Co-financed by the European Regional Development Fund

Inspire Policy Making with Territorial Evidence

FINAL REPORT //

Annex 6. Interregional flows of tourism in Europe

IRiE – Interregional Relations in Europe

Annex 6 // October 2021

This FINAL REPORT is conducted within the framework of the ESPON 2020 Cooperation Programme, partly financed by the European Regional Development Fund.

The ESPON EGTC is the Single Beneficiary of the ESPON 2020 Cooperation Programme. The Single Operation within the programme is implemented by the ESPON EGTC and co-financed by the European Regional Development Fund, the EU Member States and the Partner States, Iceland, Liechtenstein, Norway, Switzerland and the United Kingdom.

Coordination:
Nicolas Rossignol, Head of Unit for Evidence and Outreach (ESPON EGTC); Marta Roca, Financial Expert (ESPON EGTC); Xabier Velasco Echeverría, Head of Territorial Observatory of Navarra (NASUVINSA, lead partner)

Communication:
Nikos Lampropoulos, Project Expert - Press and Media Activities (ESPON EGTC); Sheila Izquieta Rojano, NASUVINSA (Spain)

Authors
Denis Cerić, Institute of Geography and Spatial Organization, Polish Academy of Sciences (IGSO PAS); Dorota Ciołek, Institute of Geography and Spatial Organization, Polish Academy of Sciences (IGSO PAS); Konrad Czapiewski, Institute of Geography and Spatial Organization, Polish Academy of Sciences (IGSO PAS); Marcin Mazur, Institute of Geography and Spatial Organization, Polish Academy of Sciences (IGSO PAS)

Advisory group
Marie Lorraine Dangeard, National Agency for Territorial Cohesion, France; Margarita Golovko, Regional Development Department, Ministry of Finance, Estonia; Ivan Lajtman, Directorate for Regional Development, Ministry of Regional Development and EU Funds, Croatia.

Acknowledgements
The authors would like to thank the members of the ESPON IRiE consortium, the members of the Advisory Group, and the participants in the Stakeholders' Workshop for the insightful comments and suggestions.

Information on ESPON and its projects can be found at www.espon.eu.
The website provides the possibility to download and examine the most recent documents produced by finalised and ongoing ESPON projects.

Layout and graphic design by BGRAPHIC, Denmark

Contact: info@espon.eu

Co-financed by the European Regional Development Fund

Inspire Policy Making with Territorial Evidence

FINAL REPORT //

Annex 6. Interregional flows of tourism in Europe

IRiE – Interregional Relations in Europe

Annex // October 2021

Disclaimer

This document is an interim report.

The information contained herein is subject to change and does not commit the ESPON EGTC and the countries participating in the ESPON 2020 Cooperation Programme.

The final version of the report will be published as soon as approved.

Table of contents

List of figures and tables

List of figures

List of maps

List of tables

Abbreviations

COVID-19	Coronavirus disease 2019
C2C	Country-to-Country
DM	Matrix of Distances
EEA	European Environment Agency
EFTA	European Free Trade Association
ESPON	European Territorial Observatory Network
EU	European Union
EUROSTAT	European Statistical Office
GDP	Gross Domestic Product
IGSO-PAS	Institute of Geography and Spatial Organisation Polish Academy of Sciences
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Co-operation and Development
PPS	Purchasing Power Standards
R2R	Region-to-Region
RLDR	Relative Level of Detail Ratio
UK	United Kingdom
UNWTO	United Nations World Tourism Organization

Abstract

This document contains the methodology and first results for Task 1.3A of the IRIE project, on interregional tourism flows.

In the methodological chapter, we have described the entire procedure for data collection and detailed the methodology we have used to compensate for data gaps in C2C matrix and disaggregate the C2C matrix to R2R scale. Further on, we have developed a set of indexes to describe tourism flows as well as a synthesis by means of three simple bi-dimensional typologies. The methodological chapter lays out in addition our explanatory-factors methodology for interregional tourism flows.

The section on Country-to-Country flows diagnoses flows between all countries. Several tourism countries dominate: namely Germany, Spain, France, the UK, the Netherlands, and to a slightly lesser extent Portugal, Poland, Greece, Denmark, and Croatia. In addition, we analyzed the balance of arrivals and departures from each country. We conclude with individual analyses of researched countries (EU27 + UK + EFTA) to determine the relevance of other countries as tourist destinations and tourist origins.

The section on Region-to-Region flows distinguish 5 dimensions of flow analysis: intensity (size), connectivity, balance (comparison of outflow and inflow), concentration (dispersion of senders and receivers), and distance (how far they travel, and the extent to which distance determines their size). These are also the 5 basic groups of indicators, which we have analysed and described, while providing maps to illustrate the most important issues of tourism flow within the entire researched period: 2010-2018. We have presented the typologies — synthesis of the indicators, confronting the indicator for intensity of tourism flows with those for balance, concentration, and distance of tourism flows. This section ends with the explanatory factors analysis.

Keywords

tourism flows, interregional tourism, tourists

Highlights

Data

- Accessible data on tourism flows is incoherent, because of differing definitional bases and data-gathering methodologies among statistical agencies.
- There are two complex sources of data on tourism flows within the researched area (EU27 + UK + EFTA) at the country-to-country level (EUROSTAT and UNWTO), both of them incomplete.
- Data on interregional tourist movement is delivered by statistical agencies of particular countries, but they are too incoherent and incomparable across countries to serve as reliable sources of information on the international scale.

  The most suitable data source for R2R matrix estimation is EUROSTAT's stocks of yearly domestic arrivals to NUTS 2.

Methodology

- To gather complex information on spatial and temporal allocation of tourism flows within the researched area, two procedures are needed: for the completion data gaps in the C2C matrix and for the disaggregation of the C2C matrix to the R2R scale of detail.
- We have implemented the following methods of C2C matrix completion (in order of decreasing priority): 1) cross-referencing of indexes on tourist movement delivered by UNWTO; 2) interpolation or extrapolation of temporal rows; 3) analysis of total tourist movement dynamics for temporal rows completion; 4) harmonization of data derived from different sources, by use of RLDR; 5) gravity-model analysis, by use of GDP PPS, number of arrivals with accommodation and Matrix of Distances.
- We have used analysis of the gravity model at the research-area scale to disaggregate these cells within C2C matrix, which are related to international movement. To complete the cells for domestic movement, we have disaggregated regional stocks of domestic arrivals at NUTS 2 through the gravity model, adjusted to national specificity of the function of distance.
- We have developed a set of indexes to describe tourism flows in four dimensions (intensity, balance, concentration, and distance impact), as well as a synthesis by means of three simple bi-dimensional typologies.
- The addition of a dozen or so explanatory variables in our analysis of explanatory factors did not produce a significant explanatory benefit. By emphasizing the importance of other factors, however, they did enable us to significantly reduce the pure influence of GDP, population size, and distance.

New territorial evidence

- Observed tourism flows between all researched countries in the period 2010-2018 had high inertia in time. This means that the dominant directions of tourism trips of the residents of a given country did not change substantially.
- It is possible to indicate a number of countries that are attractive for residents of at least several countries — Austria, Germany, Greece, Spain, France, Croatia, Italy, and Portugal. These are countries located in the Mediterranean (summer tourism) and in the Alps (winter tourism).
- There is a whole range of regional attractiveness, which means that a given country is a tourism destination for people from neighbouring countries — for example, Danes and Norwegians travel to Sweden, Bulgarians and Hungarians to Romania, and Lithuanians and Estonians to Latvia.

- Our results directly indicate the potential direction of tourism marketing for individual countries — either to strengthen the dominant destinations or to open up new destinations.
- Many interregional flows happen within the Mediterranean region's largest countries (such as France, Spain, and Italy). With less intensity, large interregional movements of tourists within the same country occur also in Scandinavian countries (Norway, Sweden, and Finland), Ireland, the Netherlands, Portugal, Poland, Switzerland, Germany, Romania, Bulgaria, and Greece.
- The north-south axis generally dominates international interregional tourism flows, as between regions of Central Europe and Adriatic Croatia, between regions of Finland, Estonia, Latvia, and Lithuania, and between Germany and Austria, France and Spain, France and Italy, and regions of the UK with NUTS 2 regions in France and Spain.
- Our analysis of tourist intensity showed that tourism's impact is greatest in the mountainous parts of Austria, Switzerland, and Italy (Alps, ski resorts) and in the NUTS 2 region of Adriatic Croatia, situated along the nearest Mediterranean coast to the European continent (and tourism demand markets). Other areas with tourist intensity values of 10 to 20 cover the southernmost region of Portugal, the Balearic Islands, Corsica, the Greek islands (in the Ionian and Aegean seas), southern Norway, northern Denmark, the Finnish autonomous region of the Åland Islands, and the southernmost NUTS 2 regions of the Netherlands and Belgium.
- The greatest imbalance in tourism flows within regions where inflow dominates was observed along the Mediterranean and Baltic seas, more precisely in the Mediterranean coastal regions of Spain, France, and Croatia, and in Denmark, northeastern Germany, and northwestern Poland (in the Baltic region).
- Inflows to the regions situated in the northern and eastern parts of Europe consist of tourists from a rather small area, mainly domestic, while the inflow is dispersed to the tourism destinations of southern Europe. Outflows are the other way around: tourists from northern and eastern Europe visit various regions, while those from southern Europe do not move to dispersed areas outside the Mediterranean zone.
- As for tourism-flow distance, tourism in northern Europe remained largely within the region, with few tourists travelling from faraway places to regions of northern Scandinavia, the northern UK, Estonia, or Latvia. From the other side, inhabitants of mentioned regions would travel longer distances to reach their tourism destination.
- Within the explanatory analysis, we observed a positive impact from such variables as the share of naturally valuable areas, the difference in rainfall between origin and destination regions, the lack of linguistic differences between two regions, and the choice of islands as a travel destination. Additionally, we should point out that the number of tourists was correlated with the level of education among the inhabitants — namely, with the participation of people with the highest level of education — and, to a lesser extent, with the general level of employment.

1 Introduction

1.1 Background

The World Tourism Organization (UNWTO) has announced that worldwide tourism flows increased continuously over the decade before the COVID-19 pandemic. In 2017, worldwide tourist arrivals reached 1,323 million, for an average annual growth rate of 7% since 2009. Some 51% of all tourist arrivals were in Europe, which witnessed an 8% expansion relative to 2016 (UNWTO, 2017). Tourism represents 10% of the European Union's GDP, 12 million people are employed in the tourism sector (9% of total EU employment), and 22% of total service exports are ascribed to tourism (UNWTO, 2018).

According to EUROSTAT (EUROSTAT, 2021), three types of tourism flows can be distinguished: domestic (within the country of residence), outbound (out of the country of residence), and inbound (to a country other than the country of residence). Until recently, the main source for European statistics on inbound tourism was statistics on arrivals and nights spent by non-residents at tourist accommodation establishments. The EUROSTAT methodological approach is based on the simple fact that an outbound flow for one country represents an inbound flow for the country visited. Combining all outbound trips made by Europeans to a given country (as their main destination) yields an estimate of intra-EU inbound tourism flows into said country. However, this approach is new, does not cover the period in which the ESPON IRiE project is interested (2010-2018), or covers the inter-regional flows within or between countries.

In addition to providing statistical data on international tourism arrivals — for better understanding of tourism flows between source markets and destinations (countries) — the World Tourism Organization (UNWTO) together with the company Telefonica has launched an interactive online dashboard on source markets and destinations, which includes arrivals and overnight stays by source market, top 10 destinations by source market, arrivals and overnight stays by destinations, top 10 source markets by destination, and heatmaps (UNWTO, 2021). Although, the UNWTO provides data on yearly international tourism from 2010 to 2018, it does not cover tourism flows within regions smaller than countries.

Access to inter-regional tourism flows within and between countries is a statistical rarity worldwide. The ESPON IRiE project has mined all the national statistical offices of the EU27 + UK + EFTA countries for data on tourism flows on the C2C and R2R level for each country. The data provided (or accessed online) do not cover altogether the time and space in which the ESPON IRiE project is interested, and therefore preclude continuous analysis over long periods of time. Use of the collected data therefore requires a complex methodology, which includes estimation of the gaps for C2C matrix and disaggregation data for R2R matrix, as detailed in the methodology chapter of this report.

1.2 Research need

Tourist arrivals and tourism revenues have been extensively studied to evaluate international tourist flows, whereas the structure and evolution of these flows have received less attention (Shao et al., 2020).

Since official data on R2R tourism flows within the research area (EU27 + UK + EFTA) do not exist, their generation is a must — they are of practical use for both tourism-sector workers and policymakers. They enable these groups to target their marketing strategies or improve tourism policies for their respective territories.

1.3 Objective

The ESPON IRiE project aims to generate new data and find new territorial evidences in interregional relations within the EU27, the UK, and the EFTA countries. We have therefore produced and analysed interregional tourism flows (Task 1.3.a). Our generated database includes C2C and R2R tourism flows for the period 2010-2018.

We have generated C2C and R2R matrices (see Chapter 2 for details) and described both C2C and R2R tourism flows in the research area (Chapter 3). Our objective was also to pinpoint key questions for subsequent tasks (Pan-European systemic analysis, Scenarios and Policy Implications – see Chapter 4) and to provide recommendations for data providers to improve data quality (see Chapter 5).

2 Methodology

2.1 Tourism data harmonization strategy

As ‘tourist movement’ is a very imprecise term and the boundary conditions of this phenomenon seems to be very blurred, all researches on tourism geography need to face a key definitional challenge at the beginning. Such challenge gains even more importance in each case when the tourism movement needs to be quantified by data gathered with use of some concrete methodology. Unfortunately, it differs significantly across agencies and surveys. We applied a broad definition based on the necessity of accommodation by person, who arrives to some area because of any tourist purpose as principle. Although we are fully conscious of imperfectness of such theoretical background, it seems to be the most relevant from the practical perspective, after taking into account all constraints of data spatial and temporal coverage.

During the process of data mining, we realized that two basic sources of data on the size and spatial allocation of tourism flow among 32 countries being in 2018 EU or EFTA member in the period of 2010-2018, can be taken into consideration as complex and reliable enough for the purpose of the project: EUROSTAT (https://ec.europa.eu/eurostat/web/main/data/database) and UNWTO (https://www.unwto.org/unwto-tourism-dashboard). Unfortunately, both deliver information at the national level of spatial detail and are incomplete in spatial and temporal dimensions to some extent. This section provides an outlook of strategy applied to harmonization and completing of data on tourism flow. It consists of two main methodological procedures to be conducted subsequently: **estimating data gaps for C2C matrix** and **disaggregating C2C matrix to R2R level of spatial detail**. The initial research allowed to identify several methodological challenges to be faced. At the preliminary step, they have been ordered and assigned to the following major groups:

a) incomparability of data delivered by different sources (diverse definitions, methodologies of sampling and data gathering etc.);
b) recognition and assessment of the broad spectrum of opportunities regarding variables to be applied as estimators and methods of completing data gaps within C2C matrix;
c) optimization of C2C matrix spatial disaggregation method;
d) selection of optimal variables to be applied within the statistical model to estimate spatial allocation of tourism flow volume.

Afterwards, **the levels of reliability and priorities** have been assigned to general approaches as well as to each individual solution within. The final hierarchy in this regard has been composed by **four main principles,** where each is subordinated to the previous one**.** They **has been accepted based on expert knowledge on the domain of tourism geography and initial recognition of empirical data**, and applied consequently every time when choice between parallel methodological paths of data processing existed.

The first, **superior rule of priorities** is that **applying** external, **non-tourist estimators shall be avoided** as long as possible. It is a direct outcome of the ESPON IRIE project general demand, where explanatory factors of tourism movement spatial and temporal variability are an object of in-depth investigation within the framework of the other tasks. Therefore, significant scale of applying such estimators for data completing would lead to the obvious *idem per idem* error of explanation (lat. *circulus in definiendo*).

The second assumed **overriding rule** is, that interpolation or extrapolation of **data based on temporal rows and dynamics leads to more reliable results of estimation than identifying regularities in spatial dimension**. This rule were results from the observation, that total volume of tourist movement is resultant of diverse elements of its heterogeneous internal structure, which is significantly set in territorial specificity of both, tourists’ origin and tourist attractions, as well as of each individual combination of origin and attractions' type. On the other hand, changes of total volume of tourist movement in temporal aspect shows significant dose of continuity.

The third general overriding rule of priorities which has been assumed, is as follows: **combining and mixing data, which origins from different data sources, shall be avoided when possible**. The heterogeneity and incoherence of data gathered with use of various methodological assumptions has been assessed as relatively high. Moreover, very different types of tourism embraced under 'total tourism interregional flow volume' and their significance is strongly determined by national and regional specificity of both, tourist origin and attraction. Thus it needs to be emphasized, that the problem of incomparable data derived from different sources affects not only a scale of tourist movement registration, but also bias of its spatial structure. After taking into account all abovementioned conditions, we decided to seek in each case when only possible to use any information on tourism movement derived from the same data source when estimating and completing data gap instead of applying any kind of data sources merging. Furtherly, we clearly acknowledged any constraints of such approach and incorporate them into our methodology as well as into the system of metadata description.

The deeper insight and comparison of two data sources, which has been indicated as worth to consider as principal for the project purpose, has evidenced that data reported by UNWTO shall be perceived as advantageous in relation to data presented by EUROSTAT. Three arguments for such statement can be raised. Firstly, despite a percentage of 992 cells (32 countries of origin x 32 countries of destination – 32 domestic flows of main diagonal) being the data gaps within C2C matrix filled based on a raw data delivered by UNWTO for each particular year of 2010-2018 period is much higher (see table 1), a simultaneously delivered complementary indexes on tourist flow allow to model the co-occurrence of tourists counted as defined here and complete most of existing data gaps with use of this single source only and without need for any variables, which would go beyond tourism movement characteristics. It allows to largely limit necessity for supporting the procedure of C2C matrix completing by means of other data sources exploitation. Secondly, a significantly higher numbers of tourists evidenced by UNWTO (see table 2) can be perceived as indicating the advantage of this source regarding reliability and accuracy of data gathered as well. Thirdly, assessing a sample of intentionally selected spatial case studies, for which data were delivered simultaneously by both sources being under comparison, by experts on tourism in each particular geographical area has been evidenced, that UNWTO is evidencing predominant directions of tourism flows and proportion of the numbers more adequately than EUROSTAT.

Table 1: A percentage of data gaps in the 2010-2018 period and in each individual year, according to the data origin

	2010-2018	2010	2011	2012	2013	2014	2015	2016	2017	2018
EUROSTAT, raw	54.28	49.50	44.86	58.87	55.44	57.96	57.56	56.05	55.14	53.13
UNWTO, raw	80.21	80.54	80.44	79.54	79.54	79.33	79.33	80.85	81.15	81.15
UNWTO, raw or modelled	9.48	14.62	14.62	10.58	8.47	7.36	7.36	7.36	7.46	7.46

Table 2: A comparison of tourism volume in the 2010-2018 period and in each individual year, according to the data source. Sum of common cells, where data according to both sources is available.

