UCL

WORKING PAPERS SERIES

Paper 175 - Dec 11

A multi-level spatial interaction modelling framework for estimating inter-regional migration in Europe

A multi-level spatial interaction modelling framework for estimating inter-regional migration in Europe

Adam Dennett and Alan Wilson

Centre for Advanced Spatial Analysis,

University College London

Abstract

This paper presents a new spatial interaction modelling framework for estimating sub-national, international migration flows. A new family of models is introduced and exemplified for a sample system before issues of parameter calibration and model inputs are discussed using examples from Europe. Sub-optimum models are used to explore model assumptions and the accuracy of flow predictions across the European system, before we present the results of the optimum model and exemplify some important inter-regional flows which emerge from the model predictions.

Our full contact details are:

Mail address:

Centre for Advanced Spatial Analysis

University College London

Gower Street

London

WC1E 6BT

Physical address:

Centre for Advanced Spatial Analysis

1st Floor

90 Tottenham Court Road

London

W1T 4TJ

Email:

Adam Dennett

a.dennett@ucl.ac.uk

Alan Wilson

a.g.wilson@ucl.ac.uk

Adam Dennett is contact author

Contents

List of Tables

List of Figures

1 Introduction

Understanding migration is one of the enduring challenges facing geographers and demographers worldwide. The challenge persists, thanks to the range of territories and geographical scales of interest, the difficulty in dealing with inconsistent definitions of migrants and migration events, the variable (and often poor) quality of data and the large and sometimes complex array of tools available. Whilst an understanding of migration patterns and processes at the global scale presents possibly the largest challenge, in Europe we still know far less about the movements of people within the Union than may be expected given the continued desire for knowledge about population change and the amount of demographic data made available from member countries. (Poulain et al. 2006). Acknowledging this, a number of recent projects have made attempts to address some of the limitations of (intra-) European migration data. Against a background of varying migrant definitions, inconsistent data relating to the same flows collected for origins and destinations, and incomplete matrices, the MIMOSA (Modelling migration and migrant populations) project (Raymer and Abel 2008) produced a series of inter-country migration estimates for years between 2002 and 2006 through harmonising available data and using a multiplicative modelling framework to model flows between countries. Following on from this, the IMEM (Integrated Modelling of European Migration) project (van der Erf et al. - http://www.nidi.nl/Pages/NID/24/842.TGFuZz1FTkc.html) is currently looking to improve upon the methodology employed in MIMOSA through a Bayesian statistical approach. Further work has also been carried out by Abel (2010) who used a negative binomial regression (spatial interaction) model to estimate inter-country flows using a suite of predictor variables.

All of these projects have limited their scope to inter-country flows, but within Europe much of the focus of the EU commission is on regional policy (http://ec.europa.eu/regional_policy/ index_en.cfm) which is intended to address the quite marked socio-economic disparities which persist between smaller zones within the Union. A recent project which had a partial focus on migration at the regional (Nomenclature of Territorial Units for Statistics level 2 - NUTS2) level in the EU was the DEMIFER project (De Beer et al. 2010). One of the outputs from this project is a set of regional population projections for four different growth/cohesion scenarios which include a model of regional in- and out-migration based upon annual transition rates (Kupiszewska and Kupiszewski 2010). Whilst in- and out-migration rates tell us something about migration at the regional level within Europe, they reveal little about the interaction between regions and the hotspots of population exchange which occur within the Union helping drive the dynamism and evolution of local population structures. Indeed our knowledge of these exchanges across the whole Union is poor.

Therefore, in this paper we propose a methodology for estimating these inter-regional flows. The work builds on previous research which has made use of variations on the entropy maximising spatial

interaction models first introduced by Wilson (1970, 1971) and used in migration research (He and Pooler 2003; Plane 1982; Stillwell 1978). A new multi-level spatial interaction model is proposed which incorporates data at both country and regional levels in Europe to produce estimates of the inter- regional inter-country flows consistent with known information at these different levels.

2 Spatial system and the modelling challenge

2006 is the year for which the maximum amount of migration data at all levels are available, and so we use this as our temporal base. The spatial system of 287 NUTS2 regions nested within 31 countries (EU 27 + Norway, Iceland and Switzerland – which will be referred to as the 'EU system' in this paper subsequently) is shown in Figure 1. Migration data for some of the flows occurring are available. These data, along with cells representing missing data can be visualised as an origin/destination matrix as shown for a sample of countries in Figure 2. The grey cells in Figure 2 represent inter-regional intra-country (internal migration) migration flow counts which are available for most counties in the system. Flows within NUTS2 regions (the white cells on the diagonal) are not included in this analysis. The internal migration data were collated for use in the ESPON funded DEMIFER project (http://www.espon.eu/main/Menu_Projects/Menu_AppliedResearch/demifer.html), although in almost all cases, these data are freely available from the Eurostat statistics database (often referred to as 'New Cronos' - http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database). Internal migration data for two countries – France and Germany – are not available on this database, and were procured separately for DEMIFER from national statistical agencies. It should be noted, though, that whilst technically European NUTS2 zones, the French overseas departments of Guadeloupe, Martinique, Reunion and French Guiana are not included. The coloured cells represent inter-country flows. Consistent estimates of international (intra-Europe) origin/destination flows have been created for the 31 countries for our year of interest by Raymer and colleagues for the MIMOSA project (Raymer and Abel 2008). gional inter-country flows consistent with known information at these different levels.

Missing data in this EU system matrix are the inter-country, inter-regional flows – for example the flows from the three zones in Country 1 to the three zones in Country 3 which sum to the 4,856 migrants we know flowed between Country 1 and Country 3 in Figure 2. The modelling challenge, therefore, is to estimate this missing data in the matrix making use of information available at both the country and regional levels. The ultimate goal is to produce a full set of inter-regional estimates which make the most use of all available flow information at all levels within the system. Therefore it will be necessary to understand the full range of the models which can be built from the elements of the migration system. In defining a suite of models, it will become apparent that some are more likely to produce better results than others in different data scenarios – the model which produces the best

2

results in this current data scenario may not be feasible to use where less data exist, and so other less-optimum models in the family might produce the next best estimates given different data availability.
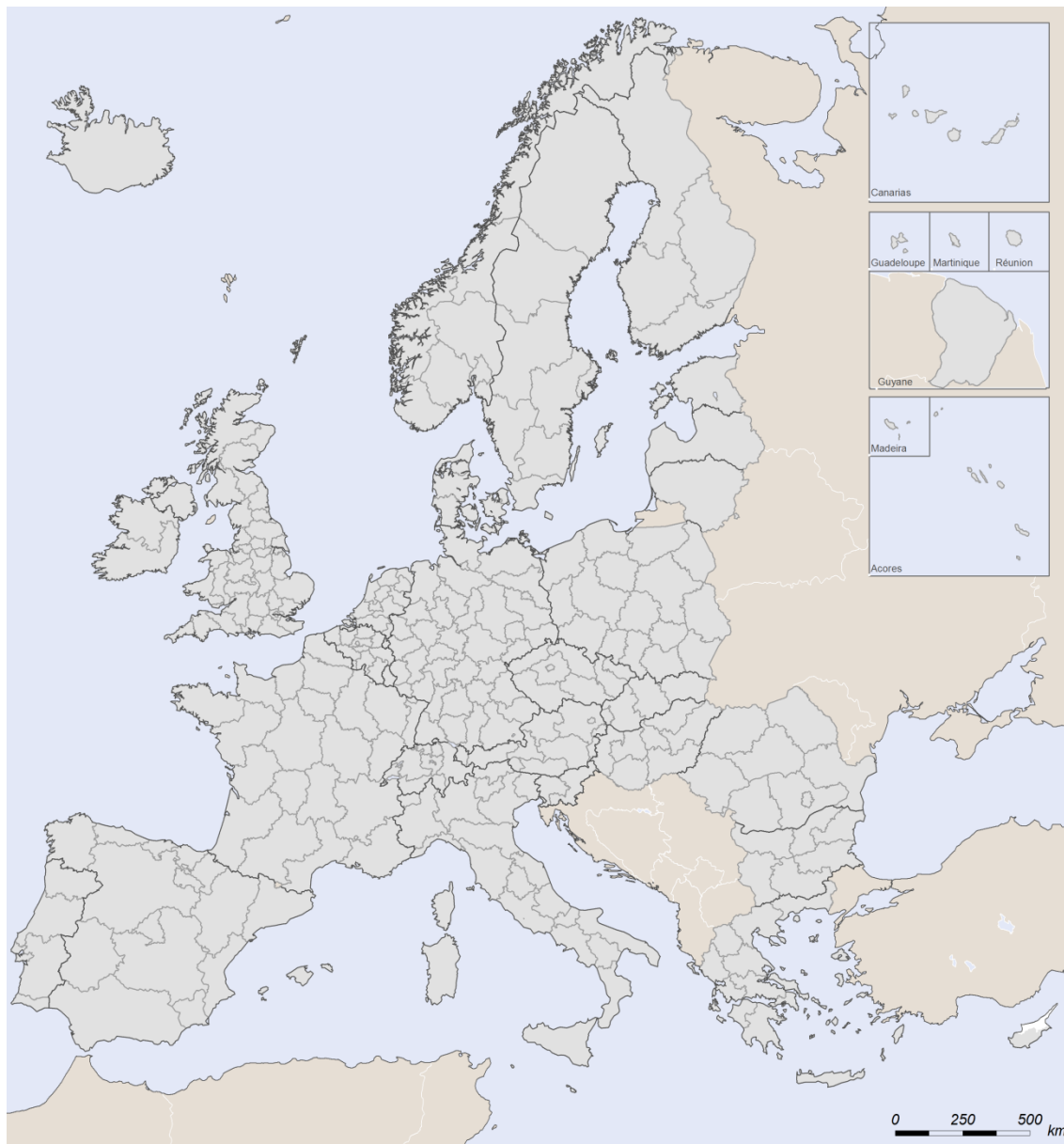
One question that arises from this challenge in the current context is whether it is feasible to treat this 287 zone EU system as a whole when it is the convention to make a distinction between 'internal migration' flows and 'international migration' flows. It could be argued that where national borders are real barriers to travel then two systems should be defined, however, in a post-Schengen Europe (Convey and Kupiszewski 1995; Kraler et al. 2006) national boundaries are not the rigid constructs (both metaphorically and physically) they once were, with flows of migrants between member countries now (in principle) as easy as flows within them. Indeed it is not uncommon for another type of human flow – daily commutes – to occur between countries such as Denmark and Sweden or

3

Luxembourg and Belgium (Mathä and Wintr 2009). With this being the case, we might expect internal migration and international migration in these areas of Europe to be virtually interchangeable in terms of, for example, the motivations for moves or the limiting factors such as distance which curtail flows. Whether this is actually the case will be explored though the modelling experiments with different models in the family detailed later in the paper.

		Country1			Country2			Country3		
		zone1	zone2	zone3	zone1	zone2	zone3	zone1	zone2	zone3
Country1	zone1	0	1131	1887		7211			4856	
	zone2	1633	0	14055						
	zone3	2301	20164	0						
Country2	zone1		9885		0	1608	328		8190	
	zone2				1252	0	1081			
	zone3				346	1332	0			
Country3	zone1		4992			4773		0	630	106
	zone2							546	0	569
	zone3							112	587	0

Figure 2 - Example migration data availability within Europe

3 Modelling methodology

To achieve the task set out in Section 2 we will make use of a variation on the doubly constrained entropy maximising spatial interaction model (Wilson 1970, 1971). Spatial interaction models (SIMs) are particularly appropriate in the context of migration where empirical studies and model experiments have demonstrated that the propensity to migrate decreases with distance (Boyle et al. 1998; Flowerdew 2010; Fotheringham et al. 2004; He and Pooler 2003; Singleton et al. 2010; Stillwell 1978; Taylor 1983). Indeed, Olsson (1970 p223) notes that *"Under the umbrella of spatial interaction and distance decay, it has been possible to accommodate most model work in transportation, migration, commuting and diffusion"*.

If T is the number of migrant *transitions,* (Rees 1977), let capital letters such as I and J denote countries and let lower case such as i and j denote NUTS2 regions within a country. Then let T^{IJ} be the number of migrants from country I to country J in some time period, say t to $t+1$ (which we will leave implicit for ease of notation). Then we can denote by T_{ij}^{IJ} the number of migrants from region i in I to region j in J. For convenience we denote all the migration flows by T, but the different of subscripts and superscripts indicate the different geographical levels in the system. This notation implies that we number the NUTS2 zones from $1, \ldots \ldots n$ for country I rather than numbering them consecutively for the whole system.

The available data described in Figure 2 can then be shown as in Figure 3. We have inter-regional, intra-country data for each country - T_{ij}^{IJ} where $I = J$. These internal migration flows could also be

described with the notation T_{ij}^{II} to distinguish them from inter-country inter-regional flows. Intra-regional flows - T_{ii}^{II} - are not available. At the country level, inter-country flows T^{IJ} are available.

		Country1			Country2			Country3		
		zone1	zone2	zone3	zone1	zone2	zone3	zone1	zone2	zone3
Country1	zone1	T^{11}_{11}	T^{11}_{12}	T^{11}_{13}		T^{12}			T^{13}	
	zone2	T^{11}_{21}	T^{11}_{22}	T^{11}_{23}						
	zone3	T^{11}_{31}	T^{11}_{32}	T^{11}_{33}						
Country2	zone1		T^{21}		T^{22}_{11}	T^{22}_{12}	T^{22}_{13}		T^{23}	
	zone2				T^{22}_{21}	T^{22}_{22}	T^{22}_{23}			
	zone3				T^{22}_{31}	T^{22}_{32}	T^{22}_{33}			
Country3	zone1		T^{31}			T^{32}		T^{33}_{11}	T^{33}_{12}	T^{33}_{13}
	zone2							T^{33}_{21}	T^{33}_{22}	T^{33}_{23}
	zone3							T^{33}_{31}	T^{33}_{32}	T^{33}_{33}

Figure 3 – Sample system in Figure 2 using defined notation.

