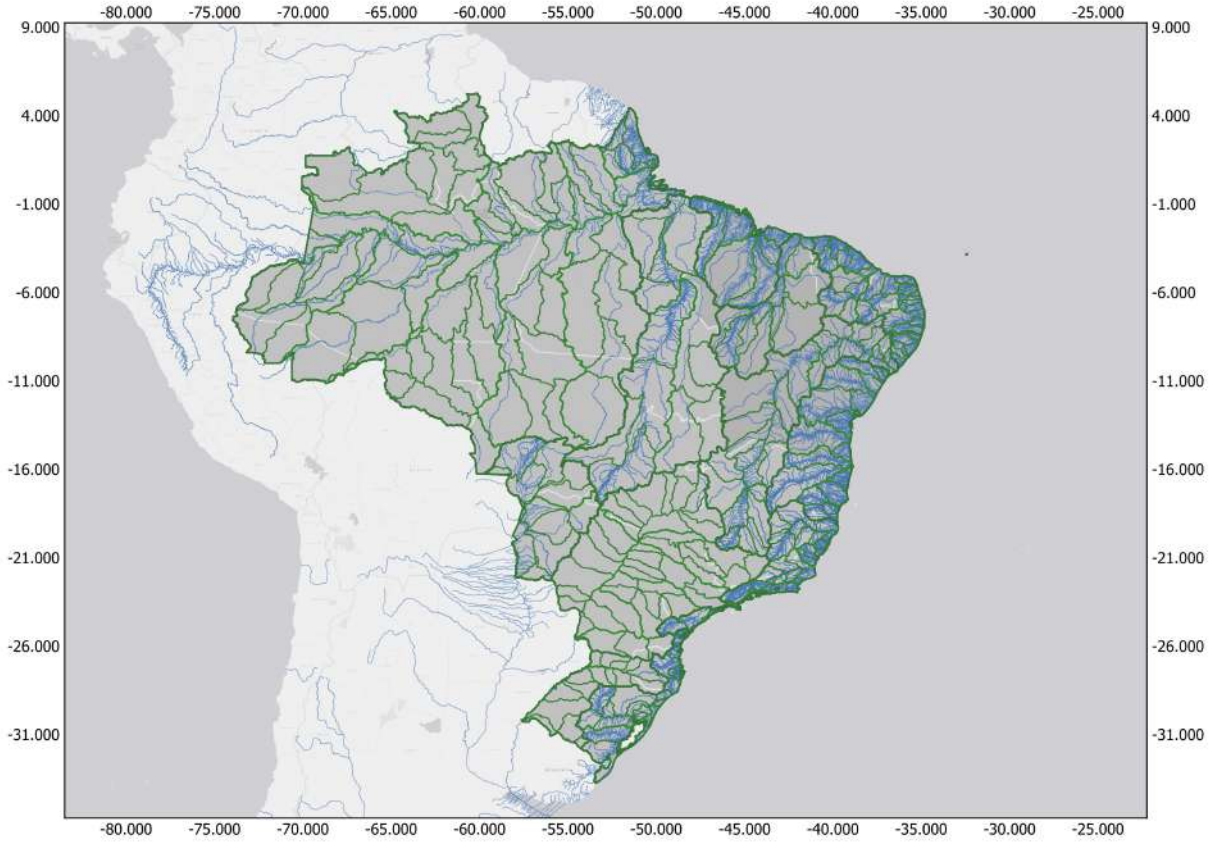
3 Data and Methodology

3 Data and Methodology

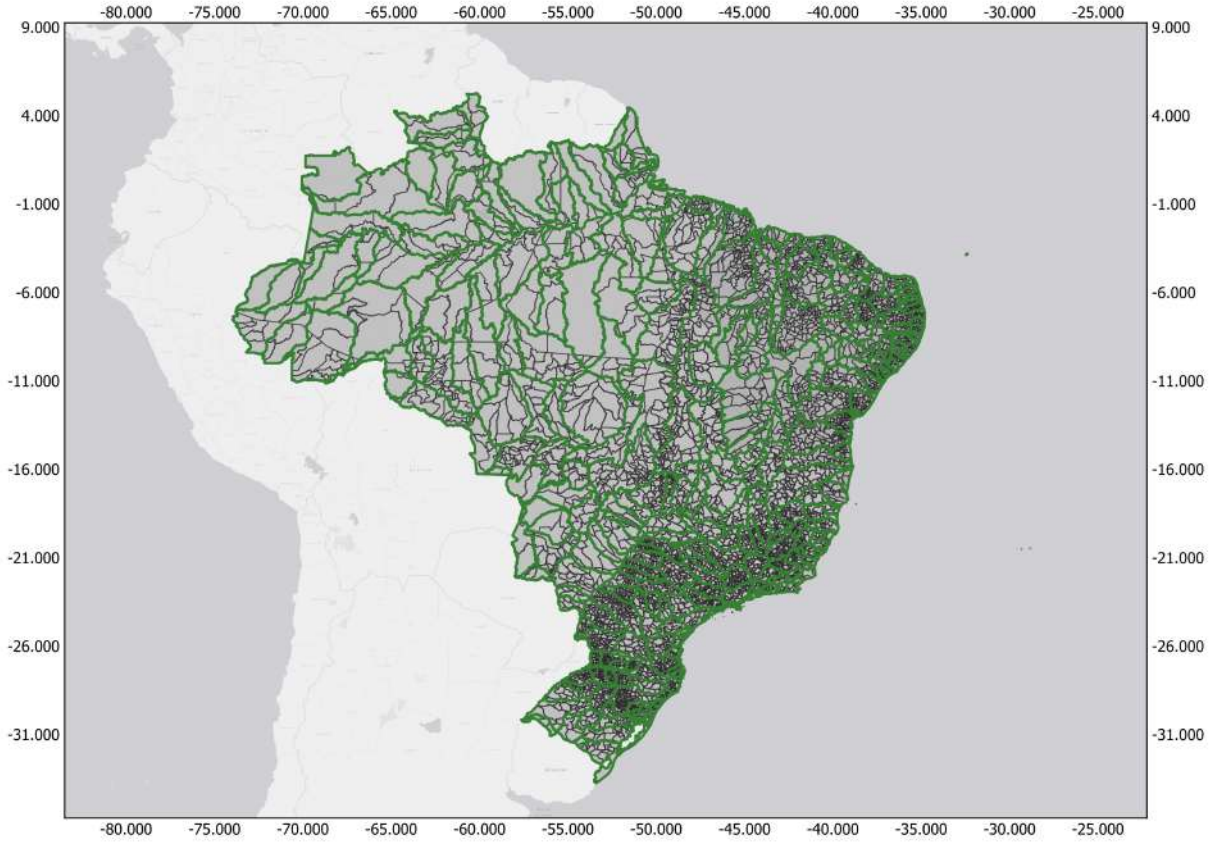
3 Data and Methodology

3 Data and Methodology

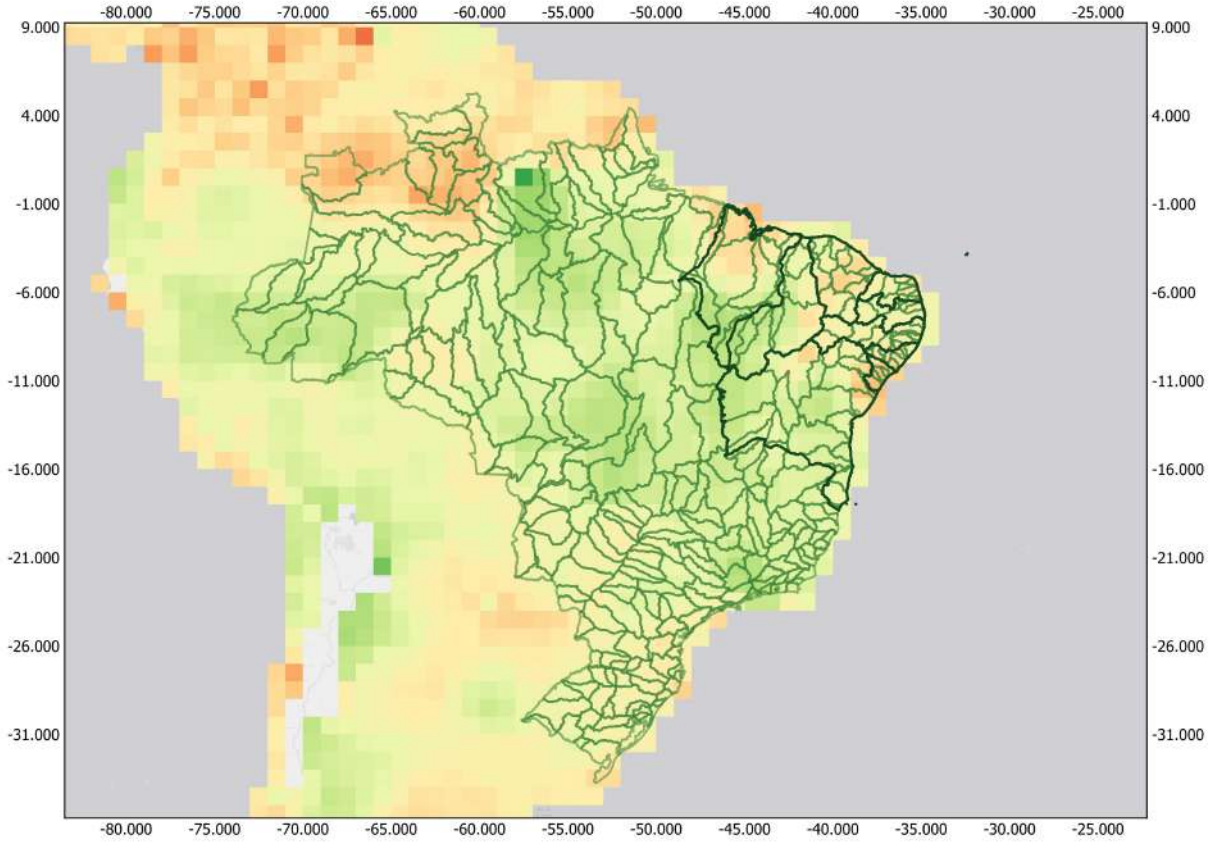
3 Data and Methodology