	2010-2018	2010	2011	2012	2013
EUROSTAT, raw	779,488,962	77,560,360	83,273,724	91,482,430	93,558,350
UNWTO, raw	1,288,901,135	110,283,112	124,600,259	148,865,429	155,540,946
Common cells	988	106	123	107	115
Percentage of common cells	11.07	10.69	12.40	10.79	11.59
Ratio for raw data	1.654	1.422	1.496	1.627	1.663
EUROSTAT, raw	1,965,900,549	188,744,035	194,564,558	222,609,787	232,491,719

UNWTO, raw or modelled	4,339,753,885	348,068,088	372,856,008	461,712,856	501,506,832
Common cells	3,855	438	479	391	425
Percentage of common cells	43.18	44.15	48.29	39.42	42.84
Ratio for raw or modelled data	2.208	1.844	1.916	2.074	2.157
	2014	**2015**	**2016**	**2017**	**2018**
EUROSTAT, raw	77,870,764	79,134,482	83,253,549	93,786,808	99,568,495
UNWTO, raw	137,001,790	143,520,959	146,098,021	158,356,692	164,633,928
Common cells	107	111	105	107	107
Percentage of common cells	10.79	11.19	10.58	10.79	10.79
Ratio for raw data	1.759	1.814	1.755	1.688	1.653
EUROSTAT, raw	202,368,502	206,004,582	218,894,841	239,613,432	260,609,093
UNWTO, raw or modelled	461,869,376	490,684,618	524,533,341	569,116,368	609,406,400
Common cells	405	412	424	431	450
Percentage of common cells	40.83	41.53	42.74	43.45	45.36
Ratio for raw or modelled data	2.282	2.382	2.396	2.375	2.338

On the other hand, EUROSTAT occurred as an indispensable source of data during the procedure of disaggregating C2C matrix to R2R level for cells of domestic tourism. Even though this data contains domestic arrivals according to NUTS 2 instead of wholesome vector parameters, it is still the only complex source of information on the spatial allocation of domestic tourist movement between NUTS 2 within each individual country. Such information is not specified in the matrix of international movement, neither by statistics on total domestic tourism flows – due to MAUP (modifiable areal unit problem) and the need for taking into account the internal movement within each particular NUTS 2 (Viegas et al. 2009).

The fourth assumed **overriding rule** is, that any **raw data** on tourism movement, even though expressed as different index, **is a more accurate estimation** of tourism flow volume change in temporal or in spatial dimension, **than any simple extension** of row or tendency **based on statistical basis exclusively**. The rationale for such statement is related principally to willingness for exploitation of any accessible empirical component of relevant information, which would be ignored otherwise.

2.2 Estimating data gaps for C2C matrix and disaggregating data for R2R matrix

The priorities, as expressed within abovementioned overriding rules, connected with insufficient empirical information to complete the full C2C matrix with respect to all of them simultaneously, imply the need for hierarchically ordered steps of procedure to be implemented. The basis for this order was the assessment of alternative approaches to data estimating according to number and position of overriding rules to be faced. The more advantageous approach has been substituted by less favourable exclusively in case of definitely exploited opportunities for pushing forward the first of them.

Finally, the following hierarchic order of approaches to C2C matrix completing, and methods related, has been implemented (in order of decreasing priority):

0. cross-reference of indexes on tourist movement delivered by UN-WTO;

1.1. interpolation or extrapolation of temporal rows within UNWTO or EUROSTAT data;

1.2. analysis of total tourist movement dynamics based on UNWTO and on EUROSTAT data;

2. harmonization of data derived from different sources, by use of RLDR;
4. model of gravity analysis, by use of: GDP PPS, number of arrivals with accommodation and DM.

Particular basic numbers of the list above are related directly to the number of overriding rules, that need to be omitted.

Step 0 relies on searching for co-occurrence of diverse indexes expressing tourism movement, which are delivered by UNWTO (e.g. excursions, without accommodation). If there is no data on index of TFR, perceived as relevant for the definition of tourism flow taken within IRIE, for a given relation in a given year, but data on any other index on tourist movement exists, a model of their co-occurrence is applied then,

Within step 1.1. an interpolation (in case of data availability for at least one year before and at least one year after a given data gap) or an extrapolation (in case of data availability for only earlier than a year of given data gap or for the years later only) of temporal rows is applied when data for a cell of a given data gap for at least two other years of 2010-2018 period available. A geometric row has been constructed, posing link between one the last accessible value from previous period and one the earliest accessible value from the forthcoming period (interpolation) or based on two the closest years with accessible values (extrapolation). It excludes eventual result of negative values.

A step 1.2. has been implemented, when data in a given cell of matrix was available only for one year of 2010-2018 period. The values of ratio on total flow volume between a year of estimation ($y\ est$) and the year of the only known value ($y\ known$) is calculated, based on these cells only, where data for both years of a given pair was available before step 1 (data of higher reliability than just calculated). Finally, the ratio has been applied for estimation of cell value according to following formula:

$$Flow_{i,j,y\ est} = Flow_{i,j,y\ known} \times \frac{Flow\ total_{y\ est}}{Flow\ total_{y\ known}}$$

At the step 2, an analogous solution has been implemented for the cells, where value for any of years within 2010-2018 period is not delivered by UNWTO and can not be estimated at step 0, but it would be available as raw data or could be estimated at step 1, when EUROSTAT data added. In such cases, the ratio is called *Relative Level of Detail Ratio* (RLDR) and is applied to compensate difference in a total flow volume evidenced by different data sources. The essential exception is, that for the best possible comparability, only cells of raw data delivered by both of data sources are taken into account for the estimation of total flow volume proportion. Finally the following formula has been implemented for estimation of data gaps within step 2:

$$Flow\ est_{i,j} = Flow_{i,j,source\ known} \times RLDR = Flow_{i,j,eurostat} \times \frac{Flow\ total_{UNWTO}}{Flow\ total_{eurostat}}$$

The last step, as well as the procedure of data disaggregation, is based on the general law of gravity, where the power of interaction between two points placed in a geographical space is proportional to the result of multiplication of pushing and of attracting force, modified by some function of distance. In this case, as using external variables for data gaps completing is exceptional and applied when necessary only, the simplest model has been constructed, with use of the least controversial variables only. Finally, the approximation of pushing force (potential tourists) has been posed by GDP expressed in PPS in a given year, and the approximation of attracting force by stock of tourists making use of accommodation facilities of a given region in a given year. The matrix of distances has been filled by calculation of orthodromic distances between centroids of a given regions weighted by population of municipalities. To limit impact of enclave-regions for shortened distance between centroids, distance between centroids has been substituted by the value of average distance within circle of the same area as the larger region of a given pair (max$\left(\sqrt{\frac{2\times Area_i}{\pi}}; \sqrt{\frac{2\times Area_j}{\pi}}\right)$), when it only occurs longer. The power function has been taken as a function of distance, instead of usually applied exponential, because a negative impact of the distance on flow volume in case of tourism is not an obvious result and should not be assumed *a priori*. The exact function of distance has been established as an regression by the Least Squares Method, using a full set of all already known values within all yearly matrices. Finally, the following model of estimation has been applied:

$$Flow\ est_{i,j} = 9{,}179 \times 10^{7} \times GDP\ PPS_i \times Pop_j \times 294{,}233 \times {dist_{i,j}}^{-0{,}802}.$$

A negative value of exponent within a formula of general model indicates a decrease of flow volume along distance.

The analysis of the model of gravity at the researched area scale has been applied also for dis-aggregation of these cells within C2C matrix, which are related to international movement. For completing the cells related to domestic movement, the regional stocks of domestic arrivals at NUTS 2 has been allocated among domestic regions of origin, by use of the model of gravity built on C2C matrix and adjusted to national specificity of the function of distance.

2.3 Measuring tourism flows

To measure and describe the tourism flows within the research area comprehensively and complementarily, a system of intentionally constructed indexes, ordered by use of overarching four dimensional framework (intensity, balance, concentration and distance impact) has been accepted initially, with full aware of demand for their theoretical independency. They are as follows:

Intensity dimension:

- Intensity index calculated for a pair of regions i and j – relation (i,j):

$$Int_{i,j} = Flow_{i,j} + Flow_{j,i}$$

or calculated for a spatial unit i:

$$Int_i = \sum_{j=1}^{296} Flow_{i,j} + Flow_{j,i}$$

- Weighted intensity index for spatial unit i:

$$Wint_i = \sum_{j=1}^{296} \frac{Flow_{i,j} + Flow_{j,i}}{Pop_i}$$

Balance dimension:

- Balance index for spatial unit i:

$$B_i = \frac{\sum_{j=1}^{296} Flow_{j,i} - Flow_{i,j}}{max\left(\sum_{j=1}^{296} Flow_{i,j}\,;\sum_{j=1}^{296} Flow_{j,i}\right)}$$

- Unbalanced volume index for spatial unit i:

$$BvAsym_i = \sum_{j=1}^{296} Flow_{j,i} - Flow_{i,j}$$

- Average relation asymmetry for spatial unit i:

$$BrAsym_i = \sum_{j=1}^{296} \frac{\left|Flow_{j,i} - Flow_{i,j}\right|}{296 \times max\left(Flow_{i,j}; Flow_{j,i}\right)}$$

Concentration dimension:

- Concentration per area index for spatial unit i, related to total volume of flow:

$$Garea_i = \frac{\sum_{j=1}^{296}\sum_{k=1}^{296} Area_j \times Area_k \times \left|\frac{\left(Flow_{i,j} + Flow_{j,i}\right)}{Area_j} - \frac{\left(Flow_{i,k} + Flow_{k,i}\right)}{Area_k}\right|}{2 \times \left(\sum_{j=1}^{296} Area_j\right)^2 \times \frac{\sum_{j=1}^{296} Area_j \times \left(Flow_{i,j} + Flow_{j,i}\right)}{\sum_{j=1}^{296} Area_j}}$$

to inflow extracted only:

$$Garea_i = \frac{\sum_{j=1}^{296}\sum_{k=1}^{296} Area_j \times Area_k \times \left|\frac{Flow_{j,i}}{Area_j} - \frac{Flow_{k,i}}{Area_k}\right|}{2 \times \left(\sum_{j=1}^{296} Area_j\right)^2 \times \frac{\sum_{j=1}^{296} Area_j \times Flow_{j,i}}{\sum_{j=1}^{296} Area_j}}$$

or to outflow extracted only:

$$Garea_i = \frac{\sum_{j=1}^{296}\sum_{k=1}^{296} Area_j \times Area_k \times \left|\frac{\left(Flow_{i,j}\right)}{Area_j} - \frac{\left(Flow_{i,k}\right)}{Area_k}\right|}{2 \times \left(\sum_{j=1}^{296} Area_j\right)^2 \times \frac{\sum_{j=1}^{296} Area_j \times \left(Flow_{i,j}\right)}{\sum_{j=1}^{296} Area_j}}$$

- Concentration per population index for spatial unit i, related to total volume of flow:

$$Gpop_i = \frac{\sum_{j=1}^{296}\sum_{k=1}^{296} Pop_j \times Pop_k \times \left|\frac{\left(Flow_{i,j} + Flow_{j,i}\right)}{Pop_j} - \frac{\left(Flow_{i,k} + Flow_{k,i}\right)}{Pop_k}\right|}{2 \times \left(\sum_{j=1}^{296} Pop_j\right)^2 \times \frac{\sum_{j=1}^{296} Pop_j \times \left(Flow_{i,j} + Flow_{j,i}\right)}{\sum_{j=1}^{296} Pop_j}}$$

to inflow extracted only:

$$Gpop_i = \frac{\sum_{j=1}^{296}\sum_{k=1}^{296} Pop_j \times Pop_k \times \left|\frac{Flow_{j,i}}{Pop_j} - \frac{Flow_{k,i}}{Pop_k}\right|}{2 \times \left(\sum_{j=1}^{296} Pop_j\right)^2 \times \frac{\sum_{j=1}^{296} Pop_j \times Flow_{j,i}}{\sum_{j=1}^{296} Pop_j}}$$

or to outflow extracted only:

$$Gpop_i = \frac{\sum_{j=1}^{296}\sum_{k=1}^{296} Pop_j \times Pop_k \times \left|\frac{Flow_{i,j}}{Pop_j} - \frac{Flow_{i,k}}{Pop_k}\right|}{2 \times \left(\sum_{j=1}^{296} Pop_j\right)^2 \times \frac{\sum_{j=1}^{296} Pop_j \times Flow_{i,j}}{\sum_{j=1}^{296} Pop_j}}$$

Distance impact dimension:

- Average distance for spatial unit i, related to total volume of flow:

$$Dist_i = \frac{\sum_{j=1}^{296} dist_{i,j} \times \left(Flow_{i,j} + Flow_{j,i}\right)}{\sum_{j=1}^{296} Flow_{i,j} + Flow_{j,i}}$$

to inflow extracted only:

$$Dist_i = \frac{\sum_{j=1}^{296} dist_{i,j} \times Flow_{j,i}}{\sum_{j=1}^{296} Flow_{j,i}}$$

or to outflow extracted only:

$$Dist_i = \frac{\sum_{j=1}^{296} dist_{i,j} \times Flow_{i,j}}{\sum_{j=1}^{296} Flow_{i,j}}$$

- Distance dependence index for spatial unit i, related to total volume of flow:

$$DistR^2{}_i = \frac{\sum_{j=1}^{296}\left(Flow\ est_{i,j} + Flow\ est_{j,i} - \frac{\sum_{j=1}^{296} Flow_{i,j} + Flow_{j,i}}{296}\right)^2}{\sum_{j=1}^{296}\left(Flow_{i,j} + Flow_{j,i} - \frac{\sum_{j=1}^{296} Flow_{i,j} + Flow_{j,i}}{296}\right)^2}$$

or to inflow extracted only:

$$DistR^2{}_i = \frac{\sum_{j=1}^{296}\left(Flow\ est_{j,i} - \frac{\sum_{j=1}^{296} Flow_{j,i}}{296}\right)^2}{\sum_{j=1}^{296}\left(Flow_{j,i} - \frac{\sum_{j=1}^{296} Flow_{j,i}}{296}\right)^2}$$

or to outflow extracted only:

$$DistR^2{}_i = \frac{\sum_{j=1}^{296}\left(Flow\ est_{i,j} - \frac{\sum_{j=1}^{296} Flow_{i,j}}{296}\right)^2}{\sum_{j=1}^{296}\left(Flow_{i,j} - \frac{\sum_{j=1}^{296} Flow_{i,j}}{296}\right)^2}$$

For the purposes of presentation of balance’s dynamics of tourism flow at the level of whole R2R matrix, the another index has been developed:

$$B_{total} = \sum_{i=1}^{297}\sum_{j=1}^{297}\frac{\left|Flow_{j,i} - Flow_{i,j}\right| \times \left(Flow_{i,j} + Flow_{j,i}\right)}{max\left(Flow_{i,j}; Flow_{j,i}\right) \times 2 \times \sum_{i=1}^{297}\sum_{j=1}^{297} Flow_{i,j}}$$

The index of intensity is expressing volume of flow within a given relation of for all relations of a given spatial unit in an absolute units of phenomenon (in case of tourism - persons). Weighted intensity is a value of the intensity index, related to the population of the given region. It supplements the previous index by the consideration of the context of demographic size of the region. Much smaller volume of flows can have relatively higher importance in case of small region.

Also three indexes of balance are complementary regarding cognitive added value. Balance index indicates if the regions is a net-sender or net-receiver of flow and in how much disproportion between both directions exists. The value range is from -1 to 1, where zero means perfectly balanced flow, -1 means outflow only and 1 means inflow only. The index of balance volume is taking into account the size of a given region’s flow and it indicates how much volume of flow is generated or absorbed there. Even largely unbalanced saldo can not generate much surplus or deficit of flow in the whole network if the region is of small size or low weighted intensity. And opposite, regions of the largest agglomerations and of the highest weighted intensity of flow can generate large unbalanced volume of flow even though value of balance index around zero. A complementary role of the average relation asymmetry relies on taking into account balance of each particular relation of the region, even though they can be balanced per saldo. The value range is from 0 to 1, where zero means all of region’s relations perfectly balanced and 1 means all region’s relations in one direction only. This index can indicate “an intermediate regions”, where inflow and outflow is balanced while regions of inflow and regions of outflow are different.

As balance index calculated for the whole matrix of internal flow within researched area is equal zero by definition, a measure based on balance of each individual relationship instead of it total *per saldo* is more appropriate. Moreover, particular relationships are largely diverse according to intensity index, thus taking into account disparities in their impact on the total spatial structure of tourism flow by any kind of weighting is necessary.

Indexes of concentration are using different versions of Gini coefficient, the most commonly used measure of concentration, weighted geographically (by area of particular spatial units) or by population. It is also complementary approach. A concentration per area is indicating pure flow concentration, which is a natural consequence of population distribution inequality to some extent. On the other hand, concentration per population is taking into account this condition and is indicating some kind of concetration’s “added value”.

An average distance index express a distance travelled by average tourist of the region, arriving to- or leaving from- the region. It indicates an empirical range of a given region’s impact. On the other hand, a distance dependency index indicates a degree of distance impact, regardless of its character (increase, decrease, dynamics etc.). The value of 0 is limiting a range of variability of one side and indicates a lack of any co-occurrence between volume of flow to other regions and distance to them, and 1 is limiting the range of values top and indicates existing function of distance, which describes full variability of flow volume.

An in-depth analysis of four dimensions has been summarized by means of three simple bi-dimensional typologies. They have been developed by classification of one the most representative index for each individual dimension: $Wint_i$ (intensity dimension), B_i (balance dimension), $Garea_i$ for flow total volume (concentration dimension) and $Dist_i$ for flow total volume (dimension of distance impact). Each of these indexes has been classified within the framework of three levels.

The thresholds were determined by statistical distribution of empirical data set, where the equal representation is assumed in case of normal distribution (thus $\pm 0{,}431\sigma$ from the mean). Such solution does not imply equal representation of each of three index classes in practice, but returns the added cognitive value of distribution's skewness display. A given pair of classified indexes has been combined furtherly and, finally, each particular combination of classes has been named as appropriate type, by use of terminology expressed at qualitative or/and ordinal scale.

In opposition to individually set of descriptive analysis framework, due to comparability purposes, the set of indexes applied for typology of NUTS 2 according to tourism flow is conventional and the same as for each of other flows' individual typology developed under task 1:

- intensity index;
- weighted intensity index;
- connectivity index;
- selectivity index;
- external influence index;
- Send-Receive balance index.

Three indexes listed as the first can be applied either, as related to total volume of flow or only to inflow/outflow extracted.