The row and column totals are known for the T_{ij}^{II} elements, i.e. at the NUTS2 level, and also for the T^{IJ} inter-country levels. Let these be M_i^I and N_j^J and O^I and D^J respectively so that:

$$M_i^I = \sum_j T_{ij}^{II} = \sum_j T_{ij}^{IJ}, I = J \tag{1}$$

$$N_j^J = \sum_i T_{ij}^{II} = \sum_i T_{ij}^{IJ}, I = J \tag{2}$$

$$O^I = \sum_J T^{IJ}, J \neq I \tag{3}$$

$$D^J = \sum_I T^{IJ}, I \neq J \tag{4}$$

These row and column totals are depicted in expanded versions of Figures Figure 2 and Figure 3, shown in Figures Figure 4a and b. Note that the O^I and D^J totals do not include intra country data contained in the M_i^I and N_j^J totals – consistent with the common practice of not including intra-country flows in international migration analysis. Internal migration data are assumed to be consistent such that:

$$\sum_i M_i^I = \sum_j N_j^J = \sum_{ij} T_{ij}^{II} \tag{5}$$

		Country1				Country2				Country3				
		zone1	zone2	zone3		zone1	zone2	zone3		zone1	zone2	zone3		
Country1	zone1	0	1131	1887	3018									
	zone2	1633	0	14055	15688		7211				4856			12067
	zone3	2301	20164	0	22465									
		3934	21295	15942	41171									
Country2	zone1					0	1608	328	1936					
	zone2		9885			1252	0	1081	2333		8190			18075
	zone3					346	1332	0	1678					
						1598	2940	1409	5947					
Country3	zone1									0	630	106	736	
	zone2		4992				4773			546	0	569	1115	9765
	zone3									112	587	0	699	
										658	1217	675	2550	
			14877				11984				13046			

a)

		Country1				Country2				Country3				
		zone1	zone2	zone3		zone1	zone2	zone3		zone1	zone2	zone3		
Country1	zone1	T^{11}_{11}	T^{11}_{12}	T^{11}_{13}	M^1_1									
	zone2	T^{11}_{21}	T^{11}_{22}	T^{11}_{23}	M^1_2		T^{12}				T^{13}			O^1
	zone3	T^{11}_{31}	T^{11}_{32}	T^{11}_{33}	M^1_3									
		N^1_1	N^1_2	N^1_3										
Country2	zone1					T^{22}_{11}	T^{22}_{12}	T^{22}_{13}	M^2_1					
	zone2		T^{21}			T^{22}_{21}	T^{22}_{22}	T^{22}_{23}	M^2_2		T^{23}			O^2
	zone3					T^{22}_{31}	T^{22}_{32}	T^{22}_{33}	M^2_3					
						N^2_1	N^2_2	N^2_3						
Country3	zone1									T^{33}_{11}	T^{33}_{12}	T^{33}_{13}	M^3_1	
	zone2		T^{31}				T^{32}			T^{33}_{21}	T^{33}_{22}	T^{33}_{23}	M^3_2	O^3
	zone3									T^{33}_{31}	T^{33}_{32}	T^{33}_{33}	M^3_3	
										N^3_1	N^3_2	N^3_3		
			D^1				D^2				D^3			

b)

Figure 4 - Expanded sample system with margins and sub-margins

The sample data shown in Figure 4a and b represent the information we currently have about our system of interest. The formulation thus far implies that we are not seeking to model flows at the NUTS2 level within each country, I (we have these data) and to and from other countries, J, $J \neq I$. The ultimate modelling goal, however, *is* to estimate these inter-country regional level flows, effectively filling all T^{IJ}_{ij} interior cells in the matrix.

In order to model these NUTS2 level flows between countries we introduce another element of notation: T^I_{iJ} and T^J_{Ij} are, respectively, the out-migration flows from NUTS2 i in country I to country J ($\neq I$) and the in-migration flows to NUTS2 j in country J ($\neq I$) from country I. T^I_{iJ} and T^J_{Ij} can be viewed as table sub-margins and are equivalent to M^I_i and N^J_j (where the country subscripts are dropped as flows are internal) so that:

$$T_{iJ}^I = \sum_{j\in J} T_{ij}^{IJ} \tag{6}$$

$$T_{Ij}^J = \sum_{i\in I} T_{ij}^{IJ} \tag{7}$$

Then T^{IJ} in (3) and (4), for $I \neq J$, would be given by:

$$T^{IJ} = \sum_{i\in I} T_{iJ}^I = \sum_{j\in J} T_{Ij}^J = \sum_{i\in I}\sum_{j\in J} T_{ij}^{IJ} \tag{8}$$

These sub-margin elements are shown in Figure 5a and b. In addition to these new sub-margins, two new row and column margins can also be calculated. O_i^I and D_j^J and directly related to O^I and D^J in that:

$$O_i^I = \sum_J T_{iJ}^I \tag{9}$$

$$O^I = \sum_{i\in I} O_i^I \tag{10}$$

$$D_j^J = \sum_I T_{Ij}^J \tag{11}$$

$$D^J = \sum_{j\in J} D_j^J \tag{12}$$

A final set of margins can be calculated for all interior cells in the matrix where:

$$P_i^I = O_i^I + M_i^I \tag{13}$$

$$Q_j^J = D_j^J + N_j^J \tag{14}$$

With a complete system description, we can then consider the variety of models which can be built. Equations (1), (2), (3), (4), (6), (7), (9), (11), (13) and (14) can provide the core constraint equations for a suite of entropy maximising models which can be used to estimate various elements and aggregations of the T_{ij}^{IJ} flows in the multi-level system matrix. We might describe this as a family of multi-level spatial interaction models (MLSIMs), with the model possibilities being:

i. Model the NUTS2 flows within each country separately – that is model T_{ij}^{II} (in which case I simply functions as a label for each country model. Equations (1) and (2) would be the accounting/constraint equations.

ii.   Model the inter-country flows, T^{IJ} separately. Equations (3) and (4) would be the accounting equations.

iii. Model asymmetric NUTS2 flows i and j in and out of each I and ($\neq I$), T_{iJ}^{I} and T_{Ij}^{J}. Three versions of the asymmetric model can be formulated.

a.   Equations (9) and (3) would hold as accounting/constraint equations for Equation (6) and Equations (11) and (4) would be the constraints for Equation (7).

b.   Known T^{IJ} flows with Equation (9) would hold as constraints for Equation (6) and known T^{IJ} flows with Equation (11) would hold as constraints for Equation 7.

c.   It would also be possible to use Equations (13) and (3) as the constraints for Equation (6) and Equations (14) and (4) as the constraints for Equation (7). This model is almost identical to a), although in this case we would also be modelling M_i^I as T_{iJ}^{I} and N_j^J as T_{Ij}^{J}.

iv.   Model T_{ij}^{IJ} for each country separately using sub-margins (6) and (7) as constraints.

v.   Model T_{ij}^{IJ} where $I \neq J$ with Equations (9) and (11) as constraints.

vi.   Model the full array of NUTS2 regions, T_{ij}^{IJ} using Equations (13) and (14) as the accounting constraints.

		Country1				Country2				Country3						
		zone1	zone2	zone3		zone1	zone2	zone3		zone1	zone2	zone3				
Country1	zone1	0	1131	1887	3018	7211			529	4856			356	885	3903	12067
	zone2	1633	0	14055	15688				2748				1850	4598	20286	
	zone3	2301	20164	0	22465				3935				2650	6584	29049	
		3934	21295	15942	41171	1181	2455	3575	7211	2513	1604	738	4856			
Country2	zone1	9885			3218	0	1608	328	1936	8190			2666	5884	7820	18075
	zone2				3878	1252	0	1081	2333				3213	7091	9424	
	zone3				2789	346	1332	0	1678				2311	5100	6778	
		818	3598	5470	9885	1598	2940	1409	5947	4239	2706	1245	8190			
Country3	zone1	4992			1441	4773			1378	0	630	106	736	2177	3554	9765
	zone2				2183				2087	546	0	569	1115	3298	5385	
	zone3				1368				1308	112	587	0	699	2067	3376	
		413	1817	2762	4992	782	1625	2366	4773	658	1217	675	2550			
		1230	5415	8232		2380	4565	3775		4897	3923	1920				
		5164	26710	24174		3561	7020	7350		7410	5527	2659				
			14877				11984				13046					

a)

		Country1				Country2				Country3						
		zone1	zone2	zone3		zone1	zone2	zone3		zone1	zone2	zone3				
Country1	zone1	T^{11}_{11}	T^{11}_{12}	T^{11}_{13}	M^1_1	T^{12}			T^1_{12}	T^{13}			T^1_{13}	O^1_1	P^1_1	O^1
	zone2	T^{11}_{21}	T^{11}_{22}	T^{11}_{23}	M^1_2				T^1_{22}				T^1_{23}	O^1_2	P^1_2	
	zone3	T^{11}_{31}	T^{11}_{32}	T^{11}_{33}	M^1_3				T^1_{32}				T^1_{33}	O^1_3	P^1_3	
		N^1_1	N^1_2	N^1_3		T^2_{11}	T^2_{12}	T^2_{13}		T^3_{11}	T^3_{12}	T^3_{13}				
Country2	zone1	T^{21}			T^2_{11}	T^{22}_{11}	T^{22}_{12}	T^{22}_{13}	M^2_1	T^{23}			T^2_{13}	O^2_1	P^2_1	O^2
	zone2				T^2_{21}	T^{22}_{21}	T^{22}_{22}	T^{22}_{23}	M^2_2				T^2_{23}	O^2_2	P^2_2	
	zone3				T^2_{31}	T^{22}_{31}	T^{22}_{32}	T^{22}_{33}	M^2_3				T^2_{33}	O^2_3	P^2_3	
		T^1_{21}	T^1_{22}	T^1_{23}		N^2_1	N^2_2	N^2_3		T^3_{21}	T^3_{22}	T^3_{23}				
Country3	zone1	T^{31}			T^3_{11}	T^{32}			T^3_{12}	T^{33}_{11}	T^{33}_{12}	T^{33}_{13}	M^3_1	O^3_1	P^3_1	O^3
	zone2				T^3_{21}				T^3_{22}	T^{33}_{21}	T^{33}_{22}	T^{33}_{23}	M^3_2	O^3_2	P^3_2	
	zone3				T^3_{31}				T^3_{32}	T^{33}_{31}	T^{33}_{32}	T^{33}_{33}	M^3_3	O^3_3	P^3_3	
		T^1_{31}	T^1_{32}	T^1_{33}		T^2_{31}	T^2_{32}	T^2_{33}		N^3_1	N^3_2	N^3_3				
		D^1_1	D^1_2	D^1_3		D^2_1	D^2_2	D^2_3		D^3_1	D^3_2	D^3_3				
		Q^1_1	Q^1_2	Q^1_3		Q^2_1	Q^2_2	Q^2_3		Q^3_1	Q^3_2	Q^3_3				
		D^1				D^2				D^3						

b)

Figure 5 - Sample system including all sub-margin and margin elements

If the accounting equations (1) to (4) are deployed as in Models (i) and (ii), this leads to the construction of doubly-constrained models for which the main task would be to identify impedance functions, associated generalised costs c_{ij}, and the model parameter values. In migration research cost is often the physical distance between places: the propensity to migrate decreases with distance and thus the cost of travel can be inferred to increase. Empirical studies have shown that this distance decay in migration propensity will often follow either a negative exponential or inverse power law (Stillwell 1978). In spatial interaction models this is represented by a parameter, β, (normally negative) which can be calibrated endogenously if data exist. In the equations which follow, we write the distance decay function, f, as exponential - $f(c_{ij}) = e^{\beta c_{ij}}$ - although it would be just as appropriate to write it as a power law - $f(c_{ij}) = c_{ij}^{\beta}$.

3.1 Model (i)

Model (i) is the most straightforward and would produce:

$$T_{ij}^{II} = A_i^I B_j^I M_i^I N_j^I e^{\beta^I c_{ij}^I} \tag{15}$$

$$A_i^I = \frac{1}{\sum_j B_j^I N_j^I e^{\beta^I c_{ij}^I}} \tag{16}$$

$$B_j^I = \frac{1}{\sum_i A_i^I M_i^I e^{\beta^I c_{ij}^I}} \tag{17}$$

where the generalised distance decay parameter β can be calibrated endogenously using T_{ij}^{II} data. An alternative version of this model could calculate origin or destination-specific β parameters:

$$T_{ij}^{II} = A_i^I B_j^I M_i^I N_j^I e^{\beta_i^I c_{ij}^I} \tag{18}$$

$$T_{ij}^{II} = A_i^I B_j^I M_i^I N_j^I e^{\beta_j^I c_{ij}^I} \tag{19}$$

3.2 Model (ii)

The inter-country, Model (ii), would be:

$$T^{IJ} = A^I B^J O^I D^J e^{\mu C^{IJ}} \tag{20}$$

where balancing factors are calculated with equivalent equations to (16) and (17).

3.3 Model (iii)

The asymmetric models in Model (iiia) would take the form:

$$T_{iJ}^I = A_i^I B^J O_i^I D^J e^{\beta_i^I c_{iJ}^I} \tag{21}$$

$$T_{Ij}^J = A^I B_j^J O^I D_j^J e^{\beta_j^J c_{Ij}^J} \tag{22}$$

With the balancing factors for (21):

$$A_i^I = \frac{1}{\sum_J B^J D^J e^{\beta_i^I c_{iJ}^I}} \tag{23}$$

$$B^J = \frac{1}{\sum_i A_i^I O_i^I e^{\beta_i^I c_{Ij}^I}} \tag{24}$$

and the balancing factors for (22):

$$A^I = \frac{1}{\sum_j B_j^J D_j^J e^{\beta_j^J c_{iJ}^J}} \tag{25}$$

$$B_j^J = \frac{1}{\sum_I A^I O^I e^{\beta_j^J c_{Ij}^J}} \tag{26}$$

Equations (21) and (22) can be visualised easily by collapsing the matrices in Figure 5a and b into just the relevant margins and sub-margins (Figures 4a and 4b, 5a and 5b). These margins then become, effectively, the i, j values in a standard two-dimensional matrix.