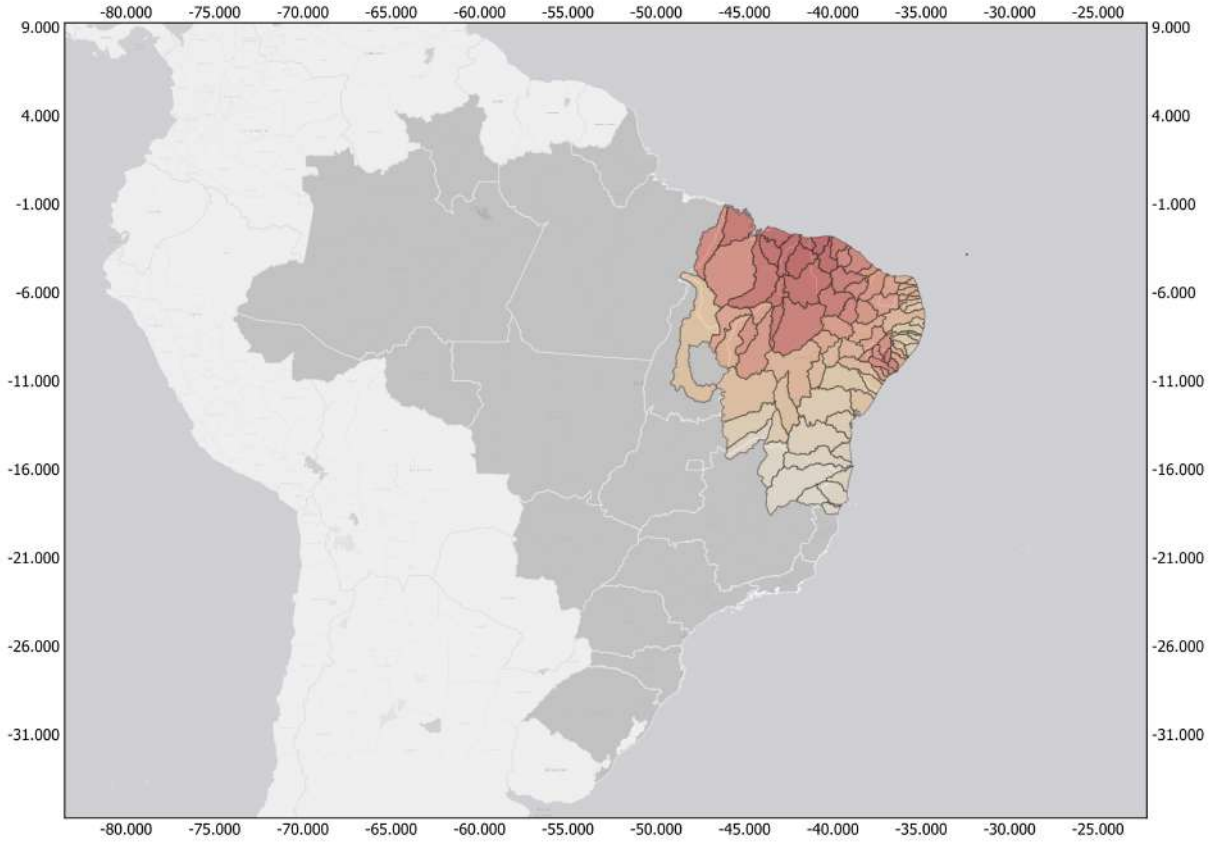
3 Data and Methodology



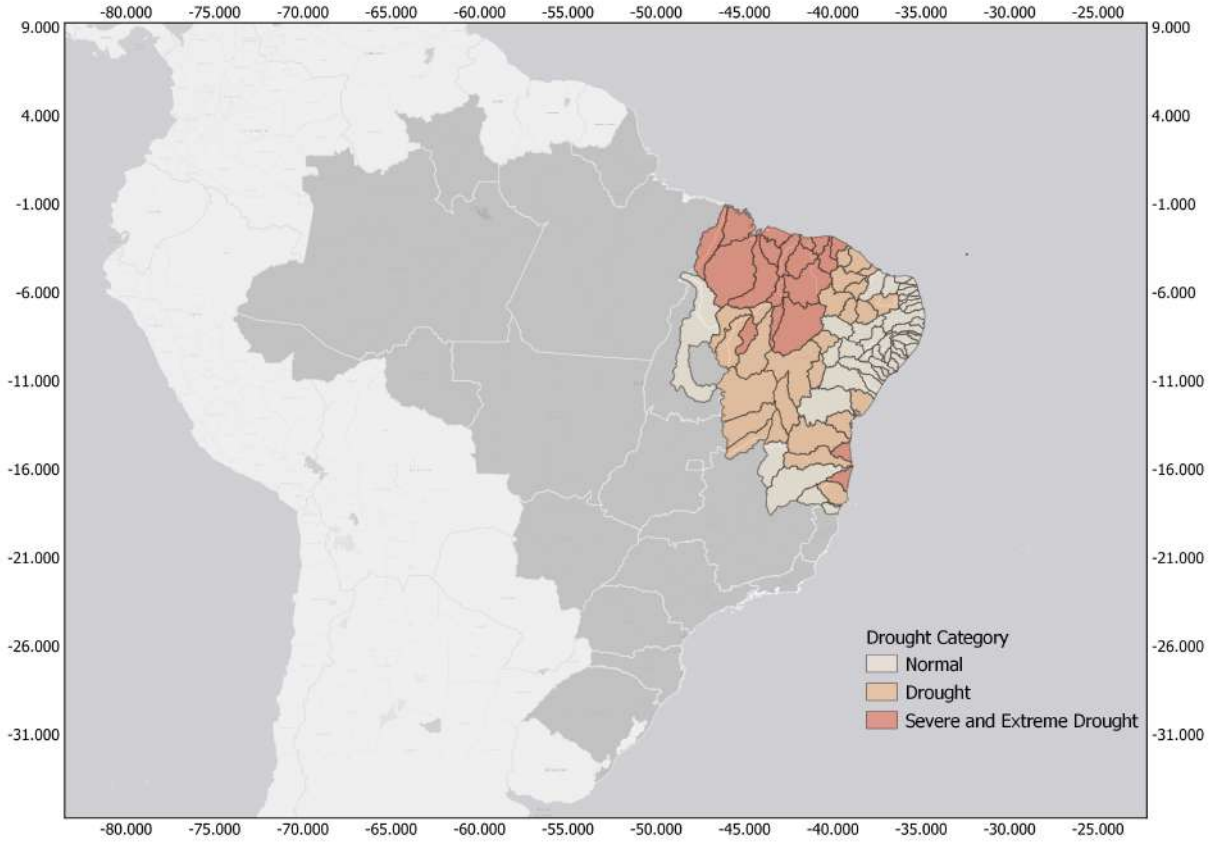
3 Data and Methodology



3 Data and Methodology



3 Data and Methodology

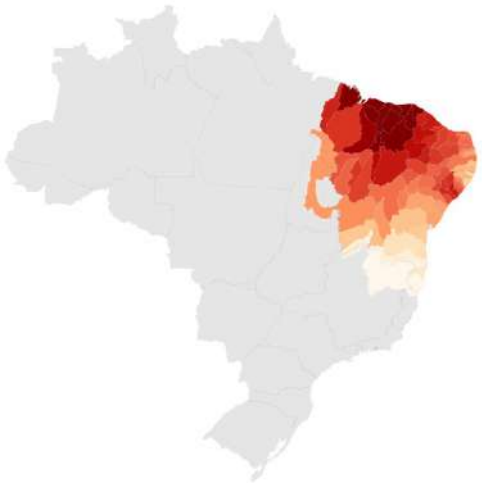


3 Data and Methodology

Time and space-wise aggregations