2.4 Explanatory factors – econometric methodology

The aim of the next stage of the study was to identify the determinants of tourist flows between all analysed regions in Europe. For this purpose, econometric modelling was conducted using panel data considering 297 NUTS 2 regions for the period 2010-2018.The starting point for the econometric model specification was the gravity model, which is used as a basic tool for the analysis of international trade (Leibenstein & Tinbergen, 1966; Baltagi, Egger & Pfaffermayr, 2003), but also migration flows (e.g. Basher & Fachin, 2008; Mayda, 2010; Molloy, Smith & Wozniak, 2011; Beine & Parsons, 2015). Classically, the role of distance and mass of both the origin region and the destination region is considered in the gravity model. While in trade analysis mass is represented by Gross Domestic Product (GDP), in models describing flows of people it is more advisable to take the population size of regions as mass. Although also many migration studies use the income maximization approach as for example, Basher & Fachin (2008), Beine & Parsons (2015), Beine et al. (2019) as well as Serlenga & Shin (2021)., that is, a more important role is attributed to the size of the economy than to the population itself. However, when analysing tourism flows, it seems that the mass of the two regions do not have the same importance. In this case, it is both the population and the economic size (wealth - GDP) of the origin economy that are more relevant. In addition to the main gravity variables, the study presented here also considers a number of factors both measurable and non-measurable that can affect tourism flows. Hence, the specification of the econometric model used can be written as follows:

$$\log(Tour_{ijt}) = \beta_0 + \beta_1 \log(GDP_{it}) + \beta_2 \log(Pop_{it}) + \beta_3 \log(dist_{ij}) + \beta_4 domestic_{ij} + \beta_5 X_{ij} + \mu_{ijt} + \varepsilon_{ijt}$$

where: the dependent variable, $Tour_{ijt}$ refers to the flow of tourists from a region i (origin) to a region j (destination) in a year t; GDP_{it} denotes the GDP of the origin region in period t while Pop_{it} denotes the population size of the origin region at period t. The geographical distance between the origin region i and the destination region j is represented by $dist_{ij}$ and is measured as the orthodromic distance between the centroids of the regions weighted by the population of the municipalities, corrected for enclaves. Then, the inter-country dummy $domestic_{ij}$ is included, which indicates if the origin region i and the destination region j belong to the same country.

All additional explanatory variables that are expected to affect the volume of tourist flows between European regions are represented by X_{ijt}. Among them some variables are measurable (quantitative), mostly changing in time, and some variables are qualitative factors represented by dummies, most often constant in time. It should be pointed out that, the inclusion of any dummy variables implies that the coefficient results must be interpreted in comparison with the excluded category of the corresponding variable. While, to facilitate interpretation and comparison of results, all the quantitative variables were included as natural logarithms. The terms μ_{ijt} represents the individual effect, specific to each pair of regions, which can be analysed as fixed over time or as random. Finally, we assume that the unexplained in the model part of the variation in tourist flow is represented by the error term ε_{ijt}. Table 3 presents the definitions and data sources of all variables that were used in the econometric modelling.

Table 3: Explanatory variables considered in the econometric analysis of region to region touristic flows

Variable	Definition	Source
$l\_Pop\_o_{it}$	log of population of the origin NUTS2 region in year t	Eurostat
$l\_GDP\_o_{it}$	log of regional gross domestic product of the origin NUTS2 region in year t - million EUR	Eurostat
l_dist_{ij}	log of the orthodromic distance between the centroids of the regions weighted by the population of the municipalities, corrected for enclaves, i.e. if the average distance of the points of a circle with an area equal to that of the larger of a given pair of regions from the centre of that circle (the average internal distance of the larger NUTS) is greater than the orthodromic distance between the centroids it replaces it.	IGSO-PAS
$domestic_{ij}$	0- international; 1 - domestic	Tourism Matrix
$l\_gdp\_pc\_o_{it}$	log of regional gross domestic product per capita (PPS per inhabitant) of the origin NUTS 2 region in year t	Eurostat
$l\_disp\_inc\_o_{it}$	log of disposable income of private households of the origin NUTS 2 region in year t	Eurostat
$l\_pop\_den\_o_{it}$	log of population density of the origin NUTS 2 region in year t	Eurostat
$l\_pop\_den\_d_{jt}$	log of population density of the destination NUTS 2 region in year t	Eurostat
$new\_eu\_o_{it}$	0 – the origin region in the "old" EU member country and EEA country; 1 – the origin region in new EU member country;	IGSO-PAS
$non\_euro\_rel_{ijt}$	0 - both regions in Eurozone countries or in the same country; 1 - at least one region in a country that is not part of the Eurozone	IGSO-PAS
$l\_natura\_d_{jt}$	log of share of NUTS2 area covered by NATURA 2000 sites – km^2	IGSO-PAS (based on EEA data)
$temp\_rel_{ij}$	difference in multi-year average temperature: average temperature in destination region minus average temperature in origin region	COPERNICUS
$precip\_rel_{ij}$	ratio of the multi-year average precipitation: average precipitation in destination region divided by average precipitation in origin region	COPERNICUS
$island\_d_j$	0 - no "island" destination region, 1 – the island destination region.	IGSO-PAS (based on Eurostat classification)
$bord\_reg\_d_j$	0 - no NUTS3 in region is border or of type "land border up to 25 km away"; 1 - at least one NUTS3 in the region is border or is of type "land border up to 25 km away	IGSO-PAS (based on Eurostat classification)
$language_{ij}$	0 - same language, 1 - same group but different language, 2 - different group. If more than one language in a region then the similarity of the most similar was taken into account.	IGSO-PAS
$metro\_o_i$	0 - no MEGA in the origin region; 1 - MEGA 1 or 2 or 3 or 4 is in the origin region	IGSO-PAS (based on ESPON 1.1.1)
$metro\_d_j$	0 - no MEGA in the destination region; 1 - MEGA 1 or 2 or 3 or 4 is in the destination region	IGSO-PAS (based on ESPON 1.1.1)
$l\_h\_edu\_o_{it}$	log of percentage of people with tertiary education in the 25-64 age group in the origin region	Eurostat
$l\_emp\_tot\_o_{it}$	log of total employment in the origin region	Eurostat

Source: authors' elaboration

Unfortunately, not for all variables defined above full sets of observations were available for all regions analyzed. Therefore, in some model specifications, the number of observations differed significantly from those where the explanatory variables were better represented.

The panel econometric model defined in this way, based on the gravity model, can be estimated by various methods (Khan & Hossain, 2010; van Bergeijk & Brakman, 2010). The choice of method depends on whether we face one of three important problems in the set of selected variables. The first potential problem arises from the fact that when we assume the presence of time-fixed individual effects for a given pair of regions there is no possibility to estimate the effects of time-constant factors, that is, many important explanatory variables in the model including the distance. The second problem is related to the potential endogeneity of explanatory factors. While the third difficulty is usual especially in the context of international trade, where multiple zeros may appear in the dataset indicating that in some years there are no trade flows between certain regions. Then it is not possible to count the logarithm from the dependent variable, so such a model cannot be estimated in a classical way. A solution to the first and second problems was first proposed by Hausman & Taylor (1980) and developed by Baltagi & Khanti-Akom (1990) and Stock, et al. (2002). The so-called Hausman-Taylor estimator for panel data with random individual effects takes into account the division of explanatory variables into time-varying and constant factors. Moreover, in both sets one can identify factors that are known to be endogenous to the dependent variable. On the other hand, regarding the third problem, the so-called zero-flows problem, Santos & Tenreyro (2006) proposed the Pseudo Poisson Maximum Likelihood (PPML) estimator, which allows to estimate the gravity model without computing the logarithm of the dependent variable. Moreover, in this model, it is possible to control the heterogeneity of the analyzed units. In the presented study, where the dependent variable was the tourist flows between all European regions, there were no zero flows (with a few minor exceptions), so the estimation of all models was carried out using the Hausman-Taylor estimator. It should also be noted that in order to obtain reliable results of the individual significance test of statistical parameters, robust standard errors were determined allowing to solve the problem of not meeting the assumptions about homoscedasticity and lack of autocorrelation in error term.

In order to verify if the identified explanatory factors are the same for all types of regions, in the next step of the analysis, econometric modelling of the model defined above was conducted for different groups of regions. The divisions were applied in which the following groups of relationships were included:

- **destinations** are mountain regions and non-mountain regions;
- **destinations** are coastal regions and non-maritime regions;
- **origins** are coastal regions and non-coastal regions;
- **destinations** are regions with metropolitan areas (METRO 1 or 2 or 3 or 4) and regions without metropolitan areas;
- **origins** are regions with metropolitan areas (METRO 1 or 2 or 3 or 4) and regions without metropolitan areas.

All estimation results for each group are presented in separate tables presented in chapter *3.2.3. Analysis of the results -> Explanatory factors: drivers and barriers*.

2.5 Data creation: C2C matrices

UNWTO database

For the purpose of the ESPON IRiE project, United Nations World Tourism Organization (UNWTO) data (https://www.unwto.org/unwto-tourism-dashboard) were perceived as advantageous in relation to data presented by the second important resource EUROSTAT. There are three main reasons for this (details in methodological chapter of this report, see chapter 2.1): 1) the UNWTO data, despite higher data gaps within C2C matrix filled based on a raw data for each particular year of 2010-2018, allow to model the co-occurrence of tourists counted as defined here and complete most of existing data gaps with use of this single source only and without need for any variables, which would go beyond tourism movement characteristics, 2) UNWTO provides a significantly higher numbers of tourists, which might be perceived as reliability and accuracy of data gathered, and 3) a sample of intentionally selected spatial case studies, for which data were delivered simultaneously by both sources being under comparison, experts on tourism in each particular geographical area declared that the UNWTO is evidencing predominant directions of tourism flows and proportion of the numbers more adequately than EUROSTAT. UNWTO matrices cover the years between 2010 and 2018 (attached in the ESPON IRIE drive).

EUROSTAT database

The EUROSTAT data https://ec.europa.eu/eurostat/web/main/data/database) was an indispensable source during the procedure of disaggregating C2C matrix to R2R level for cells of domestic tourism. EUROSTAT matrices cover the years between 2010 and 2018 (attached in the ESPON IRIE drive). Despite of this overriding rule, a lack of data at R2R level caused that the procedure to disaggregate the C2C matrix relied on the application of the spatial regularities and dependencies of the tourists movement spatial pattern from some basic explanatory factors, which were evidenced at C2C level, to estimate the share of each individual pair of NUTS 2 within a given cell. More details on EUROSTAT database and creation of R2R matrices for tourism flows can be found in methodological chapter of this report.

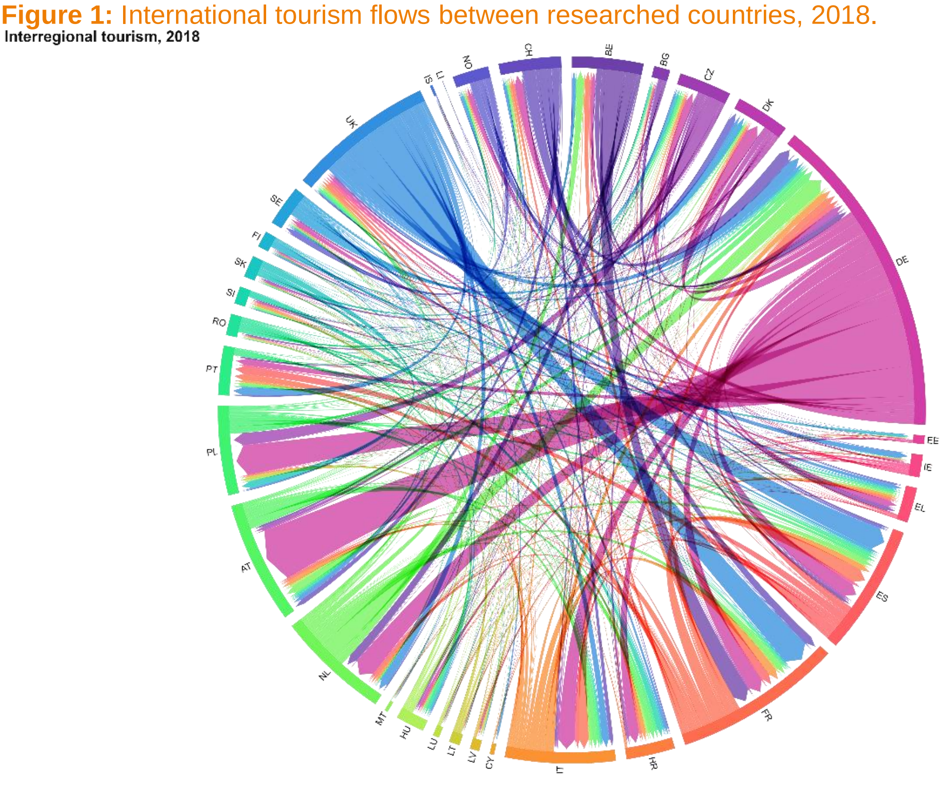
3 Results

3.1 Country-to-country flows

3.1.1 Description of results

Analysis of the results allows us to indicate the countries which have the highest tourist traffic in the total mass — for both outgoing tourists from a given country and incoming tourists to it. These are: Germany, France, Great Britain, Spain, Italy, and the Netherlands. The first five are among the largest of the researched countries, so the number of tourists was significant. Figure 1 illustrates all tourism flows between countries. On the one hand, large flows are visible — e.g. Germans to Austria and Poland, British people to France and Spain. More importantly, however, the figure shows a whole range of smaller flows connecting all countries to all countries. In the pre-COVID-19 era, then, the tourism-related mobility of residents was very high, and in addition to a number of classic and expected destinations, many tourists desired to get to know less-known countries in Europe.

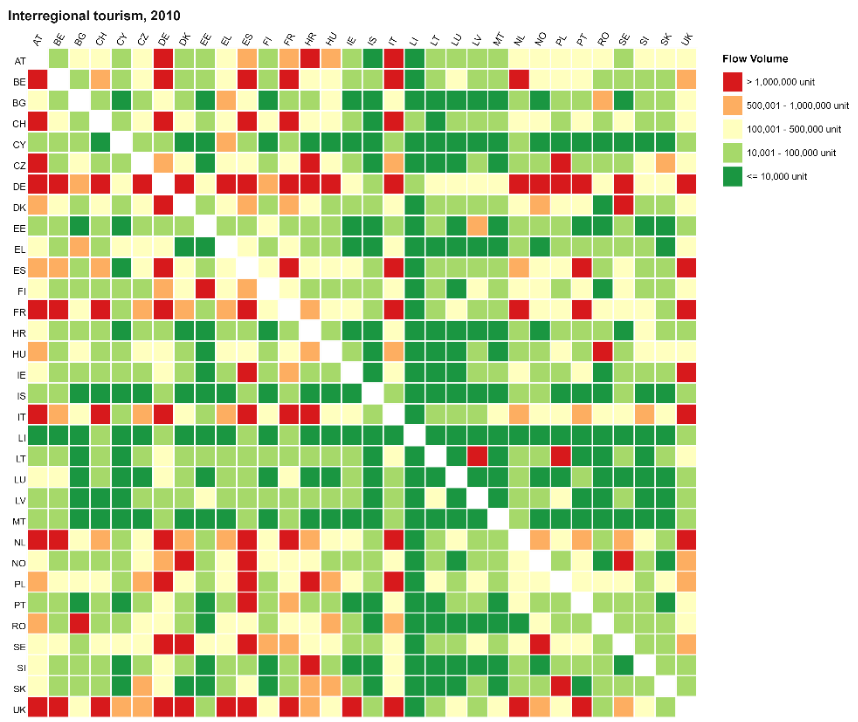
Figure 1: International tourism flows between researched countries, 2018.

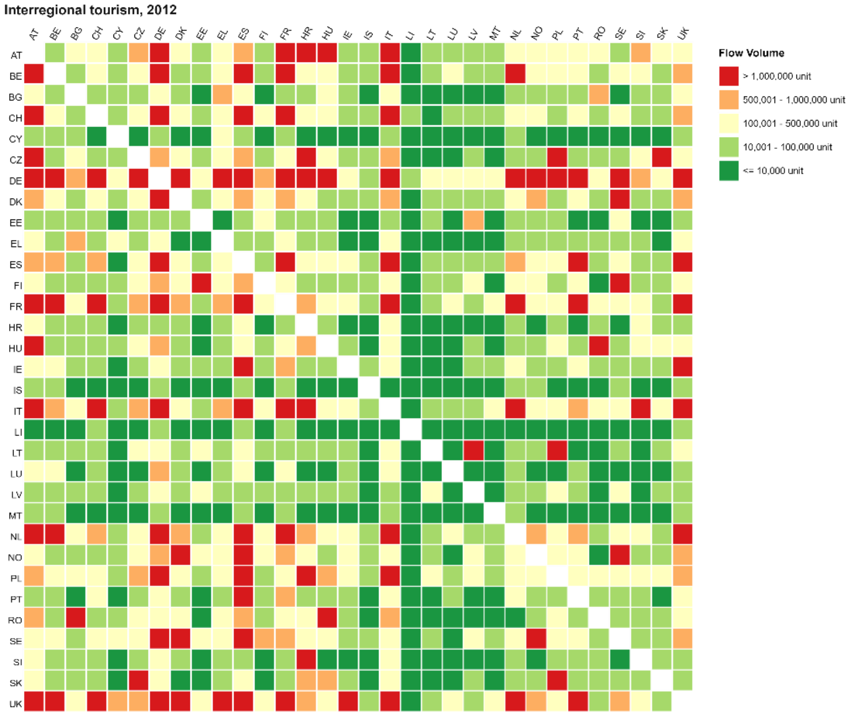


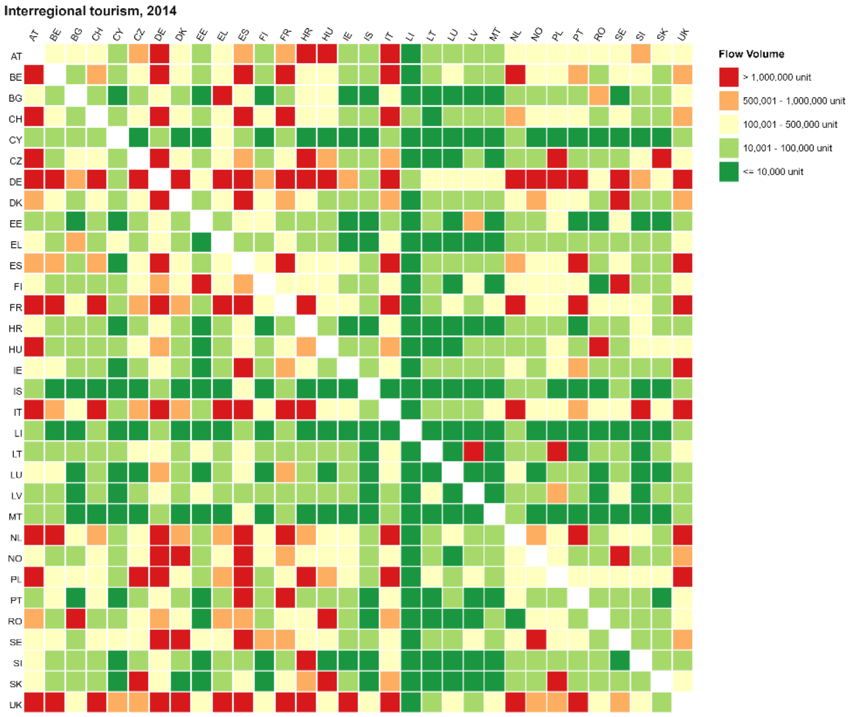
Sources: UNWTO & EUROSTAT databases / Authors' elaboration

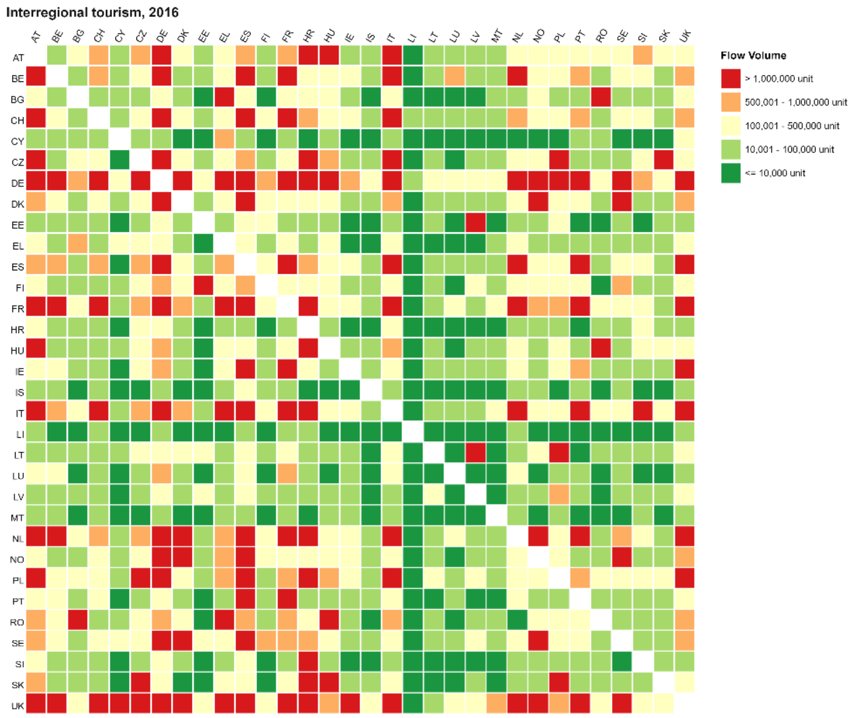
Figure 2 presents the intensity of tourist travel between all researched countries in a two-year interval in the period 2010-2018. The first conclusion that can be drawn from detailed analysis of the diagrams is the high inertia in time of the observed flows. This means that the dominant directions of tourism trips of residents of a given country did not change substantially. Secondly, a number of countries — Austria, Germany, Greece, Spain, France, Croatia, Italy, and Portugal — were attractive to residents of several countries. These are located on the Mediterranean (summer tourism) and in the Alps (winter tourism). Their attractiveness is so universal that the number of tourist arrivals did not depend on the distance to tourists' places of residence. Thirdly, there was a whole range of country attractiveness, which means that a given country tends to be a tourist destination for people from neighbouring countries — for example, Danes and Norwegians travel to Sweden, Bulgarians and Hungarians to Romania, Lithuanians and Estonians to Latvia. The results directly indicate the potential direction of tourism marketing for individual countries — either to strengthen the dominant destinations or to open up to new destinations.