	T_{iJ}^I	Country1	Country2	Country3		
	zone1		529	356	885	
Country1	zone2		2748	1850	4598	12067
	zone3		3935	2650	6584	
	zone1	3218		2666	5884	
Country2	zone2	3878		3213	7091	18075
	zone3	2789		2311	5100	
	zone1	1441	1378		2818	
Country3	zone2	2183	2087		4270	9765
	zone3	1368	1308		2677	
		14877	11984	13046	39907	

a)

b)

Figure 6 - Collapsed matrix showing only region-to-country sub-margins depicted in Figure 5

T^J_{Ij}	Country1			Country2			Country3			
	zone1	zone2	zone3	zone1	zone2	zone3	zone1	zone2	zone3	
Country1				1181	2455	3575	2513	1604	738	12067
Country2	818	3598	5470				4239	2706	1245	18075
Country3	413	1817	2762	782	1625	2366				9765
	1230	5415	8232	1963	4080	5941	6752	4310	1984	39907
		14877			11984			13046		

a)

T^J_{Ij}	Country1			Country2			Country3			
	zone1	zone2	zone3	zone1	zone2	zone3	zone1	zone2	zone3	
Country1	T^1_{11}	T^1_{12}	T^1_{13}	T^2_{11}	T^2_{12}	T^2_{13}	T^3_{11}	T^3_{12}	T^3_{13}	O^1
Country2	T^1_{21}	T^1_{22}	T^1_{23}	T^2_{21}	T^2_{22}	T^2_{23}	T^3_{21}	T^3_{22}	T^3_{23}	O^2
Country3	T^1_{31}	T^1_{32}	T^1_{33}	T^2_{31}	T^2_{32}	T^2_{33}	T^3_{31}	T^3_{32}	T^3_{33}	O^3
	D^1_1	D^1_2	D^1_3	D^2_1	D^2_2	D^2_3	D^3_1	D^3_2	D^3_3	
		D^1			D^2			D^3		

b)

Figure 7 - Collapsed matrix showing only country-to-region sub-margins depicted in Figure 5

It is important to note that whilst in the examples in Figure 6a and Figure 7a, corresponding country to country sums are equal – e.g. $\sum_i T^I_{iJ} = \sum_j T^J_{Ij}$ – as they should be, in model (iiia) the modelled values will not correspond in this way, due to the constraints used. To exemplify, consider Figure 8 and

13

Figure 9. The marginal values in these figures are almost identical to those in Figure 6a and Figure 7a (only 2 migrants are misplaced in Figure 8). The interior T_{iJ}^I and T_{Ij}^J values are quite different. In these modelled matrices, $\sum_i T_{iJ}^I \neq \sum_j T_{Ij}^J$. For example the total flows from Country1 to Country 2 in Figure 8 are 6915, whereas the total flows from Country1 to Country 2 in Figure 9 are 7776. The reason for this is that the T_{iJ}^I and T_{Ij}^J flows are only constrained to the marginal totals – either O_i^I and D^J or O^I and D_j^J respectively. In these models, T_{iJ}^I and T_{Ij}^J have multiple equilibria, only a small number of which result in $\sum_i T_{iJ}^I = \sum_j T_{Ij}^J$. This has implications for Model (iv) in our suite of models.

T^I_{iJ}		Country1	Country2	Country3		
Country1	zone1	0	682	203	885	
	zone2	0	2441	2157	4598	12066
	zone3	0	3792	2792	6584	
Country2	zone1	2940	0	2944	5884	
	zone2	4461	0	2630	7091	18076
	zone3	2780	0	2320	5100	
Country3	zone1	1173	1645	0	2818	
	zone2	2086	2184	0	4270	9765
	zone3	1437	1240	0	2677	
		14877	11984	13046	39907	

Figure 8 - T_{iJ}^I values modelled using the entropy maximising model in (21)

T^J_{Ij}	Country1			Country2			Country3			
	zone1	zone2	zone3	zone1	zone2	zone3	zone1	zone2	zone3	
Country1	0	0	0	760	2968	4048	910	2459	922	12067
Country2	610	3820	4890	0	0	0	5842	1851	1062	18075
Country3	620	1595	3342	1203	1112	1893	0	0	0	9765
	1230	5415	8232	1963	4080	5941	6752	4310	1984	39907
		14877			11984			13046		

Figure 9 - T_{Ij}^J values modelled using the entropy maximising model in (22)

3.4 Model (iv)

Model (iv) takes T_{iJ}^I and T_{Ij}^J as constraints, with the doubly constrained version of the model defined as:

$$T_{ij}^{IJ} = A_{iJ}^I B_{Ij}^J T_{iJ}^I \, T_{Ij}^J e^{\beta_i^I c_{ij}^I} \qquad (27)$$

14

$$T_{ij}^{IJ} = A_{iJ}^I B_{Ij}^J T_{iJ}^I T_{Ij}^J e^{\beta_j^J c_{ij}^J} \tag{28}$$

With the balancing factors for (27):

$$A_{iJ}^I = \frac{1}{\sum_{j \in J} B_{Ij}^J T_{Ij}^J e^{\beta_i^I c_{ij}^I}} \tag{29}$$

$$B_{Ij}^J = \frac{1}{\sum_{i \in I} A_{iJ}^I T_{iJ}^I e^{\beta_i^I c_{ij}^I}} \tag{30}$$

If $\sum_i T_{iJ}^I = \sum_j T_{Ij}^J$ then it is possible to solve Equations (27) and (28) – the iterative procedure which calculates the A_{iJ}^I and B_{Ij}^J balancing factors is able to converge when $\sum_i T_{iJ}^I$ and its corresponding sub-margin $\sum_j T_{Ij}^J$ are the same value. If T_{iJ}^I and T_{Ij}^J values are estimated using the entropy maximising procedure described in Equations (21) and (22), then $\sum_i T_{iJ}^I \neq \sum_j T_{Ij}^J$, meaning that the iterative balancing factor routine will not converge and Equations (27) and (28) cannot be solved.

One solution to this issue is to estimate T_{iJ}^I and T_{Ij}^J using a method other than the entropy maximising model described. As already noted, T_{iJ}^I and T_{Ij}^J are equivalent to M_i^I and N_j^J. In this system we already know the values of M_i^I and N_j^J from the T_{ij}^{II} internal migration data available. Given this information the following equations can be used to estimate T_{iJ}^I and T_{Ij}^J:

$$T_{iJ}^I = \left(\frac{M_i^I}{T^{II}}\right) T^{IJ} \tag{31}$$

$$T_{Ij}^J = \left(\frac{N_j^J}{T^{II}}\right) T^{IJ} \tag{32}$$

Where these T_{iJ}^I and T_{Ij}^J estimates are constrained to the corresponding T^{IJ} values, $\sum_i T_{iJ}^I = \sum_j T_{Ij}^J$ and thus it is possible to solve Equations (27) and (28).

There is, however, an entropy maximising solution to this issue as well. In Model (iiib) the constraints used to estimate T_{iJ}^I and T_{Ij}^J are not the matrix margins as shown in Figure 6 and Figure 7. By using these margins in (iiia) we are not taking advantage of all known information in the system. As T^{IJ} flows are known, a combination of matrix margins and known interior T^{IJ} values can be used as constraints, thus the equations for T_{iJ}^I and T_{Ij}^J become:

$$T_{iJ}^I = A_i^I X^{IJ} O_i^I T^{IJ} e^{\beta_i^I c_{iJ}^I} \tag{33}$$

$$T_{Ij}^J = Y^{IJ} B_j^J T^{IJ} D_j^J e^{\beta_j^J c_{Ij}^J} \tag{34}$$

with the balancing factors for (33) calculated:

$$A_i^I = \frac{1}{\sum_J X^{IJ} T^{IJ} e^{\beta_i^I c_{iJ}^I}} \tag{35}$$

$$X^{IJ} = \frac{1}{\sum_i A_i^I O_i^I e^{\beta_i^I c_{Ij}^I}} \tag{36}$$

and the balancing factors for (34):

$$Y^{IJ} = \frac{1}{\sum_j B_j^J D_j^J e^{\beta_j^J c_{iJ}^J}} \tag{37}$$

$$B_j^J = \frac{1}{\sum_J Y^{IJ} T^{IJ} e^{\beta_j^J c_{Ij}^J}} \tag{38}$$

In constraining T_{iJ}^I and T_{Ij}^J to T^{IJ} flows, $\sum_i T_{iJ}^I = \sum_j T_{Ij}^J$. This means that when Equations (33) and (34) are used as inputs into (27) and (28) in model (iv), the balancing factors will always converge and the equations can be solved. Model (iv) represents the T_{ij}^{IJ} estimates which will adhere most closely to the known information about the system, and as such might be described as the *optimum* model for the EU system in this study.

3.5 Model (v)

If Model (iv) is the optimum model, then Models (v) and (vi) which produce alternative T_{ij}^{IJ} estimates using less information might be described as being *suboptimal*. Model (v) will only produce T_{ij}^{IJ} estimates where T^{IJ}, $i \neq j$. This model can be written:

$$T_{ij}^{IJ} = A_i^I B_j^J O_i^I D_j^J e^{\beta_i^I c_{ij}^I} \tag{39}$$

where:

$$A_i^I = \frac{1}{\sum_j B_j^J D_j^J e^{\beta_i^I c_{ij}^I}} \tag{40}$$

$$B_j^J = \frac{1}{\sum_i A_i^I O_i^I e^{\beta_i^I c_{ij}}} \qquad (41)$$

In this model, O_i^I and D_j^J can be estimated in exactly the same way as T_{ij}^I and T_{Ij}^J in Equations (31) and (32), so:

$$O_i^I = \left(\frac{M_i^I}{T^{II}}\right) O^I \qquad (42)$$

$$D_j^J = \left(\frac{N_j^J}{T^{II}}\right) D^J \qquad (43)$$

The T_{ij}^{IJ} estimates in model (v) will not adhere as closely to known T^{IJ} values as those in Model (iv), as the constraints are the outer margins on the expanded matrix shown in Figure 5.

3.6 Model (vi)

Finally, Model (vi) models the whole T_{ij}^{IJ} matrix, including T_{ij}^{II} flows. This model (with an origin-specific distance decay parameter) takes the form:

$$T_{ij}^{IJ} = A_i^I B_j^J P_i^I Q_j^J e^{\beta_i^I c_{ij}^I} \qquad (44)$$

where:

$$A_i^I = \frac{1}{\sum_j B_j^J Q_j^J e^{\beta_i^I c_{ij}}} \qquad (45)$$

$$B_j^J = \frac{1}{\sum_i A_i^I P_i^I e^{\beta_i^I c_{ij}^I}} \qquad (46)$$

with the P_i^I and Q_j^J constraints calculated as in Equations (13) and (14).

This new family of doubly constrained multi-level spatial interaction models allows estimates of a full matrix of 287 x 287 flows within the defined European system to be made. Whilst Model (iv) defined in Equations (33) and (34) will produce estimates which are forced to adhere most closely to the known information in the system, other models in the family, which by definition will produce results constrained to less information, will allow us to examine features of the European migration system which do not fit our model assumptions. In doing this we might, for example, be able to identify areas where it would be prudent to adjust the cost proxy in order to distribute migrant flows more effectively within the system without the 'helping hand' that constraints give, or indeed answer the question posed in the introduction to this paper relating to whether it is feasible to treat the European system as, effectively, an internal migration system where national boundaries have little influence on

migration flows. First, however, a number of technical challenges relating to the implementation of the models need to be overcome.

4   Model parameter calibration

All of the models described in the MLSIM family make use of a calibrated distance decay parameter (or parameters), but in making use of such a parameter, a number of problems present themselves. Firstly, calibration can only be carried out using known data within the system – therefore the β parameter(s) will have to be calibrated using either T_{ij}^{II} flows of T^{IJ} flows. This means that, potentially, these parameters may not be completely appropriate for T_{ij}^{IJ} flows. In the absence of other means of estimating appropriate parameters, however, it could be argued this is the best option available at this time, and so it is the option we will have to take.

Accepting that available observed data will be used to calibrate the best-fit parameter(s), the next issue relates to the method used to carry out the calculation. Distance decay parameters in spatial interaction models have historically been calibrated using maximum likelihood techniques employed in computer algorithms – these commonly use iterative procedures to search for the 'best-fit' between the estimates created by the model and the sample data. As an aside, whilst standard iterative procedures are most frequently used in this type of modelling, it should be noted that a significant amount of work has been carried out by Openshaw and colleagues on the calibration of spatial interaction models using genetic algorithms (Diplock and Openshaw 1996; Openshaw 1998); an approach perhaps operationalised most recently by Harland (2008) – we will not explore these methods here, but will use a conventional iterative approach. Batty and Mackie (1972) discuss a range of maximum likelihood calibration methods, but the Newton-Raphson search algorithm has been shown to perform better than most and has been adopted in both the SIMODEL computer program developed by Williams and Fotheringham (1984) and the IMP program developed by Stillwell (1978); both Fortran programs using the search routine to find the parameter estimates which minimise the divergence between the mean value of the total distance travelled in the observed and modelled flow matrices – an approach also used by Pooler (1994). Thanks to its successful implementation in SIMs for migration analysis, the Newton-Raphson algorithm is the one that we choose to use here.

Initially two versions of the doubly constrained model were run to calibrate a best-fit general distance decay β parameter for the whole system. The results of these models are shown in Table 1 and are contrasted with a more basic singly constrained model for comparison. Here a selection of goodness-of-fit (GOF) statistics are displayed – the coefficient of determination (R^2), the square root of the mean squared error (SRMSE), the sum of the squared deviations and the percentage of misallocated flows – although they all display very similar findings. It is clear that the doubly constrained model with the inverse power function applied to the distance matrix produces the best fit to the original

data, with an R^2 of some 87%. This compares to an R^2 of 72% for the negative exponential function and 62% for the reference production constrained model.

Table 1 - Goodness-of-fit statistics for T_{ij}^{II} model experiments

Model equation	β	R^2	SRMSE	Sum Sq Dev	% Misallocated
$T_{ij}^{II} = A_i^I B_j^J M_i^I N_j^J e^{\beta c_{ij}^I}$	-4.2986	0.718	39.393	10,456,839,051	21.554
$T_{ij}^{II} = A_i^I B_j^J M_i^I N_j^J c_{ij}^{\beta}$	-0.9136	0.865	27.992	5,280,098,085	17.008
$T_{ij}^{II} = A_i^I M_i^I c_{ij}^{\beta}$	-1.2201	0.623	45.457	13,886,764,628	28.131

The question that follows is: should this overall distance decay parameter be used as the distance decay input to the estimation model? If this parameter is representative of the whole system, then it could be argued that it could. To test this, a T_{ij}^{II} model with an inverse power distance decay function (akin to that in the second row of Table 1) was run separately for each of the 21 countries in the system comprised of more than a single zone in order to calibrate a series of β^I parameters. The results of these experiments are shown in Table 2.

In this instance we chose the inverse power distance decay function as it was the best performing function in the T_{ij}^{II} experiment. Serendipitously, the power function is scale independent whereas the exponential function is not (Fotheringham and O'Kelly 1989), meaning we are able to directly compare the β^I parameters directly. In Table 2 we use the R^2 value as our measure of goodness-of-fit. We are aware that there has been some debate over which is the most appropriate metric to use (Knudsen and Fotheringham 1986), however R^2 is commonly used and for comparative proposes the choice of statistic has little relevance to the outcome. A number of points can be made about the results displayed in Table 2. Firstly the countries are ranked according to their goodness-of-fit and we can observe that around half of the list have R^2 values over 90%, with Finland, Sweden and Austria ranked the highest – Finland with an exceptionally high R^2. It is clear, however, there is considerable variation in the β^I parameters for each country. This would suggest that it may not be ideal to use the generalised β parameter to model flows for the whole EU system. Furthermore, the reliability of some of the β^I parameters can be called into question with particularly low R^2 values for Spain and France – countries which exhibit positive β^I parameter values. The exact way in which these parameters can be understood has been questioned (Fotheringham 1981), however one interpretation is that the value can be read behaviourally and the number is an index of the deterrent to migration, with high negative values representing distance being a strong deterrent to migration and low negative values inferring that distance is a weak deterrent. Positive values in this context would indicate that distance is an

attraction to interaction – i.e. the further away origins and destinations, the more likely migration is to occur. Clearly this is unlikely to be the case across the whole of Spain and France.