1983-01

1983-12



1985-12

Darker colours represent more severe drought.

3 Data and Methodology

Time and space-wise aggregations

1983-01

1983-12

1985-12

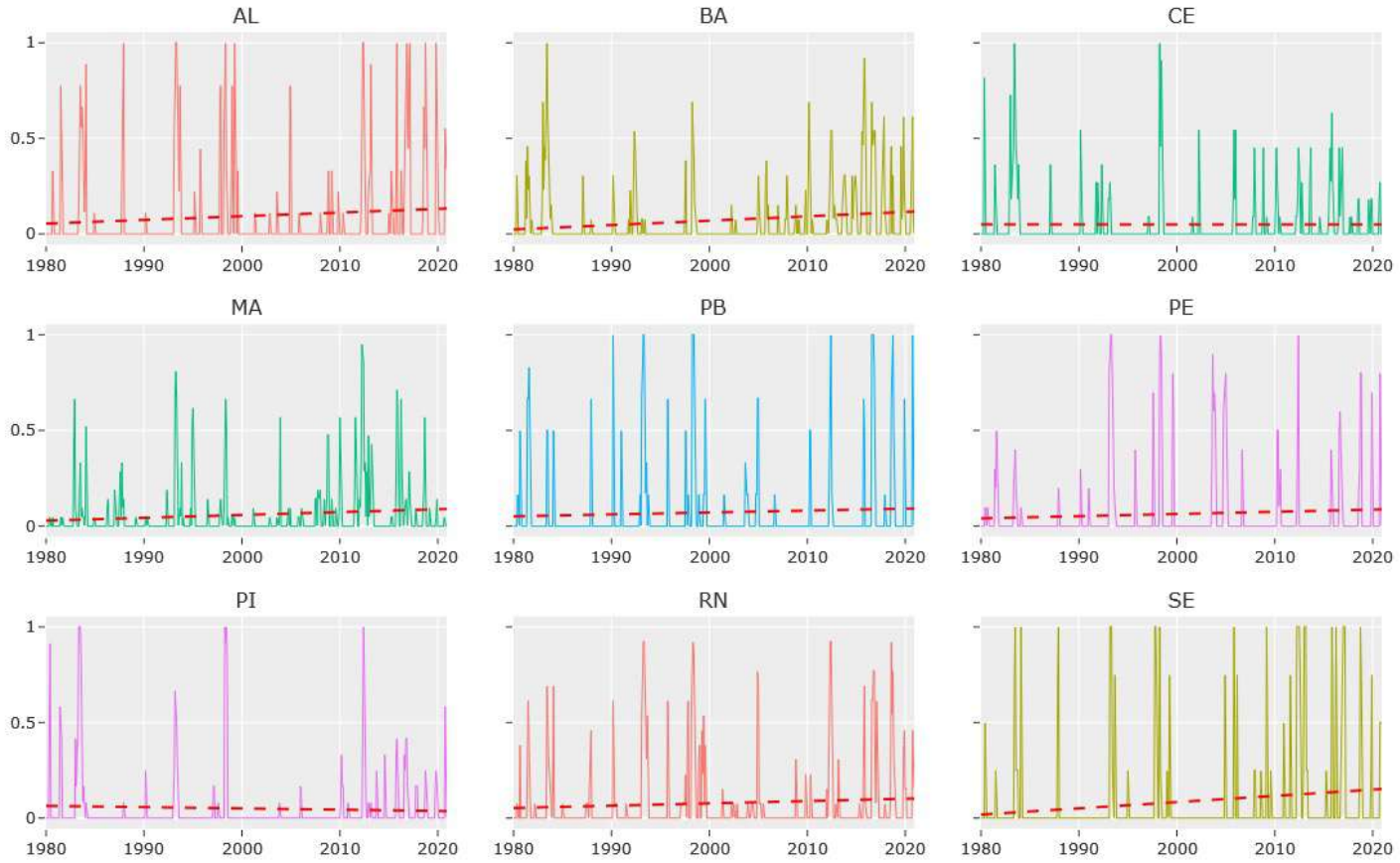
Darker colours represent more severe drought.