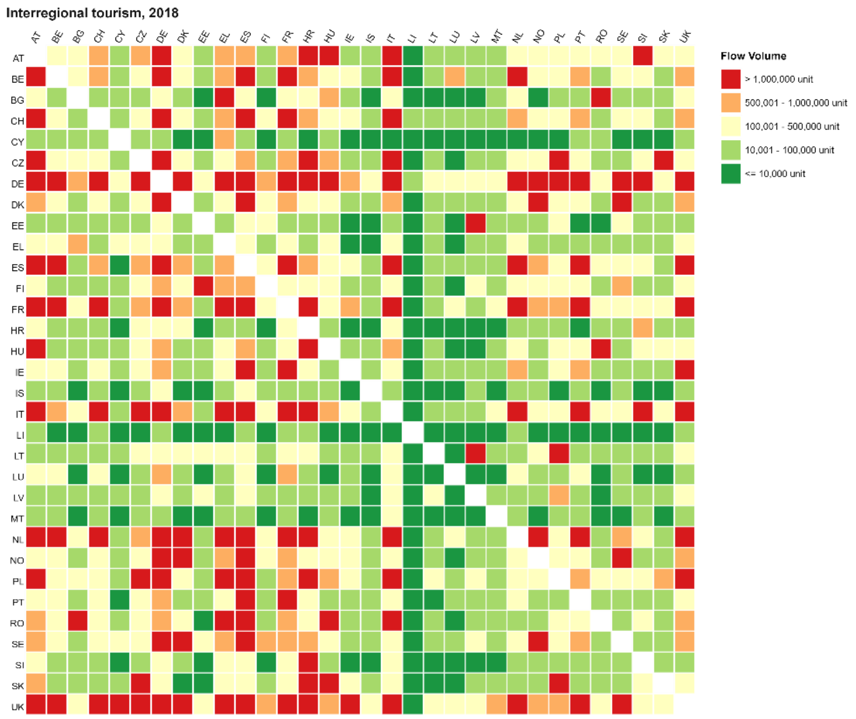
Figure 2: Matrix of the intensity of tourism flows between researched countries, 2010, 2012, 2014, 2016, and 2018.




Interregional tourism, 2012
AT BE BG CH CY CZ DE DK EE EL ES FI FR HR HU IE IS IT LI LT LU LV MT NL NO PL PT RO SE SI SK UK
Flow Volume
> 1,000,000 unit
500,001 - 1,000,000 unit
100,001 - 500,000 unit
10,001 - 100,000 unit
<= 10,000 unit


Interregional tourism, 2014
AT BE BG CH CY CZ DE DK EE EL ES FI FR HR HU IE IS IT LI LT LU LV MT NL NO PL PT RO SE SI SK UK
Flow Volume
> 1,000,000 unit
500,001 - 1,000,000 unit
100,001 - 500,000 unit
10,001 - 100,000 unit
<= 10,000 unit

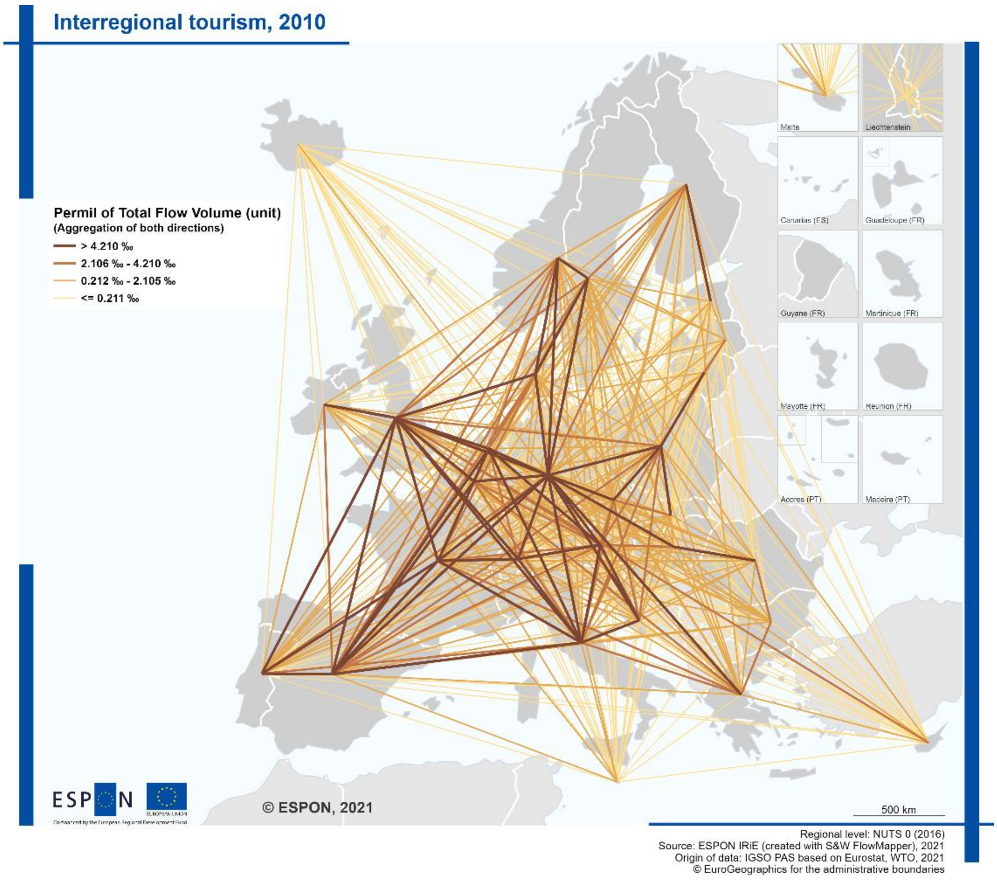




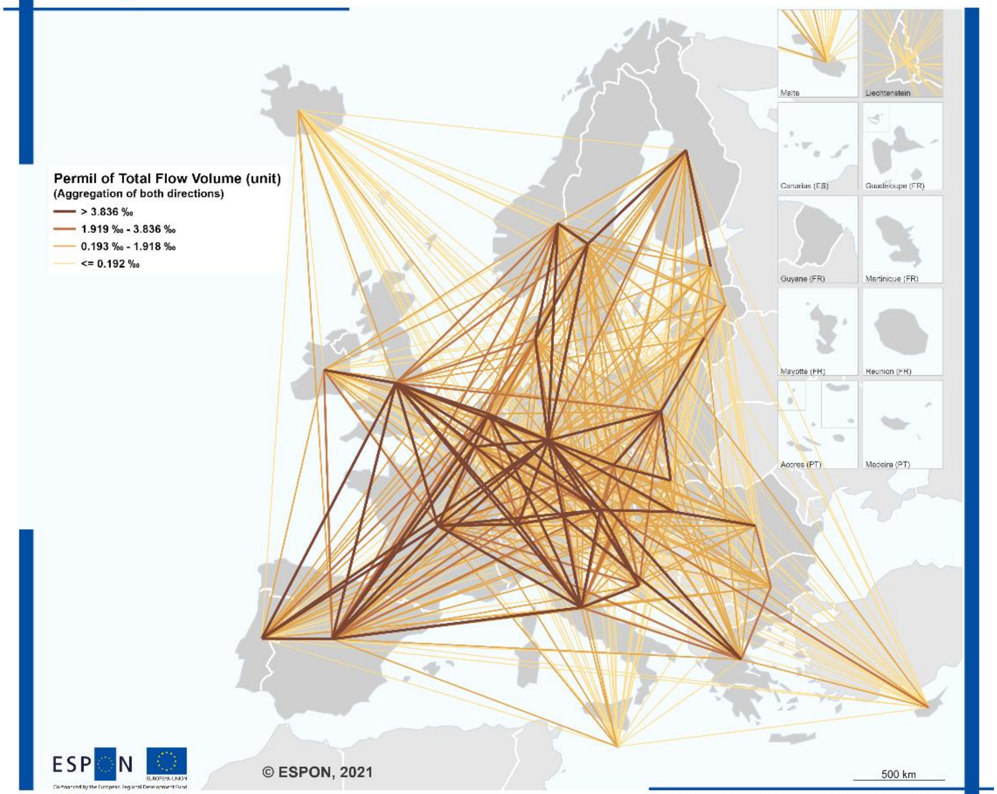
Sources: UNWTO & EUROSTAT databases / Authors' elaboration

Our analyses above are presented also as maps for the period 2010-2018. They show half of the most intensive tourist relations — expressed in number of tourists travelling between pairs of countries. As indicated earlier, we can first observe the existence of several dominant tourism countries: Germany, Spain, France, the UK, the Netherlands, and to a slightly lesser extent Portugal, Poland, Greece, Denmark, and Croatia. Secondly, as our analyses are expressed in relative values (share in the mass of tourism), it is easy to see great spatial and temporal inertia in the observed picture. However, a comparison of the same relations in absolute values indicates a permanent increase in the value of inflows in the analysed period. Thirdly, one can again see the formation of neighbouring arrangements of tourism flows: e.g. Poland, Czechia, Slovakia. Finally, it is worth mentioning Malta, which, because of its small population and area, cannot compare in observed tourist flows with, for example, neighbouring Italy, but flows to Malta have been clearly captured in our cartographic images.

Map 1: All tourist connections between researched countries (number of tourists in both directions), 2010, 2012, 2014, 2016, and 2018.

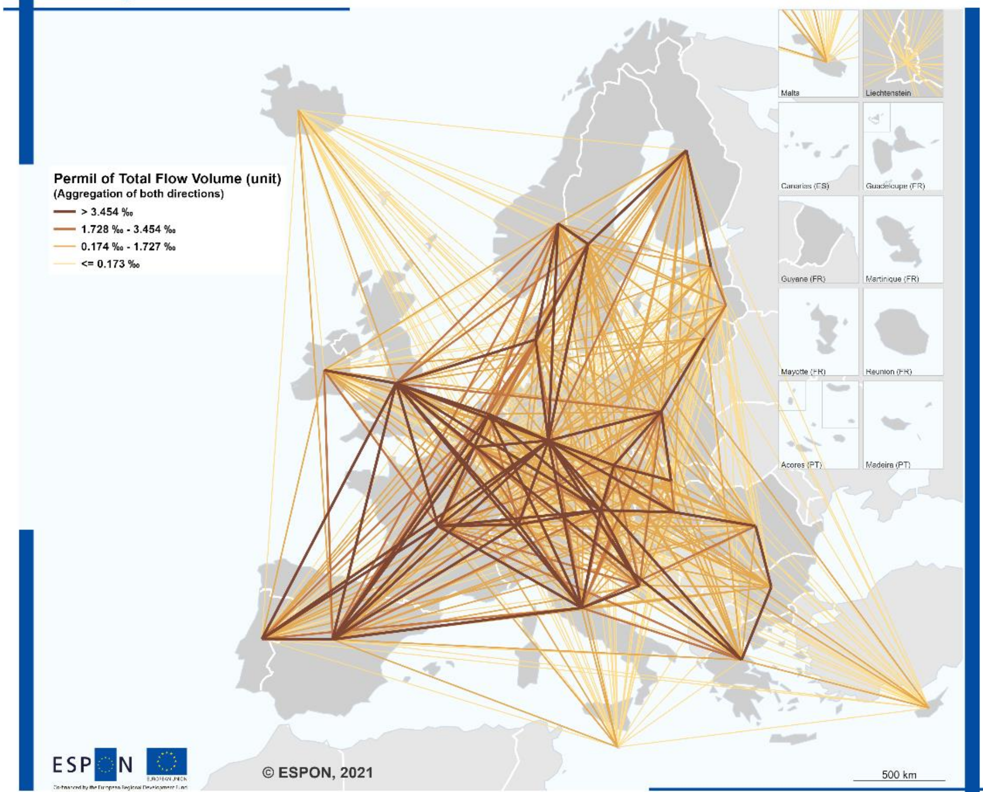


Interregional tourism, 2012

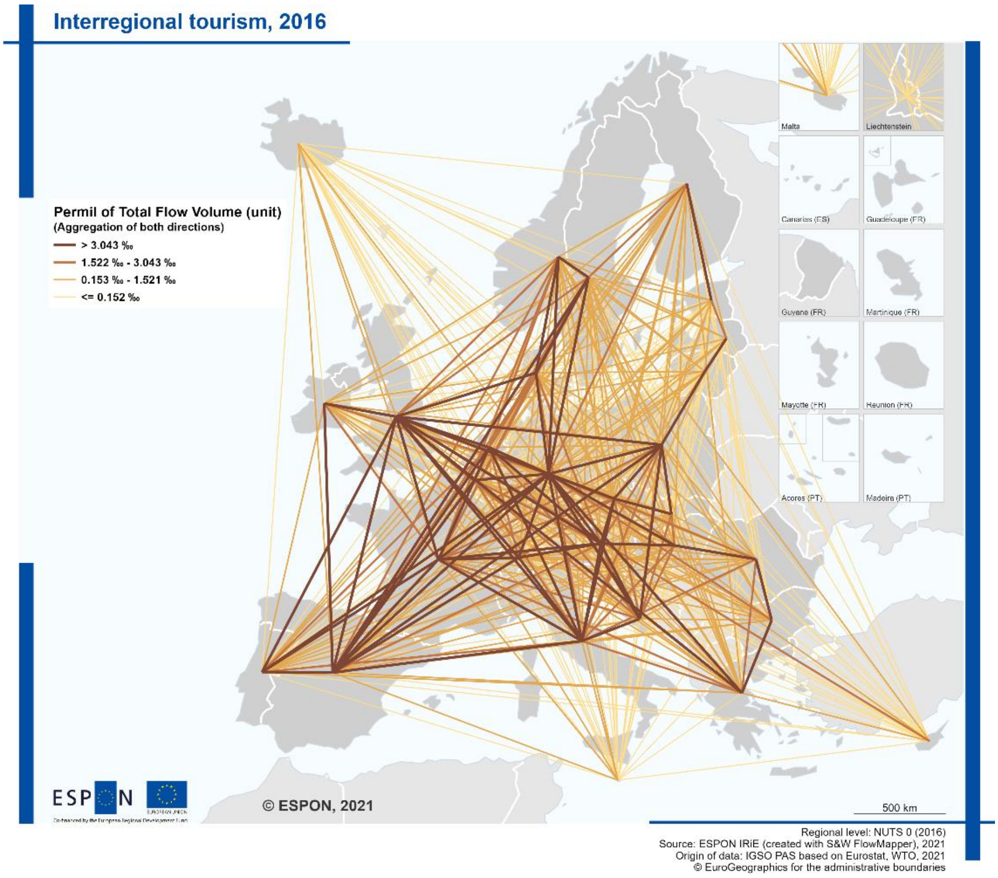


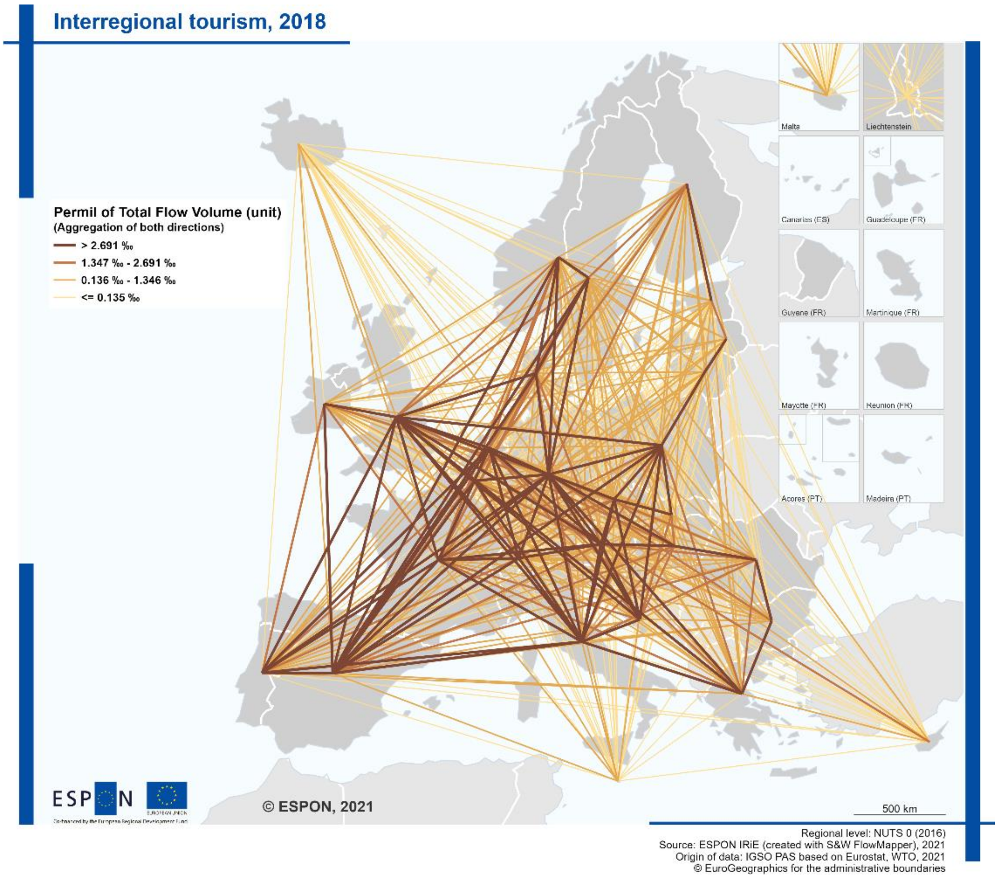
Regional level: NUTS 0 (2016)
Source: ESPON IRiE (created with S&W FlowMapper), 2021
Origin of data: IGSO PAS based on Eurostat, WTO, 2021
© EuroGeographics for the administrative boundaries