Table 2 – Goodness-of-fit statistics for inter-regional migration data modelled with a doubly constrained model with a power distance decay β parameter

Country Code	Country	R^2	β^I (power function)
FI	Finland	0.996	-0.754
SE	Sweden	0.974	-0.771
AT	Austria	0.972	-0.747
HU	Hungary	0.963	-0.567
SK	Slovakia	0.948	-0.773
NL	Netherlands	0.936	-1.279
DK	Denmark	0.930	-0.969
NO	Norway	0.919	-0.814
BG	Bulgaria	0.901	-0.825
CZ	Czech Republic	0.889	-0.807
UK	United Kingdom	0.884	-0.927
PL	Poland	0.877	-1.068
CH	Switzerland	0.788	-0.867
BE	Belgium	0.772	-1.049
RO	Romania	0.745	-0.763
DE	Germany	0.715	-0.760
IT	Italy	0.699	-0.718
ES	Spain	0.621	0.154
FR	France	0.549	1.093

Given this evidence, generalised distance decay parameters are currently poor candidates for inputs into an estimation model for the whole of Europe. A potential solution, therefore, would be to use distance decay parameters which are specific to each NUTS2 zone - a technique first outlined by Stillwell (1978). This returns us to Model (i) and Equations (18) and (19).

The GOF statistics for Model (i) – taken for all internal migration flows in the system rather than for each separate country) – are shown in Table 3. Evidently these models provide much better fits than the generalised parameter models, with R^2 values around 93%. A geography to these distance decay parameters can be observed, with the frictional effects of distance operating very differently for in- and out-migration flows across the EU system, as is shown in Figures Figure 10 and Figure 11. It should be noted that the nature of the algorithm used to carry out this calibration means that where it is not possible to calculate a zone-specific distance decay parameter (for example in those countries where T_{ij}^{II} data do not exist such as Greece), a generalised distance decay parameter which is

calculated for the whole system prior to zone specific calibration is allocated. Given the results of these experiments it is these origin and destination specific parameters calibrated on internal migration data which will be used as distance decay inputs into our later estimation models.

Table 3 - Goodness-of-fit statistics for Model (i) with β_i^I and β_j^J parameters

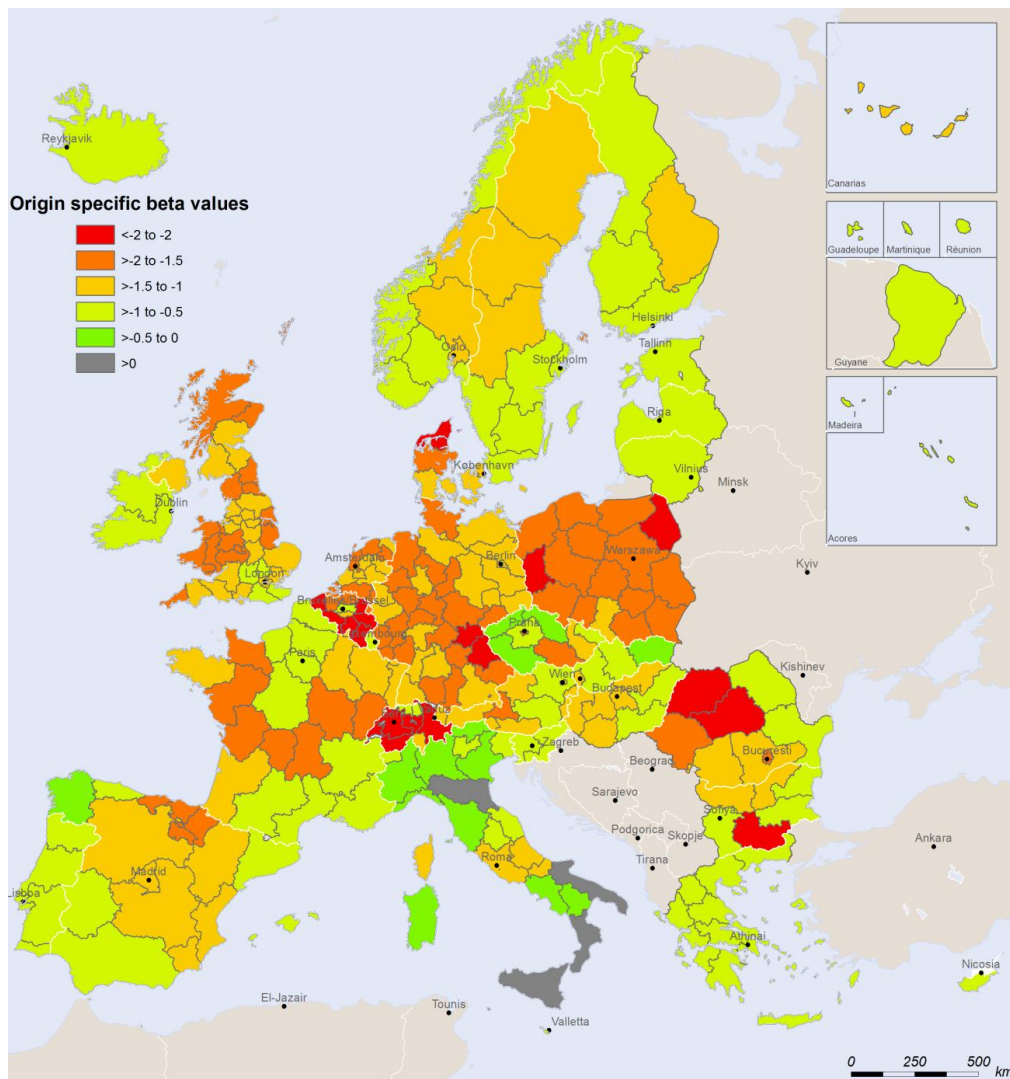| Model equation	R^2	SRMSE	Sum Sq Dev	% Misallocated
$T_{ij}^{II} = A_i^I B_j^J M_i^I N_j^J e^{\beta_i^I c_{ij}^I}$	0.928	19.802	2,642,462,153	12.284
$T_{ij}^{II} = A_i^I B_j^J M_i^I N_j^J e^{\beta_j^J c_{ij}^J}$	0.931	19.582	2,583,959,209	12.163

Figure 10 - β_i^I values calibrated on inter-regional, intra-country migration data, 2006
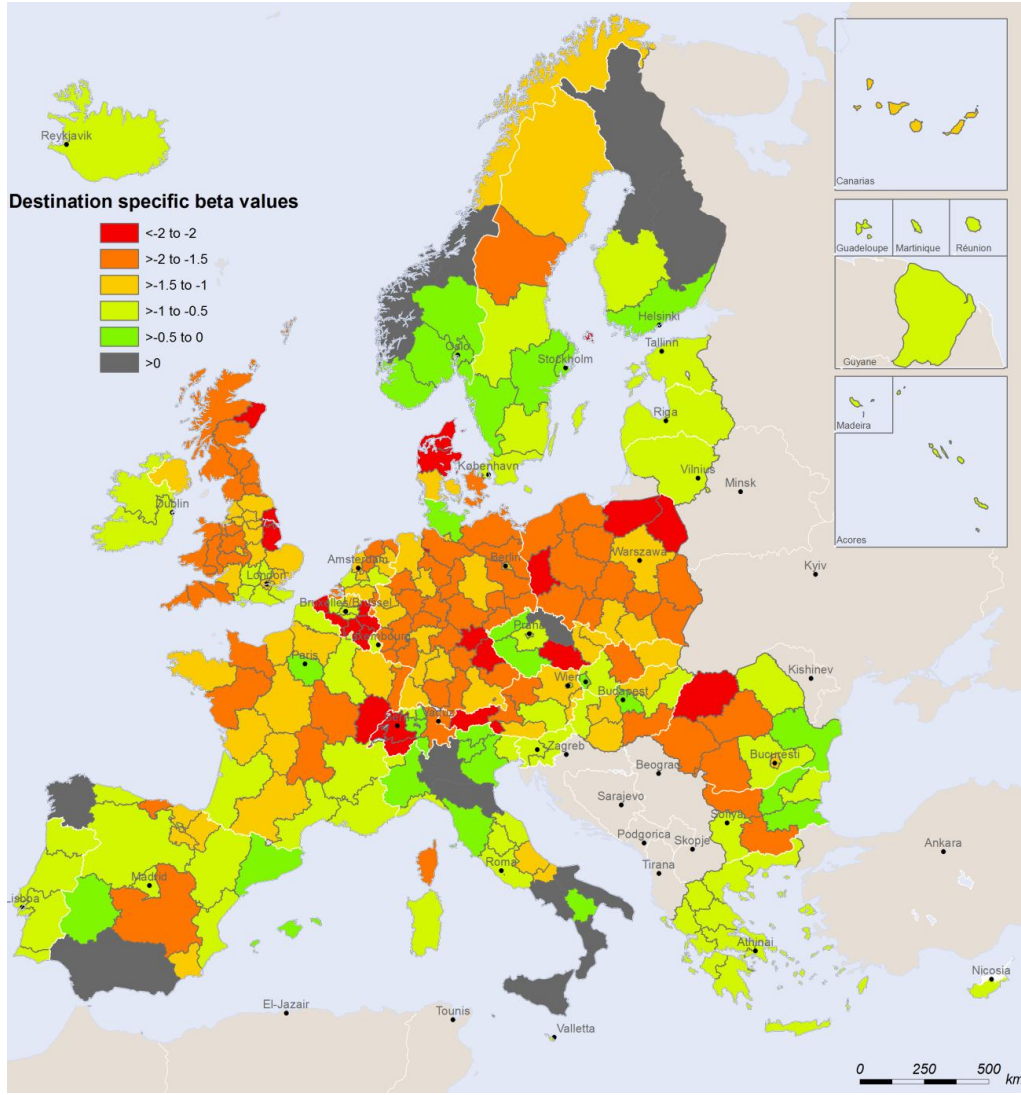
Figure 11 - β_j^J values calibrated on inter-regional, intra-country migration data, 2006

5 Model experiments

5.1 Estimating margin constraints

The sub-section of the MLSIM family of models outlined in Section 3 which used to estimate T_{ij}^{IJ} flows, all require some inputs which are not available directly from the data to hand. In addition to the distance decay parameters that will be calibrated only on internal migration data, Models (iiia), (iiib), (iv) (v) and (vi) make use (directly and indirectly) of O_i^I and D_j^J margins. Consequently sub-models are required to make estimates of these data. Where $\sum_{i \in I} O_i^I = O^I$ and $\sum_{j \in J} D_j^J = D^J$, it follows that it should be feasible to estimate the NUTS2 level O_i^I and D_j^J margins from the country level O^I and D^J margins, given the appropriate ratio values. But which are the appropriate ratios to use?

As information at the internal migration T_{ij}^{II} level is complete, it might be possible to use the distribution of internal migrants to estimate the distribution of international migrants such that:

$$O_i^I = \left(\frac{M_i^I}{O^I}\right) O^I \tag{47}$$

$$D_j^J = \left(\frac{N_j^J}{D^J}\right) D^J \tag{48}$$
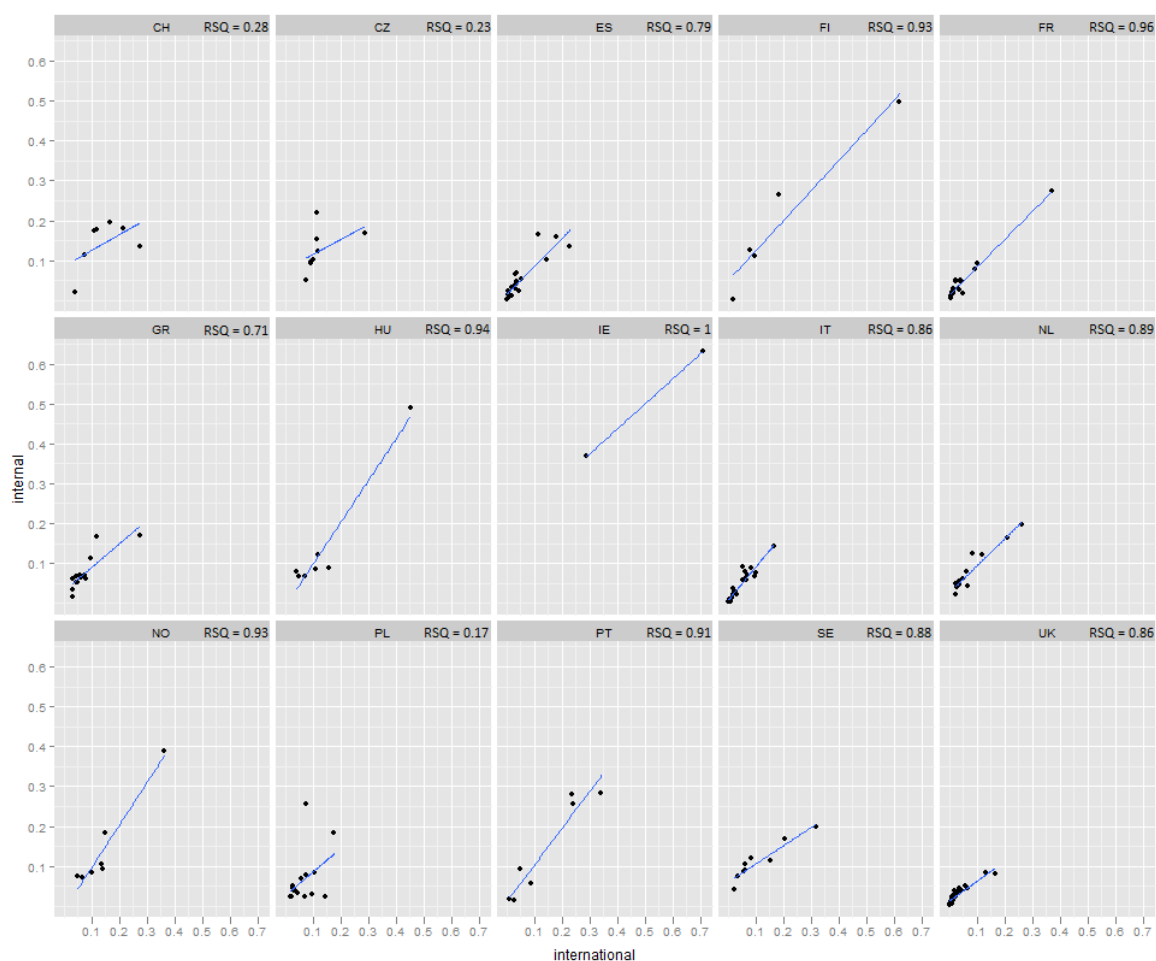
Source: Eurostat, Table cens_ramigr