3 Data and Methodology

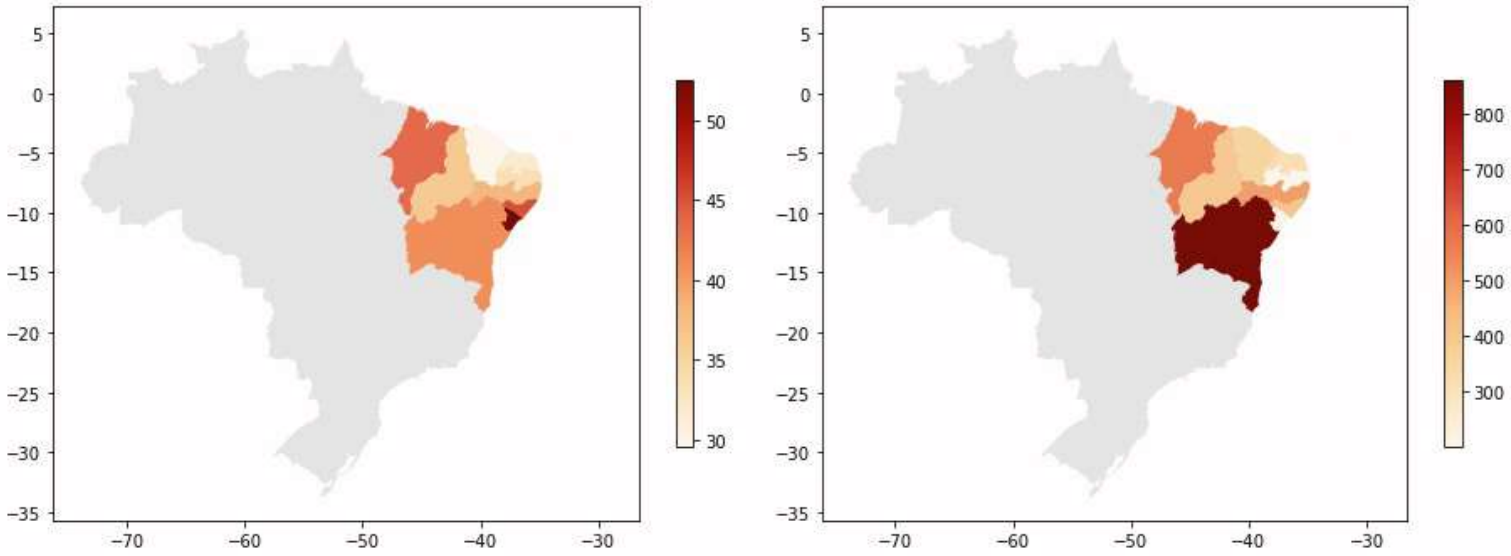
Drought evolution

3 Data and Methodology

Normalized drought events over time



3 Data and Methodology



Drought indicators

UF	Microbasins	Moderate Drought	Extreme Drought	Moderate Drought (normalized)	Extreme Drought (normalized)
MA	13	567	195	43.62	15.00
PI	11	399	132	36.27	12.00
CE	12	355	112	29.58	9.33
RN	10	318	159	31.80	15.90
PB	6	202	75	33.67	12.50
PE	13	497	195	38.23	15.00
AL	9	407	153	45.22	17.00
SE	4	210	75	52.50	18.75
BA	21	859	330	40.90	15.71

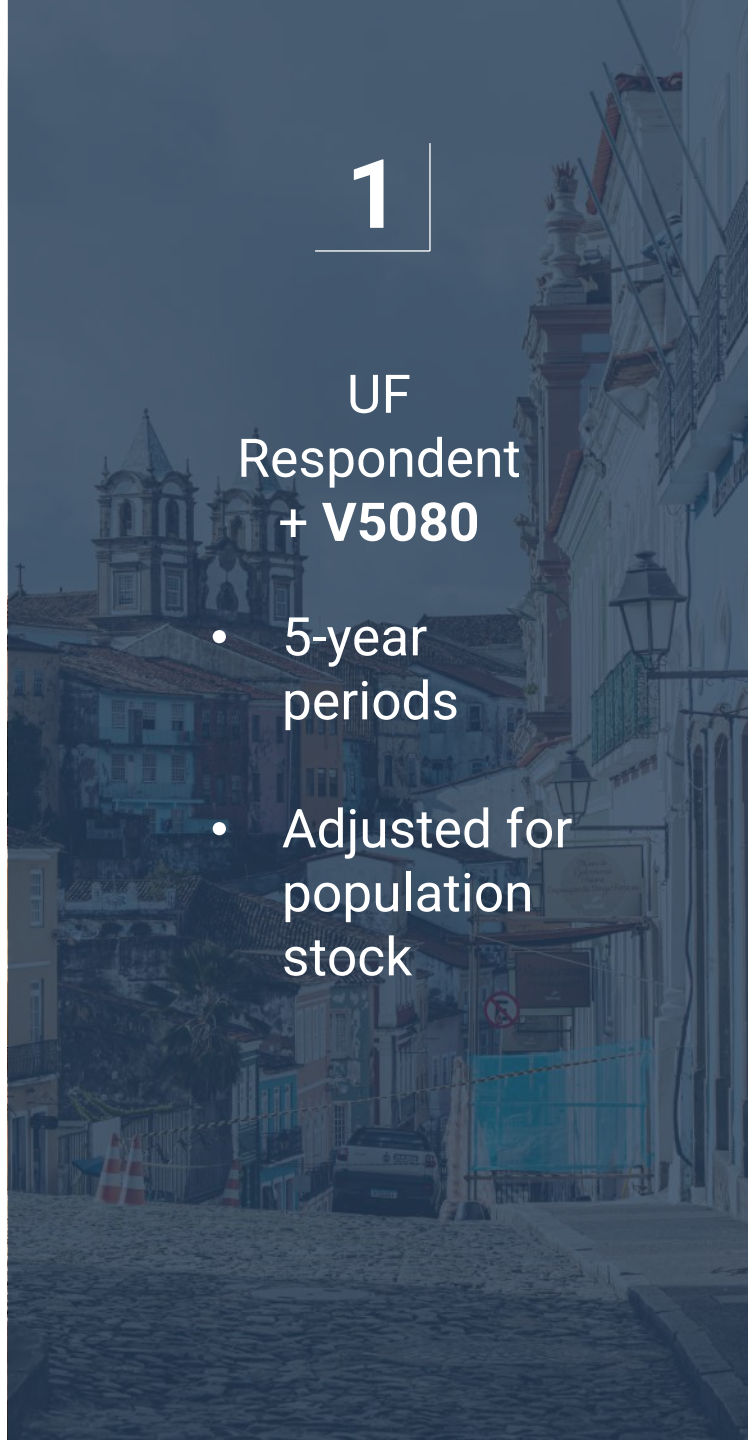
4 Results

Overlapping periods

5-year migration flows (Northeast - all regions)

Period	Out-migration NE	Out-migration NE (per 1,000 population)	In-migration NE	Net migration NE	Net migration NE (per 1,000 population)
1996 - 2001	937,801	20.85	750,373	-187,428	-4.17
1997 - 2002	945,711	20.86	854,259	-91,452	-2.02
1998 - 2003	935,715	20.43	823,117	-112,598	-2.46
1999 - 2004	938,996	20.29	864,066	-74,930	-1.62
2000 - 2005	921,101	19.69	889,396	-31,705	-0.68
2001 - 2006	899,552	18.61	863,932	-35,620	-0.74
2002 - 2007	660,103	13.51	588,014	-72,089	-1.48
2003 - 2008	728,909	14.77	549,904	-179,005	-3.63
2004 - 2009	742,869	14.73	542,907	-199,962	-3.97
2006 - 2011	857,048	16.61	695,215	-161,833	-3.14
2007 - 2012	835,266	16.21	596,496	-238,770	-4.63
2008 - 2013	742,311	13.98	523,259	-219,052	-4.13
2009 - 2014	848,566	15.83	607,114	-241,452	-4.51
2010 - 2015	561,780	10.58	413,449	-148,331	-2.79