Interregional tourism, 2014



Regional level: NUTS 0 (2016)
Source: ESPON IRiE (created with S&W FlowMapper), 2021
Origin of data: IGSO PAS based on Eurostat, WTO, 2021
© EuroGeographics for the administrative boundaries

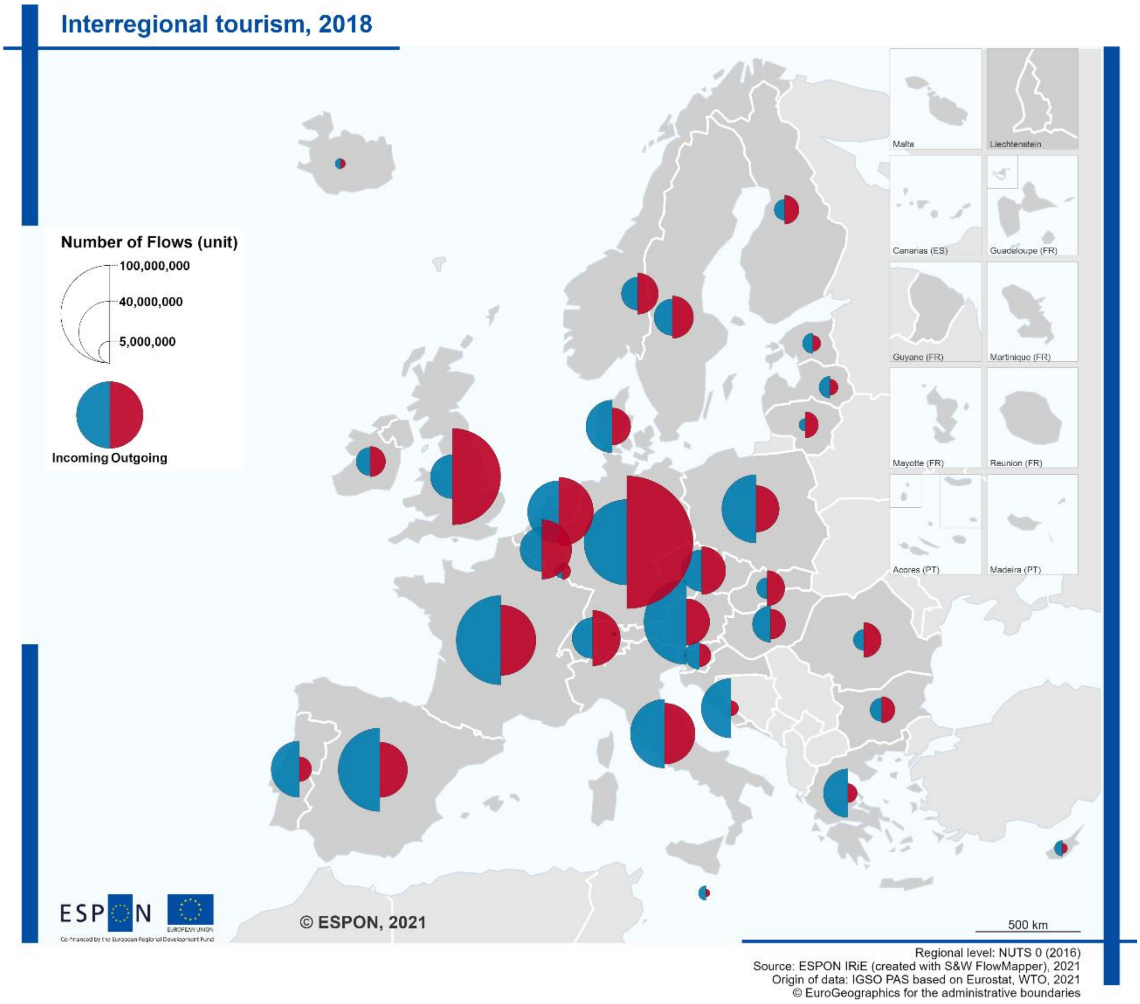




Sources: UNWTO & EUROSTAT databases / Authors' elaboration

It is also important not only to look at the sum of tourism flows between countries, but also to determine the balance of flows, which indicates the most popular countries among tourists (attractiveness for foreigners), and also indicates which countries residents prefer as travel destinations. In most cases, we observed a positive balance is in Mediterranean countries, while in Western and Northern European countries this balance took negative values. This is linked also to high levels of disposable income in households, which results in a propensity to travel abroad for tourism.

Map 2: Differences between the number of incoming and outgoing tourists in researched countries, 2018.



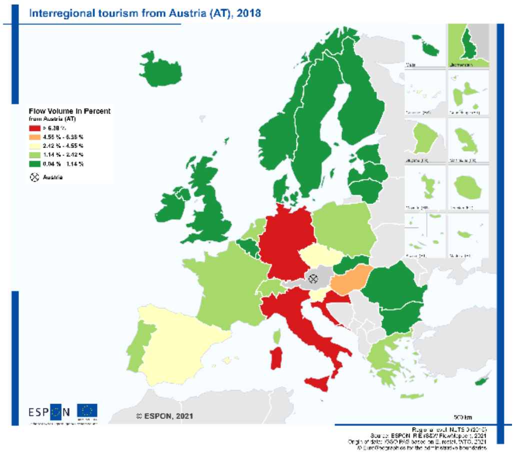
Sources: UNWTO & EUROSTAT databases / Authors' elaboration

The maps of Map 3 illustrate for all researched countries the importance of other countries as a tourist destination and origin. Countries in red are either the main tourist destination from the analysed country or the origin of the largest group of tourists. Countries in green, on the other hand, have insignificant mutual tourist traffic. Analysis of the observed systems indicates a huge combination of possible tourist connections, which, however, in the final system can be reduced to one universal model: global connections to the largest tourist dominators of the researched area are slightly corrected by the formation neighbouring regional systems.

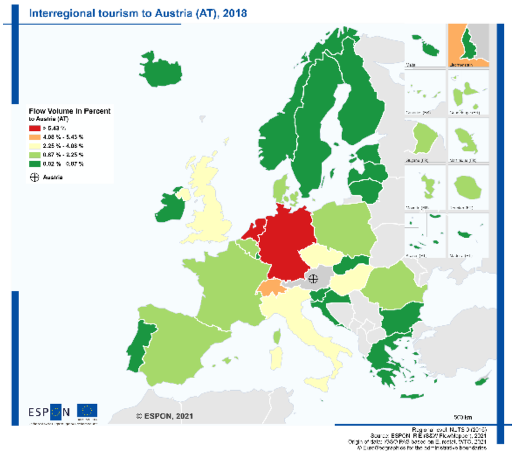
Map 3: Interregional tourism flows from and to all researched countries, 2018.

Tourism flows from researched countries **Tourism flows to researched countries**

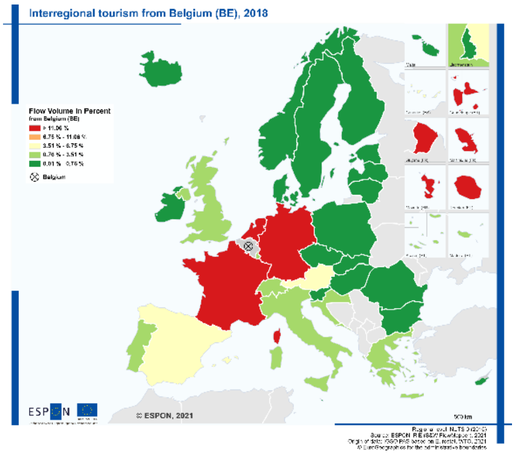
Interregional tourism from Austria (AT), 2018



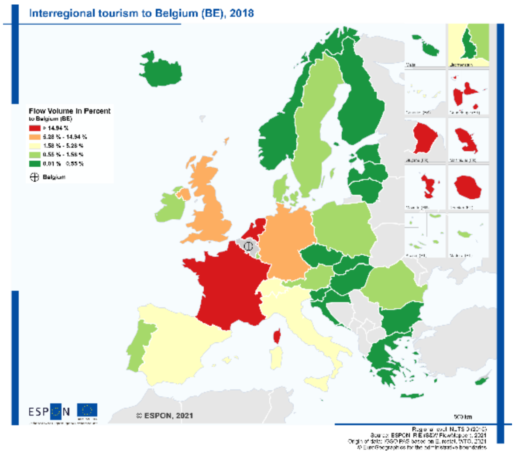
Interregional tourism to Austria (AT), 2018



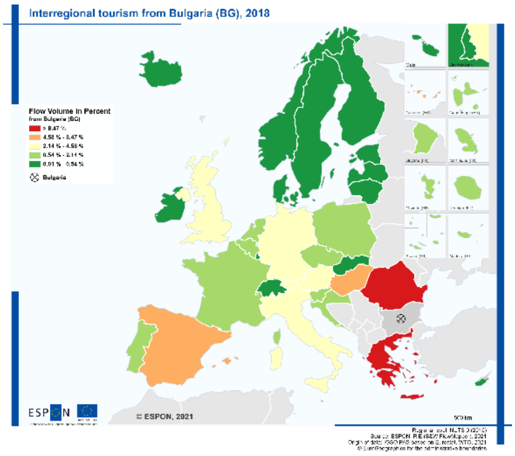
Interregional tourism from Belgium (BE), 2018



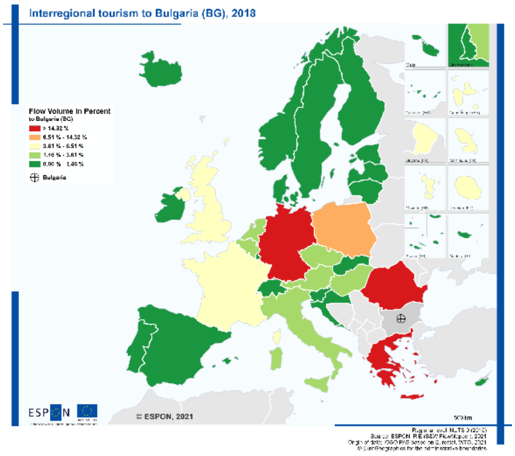
Interregional tourism to Belgium (BE), 2018



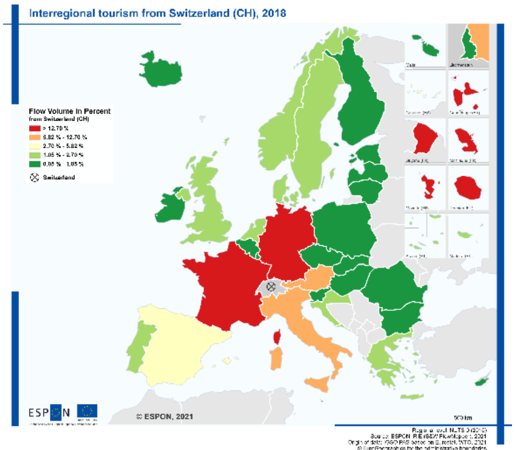
Interregional tourism from Bulgaria (BG), 2018



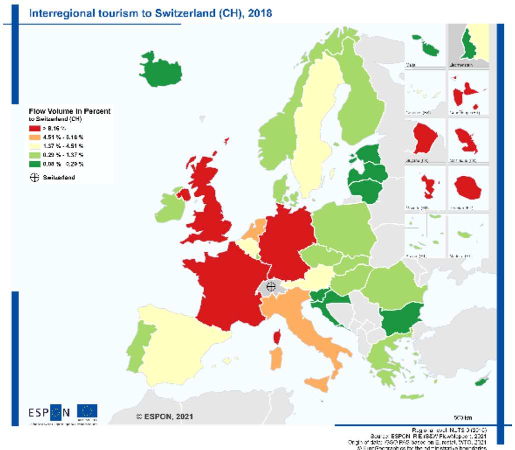
Interregional tourism to Bulgaria (BG), 2018



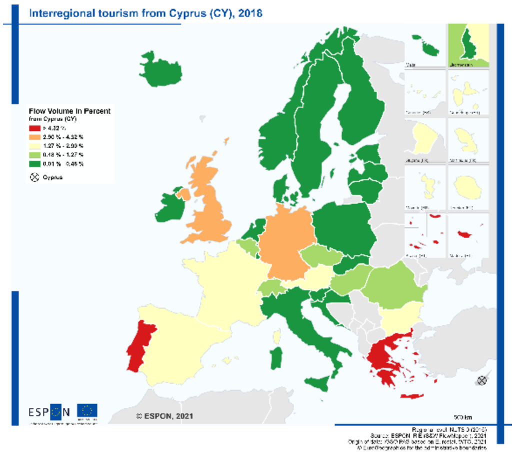
Interregional tourism from Switzerland (CH), 2018



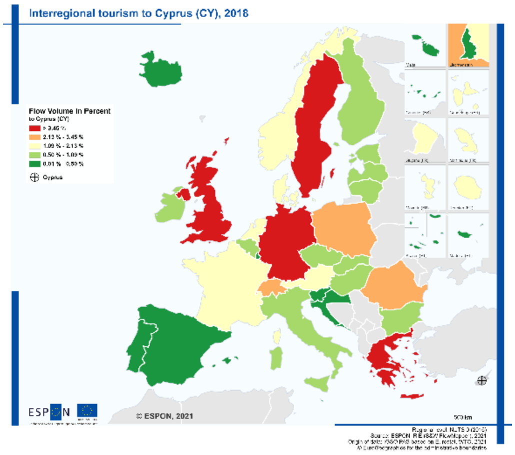
Interregional tourism to Switzerland (CH), 2018



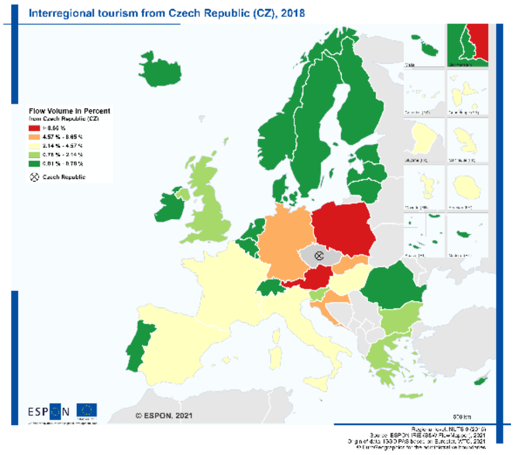
Interregional tourism from Cyprus (CY), 2018



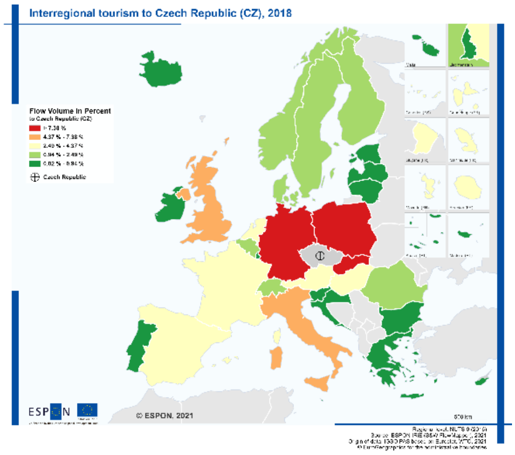
Interregional tourism to Cyprus (CY), 2018



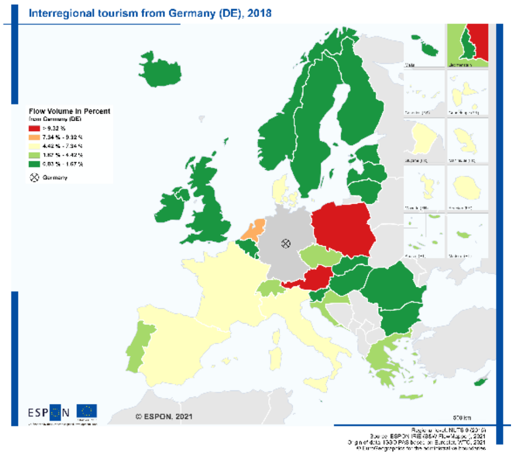
Interregional tourism from Czech Republic (CZ), 2018



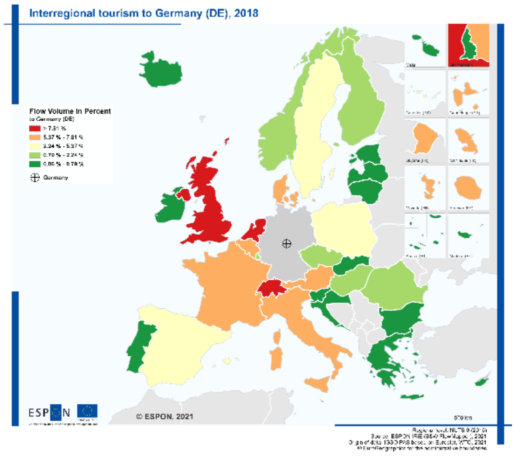
Interregional tourism to Czech Republic (CZ), 2018



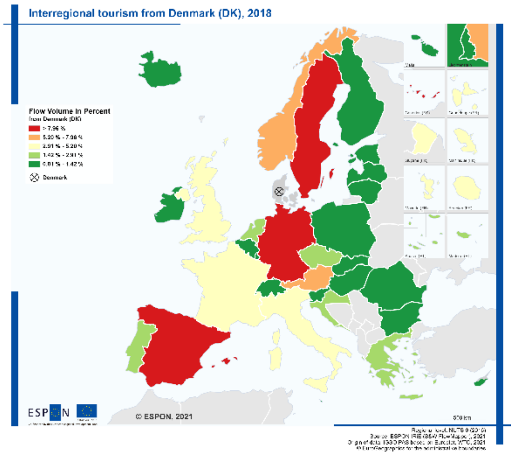
Interregional tourism from Germany (DE), 2018



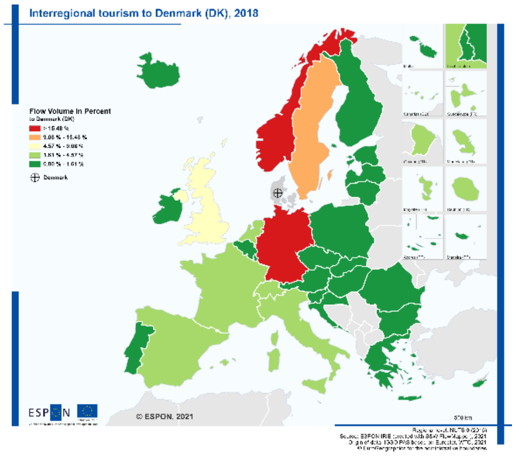
Interregional tourism to Germany (DE), 2018