Figure 12 – Correlation between internal ('Place of residence changed outside the NUTS3 area') and international ('Place of residence changed from outside the declaring country') migrant distributions for NUTS2 regions, selected EU countries, 2001

The assumption here is that the distribution of internal in- and out-migrants within countries is the same as the distribution of immigrants and emigrants moving between countries. But can internal

migrant distributions be used to estimate distributions of international migrants within countries accurately? We might expect, for example, capital cities to dominate these distributions with larger urban areas also providing significant origins and destinations at both levels. Is this the case in reality? Figure 12 shows the comparable distributions of internal and international migration for a selection of European countries at NUTS2 level (all countries where comparable data exists at this level), taken from Census data from the 2000-01 census round and compiled by Eurostat. Broadly speaking there are positive correlations between internal and international migration distributions, although there are some noticeable differences in the correlation coefficients denoted by the R^2 values (and the scatter plots). For most countries in the selection, R^2 values are over 80%, indicating that internal migration distributions are reasonably good predictors of international migration distributions. For some countries, though, this predictive relationship is weak. Poland, for example, has an R^2 value of only 17%, with the Czech Republic (23%) and Switzerland (28%) not faring much better. The reasons for the lack of correlation in these countries are difficult to ascribe, but differences in the perceived attractiveness of particular destinations to internal and international migrants will affect the correlations. Studying Figure 12, the scatter plots show that there is very little pattern in the association between internal in-migration and immigration in Poland, although examining Switzerland and the Czech Republic, it appears that were it not for one or two outliers in the scatter plots, the correlation would be far stronger. Through mapping the differences between internal and international migrant distributions it is possible to interrogate these and other outliers a little further.

Figure 13 maps the distribution of the differences between the regional shares of internal and international (in-)migration across NUTS2 zones in Europe (where data are available). A number of points should be made about this map. Firstly, all yellow zones signify less than a 1% deviation between the distribution of internal and international migrants – these zones include much of the UK, and large parts of France, Italy, the Czech Republic, Poland and Greece. In these areas internal migration distributions can be seen to be good predictors of international migration distributions. Secondly, zones in light orange and light green show only up to a 3% deviation – these include most of the rest of France, a number of regions in Scandinavia, the Netherlands, Poland, the UK, Italy and Greece. Perhaps the most important point of note, however, which becomes very apparent when examining Figure 13, is that there appears to be a 'capital city effect'. The regions containing London, Paris, Madrid, Rome, Amsterdam, Stockholm, Helsinki, Prague, Lisbon and Dublin all exhibit a noticeably higher (average 8.7%) proportion of the national share of international immigrants compared to the national share of internal in-migrants. Some capital cities go against this pattern, although Bern can probably be discounted as in terms of city status, Zurich (which matches this trend) could be argued to be a city of more importance within Switzerland. Oslo, Athens and Budapest have lower proportions of international immigrants than internal in-migrants, but the city region where a very large trend in the opposite direct occurs is Warsaw in Poland. Here the proportion of internal

migrants to Warsaw is over 18% higher than the proportion of international migrants. That Warsaw is an attractive destination for internal migrants would not be surprising, but why it accounts for a much larger proportion of these migrants compared to international migrants is unclear without further investigation of the particular motivations of migrants in Poland.
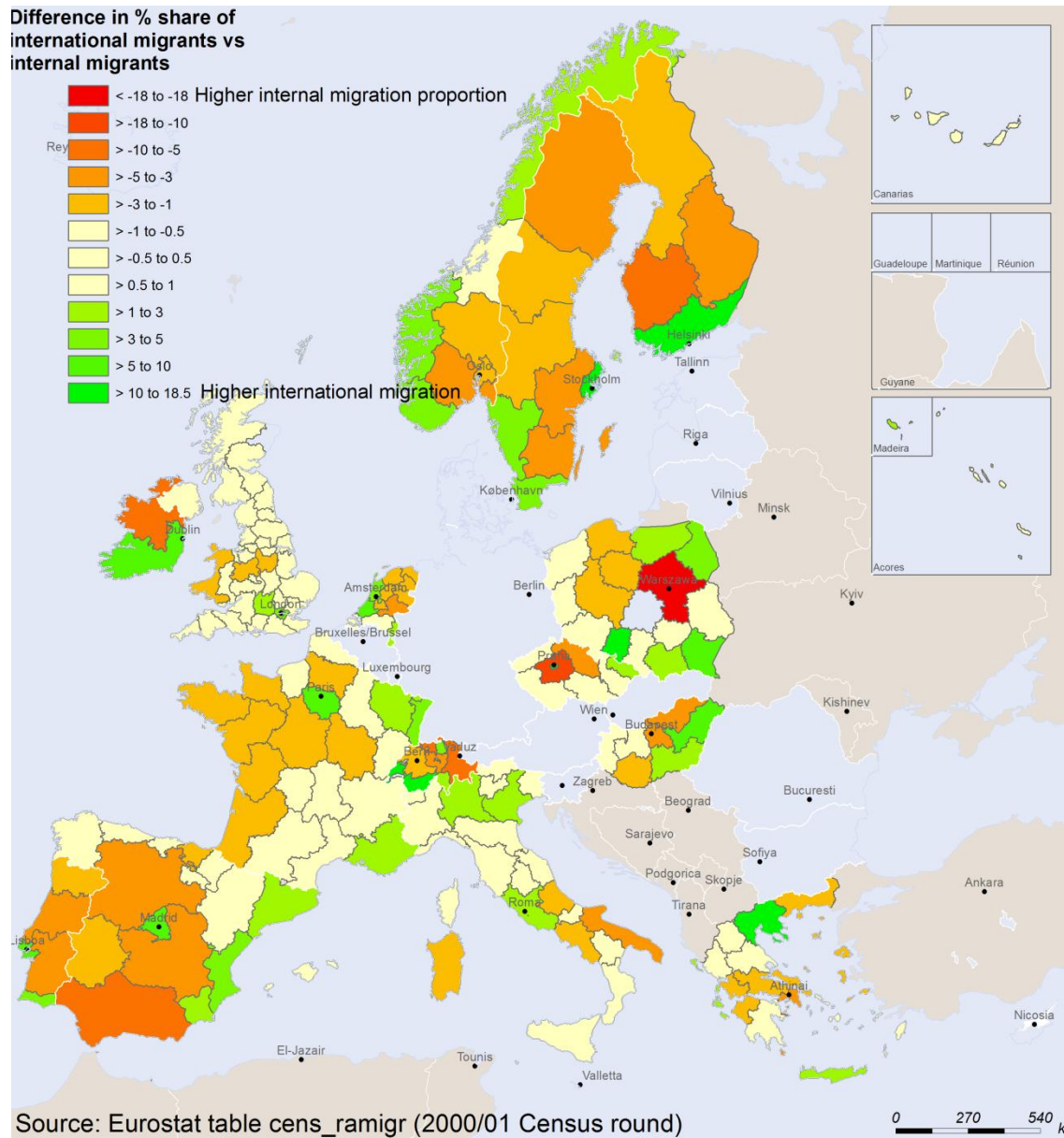
Figure 13 – Distribution of NUTS 2 regions where shares of internal and international in-migrants differ, selected EU countries, 2001

Based on this it could be argued that if this capital city effect could be accounted for consistently, and the proportions of migrants associated with other regions in the country adjusted accordingly, then internal migration distributions could be used to make international migration D_j^J margin estimates relatively reliably, assuming that these associations hold over time.

Incidentally, the time dimension provides us with another option for modelling the sub-national distributions of international migrants. Where decennial census (or other periodic) data can provide sub-national immigrant distributions, if country level immigrant data are available, sub-national distributions can be estimated with the formula:

$$D_j^{Jt+n} = \left(\frac{D_j^{Jt}}{D^{Jt}}\right) D^{Jt+n} \qquad (49)$$

Even if more up-to-date national data are not available, an assumption could be made that these ratios hold over time so that:

$$D_j^{Jt+n} = \left(\frac{D_j^{Jt}}{D^{Jt}}\right) D^{Jt} \qquad (50)$$

Returning to Equations (47) and (48), unfortunately the nature of the data collated by Eurostat means that it is not possible to assess whether emigrant distributions also follow the distributions of internal out-migrants (these data are census/population register data relating to resident populations in recording countries and therefore cannot contain emigrant data). Given the high degree of association between internal migration in- and out-migration distributions (Figure 14) it might be reasonable to use international immigrant distributions to estimate international emigrant distributions, but the capital city effect would need to be explored before this could be done with confidence. Here our concern is to present a general methodology for estimating the full EU matrix of NUTS2 flows and so we will not dwell on this element of the estimation process at this stage, although it should be stressed that the estimation of O_i^I and D_j^J marginal values will have an important bearing on reliability of the final modelled outputs.

As a consequence of the data to hand and the investigations of internal/international migration associations, at this stage internal migration distributions will be used to estimate O_i^I and D_j^J marginal values for the model as in Equations (47) and (48), but we recognise that this is an area of the methodology which could be improved in the future.
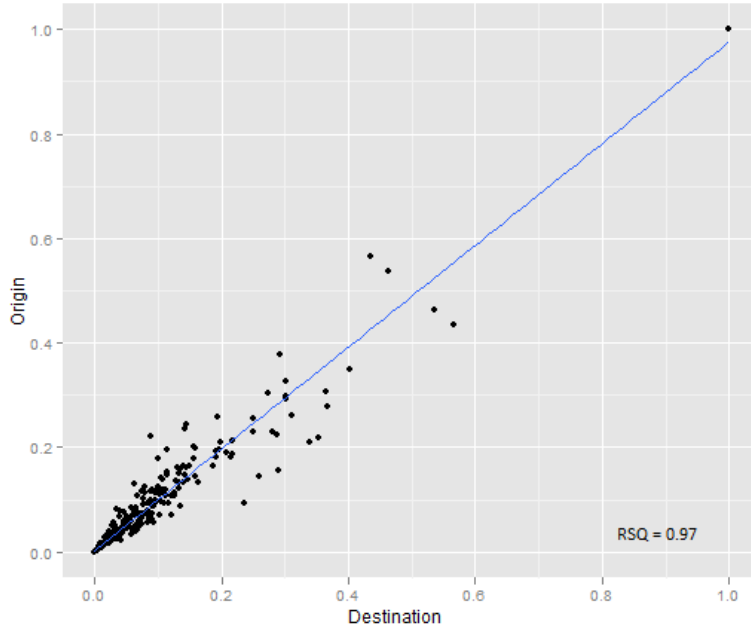
Figure 14 – Correlation between the NUTS2 regional share of internal in- and out-migration flows across EU countries, 2006

5.2 Suboptimal models for exploring the EU system

5.2.1 Examining internal migration flows

Once O_i^I and D_j^J margins have been estimated and an initial calibration routine has been run using Model (i) to calibrate β_i^I and β_j^J values from T_{ij}^{II} data, it is possible to use these data as inputs into Models (iiia), (iiib), (iv) (v) and (vi). Model (iv), takes in T_{iJ}^I and T_{Ij}^J inputs from Model (iiib) and (as previously described) can be viewed as the *optimum* model as any outputs will be constrained to known T^{IJ} flows and estimated O_i^I and D_j^J margins (where T_{iJ}^I and T_{Ij}^J estimates used these constraints). Models (v) and (vi), in contrast, are *suboptimal* as estimates will not be constrained to T^{IJ} flows, only O_i^I and D_j^J, or P_i^I and Q_j^J margins. Running suboptimal models is an important part of the model building process as they allow us to explore the reliability of some of the general model assumptions. All of the models in the multi-level family use distance as the proxy for estimating flows, with the constraints acting to force these estimates to conform to known information about the system. Where suboptimal constraints mean the distance proxy has more influence on the final model results, we are able to examine, through comparing model outputs with other data, where in Europe distance (and the decay assumptions developed from internal migration data) is a less successful proxy for estimating migration flows.

Model (v) can be run using both β_i^I (origin specific) and β_j^J (destination specific) parameters – the former taking the form of Equation (39), the latter being written:

$$T_{ij}^{IJ} = A_i^I B_j^J O_i^I D_j^J e^{\beta_j^J c_{ij}^J} \qquad (51)$$

with balancing factors equivalent to Equations (40) and (41)

Similarly, model (vi) can be run as in Equation (44) with origin and destination specific parameters, with the destination specific version taking the form:

$$T_{ij}^{IJ} = A_i^I B_j^J P_i^I Q_j^J e^{\beta_j^J c_{ij}^J} \qquad (52)$$

With Model (vi) estimating all T_{ij}^{IJ} cells from P_i^I and Q_j^J margins, the estimates from this model differ noticeably from those from Model (v) where T_{ij}^{II} internal migration flows are estimated separately, as in Model (i), before being added to T_{ij}^{IJ} flow estimates which are constrained to O_i^I and D_j^J margins. As may be expected, in Model (vi), where the T_{ij}^{IJ} estimates are the internal migration T_{ij}^{II} flows within countries, the estimates are inaccurate as there are no constraints operating at the intra-country level. To exemplify this, consider Table 4.

Table 4 compares the modelled internal migration flows for Austria (Model vi) with the observed values taken from the Eurostat database. The negative values in this table indicate that the modelled internal migration flows in all cases are significant underestimates (and it is a similar story when destination specific β_j^J parameters are used – table not shown). These Model (vi) residuals can be contrasted with the Model (i/v) residuals for the same cells where internal migration constraints are applied (Table 5). The constraints in this case mean that all residuals sum to 0 across both origins and destinations leading to a much lower overall absolute error across all estimates (16,825 migrants in the T_{ij}^{II} table vs. 75,808 migrants in the T_{ij}^{IJ} table). The same T_{ij}^{IJ} underestimation is present for most origin/destination flows in all EU countries, as demonstrated in Table 6. Here it is shown that in most cases the underestimate of the model can be measured in the thousands.