Mariana Smiley
Bahia, Brazil



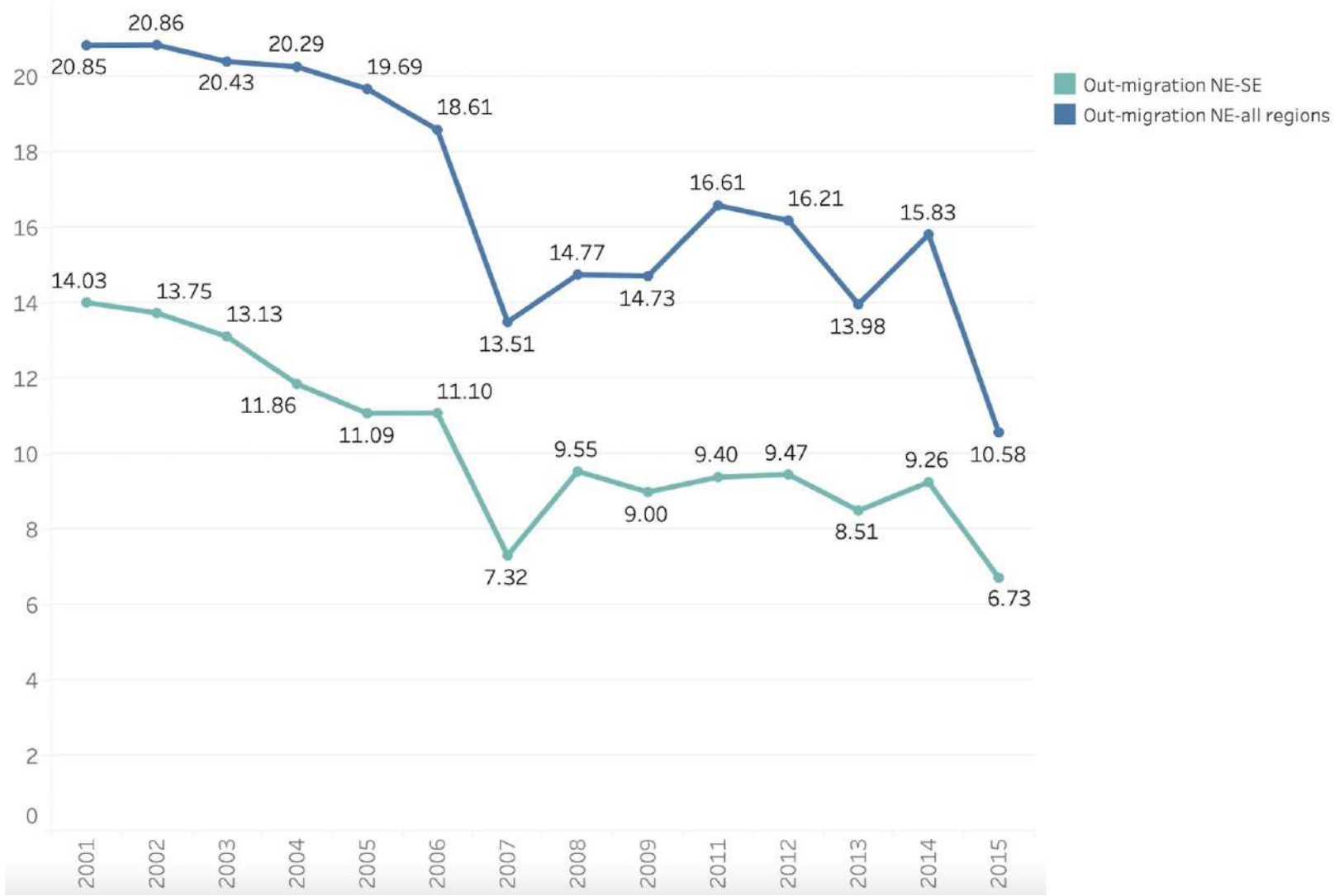
1

UF Respondent + **V5080**

- 5-year periods
- Adjusted for population stock

4 Results

Overlapping periods – Out-migration Northeast



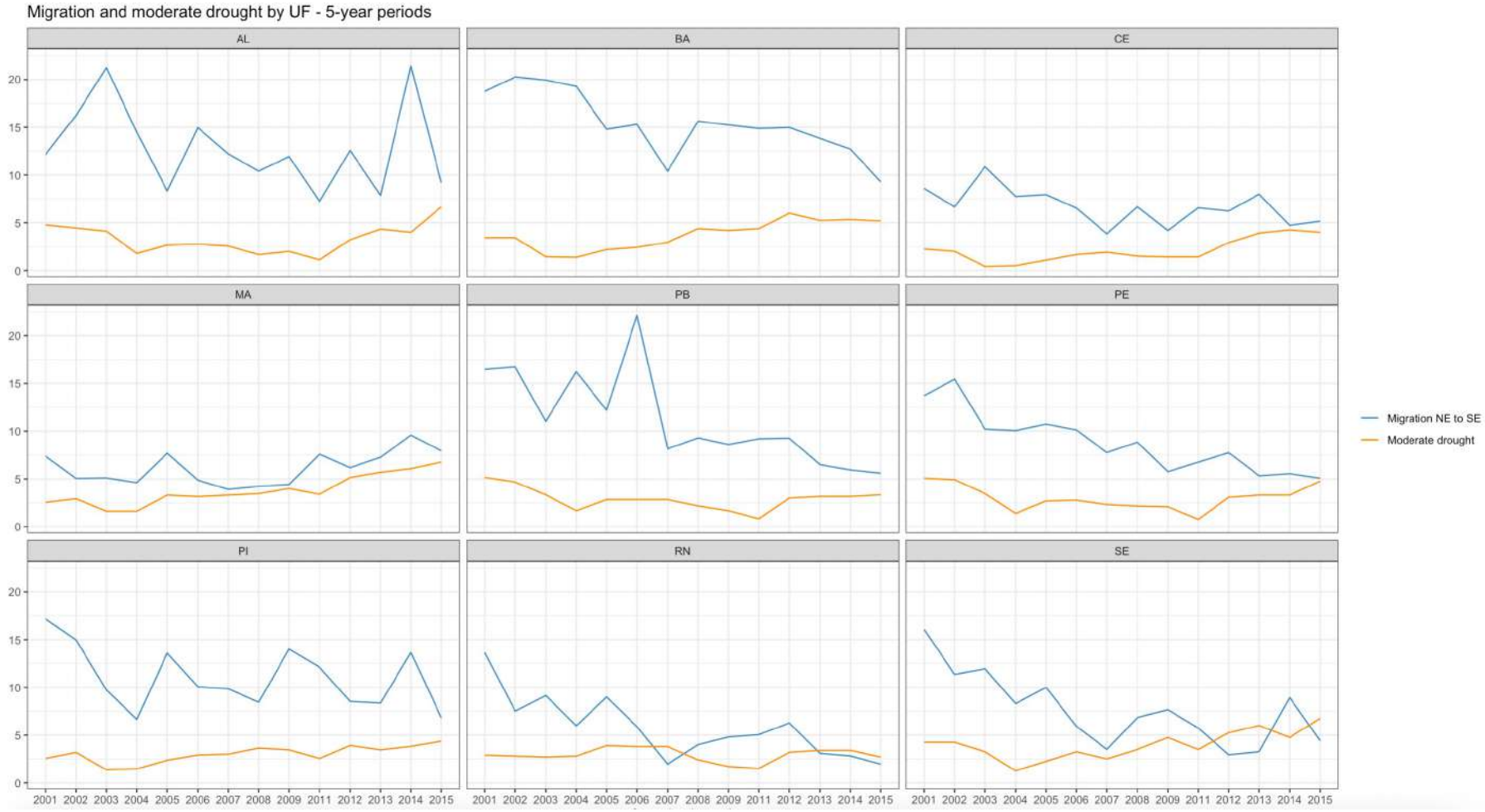
1

UF Respondent **+ V5080**

- 5-year periods
- Adjusted for population stock

4 Results

Overlapping periods – Northeast to the Southeast



1

UF Respondent + **V5080**

- 5-year periods
- Adjusted for population stock

4 Results

Non-overlapping periods

Migration from the Northeast to all regions

5-year migration flows (Northeast - all regions)