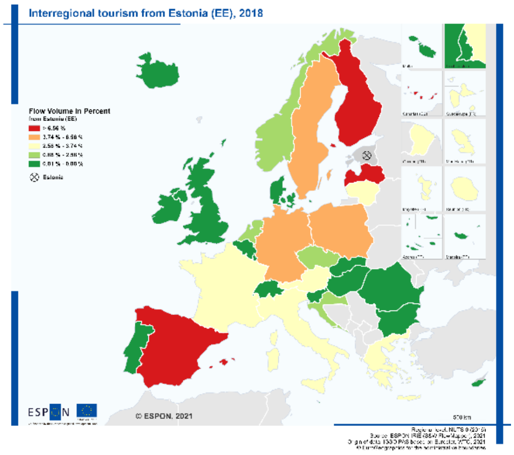
Interregional tourism from Denmark (DK), 2018



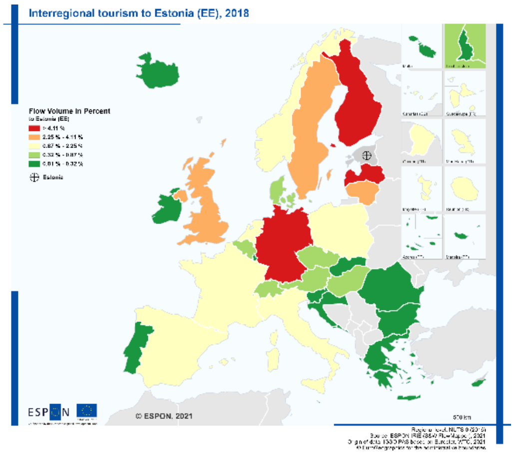
Interregional tourism to Denmark (DK), 2018



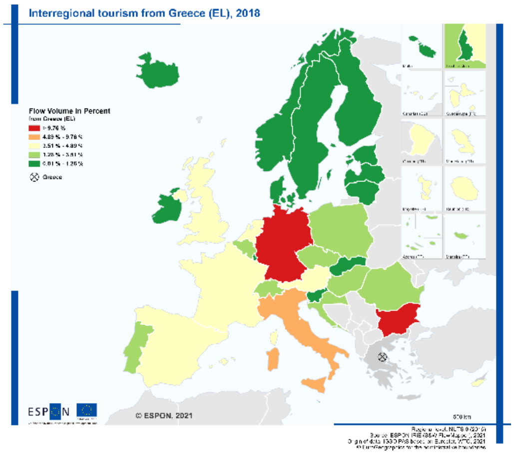
Interregional tourism from Estonia (EE), 2018



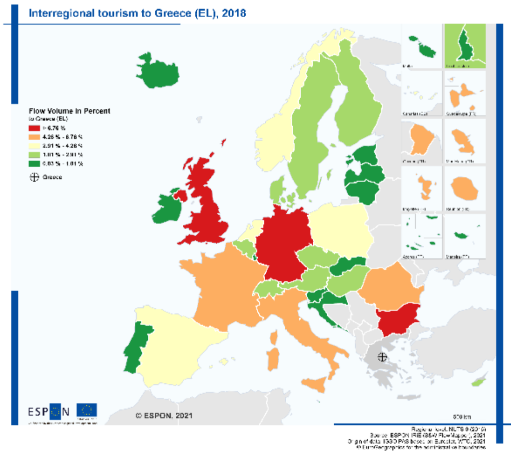
Interregional tourism to Estonia (EE), 2018



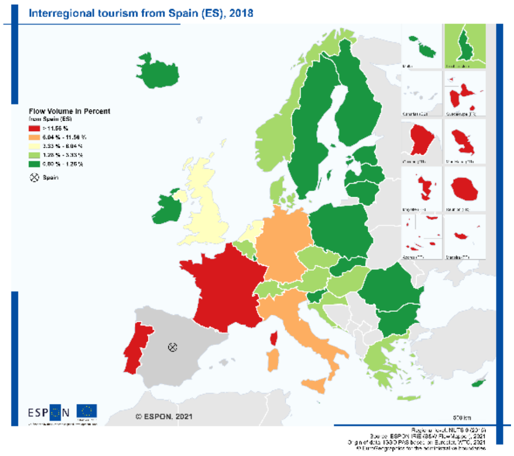
Interregional tourism from Greece (EL), 2018



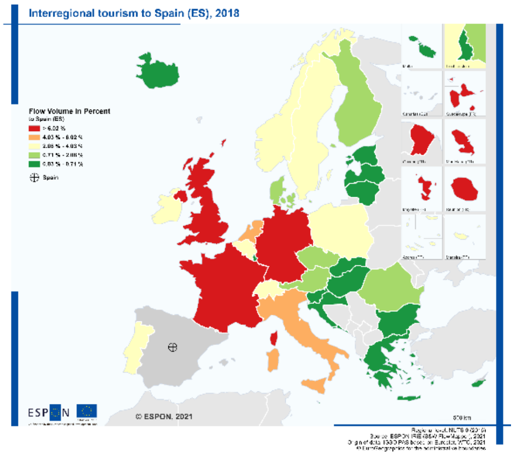
Interregional tourism to Greece (EL), 2018



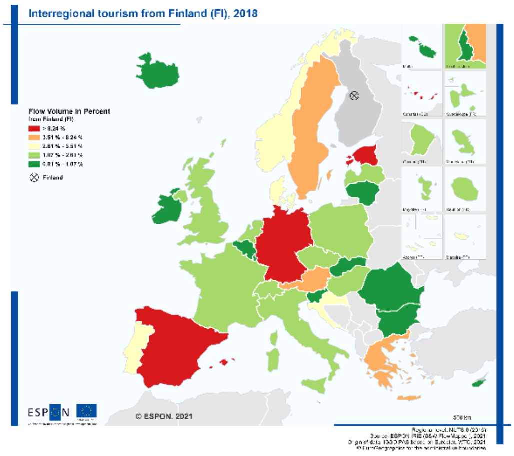
Interregional tourism from Spain (ES), 2018



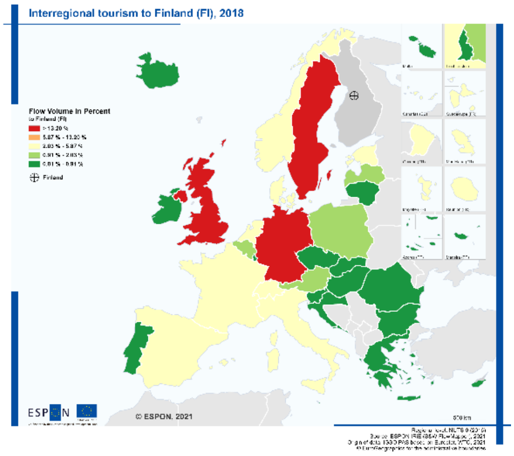
Interregional tourism to Spain (ES), 2018

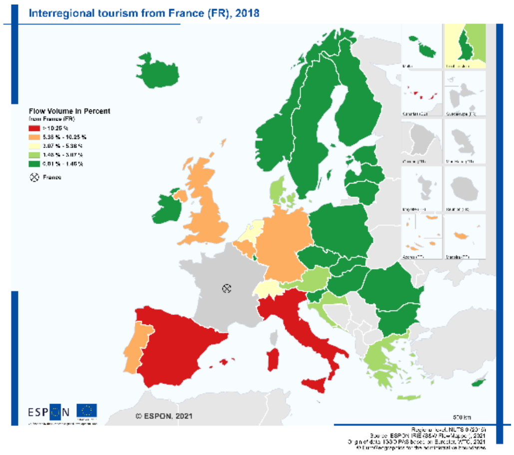


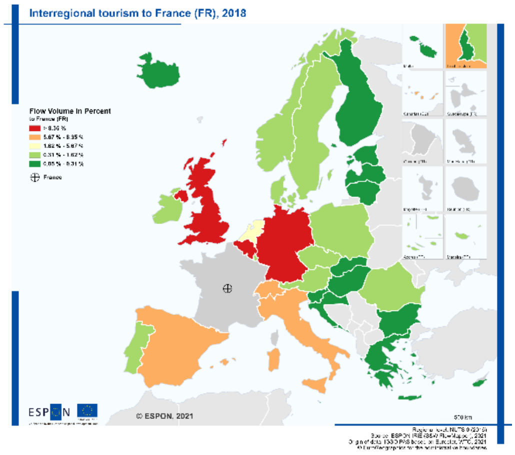
Interregional tourism from Finland (FI), 2018

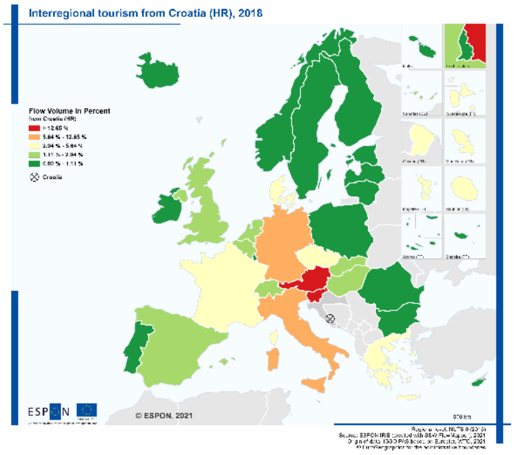


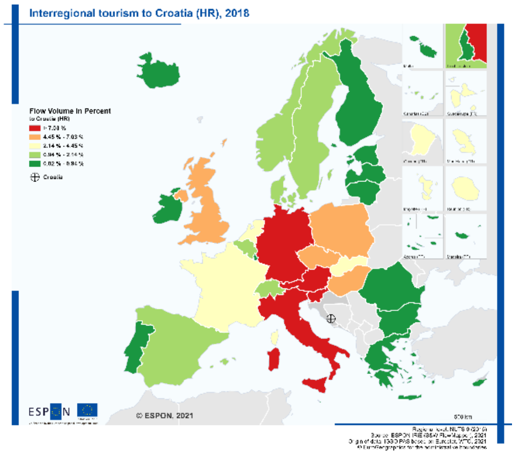
Interregional tourism to Finland (FI), 2018

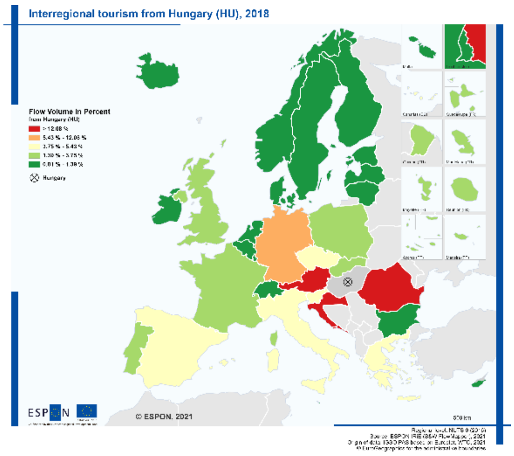


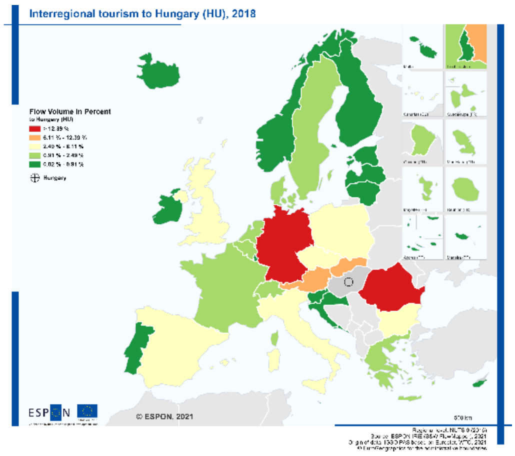

Interregional tourism from France (FR), 2018
Flow Volume in Percent
from France (FR)
> 10.25 %
5.38 % - 10.25 %
3.07 % - 5.38 %
1.45 % - 3.07 %
0.01 % - 1.45 %
France
ESPON
© ESPON, 2021

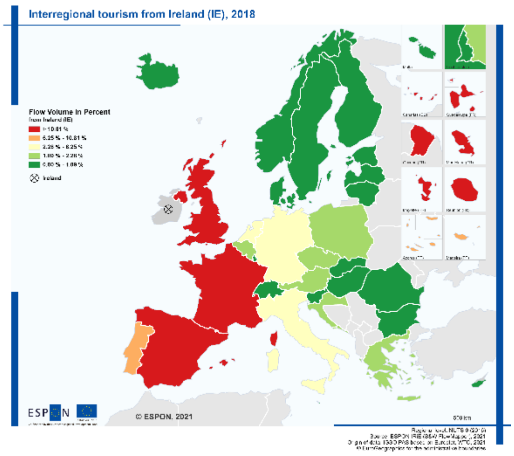

Interregional tourism to France (FR), 2018
Flow Volume in Percent
to France (FR)
> 8.35 %
5.67 % - 8.35 %
1.82 % - 5.67 %
0.31 % - 1.82 %
0.05 % - 0.31 %
France
ESPON
© ESPON, 2021

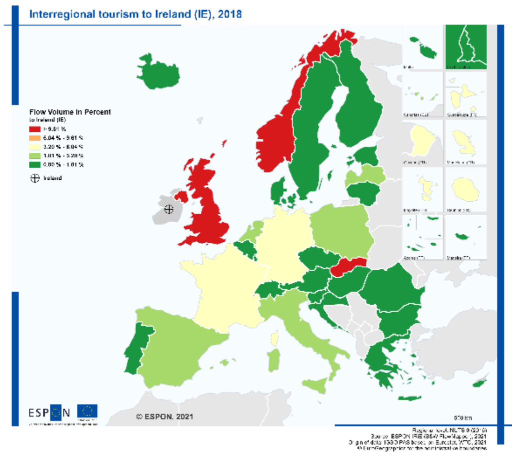

Interregional tourism from Croatia (HR), 2018
Flow Volume in Percent
from Croatia (HR)
> 12.65 %
5.84 % - 12.65 %
2.94 % - 5.84 %
1.11 % - 2.94 %
0.00 % - 1.11 %
Croatia
ESPON
© ESPON, 2021


Interregional tourism to Croatia (HR), 2018
Flow Volume in Percent
to Croatia (HR)
> 7.03 %
4.45 % - 7.03 %
2.14 % - 4.45 %
0.84 % - 2.14 %
0.02 % - 0.84 %
Croatia
ESPON
© ESPON, 2021

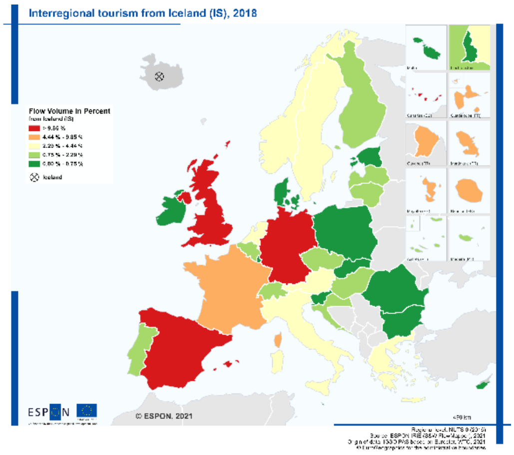

Interregional tourism from Hungary (HU), 2018
Flow Volume in Percent
from Hungary (HU)
> 12.08 %
5.43 % - 12.08 %
3.75 % - 5.43 %
1.30 % - 3.75 %
0.01 % - 1.30 %
Hungary
ESPON
© ESPON, 2021


Interregional tourism to Hungary (HU), 2018
Flow Volume in Percent
to Hungary (HU)
> 12.39 %
6.11 % - 12.39 %
2.49 % - 6.11 %
0.81 % - 2.49 %
0.02 % - 0.81 %
Hungary
ESPON
© ESPON, 2021

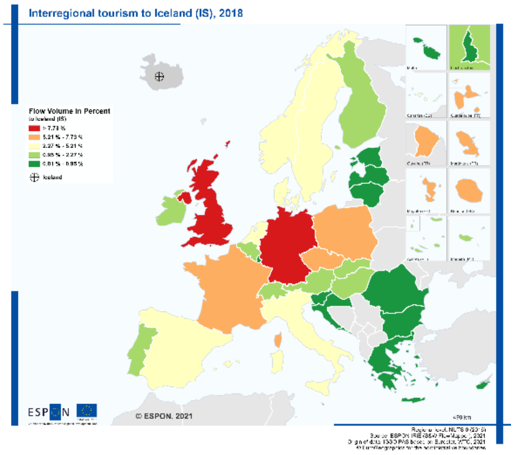

Interregional tourism from Ireland (IE), 2018
Flow Volume in Percent
from Ireland (IE)
> 10.81 %
6.25 % - 10.81 %
2.28 % - 6.25 %
1.09 % - 2.28 %
0.00 % - 1.09 %
Ireland
ESPON
© ESPON, 2021


Interregional tourism to Ireland (IE), 2018
Flow Volume in Percent
to Ireland (IE)
> 9.61 %
6.04 % - 9.61 %
3.20 % - 6.04 %
1.01 % - 3.20 %
0.00 % - 1.01 %
Ireland
ESPON
© ESPON, 2021

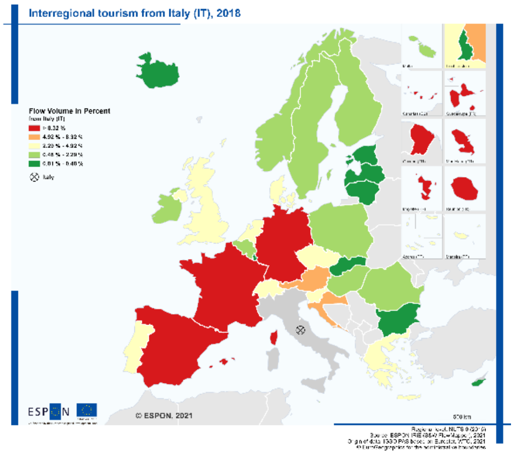
Interregional tourism from Iceland (IS), 2018



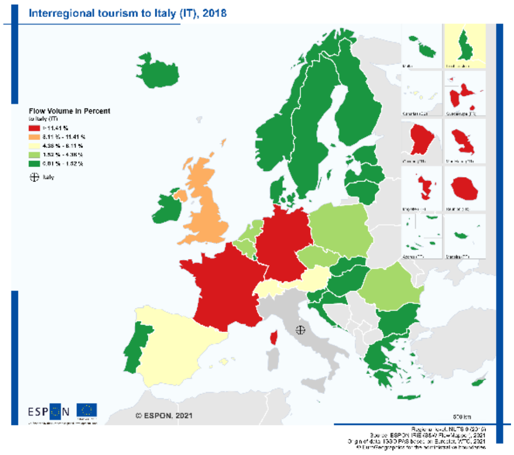
Interregional tourism to Iceland (IS), 2018



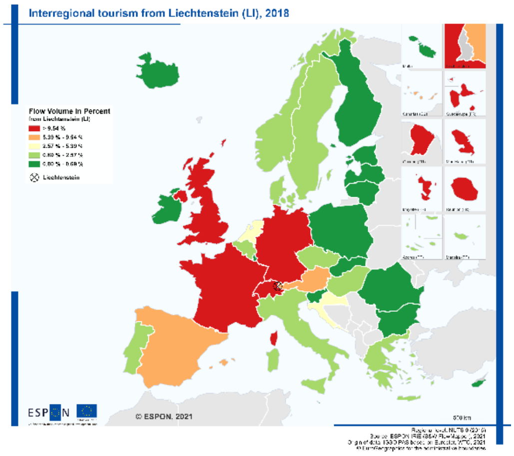
Interregional tourism from Italy (IT), 2018