These underestimates of internal migration flows in Model (vi) mean that far too many flows are being distributed internationally. This indicates that the country border effect is far stronger than the model accounts for, even when origin and destination specific distance decay parameters have been calibrated, suggesting that in reality, despite the theoretically free movement of people in the post-Schengen Europe, country effects cannot be over-stated. An interesting avenue of future research would be to see whether or not these effects are truly 'country effects' or whether other factors such as language play an equally important role. To analyse this, sub-sections of the full interaction matrix could be modelled analysed for countries which share common or similar languages – e.g. France and

Belgium, Belgium and The Netherlands, Austria and Germany, France and Switzerland, Switzerland and Germany, *etcetera*.

Table 4 – Observed vs. modelled T_{ij}^{IJ} residuals – Model (vi)

		Destination									
		AT11	AT12	AT13	AT21	AT22	AT31	AT32	AT33	AT34	O_i
Origin	AT11		-848	-1489	-49	-660	-64	-13	-35	-17	-3176
	AT12	-1457		-12119	-339	-1022	-1659	-303	-264	-150	-17313
	AT13	-1905	-16571		-972	-1472	-1698	-628	-621	-395	-24262
	AT21	-53	-224	-1458		-1413	-249	-219	-443	-110	-4168
	AT22	-700	-815	-2707	-1175		-986	-562	-403	-257	--7605
	AT31	-124	-1378	-3039	-238	-1035		-1858	-769	-266	-8706
	AT32	-14	-135	-1164	-120	-620	-1722		-565	-99	-4438
	AT33	-77	-364	-925	-461	-634	-534	-496		-552	-4043
	AT34	-30	-110	-626	-148	-319	-190	-105	-569		-2098
	D_j	-4358	-20446	-23527	-3502	-7175	-7102	-4183	-3669	-1846	

Sum absolute error = 75808.37

Table 5 – Observed vs. modelled T_{ij}^{IJ} residuals – Model (i/v)

		Destination									
		AT11	AT12	AT13	AT21	AT22	AT31	AT32	AT33	AT34	O_i
Origin	AT11		-265	462	27	-397	89	48	26	9	0
	AT12	-681		676	60	146	-501	100	159	40	0
	AT13	499	-1614	0	-204	800	319	116	143	-58	0
	AT21	106	294	-441		-407	285	301	-126	-12	0
	AT22	-235	731	217	-505		-23	-77	-11	-98	0
	AT31	175	400	187	288	247		-588	-237	-99	0
	AT32	78	258	-416	467	8	-267		-120	-8	0
	AT33	31	79	-118	-105	-240	70	59		224	0
	AT34	28	117	-194	-29	-156	28	40	165		0
	D_j	0	0	0	0	0	0	0	0	0	

Sum absolute error = 16825.34

It could be that two different distance decay parameters are required for internal and international migration flows. If this is the case, then despite the extra constraints forcing the results in Model (i/v) to conform to the known data more closely, the international estimates may still suffer from some inaccuracy. The problem faced, despite acknowledging this issue, is that with a dearth of inter-regional international migration data to call upon, it is very difficult to validate these model estimates effectively.

Table 6 - Average total error for internal migration Model (vi) T^{IJ}_{ij} estimated flows for EU countries, 2006

Country	Average error β^I_i	β^J_j	Country	Average error β^I_i	β^J_j
AT	-1053	-1009	HU	-3302	-3268
BE	-922	-937	IT	-671	-670
BG	-1375	-1445	NL	-1408	-1456
CH	-1055	-1211	NO	-1783	-1840
CZ	-1130	-1199	PL	-88	-128
DE	-393	-382	RO	-673	-852
DK	-4134	-3692	SE	-2555	-2606
ES	-871	-938	SI	-2022	-2021
FI	-3526	-3599	SK	-1179	-1087
FR	-1652	-1632	UK	-410	-402

5.2.2   Examining inter-regional, inter-country flows

Suboptimal models also provide a useful contrast with the inter-regional inter-country estimates from the optimal constrained MLSIMs derived from Model (iv) (where known T^{IJ} flows constrain the inter-regional estimates). For example, consider the two maps which show the outputs from the suboptimum Model (v) and optimum Model (iv) in Figure 15 and Figure 16. Both maps use the same scale for display, showing all large (750 migrants and above) origin/destination flows between NUTS2 regions located in different countries in the system. Immediately noticeable is the very large difference in the volume of flows displayed in each map. The suboptimal model has far fewer large flows, while the optimal model has many more. What this indicates is that the suboptimum model distributes the migrants far more evenly around the system than the optimum model, which concentrates regional flows in accordance with country level flows estimated by the MIMOSA project. Where the disparities between the two models are most obvious can be seen when regional flows are aggregated to the country level – this is shown in Table 8.

Table 7 shows the differences between the aggregated estimates from the MLSIM suboptimal β^I_i Model (v) and the MIMOSA model. Where entries in the table are shaded grey, the estimates depart by greater than or less than 2000 migrants. When examining these residuals it is worth bearing in mind that in a closed system such as this, where large over or under estimates occur in a row or column, this will affect other estimates in the same row or column. For example, this can be seen in the destination column for Germany (DE) in Table 7. Here the MLSIM very much under-estimates the number of migrants flowing from Poland into Germany when compared to the MIMOSA estimate

– some 47,842 migrants fewer. This large under-estimate has the effect of influencing over-estimates elsewhere in the column.

From Table 7 a handful of countries appear for which the two models disagree. Flows out of Germany, Poland and Romania exhibit a relatively high level of disagreement, as do flows into Germany and Spain. Individual flows which draw particular attention are those from Poland into Germany and vice versa (MLSIM under-estimates compared to MIMOSA on both counts), Romania into Spain and into Italy, as well as Bulgaria into Spain, UK into Spain and Germany into Austria (MLSIM under-estimates). Romania into Bulgaria, France into Spain, Romania into UK and Germany into UK show noticeable over-estimates for MLSIM.
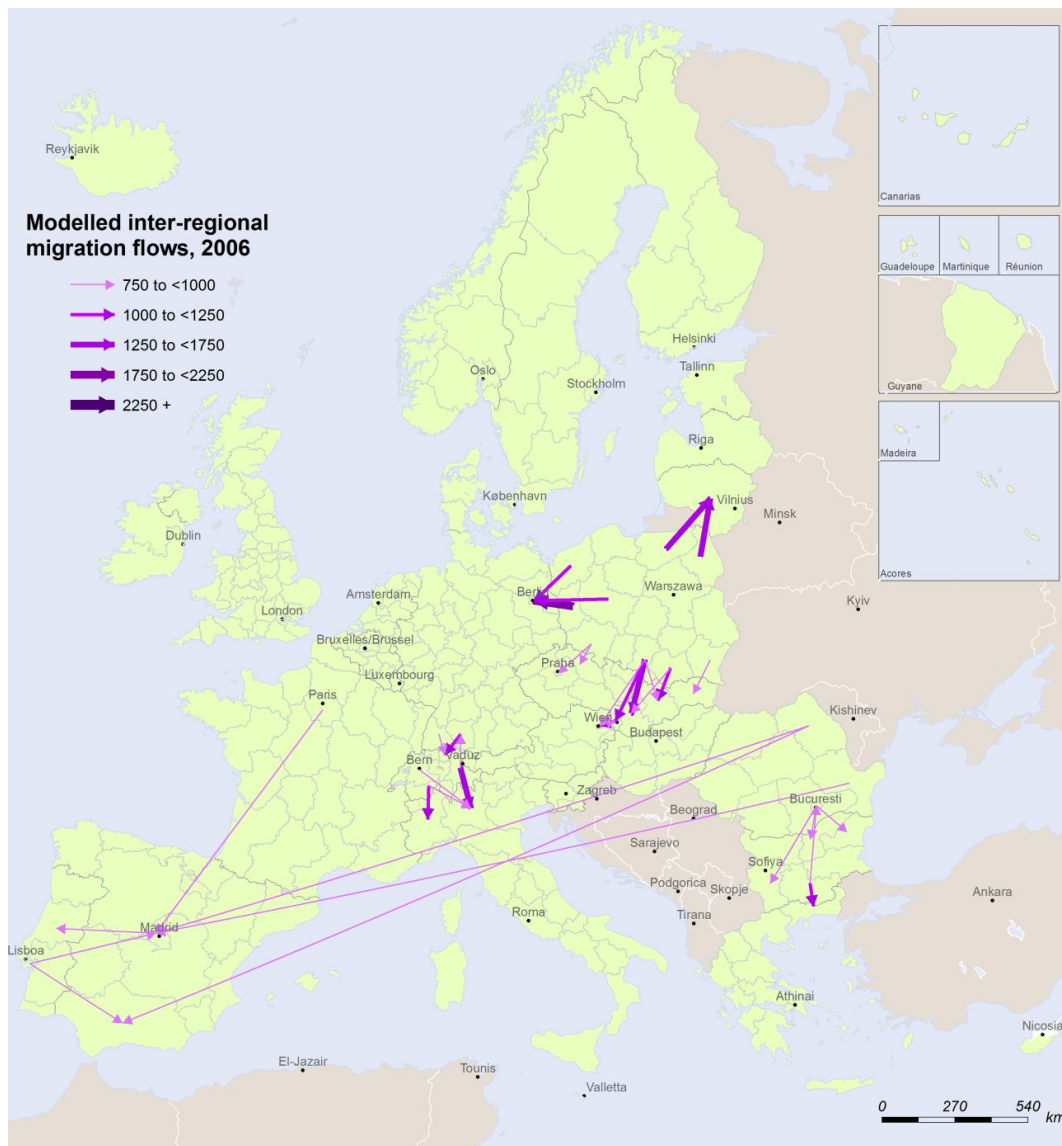
Figure 15 - Largest flows from suboptimal Model (v) with β_i^I parameters
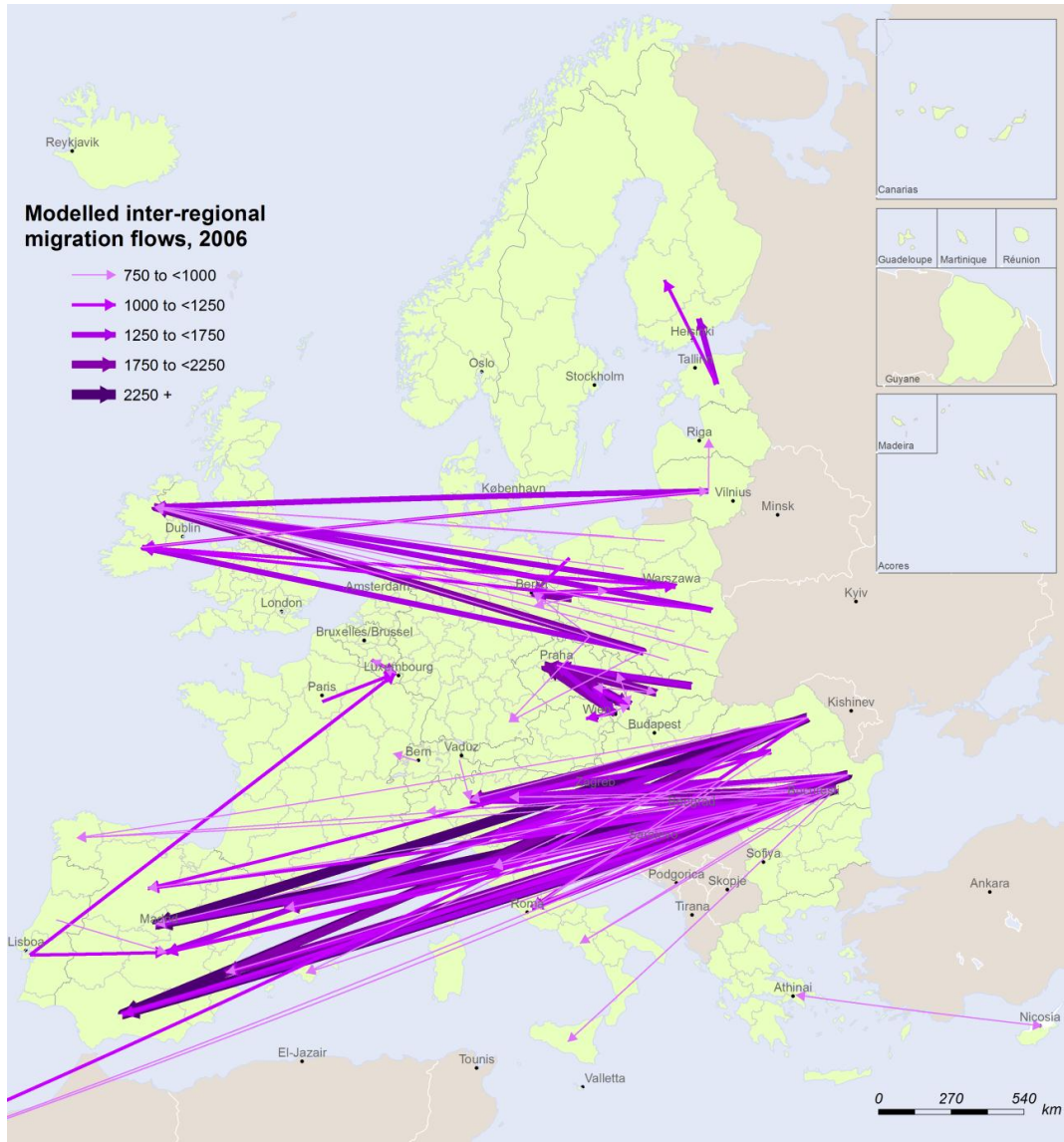
Figure 16 - Largest flows from optimum Model (iv) with β_i^I parameters

In cases where we have these large disagreements, it is where large flow volumes occur between countries not in close geographical proximity. For example, as shown in Figure 16 and Table 7, very large flow volumes occur between regions in Romania and Spain and Poland and Ireland in the MIMOSA model – where these countries are not close, the suboptimal MLSIM distributes the flows far closer to the origins. Similarly, relatively low flow volumes are shown between France and Spain in the MIMOSA data – something which runs counter to the short distances (and associated distance decay assumptions) and the large total in and out flows from these countries used to predict flows in the suboptimal MLSIM.