	1998/2003	2004/2009	2010/2015
Out-migration NE	935,715	742,869	561,780
Out-migration NE (per 1,000 population)	20.43	14.73	10.58
In-migration NE	823,117	542,907	413,449
Net migration NE	-112,598	-199,962	-148,331
Net migration NE (per 1,000 population)	-2.46	-3.97	-2.79

Out-migration from the Northeast to all regions has reduced significantly (48% adjusted to population) between 1998/2003 and 2010/2015.

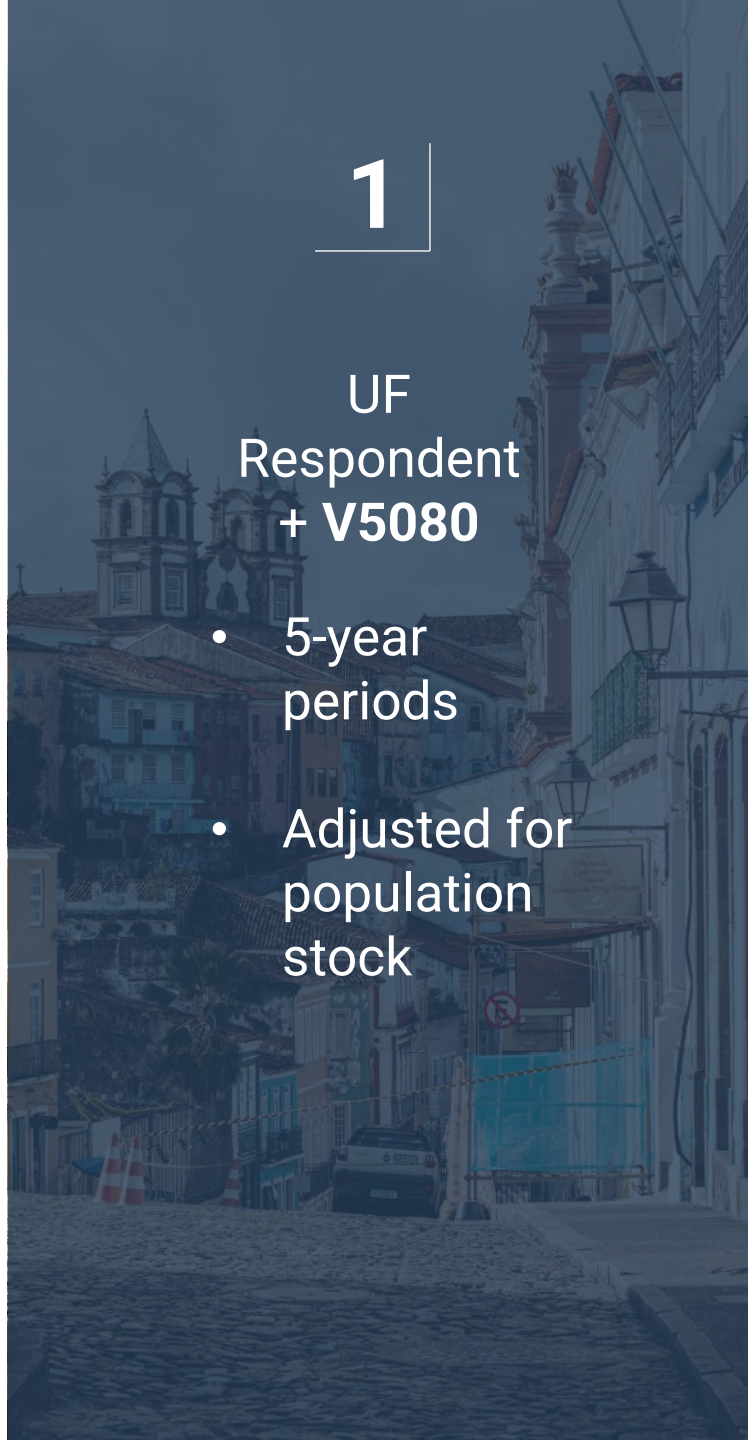
Net migration has shown a negative increase, connected to the reduction of the in-migration, but has returned to similar initial values with the reduction of out-migration in the 2010/2015 period

Migration from the Northeast to the Southeast

5-year migration flows (Northeast - Southeast)

	1998/2003	2004/2009	2010/2015
Migration NE to SE	601,331	453,693	357,132
Migration NE to SE per 1,000 pop	13.13	9.00	6.73
% of total out-migration	64.26%	61.07%	63.57%

Migration from the Northeast to the Southeast has also shown a 48% decrease, and NE-SE flows still account for a similar proportion of total out-migration.



1

UF Respondent + **V5080**

- 5-year periods
- Adjusted for population stock

4 Results

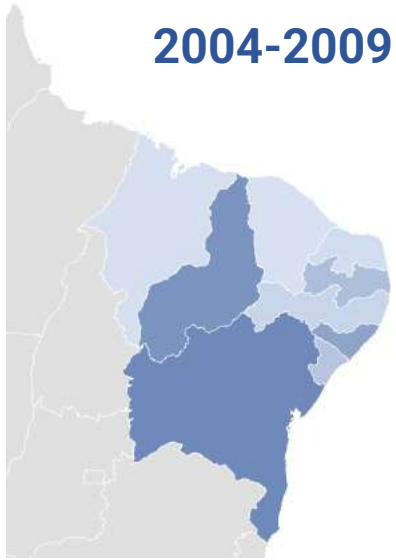
Non-overlapping periods – Migration NE-SE p. 1,000

1998-2003

1998 - 2003

UF	Out-migration to SE per 1,000 pop
AL	21.24
BA	19.94
CE	10.87
MA	5.10
PB	11.02
PE	10.22
PI	9.73
RN	9.15
SE	11.94

2004-2009



2004 - 2009

UF	Out-migration to SE per 1,000 pop
AL	11.94
BA	15.25
CE	4.19
MA	4.42
PB	8.61
PE	5.77
PI	14.01
RN	4.80
SE	7.65

2010-2015

2010 - 2015

UF	Out-migration to SE per 1,000 pop
AL	9.19
BA	9.29
CE	5.17
MA	7.94
PB	5.60
PE	5.08
PI	6.81
RN	1.96
SE	4.41

1

UF Respondent + **V5080**

- 5-year periods
- Adjusted for population stock

4 Results

Pernambuco, Brazil

Another possibility for estimating migration is using the yearly answers to the question:

"[Reference period], for how long has [person] lived, uninterruptedly, in this state/UF?"