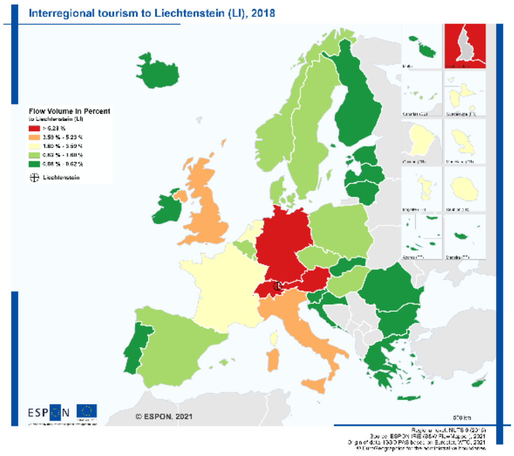
Interregional tourism to Italy (IT), 2018



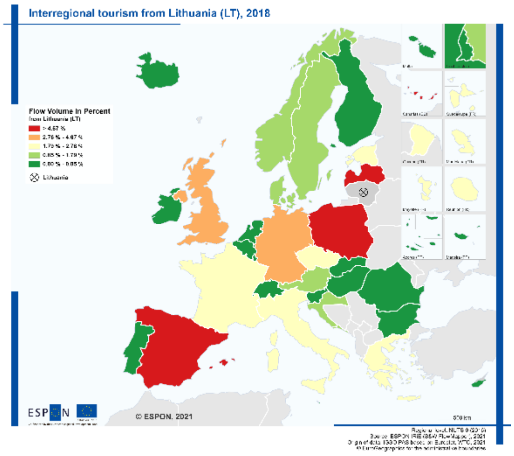
Interregional tourism from Liechtenstein (LI), 2018



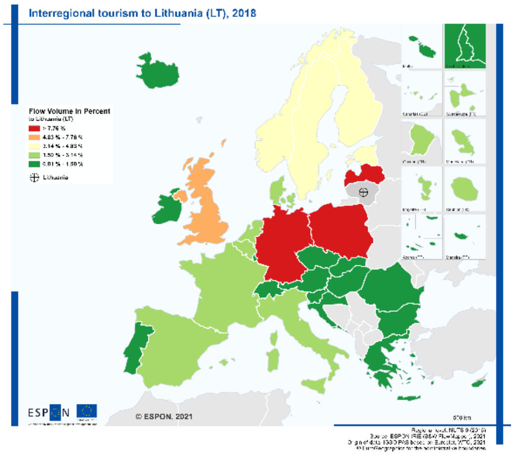
Interregional tourism to Liechtenstein (LI), 2018



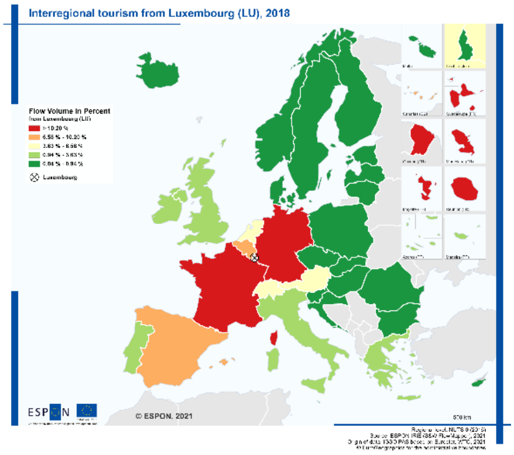
Interregional tourism from Lithuania (LT), 2018



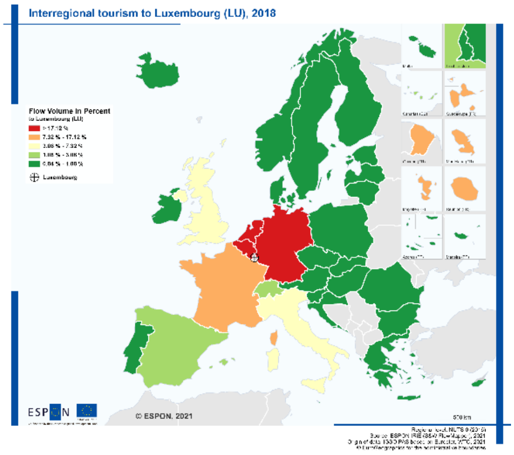
Interregional tourism to Lithuania (LT), 2018



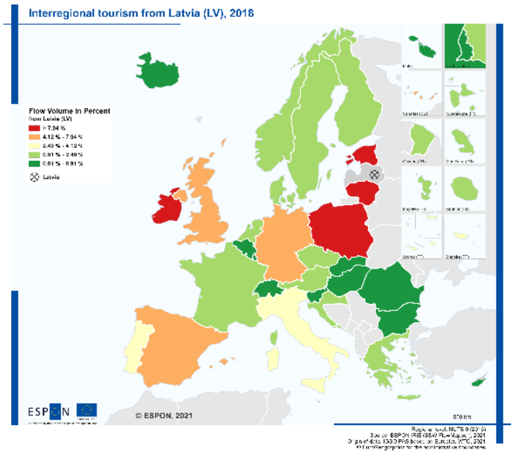
Interregional tourism from Luxembourg (LU), 2018



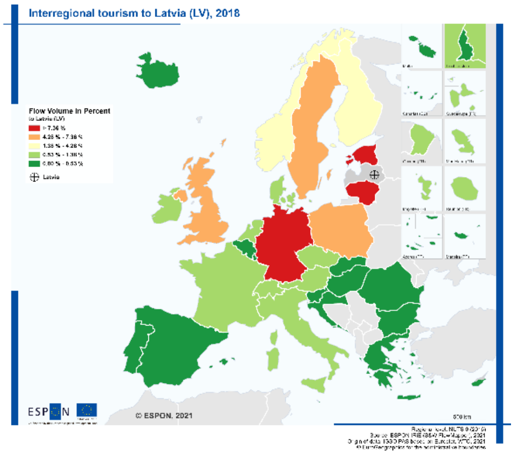
Interregional tourism to Luxembourg (LU), 2018



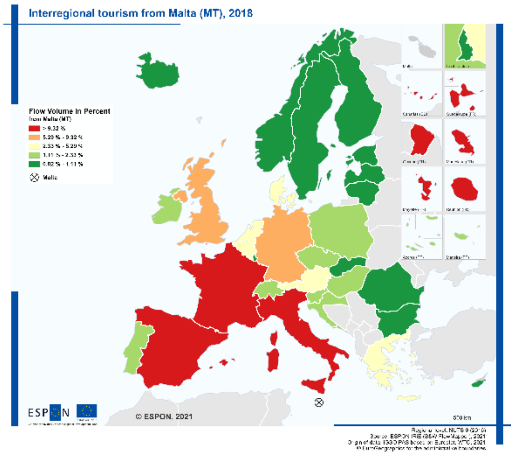
Interregional tourism from Latvia (LV), 2018



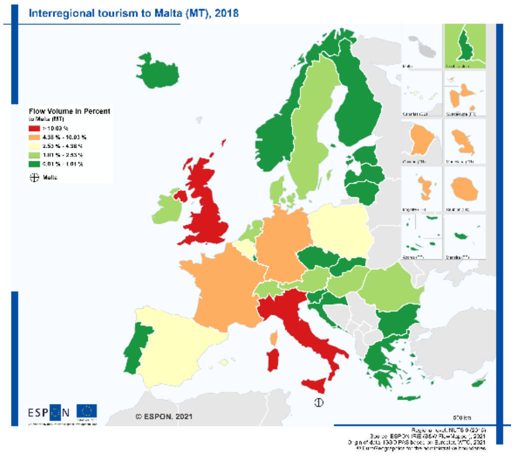
Interregional tourism to Latvia (LV), 2018



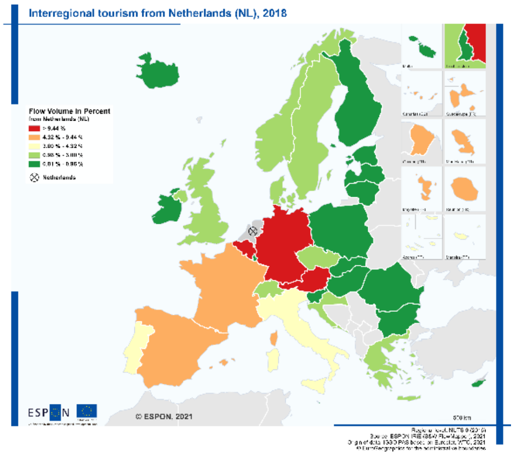
Interregional tourism from Malta (MT), 2018



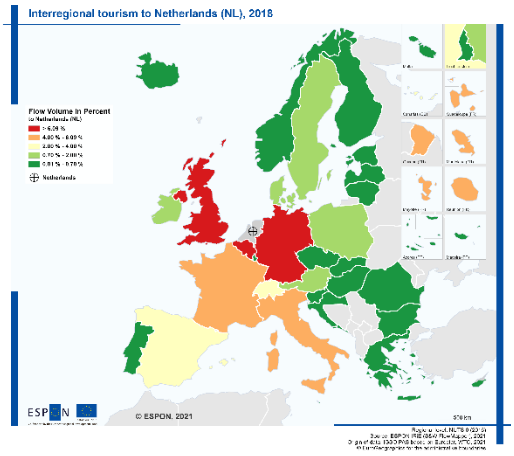
Interregional tourism to Malta (MT), 2018



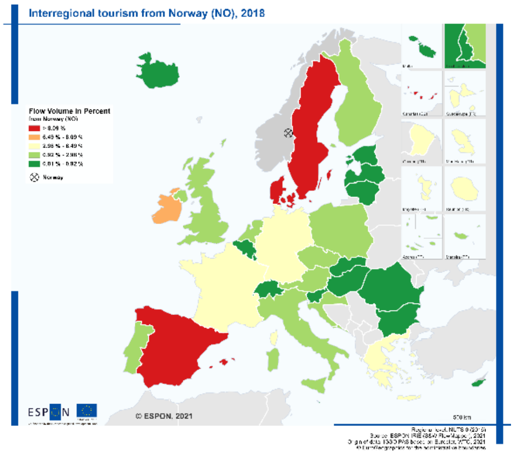
Interregional tourism from Netherlands (NL), 2018



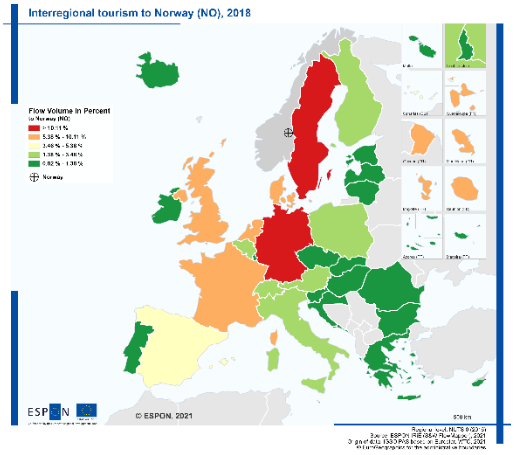
Interregional tourism to Netherlands (NL), 2018



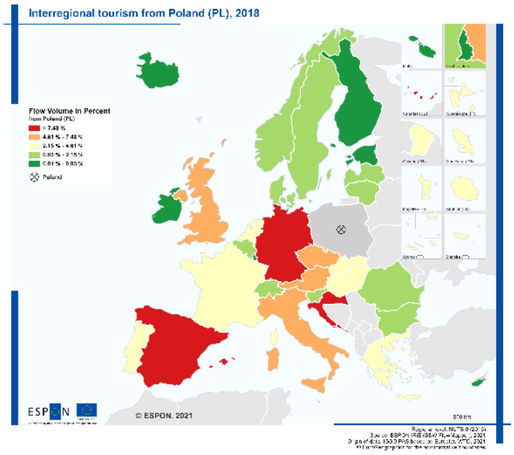
Interregional tourism from Norway (NO), 2018



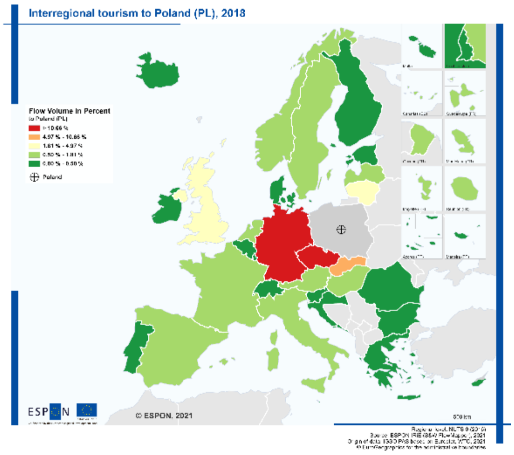
Interregional tourism to Norway (NO), 2018



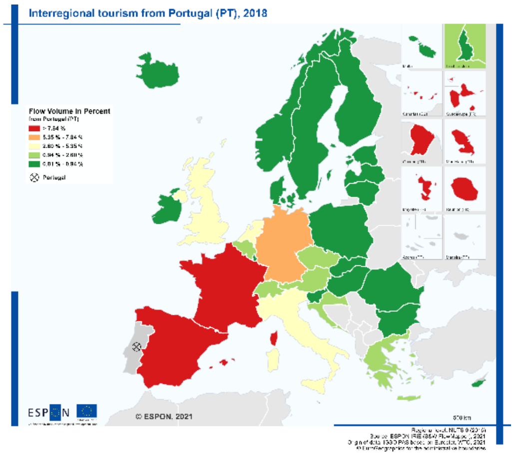
Interregional tourism from Poland (PL), 2018



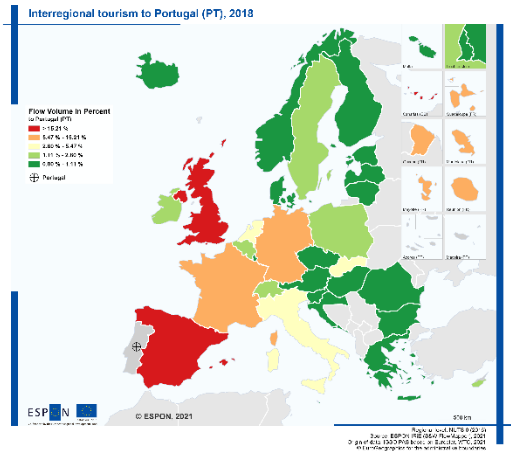
Interregional tourism to Poland (PL), 2018



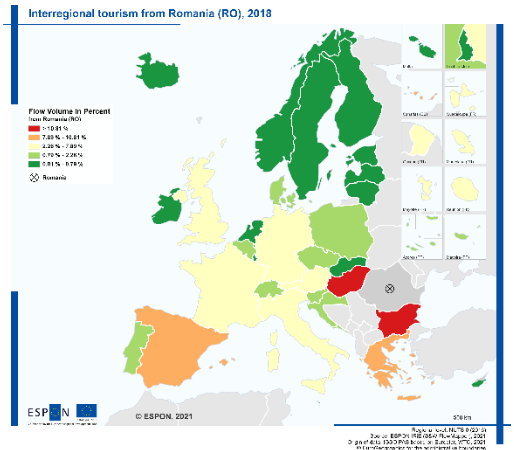
Interregional tourism from Portugal (PT), 2018



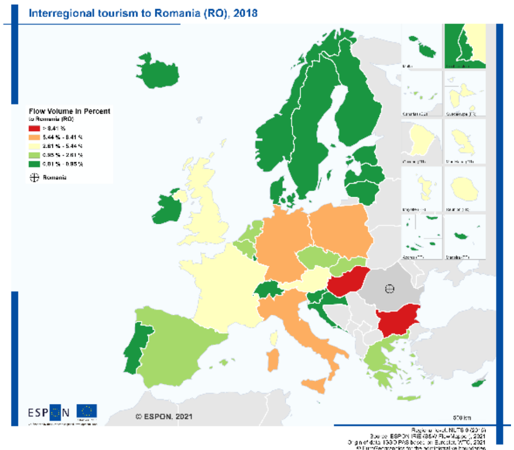
Interregional tourism to Portugal (PT), 2018



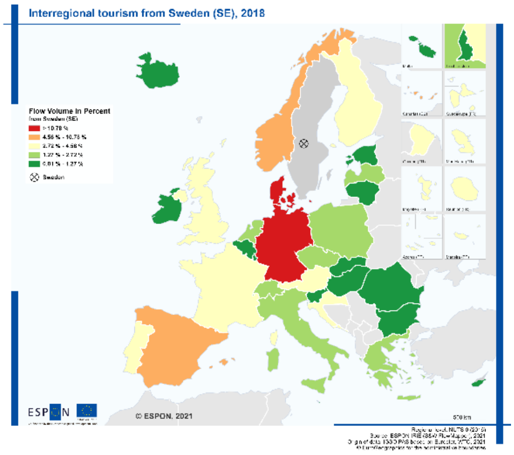
Interregional tourism from Romania (RO), 2018



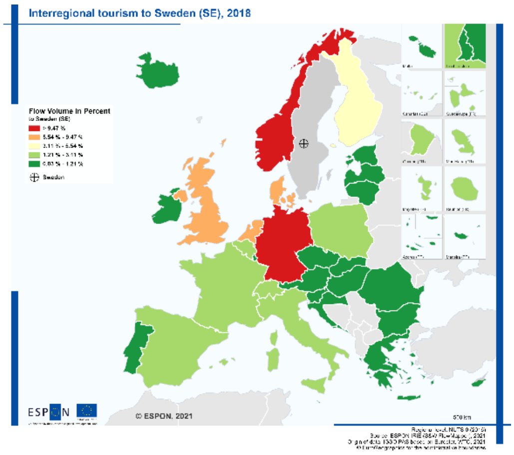
Interregional tourism to Romania (RO), 2018



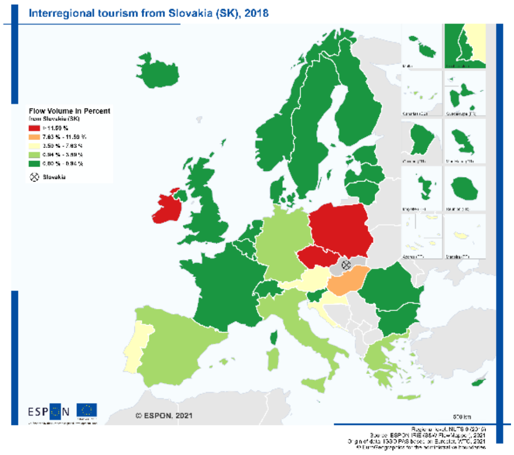
Interregional tourism from Sweden (SE), 2018



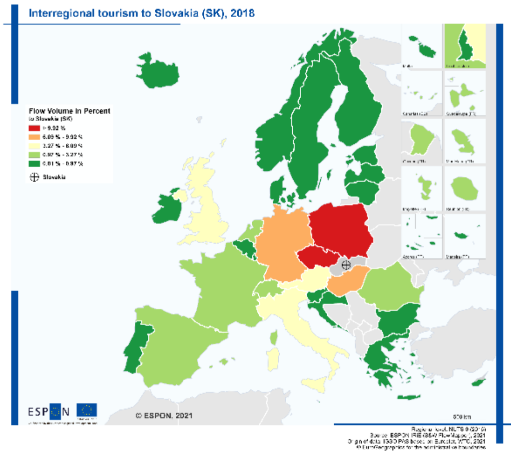
Interregional tourism to Sweden (SE), 2018