32

Table 7 - Comparison of absolute flow estimates from Model (v) β_i^I and MIMOSA models

Orig/Dest	AT	BE	BG	CH	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HU	IE	IS	IT	LI	LT	LU	LV	MT	NL	NO	PL	PT	RO	SE	SI	SK	UK
AT	0	189	-1167	-36	91	717	-2298	33	-12	2790	77	1013	-82	-2531	366	29	3823	-121	45	101	14	23	-101	174	590	-90	-3963	183	-561	-371	1069
BE	20	0	-32	660	16	75	6916	-149	-1	-225	-129	-4543	-91	-27	381	18	-916	-2	-68	-1192	-56	-4	-2534	53	-450	-19	34	-116	-88	-126	2599
BG	-612	168	0	807	40	-1226	-1682	185	38	-10539	257	706	2929	448	590	71	-11	0	64	97	62	45	-56	278	2264	299	5041	475	85	-878	57
CH	191	723	-3	0	84	283	5201	-438	11	-127	-212	-2175	52	89	453	44	-1218	365	-2	320	-13	11	-608	50	689	-65	350	-307	-332	-864	-2563
CY	170	78	-1266	232	0	72	1055	68	2	1842	92	438	-2407	-79	222	40	944	0	20	26	18	-18	69	118	534	155	-683	129	-4	113	-1980
CZ	1360	191	-4836	862	92	0	3466	165	30	3921	228	412	156	365	624	81	2339	1	55	153	32	18	6	305	4672	276	-967	513	79	-13249	-1356
DE	-10015	13325	-4160	8249	247	9374	0	3355	-102	9693	321	11655	-9217	-8921	4086	178	7057	-54	-612	3849	-308	48	11509	1057	-52765	-3372	-10615	1784	-642	-1598	16603
DK	81	125	-5	26	17	156	4416	0	-51	478	-225	21	-17	-53	280	-1136	464	-4	-82	-7	-108	-13	178	-1441	1202	7	133	-4727	5	-114	395
EE	77	71	41	161	26	61	547	-55	0	965	-3003	283	107	34	175	11	388	0	9	8	-210	2	22	-39	616	84	198	-157	10	77	-512
ES	1049	-3519	-2430	-6658	602	861	4778	-423	81	0	73	-1882	1358	268	1369	381	9299	-95	-555	351	89	144	-1610	578	3536	689	-9802	410	85	723	267
FI	-160	-39	64	-77	61	36	7	-182	-687	1034	0	266	128	10	113	23	472	0	5	21	-71	-3	-170	-167	779	131	310	-2551	10	57	579
FR	340	-5193	-278	212	223	161	2813	-543	-18	12244	60	0	-166	-208	1427	189	3044	-10	-12	-4253	-26	-73	-1283	362	-901	-3282	-190	52	-43	-696	-3930
GR	41	98	-31	495	-2627	129	-4318	0	20	3688	132	457	1378	121	515	78	1840	-1	38	49	35	18	-928	194	164	295	435	-218	38	134	-2266
HU	-2119	13	216	750	60	677	-9721	57	25	3168	169	-338	526	0	400	61	3068	0	105	101	53	27	-232	230	4069	170	-213	77	138	-767	-773
IE	130	-557	-28	-141	53	67	1519	14	-4	2757	51	-822	68	30	0	119	933	-3	-1604	6	-306	-8	-87	269	-1257	210	81	126	20	53	-1684
IS	11	39	17	69	21	22	294	-1165	-6	542	-33	173	47	17	174	0	193	0	-14	12	-7	0	-7	-178	11	32	71	-383	-1	4	47
IT	-131	-1914	357	-14077	642	354	-4029	-9	118	9909	637	-1331	1099	301	2279	390	2	-21	-34	-8	22	-67	-782	982	100	1185	627	977	-295	-502	3234
LI	-41	1	0	-5	1	2	-3	-5	0	20	1	-16	1	0	3	1	13	0	0	1	0	0	2	1	-4	1	2	4	0	-5	25
LT	108	127	26	429	103	139	-1063	-752	14	839	252	632	266	153	-2273	-99	898	0	0	64	-685	11	-14	-734	1635	111	624	-211	29	253	-879
LU	80	-1924	35	163	38	97	338	-20	-1	1377	10	-1703	86	32	238	7	453	0	-4	0	5	6	100	99	411	-860	159	80	-6	92	613
LV	81	47	57	45	-18	72	-401	-265	-494	1006	94	-295	114	36	-769	-17	373	-11	-276	7	0	-2	0	-64	778	79	255	-23	13	65	-486
MT	13	-15	6	33	0	10	85	-13	-2	340	9	-176	40	4	3	5	75	0	4	5	-3	0	-58	9	75	19	42	1	-7	-10	-494
NL	-441	-7083	-41	-351	12	-108	9747	-34	-20	-626	-137	-659	-481	-257	394	17	712	-1	-42	412	-50	-17	0	-300	-840	-789	109	-29	-302	—	1893
NO	152	126	43	230	74	113	1606	-1913	-22	913	-428	555	147	70	573	-259	828	0	-184	54	-35	-6	-193	0	383	147	279	-3498	15	-183	412
PL	8550	-2264	1653	5037	192	12671	-47842	2679	804	9215	2743	-721	184	7074	-21310	-937	6739	-20	5249	924	1554	113	-3781	-2347	0	880	13429	4126	727	18620	-24060
PT	20	140	87	484	146	167	-1293	24	23	1866	152	-3273	361	124	977	57	1584	-2	2	-2726	0	24	-1381	228	1094	1703	439	243	19	157	-1433
RO	1827	1759	11522	5509	1617	2235	9052	1351	309	-57424	1722	5063	7738	5352	3264	428	-51509	1	794	558	457	236	1665	1836	25370	1550	0	3297	602	-28	13834
SE	-191	58	123	-108	78	203	2601	-1846	-71	2196	-3256	373	-337	-9	616	-270	1361	0	-51	47	-66	-46	-252	-2902	897	174	605	0	-10	88	-1
SI	-626	-132	18	-336	16	21	-310	-9	5	465	23	121	60	28	45	6	42	-19	3	14	3	-2	-68	29	230	38	118	2	0	-158	372
SK	-388	400	295	-3395	60	-27602	874	444	96	7770	574	2092	873	-110	810	161	4814	-9	196	212	111	68	189	470	6658	608	2105	1077	128	0	418
UK	425	4961	-283	729	-1966	159	17644	-554	-84	-10099	-256	-6329	-4919	-2361	3976	323	1895	2	-3050	795	-513	-536	405	850	-538	-367	987	-687	14	-586	0

Positive values = MLSIM model estimate higher

Negative values = MIMOSA model estimate higher

6 Results

A file containing the full suite of Model (v) and Model (vi), T_{ij}^{IJ} and T_{ij}^{II} estimates is publicly available for anyone wishing to make use of the data through the following link:

http://dl.dropbox.com/u/8649795/Multilevel_SIM_Results.xlsx

6.1 Major pattern exemplification

MLSIMs offer the opportunity to examine inter-regional flows between all countries in our chosen EU system – examining all flows or even all significant flows would be an extensive task. Therefore we will take the UK as exemplification. Figure 17 depicts all flows entering UK regions from other EU regions modelled using the optimum, T^{IJ} constrained Model (iv) from the family of MLSIMs.
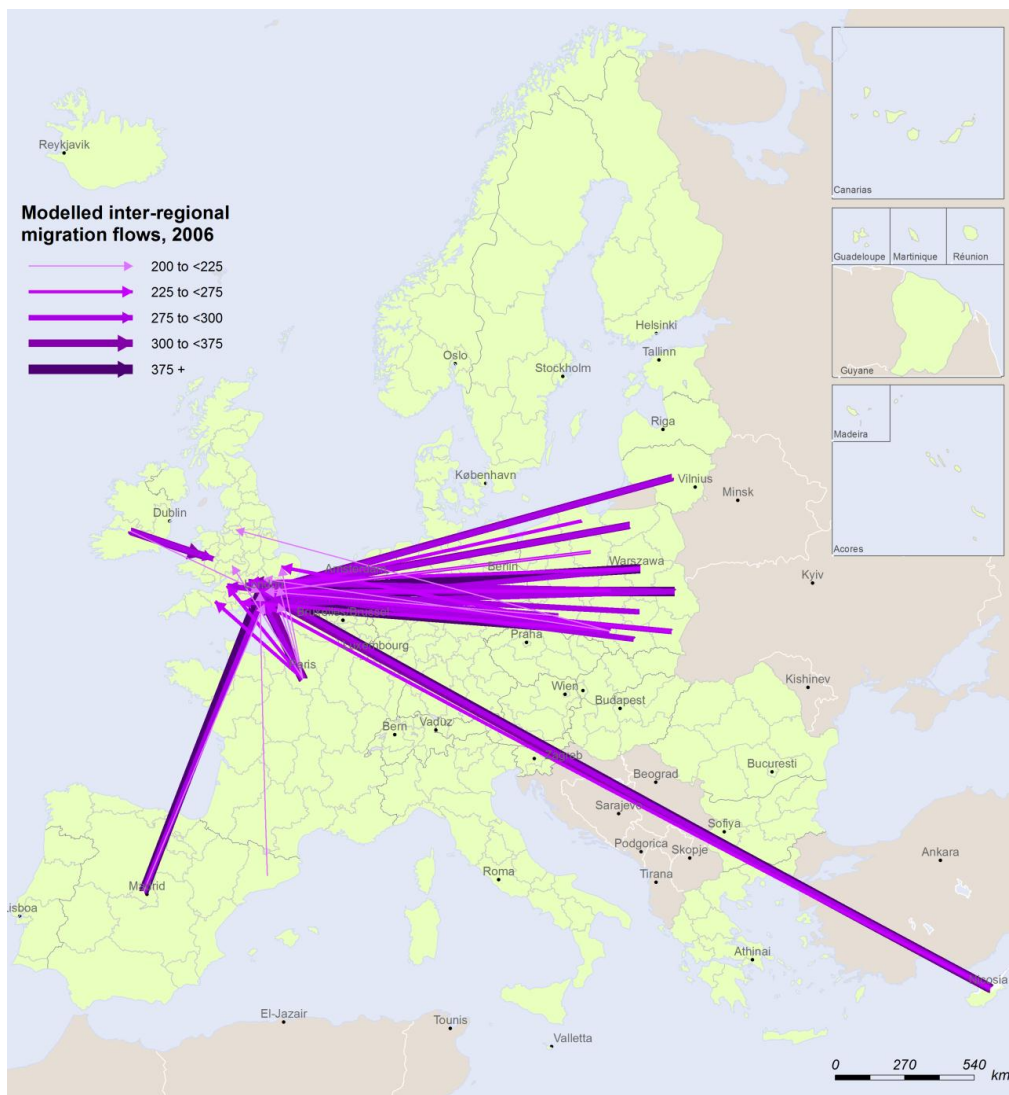
Figure 17 - Flows greater than 200 migrants entering UK regions from other EU system regions, 2006
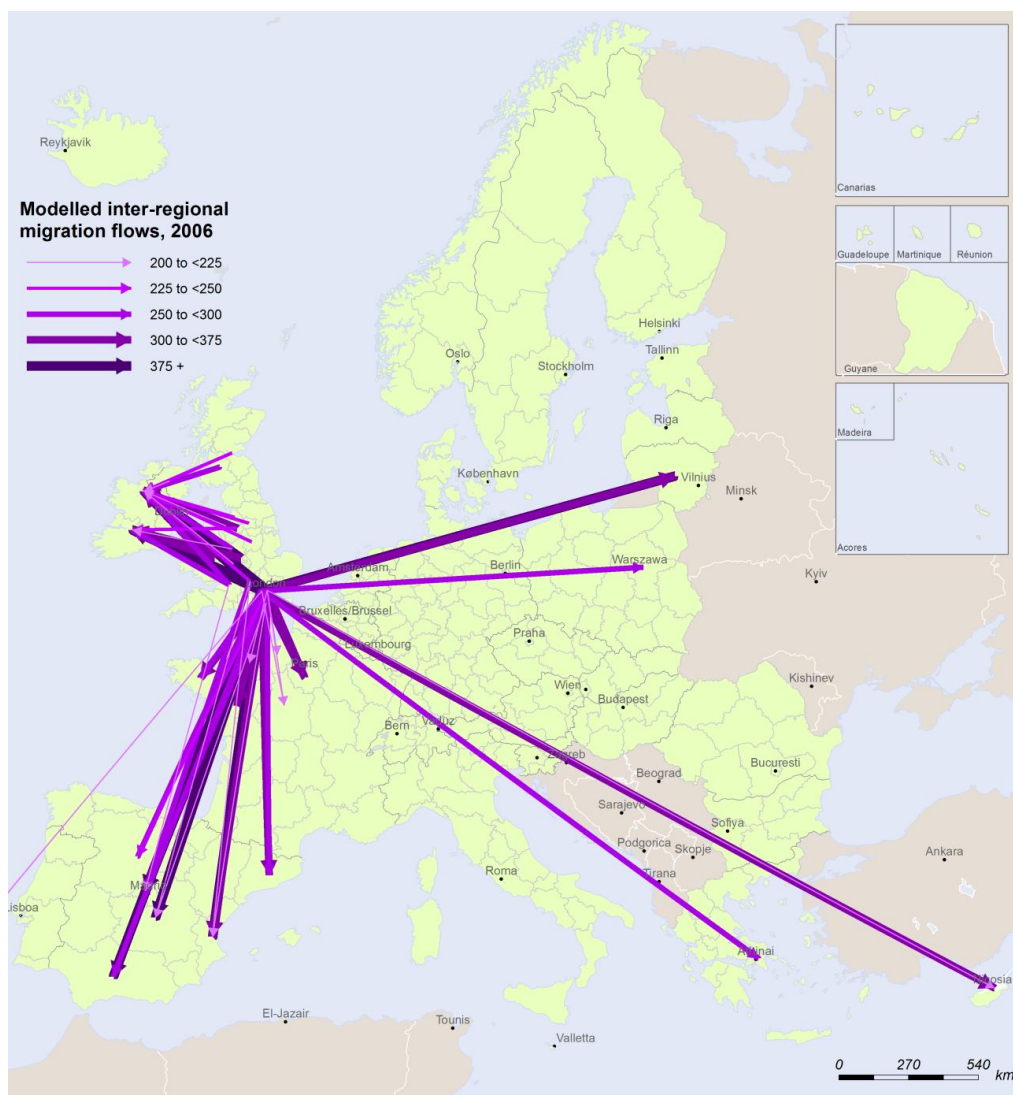
Figure 18 - Flows greater than 200 migrants leaving UK regions for other EU system regions, 2006

A number of important patterns emerge. Firstly, the importance of London and the South East corner of the UK is very apparent – nearly all flows are concentrated in this area, with only a small number entering regions containing other large cities such as Manchester and Birmingham. A large number of flows come from regions in Poland, many of which terminate in London. Interestingly, flows from Poland into East Anglia, which have gained so much media attention in the UK are picked up by the model, even at a region-to-region scale. One small caveat in relation to these flows can be made referring back to our observations about the relationship between internal and international migration distributions in Poland made in Section 5.1 – where the relationship between these flow distributions is poor in Poland, some of the precise flow volumes originating from these Polish NUTS2 regions should be treated with caution. Where these relationships are stronger in France and Spain, the large flows from other major capital cities such as Paris and Madrid can be viewed more reliably, indeed given the 'capital city effect' also noticed, these flows may even be larger in reality. High volume

35

flows are also noticeable from Cyprus, although these may well be associated with armed forces movement.