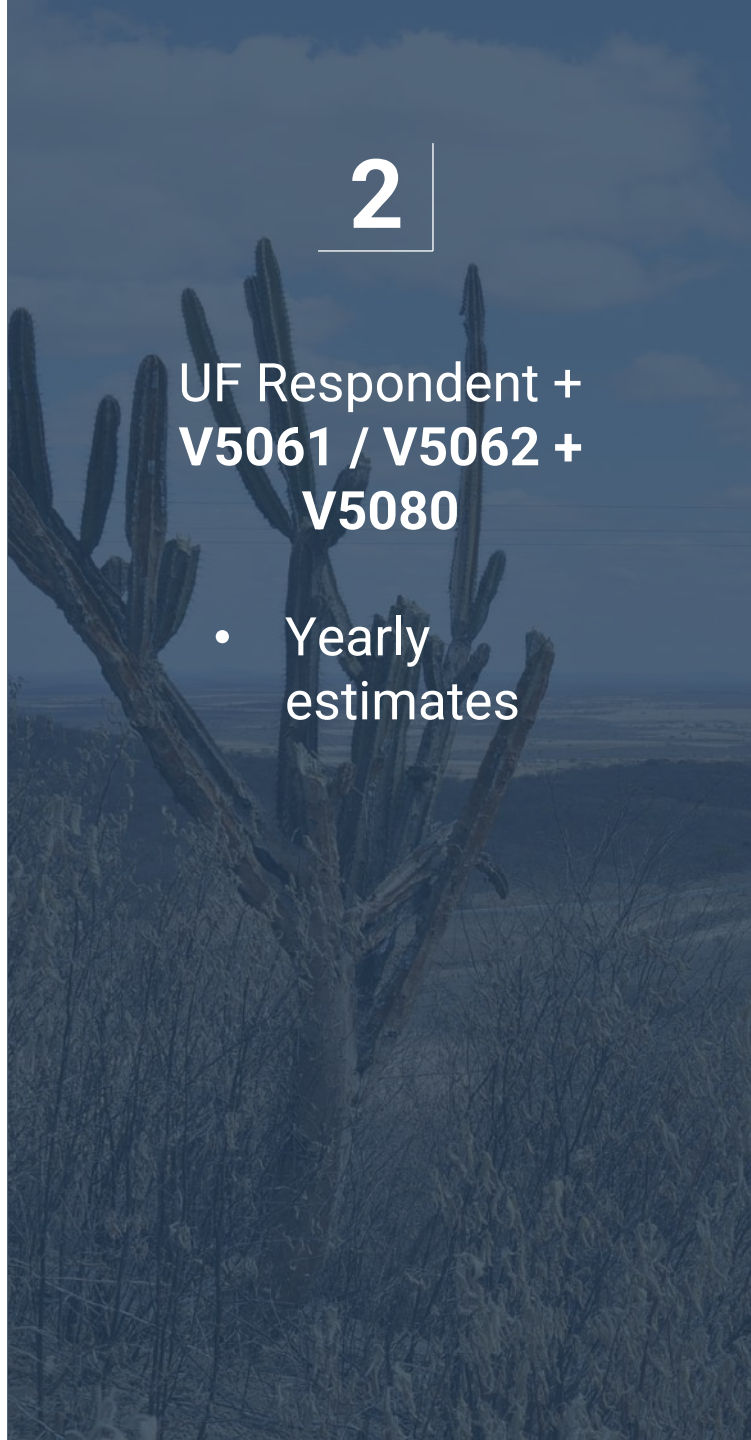
By pooling the cross-sectional datasets of the survey, we achieved different estimates based on the number of years residing in the state.

The point estimate for a specific year was calculated as a mean of the answers to different PNAD surveys.

There are substantial differences in the estimates of the two methods for the period analysed:

5-year and 1-year estimates comparison (Northeast - Southeast)

Period	Out-migration NE-SE (5-year estimate)	Out-migration NE-SE (1-year estimate)	Difference (%)
1998 - 2003	601,331	726,231	17.20
2004 - 2009	453,693	540,923	16.13
2010 - 2015	357,132	500,948	28.71