Examining the flows out of UK regions to the rest of the EU system, the South East – and especially London – predominates as with immigration. Destinations for migrants leaving the UK are quite different to the origins for those arriving in 2006. The large volumes of migration (we may assume related to retirement) can be observed flowing into Spanish regions – regions including the largest cities of Madrid, Barcelona and Valencia, as well as the Costa del Sol. Large flows can also be observed from London and other regions of the UK into Ireland – this is partially a function of Ireland consisting of only two regions and so these flows appear more concentrated, although the close ties between all countries of the UK and Ireland mean that these flows are entirely expected.

6.2 Evaluation

Evaluating the success of models such as those in the MLSIM family can be difficult task, where by definition there are no comparable data with which to validate the results. As shown in Section 5.2.2, it is possible to aggregate the NUTS2 regional level predictions from the suboptimum MLSIM Model (v) into country level predictions and compare these with the MIMOSA model predictions. This comparison can also be carried out for the recorded data for some counties collated by Eurostat. Comparing the results from two models with recorded data is an exercise which should be approached with caution. As mentioned at the beginning of this paper, the researchers involved in the MIMOSA project recognise the limitations of the estimates produced. Similarly, whilst the data from Eurostat are recorded rather than modelled and thus might be viewed as more reliable, this is not necessarily the case. Harmonised, consistent and accurate international migration data across European countries are not a reality (Poulain et al. 2006) – hence the inception of projects such as MIMOSA and IMEM.

Migration flow data for 2006 were obtained from the Eurostat database from Table migr_imm5prv (immigration by sex, age group and country of previous residence). Whilst data are available in this table for all 31 country origins, only 20 destinations are available; and of these, not all have the full set of origin/destination flows. Consequently, it is not possible to evaluate all flows, although it is possible to compare 583 origin/destination pairs.

Table 8 compares the inter-country flows from the Eurostat data with modelled data for 2006 from the MIMOSA model and from the suboptimum Model (v) β_i^I and β_j^J models. A suite of measures are used to assess the relative model performances. Taking the R^2 first, the MIMOSA model appears to perform better than the results aggregated from the regional MLSIMs, with a value of around 57% compared to 52% and 55%. This better performance would appear to hold when examining the % of misallocated flows, with MIMOSA misallocating around 10% fewer flows than the aggregated MLSIMs. However, when examining the figures for the SRMSE and the sum of the squared

deviations, it appears that the suboptimal MLSIMs fit the Eurostat data better, with lower deviations than the MIMOSA model.

Given this evidence, whilst it is difficult to say that one method produces preferable results to the other, it is possible to say that the fits to the Eurostat data are not wildly different from one another and in some sense validating the results of the suboptimal MLSIMs. As we have shown already though, there are some country pairs with high levels of interaction which are unexpected with the assumption that propensity to migrate drops off with distance.

The question is, where the two models differ, which estimate is most likely to be closer to reality? Comparing the model results with the corresponding cells in the Eurostat data, we can observe that the under-estimate of flows from Poland into Germany in the suboptimal Model (v) holds and the MIMOSA estimate is likely to be more accurate. However with only some 3,227 migrants recorded flowing from Germany into Poland by Eurostat, it might be that the lower MLSIM estimate here is more accurate. In reality, however, issues with the way in which Poland records permanent migrants (Poulain et al. 2006) means that this figure of 3,227 is very likely to be a large underestimate, and so this comparison may be misleading.

Table 8 - Comparison between flows recorded in Eurostat Table migr_imm5prv and modelled migration flows at country level

Model	R^2	SRMSE	Sum Sq Dev	% Misallocated
MIMOSA	0.571	104.271	21,176,403,803	21.863
MLSIM suboptimal model (v) β_i^I	0.518	94.637	13,303,422,410	33.148
MLSIM suboptimal model (v) β_j^J	0.550	92.409	14,067,006,845	32.505

Indeed, there are other examples where the MIMOSA model appears to perform better for some country interactions. This enhanced performance is due to the incorporation of real data where possible an offset variable in the statistical model used – what this does is enhance the performance of the model where flow estimates based on spatial structure and distance alone are not adequate. For example, the distance between Romania and Spain means that we would expect less interaction to occur between these countries, when in reality large flows occur for a variety of complex economic and socio-cultural reasons (Bleahu 2004). These influences affect the perception of cost which means that an effective estimate of this influence is not captured by distance alone. Where this is the case, and for instances where there is an absence of supplementary recorded origin/destination data to include in the model, then it will certainly be worth exploring what the real cost of migration is. This

could be achieved through solving the doubly constrained spatial interaction model equation for c_{ij} rather than M_{ij} using methods already outlined by Plane (1984).

7 Conclusions and comments on the new framework for estimating inter-regional, inter-country migration flows in Europe

In this paper we have demonstrated the utility of a new framework for estimating inter-regional migration flows in Europe. Our guiding principal was a simple one – to make use of the maximum amount of available data (embodied in the constraints imposed within the model and the parameters used to influence the patterns) to produce maximum likelihood estimates given the information available.

The Model (iv) estimates represent the 'best-guess' estimate data at this time. They embody all known information about flows into and out of countries, the behaviour of internal migrants within their home countries and the relationships between the destination preference of internal and international migrants. There are, of course, a number of areas where these estimates could be improved. Firstly, the country level international migration data constraints are themselves estimates. The data used were taken from the MIMOSA project (Raymer and Abel 2008) – data which the authors recognise the limitations of, and which will soon be superseded by improved estimates from the IMEM project mentioned in the introduction. Where these model inputs can be improved, then there will be a knock-on improvement to our own estimates. We have already acknowledged that there are issues with the methodology we employed to estimate the O_i^I and D_j^J matrix margins which formed constraints either directly or indirectly for all models. As outlined, in these estimates we have simply taken the national distributions of internal migrants to distribute international migrants. Whilst there are high correlations between these distributions for in-migration, demonstrated across Europe from Census and register data, a 'capital city effect' persists where these destinations can attract up to 10% more migrants internationally than internally. Furthermore, we have been unable to ascertain whether a similar situation exists for out-migration flows. Finally, in using distance decay parameters calibrated with internal migration data, we could be introducing error where internal migration flows, even in an open border Europe, act very differently to international flows. Our experimentation with suboptimal Model (vi) which models internal migration flows using P_i^I and Q_j^J constraints has suggested that this might be the case, with country border effects far stronger than the doubly constrained model estimates. Model (vi) does not incorporate country-level constraints in the same way as Model (v) and results in many fewer migrants than observed being distributed within-countries. Investigating these border effects in more detail would be a fruitful avenue of future research.

Comparison of the results from the Model (v) β_i^I and β_j^J (aggregated to country level) with MIMOSA and Eurostat data has shown that the models produce results which, for the most part, are comparable to both other data sets. The MLSIMs distribute migrants according to physical distance between regions in countries and the calibrated distance decay parameters associated with these distances. The disagreement in the model predictions for flows between some countries, are the consequence of a number of real flows that cannot be characterised effectively using physical distance as the distribution proxy. Whilst an issue in this piece of work, the problem offers possibilities for future research focusing on the development of new cost measures for an international migration system. Given consistent migration data at country level, rearranging the doubly constrained spatial interaction model formula will yield estimates of migrant inferred distance.

Whilst much of the discussion in this paper has focused on the suboptimal models as these allow us to explore model failings in more detail, we should finish by extolling the virtues of our optimum model, Model (vi). Model (vi) constrains inter-regional estimates to know (but also estimated) inter-country flows allowing us to explore the likely inter-regional international flows within Europe. This is an important development as for the first time we are able to examine, at a much higher resolution than previously possible, pressure points within the migration system. We can see, for example, the regions of Spain which are affected most heavily by the large influx of migrants from Romania, and the areas of Romania which are equally as affected (if not more) socially, demographically and economically by these large flows of people. Whilst even in this optimum model there are improvements that can be made, now the modelling framework is in place, when improved inputs can be supplied to the model, then improved outputs can be very easily achieved. Our also need not be limited to Europe – the framework is in place for Europe, but can easy be applied to estimate sub-national flows in a global context, opening up exciting possibilities for a more complete global understanding of migration which has not been possible before.

References

Abel, G.J. (2010), 'Estimation of international migration flow tables in Europe', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Batty, M. and Mackie, S. (1972), 'The calibration of gravity, entropy, and related models of spatial interaction', *Environment and Planning,* 4 (2), 205-33.

Bleahu, A. (2004), 'Romanian migration to Spain - motivation, networks and strategies', Institute for Quality of Life, Bucharest. (http://pdc.ceu.hu/archive/00003393/01/romanian_migration_to_spain.pdf).

Boyle, P.J., Flowerdew, R., and Shen, J. (1998), 'Modelling inter-ward migration in Hereford and Worcester: The importance of housing growth and tenure', *Regional Studies,* 32 (2), 113 - 32.

Convey, A. and Kupiszewski, M. (1995), 'Keeping up with Schengen: Migration and Policy in the European Union', *International Migration Review,* 29 (4), 939-63.

De Beer, J., Van der Gaag, N., Van der Erf, R., Bauer, R., Fassmann, H., Kupiszewska, D., Kupiszewski, M., Rees, P., Boden, P., Dennett, A., Stillwell, J., De Jong, A., Ter Veer, M., Roto, J., Van Well, L., Heins, F., Bonifazi, C., and Gesano, G. (2010), 'DEMIFER - Demographic and Migratory Flows affecting European Regions and Cities', *Applied Research Project 2013/1/3. Final Report*, ESPON and NIDI. (http://www.espon.eu/main/Menu_Projects/Menu_AppliedResearch/demifer.html).

Diplock, G. and Openshaw, S. (1996), 'Using Simple Genetic Algorithms to Calibrate Spatial Interaction Models', *Geographical Analysis,* 28 (3), 262-79.

Flowerdew, R. (2010), 'Modelling migration with poisson regression', in J. Stillwell, O. Duke-Williams, and A. Dennett (eds.), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*: IGI Global.

Fotheringham, A. S. (1981), 'Spatial Structure and Distance-Decay Parameters', *Annals of the Association of American Geographers,* 71 (3), 425-36.

Fotheringham, A. S., Rees, P., Champion, T., Kalogirou, S., and Tremayne, A. R. (2004), 'The development of a migration model for England and Wales: overview and modelling out-migration', *Environment and Planning A,* 36 (9), 1633-72.

Fotheringham, A.S. and O'Kelly, M.E. (1989), *Spatial interaction models: formulations and applications*: Kluwer Academic Publishers.

Harland, K. (2008), 'Journey to learn: geographical mobility and education provision', (University of Leeds).

He, J. and Pooler, J. (2003), 'Modeling China's Province-to-Province Migration Flows Using Spatial Interaction Model with Additional Variables', *Geographical Research Forum,* 23, 30-55.

Knudsen, D.C. and Fotheringham, A.S. (1986), 'Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling', *International Regional Science Review,* 10 (2), 127-47.

Kraler, A., Jandl, M., and Hofmann, M. (2006), 'The evolution of EU migration policy and implications for data collection', in N. Poulain, N. Perrin, and A. Singleton (eds.), *Towards the Harmonisation of European Statistics on International Migration (THESIM)*; Louvain-la-Neuve: Université Catholique de Louvain–Presses Universitaires de Louvain, 35-75.

Kupiszewska, D. and Kupiszewski, M. (2010), 'Deliverable 4 - Multilevel Scenario Model', *DEMIFER - Demographic and Migratory Flows Affecting European Regions and Cities*, ESPON & CEFMR, Warsaw. (http://www.espon.eu/export/sites/default/Documents/Projects/AppliedResearch/DEMIFER/FinalReport/DEMIFER_Deliverable_D4_final.pdf).

Mathä, T. and Wintr, L. (2009), 'Commuting flows across bordering regions: a note', *Applied Economics Letters,* 16 (7), 735 - 38.

Olsson, G. (1970), 'Explanation, Prediction, and Meaning Variance: An Assessment of Distance Interaction Models', *Economic Geography,* 46, 223-33.

Openshaw, S. (1998), 'Neural network, genetic, and fuzzy logic models of spatial interaction', *Environment and Planning A,* 30 (10), 1857-72.

Plane, D. A. (1982), 'An information theoretic approach to the estimation of migration flows', *Journal of Regional Science,* 22 (4), 441-56.

Plane, D.A. (1984), 'Migration Space: Doubly Constrained Gravity Model Mapping of Relative Interstate Separation', *Annals of the Association of American Geographers,* 74 (2), 244-56.

Pooler, J. (1994), 'An extended family of spatial interaction models', *Progress in Human Geography,* 18 (1), 17-39.

Poulain, M., Perrin, N., and Singleton, A. (eds.) (2006), *THESIM: towards harmonised European statistics on international migration* Louvain-la-Neuve: Presses universitaires de Louvain.

Raymer, J. and Abel, G. (2008), 'The MIMOSA model for estimating international migration flows in the European Union.', *Joint UNECE/Eurostat Work Session on Migration Statistics*, UNECE/Eurostat, Geneva, (Working paper 8). (http://www.unece.org/stats/documents/ece/ces/ge.10/2008/wp.8.e.pdf).

Rees, P. (1977), 'The measurement of migration from census and other sources', *Environment and Planning A,* 9, 257-80.

Singleton, A., Wilson, A., and O'Brien, O. (2010), 'Geodemographics and spatial interaction: an integrated model for higher education', *Journal of Geographical Systems*, 1-19.

Stillwell, J. (1978), 'Interzonal migration: some historical tests of spatial-interaction models', *Environment and Planning A,* 10, 1187-200.

Taylor, P.J. (1983), *Distance decay in spatial interactions*, CATMOD, 2; Norwich: Geo Books.

Williams, P.A. and Fotheringham, A.S. (1984), *The calibration of spatial interaction models by maximum likelihood estimation with program SIMODEL*: Dept. of Geography, Indiana University.

Wilson, A. (1970), *Entropy in urban and regional modelling*, eds R.J. Chorley and D.W. Harvey, Monographs in spatial and environmental systems analysis; London: Pion.

Wilson, A. (1971), 'A family of spatial interaction models, and associated developments', *Environment and Planning A,* 3, 1-32.