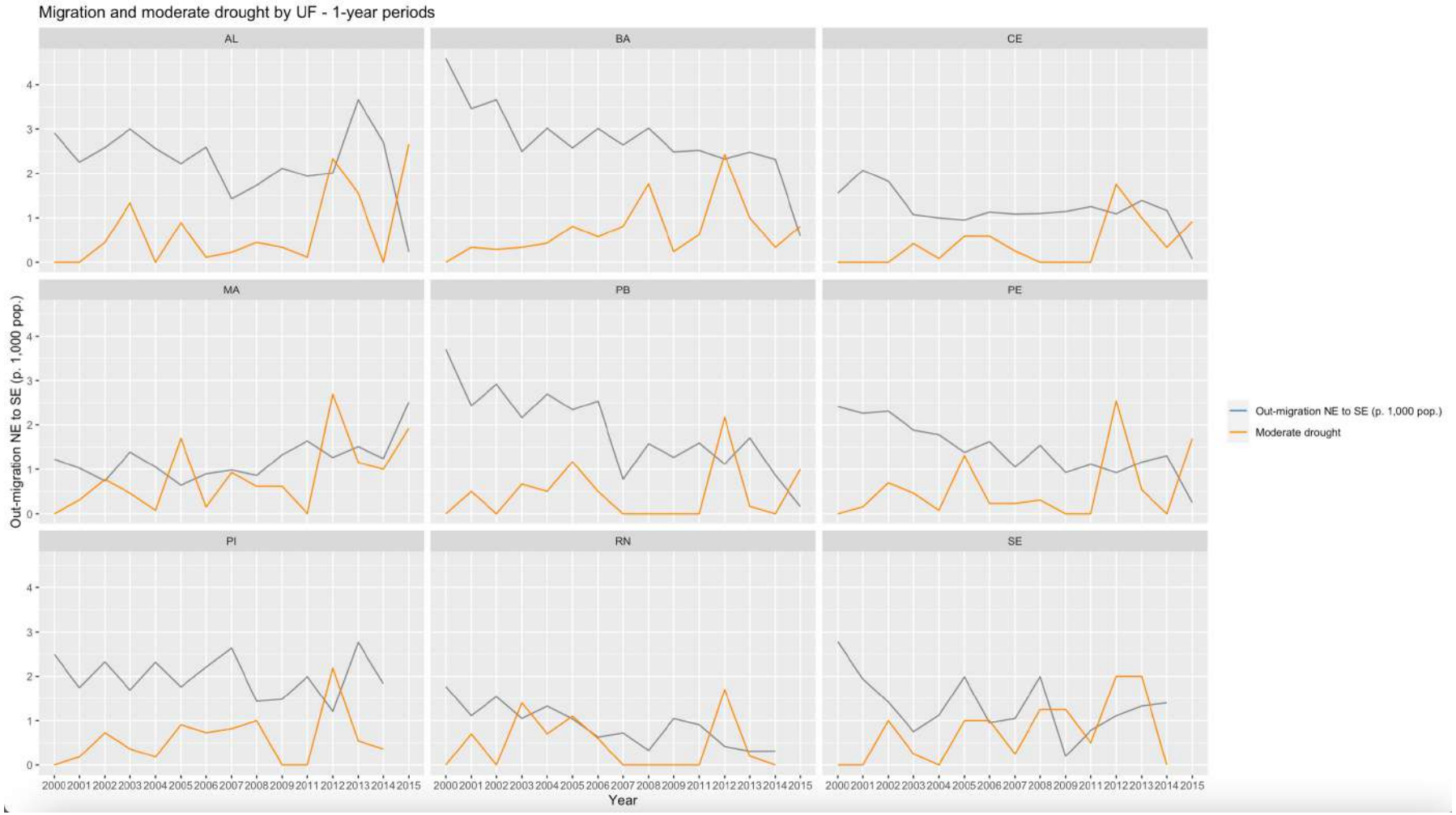


2

UF Respondent + **V5061 / V5062 + V5080**

- Yearly estimates

4 Results

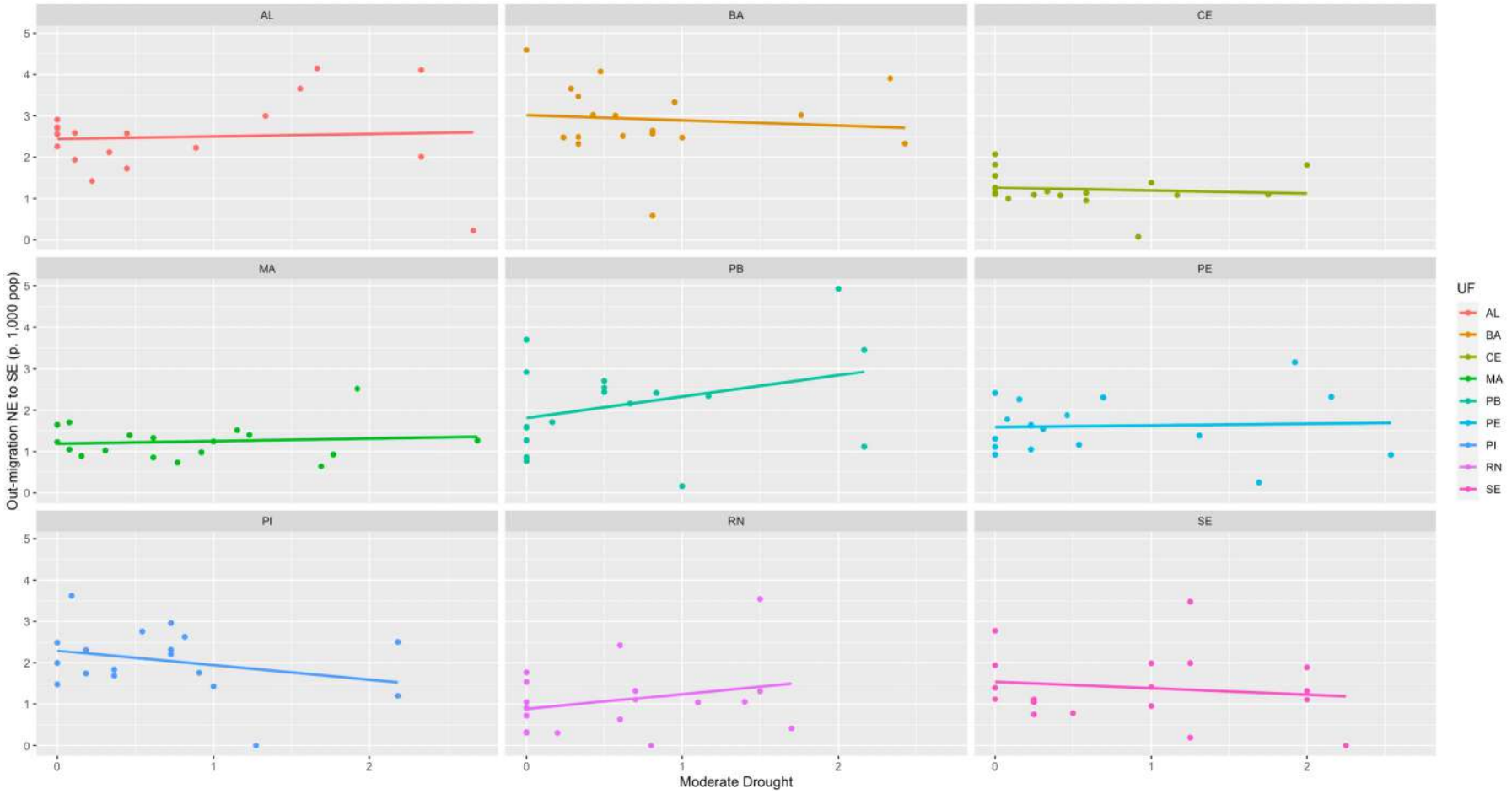


2

UF Respondent + **V5061 / V5062 + V5080**

- Yearly estimates

4 Results



2

UF Respondent + **V5061 / V5062 + V5080**

- Yearly estimates

4 Results

Table 1: One-year estimates (normalized)

	Dependent variable: Out-migration Northeast adjusted	
	(1)	(2)
Extreme Drought		0.115
		(0.307)
Moderate Drought	0.215	0.179
	(0.167)	(0.218)
Extreme Drought (t-1)		−0.365
		(0.306)
Moderate Drought (t-1)	0.298*	0.458**
	(0.166)	(0.207)
Extreme Drought (t-2)		−0.331
		(0.328)
Moderate Drought (t-2)	0.237	0.377*
	(0.189)	(0.214)
Year dummies	X	X
Observations	144	144
R^2	0.337	0.353
Adjusted R^2	0.242	0.242
Residual Std. Error	0.831 (df = 125)	0.832 (df = 122)
F Statistic	3.537*** (df = 18; 125)	3.169*** (df = 21; 122)
Note:		*p<0.1; **p<0.05; ***p<0.01

Fábio Hanashiro
Maranhão, Brazil

4 Results

Table 2: Five-year estimates (normalized)

	Dependent variable: Out-migration Northeast adjusted	
	(1)	(2)
Extreme Drought		−0.146
		(1.577)
Moderate Drought	0.221	0.268
	(0.649)	(0.863)
Extreme Drought (t-1)		0.881
		(2.016)
Moderate Drought (t-1)	0.593	0.284
	(0.880)	(1.152)
Extreme Drought (t-2)		−0.508
		(1.455)
Moderate Drought (t-2)	0.125	0.268
	(0.732)	(0.996)
Year dummies	X	X
Observations	108	108
R^2	0.197	0.199
Adjusted R^2	0.076	0.048
Residual Std. Error	4.146 (df = 93)	4.207 (df = 90)
F Statistic	1.627* (df = 14; 93)	1.318 (df = 17; 90)
Note:		*p<0.1; **p<0.05; ***p<0.01

5 Conclusions and next steps

- The main challenge for analysis combining survey microdata and meteorological big data is the **differences in time resolution**: survey microdata is collected less periodically than big data.

- Combining these two data sources requires a level of simplification (for the big data) and some degree of extrapolation (for the microdata), losing the advantages of big data and introducing a high degree of uncertainty to the microdata estimates.

- Five-year meteorological aggregations are not physically meaningful: seasonality effects are restricted to 1 year and climate effects are considered for above twenty years. However, several surveys measuring migration utilize a fixed date 5-year criteria. Although the five-year estimates for migration flows are more reliable (in the case of PNAD) than the one-year estimates, they have limitations for analysis.

- The potential for the use of big data and open data in social sciences is evident: having access to more granular data, possibilities for exploring new research questions and reducing costs of data collection. In order to leverage the potential of big data in the social sciences, traditional data collection methods (such as survey) could also consider methodological specificities of big data.

- We found no significant evidence for drought-induced migration from the Northeast to the Southeast in the period between 1996 and 2015. There is limited evidence of drought affecting UFs/states differently, which can be a future avenue for research.

2nd Research Conference on Forced Displacement
World Bank-UNHCR Joint Data Center on Forced Displacement (JDC), School of Economics at Universidad de Los Andes,
World Bank's Development Research Group

Thank you!

Marianne Bueno dos Passos Brum
marianne.brum@gmail.com

Guilherme Miranda Dutra
dutra.guilherme@gmail.com

Photo: Gabriel Tomaz, Ceará, Brazil