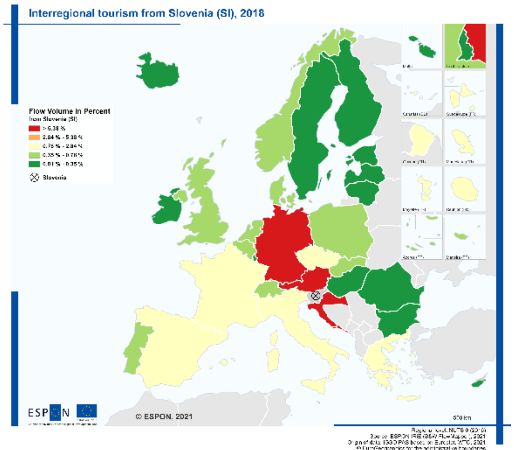
Interregional tourism from Slovakia (SK), 2018



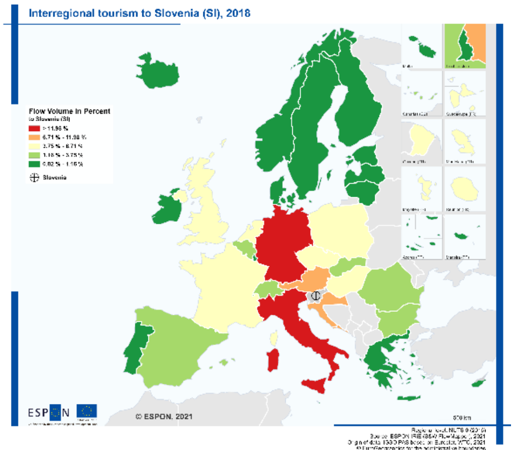
Interregional tourism to Slovakia (SK), 2018



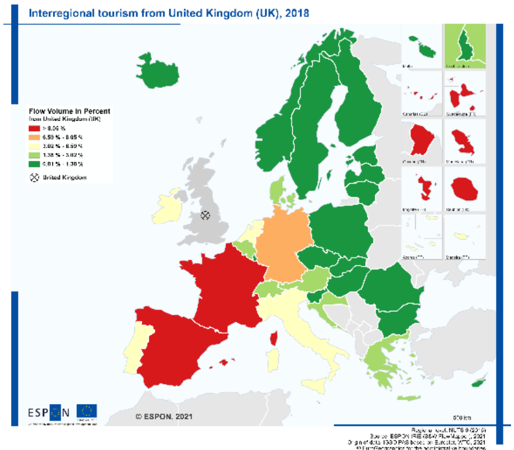
Interregional tourism from Slovenia (SI), 2018



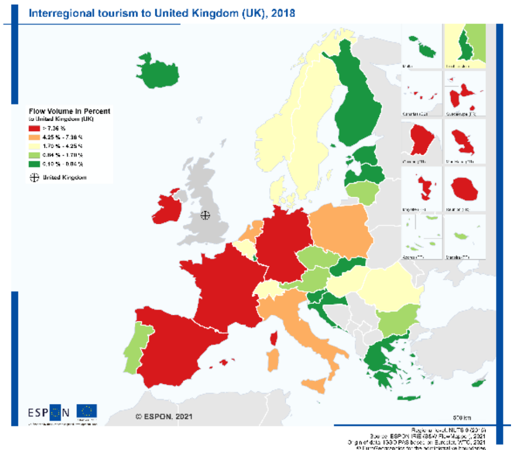
Interregional tourism to Slovenia (SI), 2018



Interregional tourism from United Kingdom (UK), 2018



Interregional tourism to United Kingdom (UK), 2018



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

3.2 Region-to-region flows

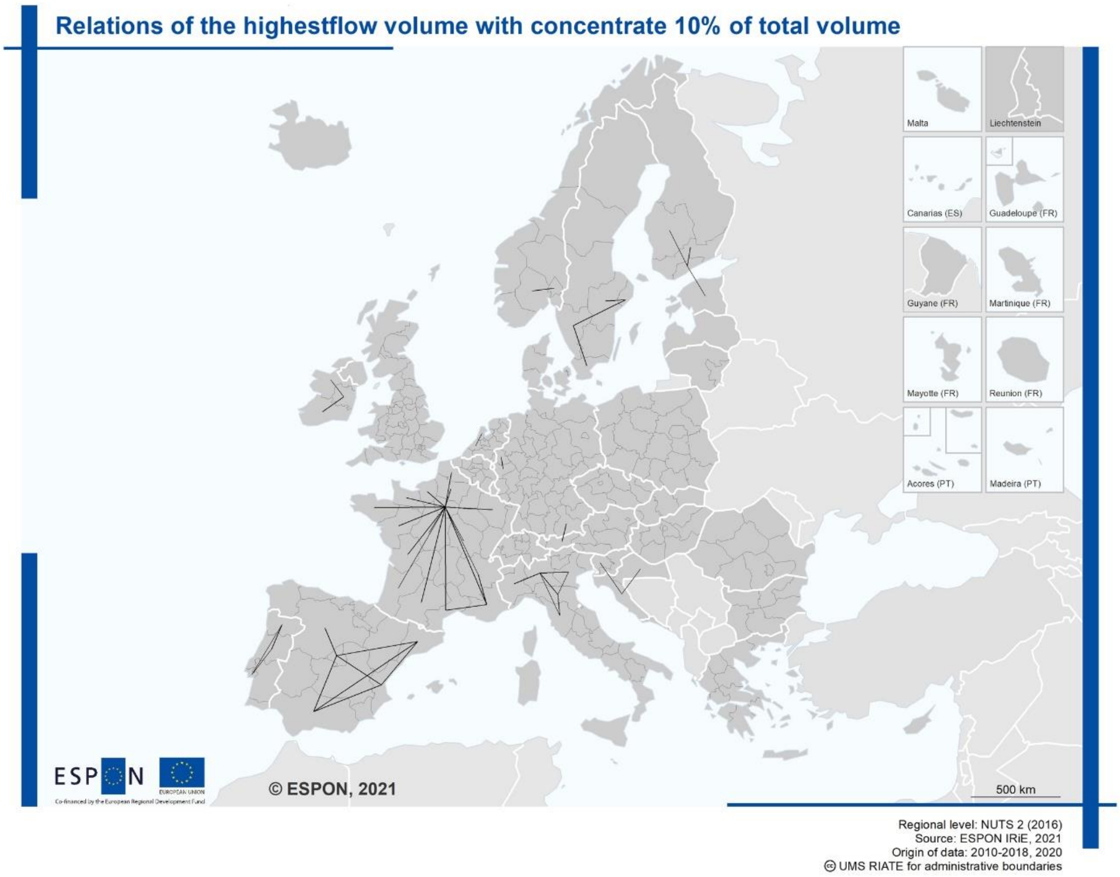
In this section, we review the main results from our new and broad dataset. Our analysis focuses on the main aspects, leaving several potential analyses for future research.

3.2.1 Description of results

Within the ESPON IRiE project we distinguish 5 dimensions of flow analysis: **intensity (size)**, **connectivity**, **balance** (comparison of outflow and inflow), **concentration** (dispersion of senders and receivers), and **distance** (how far tourists travel – the extent to which distance determines flow size). We have already mentioned, above, the 5 basic groups of indicators. In addition, we have researched changes over time, analyzing them for each indicator separately. The following maps illustrate the most important issues of tourism flow within the entire researched period 2010-2018.

Our analysis showed that the largest absolute values of the interregional tourism flows, understood as the sum of both directions, occurred within the largest countries of the Mediterranean region (France, Spain, and Italy). To lesser extent, large numbers of interregional flows were recorded in other Mediterranean countries (Portugal, Croatia), Scandinavia (Norway, Sweden, Finland), Ireland, Germany, and the Netherlands. Unlike with France, Spain, and Italy, however, the largest of these other interregional flows occurred between only two or three NUTS 2 regions within the respective countries. For the international interregional flows with the highest flow volume, concentrating 10% of total tourist flow volume, only relations between the regions pairs Slovenia-Croatia, Germany-Austria, and Finland-Estonia can be distinguished (Map 4).

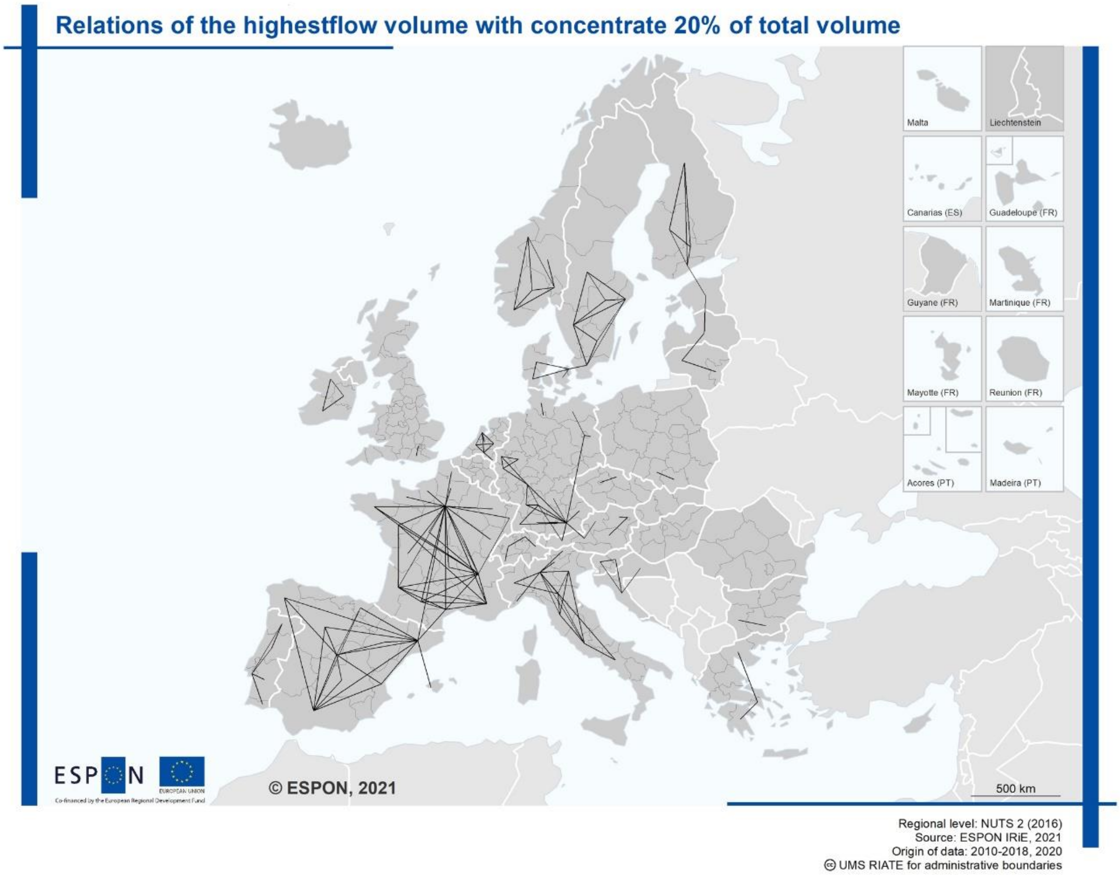
Map 4: Relations of the highest flow volume, concentrating 10% of total volume



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

Within relations of the highest flow volume -concentrating 20% of total tourist flow volume- (Map 5), again the largest interregional movement of tourists is domestic. Except for aforementioned countries where the number of interregional flows increased, domestic relations were observed also in Greece, Bulgaria, Switzerland, Austria, Czechia, Poland, Denmark, Lithuania, and the UK. For international interregional tourist flows, there were additional connections between the regions of Spain-France, Denmark-Sweden, Estonia-Latvia, and Latvia-Lithuania. In most cases, these international interregional flows existed between two neighbouring NUTS 2 regions.

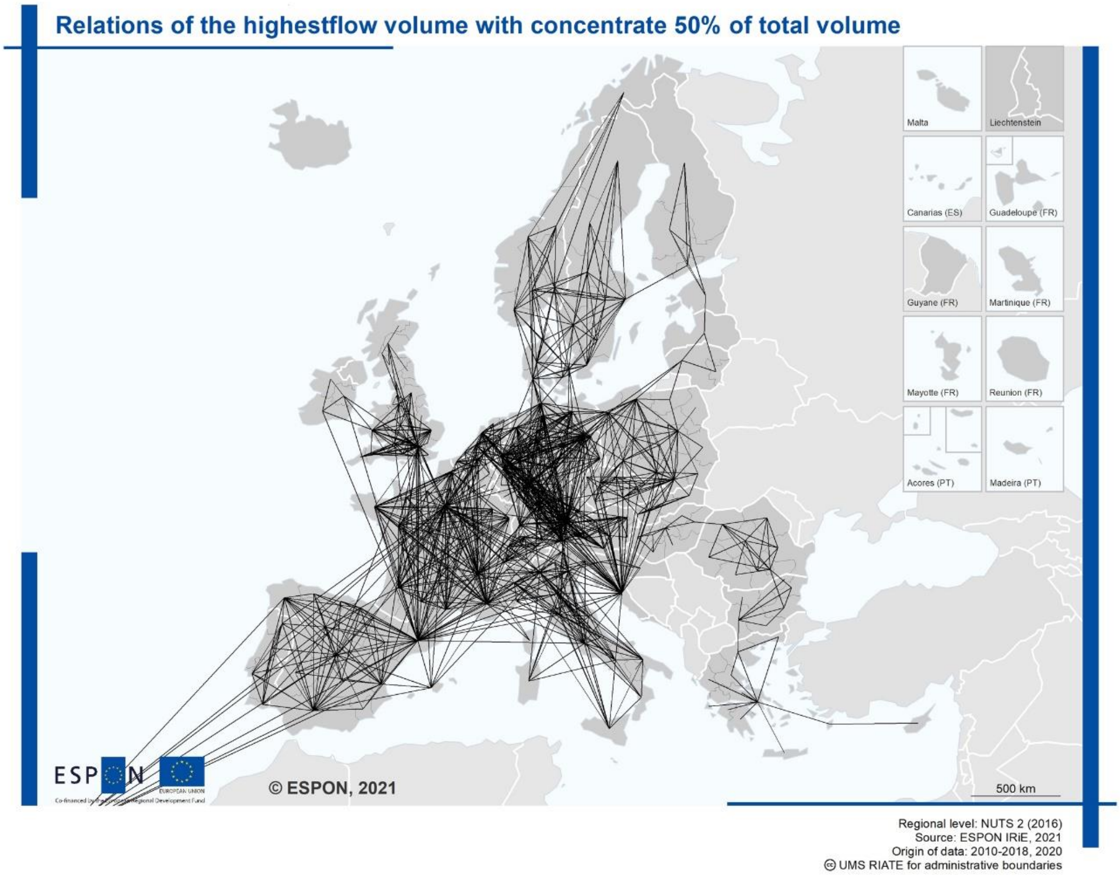
Map 5: Relations of the highest flow volume, which concentrate 20% of total volume



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

For relations that cover half of the entire tourist flow volume (Map 6), the concentration of tourism relations between the different NUTS 2 regions occurred within several large European countries, headed by Germany, France, the UK, Spain, Italy, and Poland. On a smaller scale, these regional relations within the country were observed within the regions of Scandinavia and southeastern Europe, but these concentrations dispersed to other regions, generally in neighbouring countries. For international interregional tourism flows, the north-south axis generally dominated, with tourist flows between regions of Central Europe and Adriatic Croatia, flows between the regions of Finland, Estonia, Latvia, and Lithuania, as well as flows between Germany and Austria, France and Spain, France and Italy, and regions of the UK with NUTS 2 regions in France and Spain.

Map 6: Relations of highest flow volume, which concentrate 50% of total volume

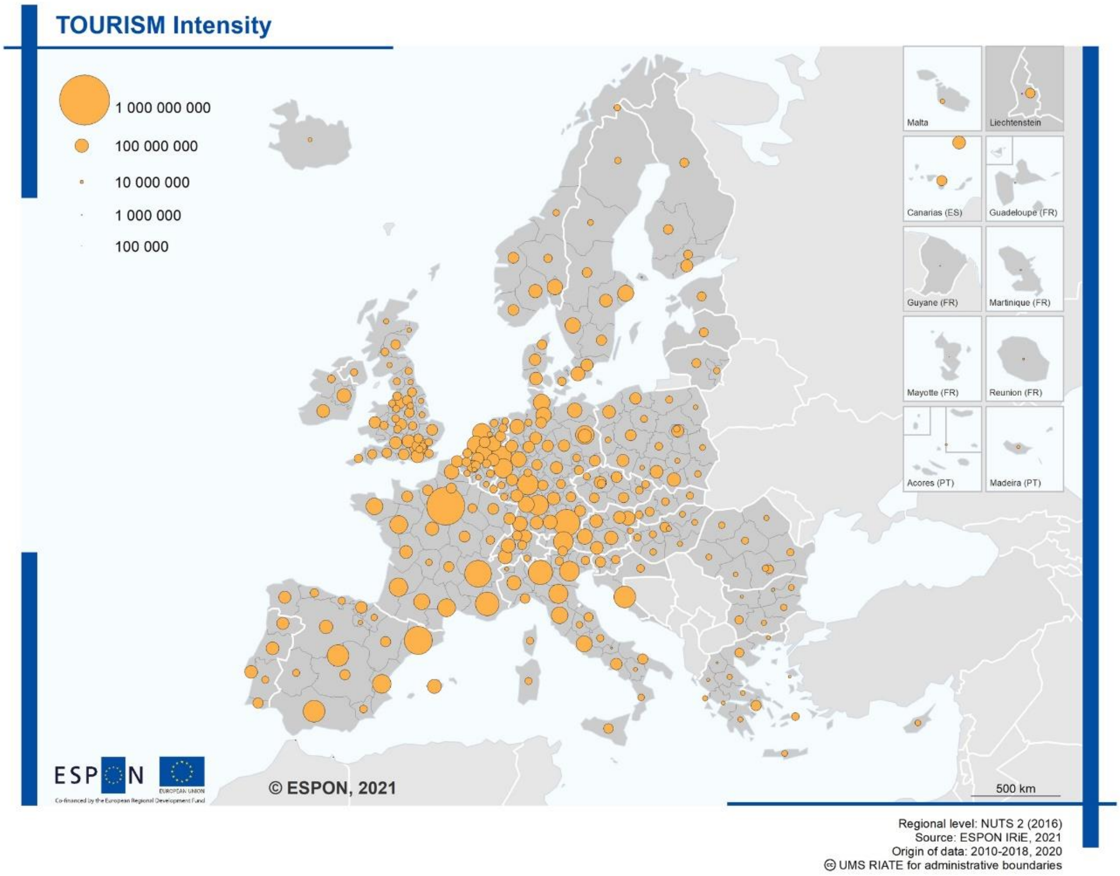


Sources: UNWTO & EUROSTAT databases / Authors' elaboration

3.2.1.1 Intensity

The intensity dimension represents the size of tourist flows in individual regions: it is the sum of the volumes for all relations of a given region. The first map (Map 7) shows NUTS 2 regions by number of flows.

Map 7: Intensity of tourist flows

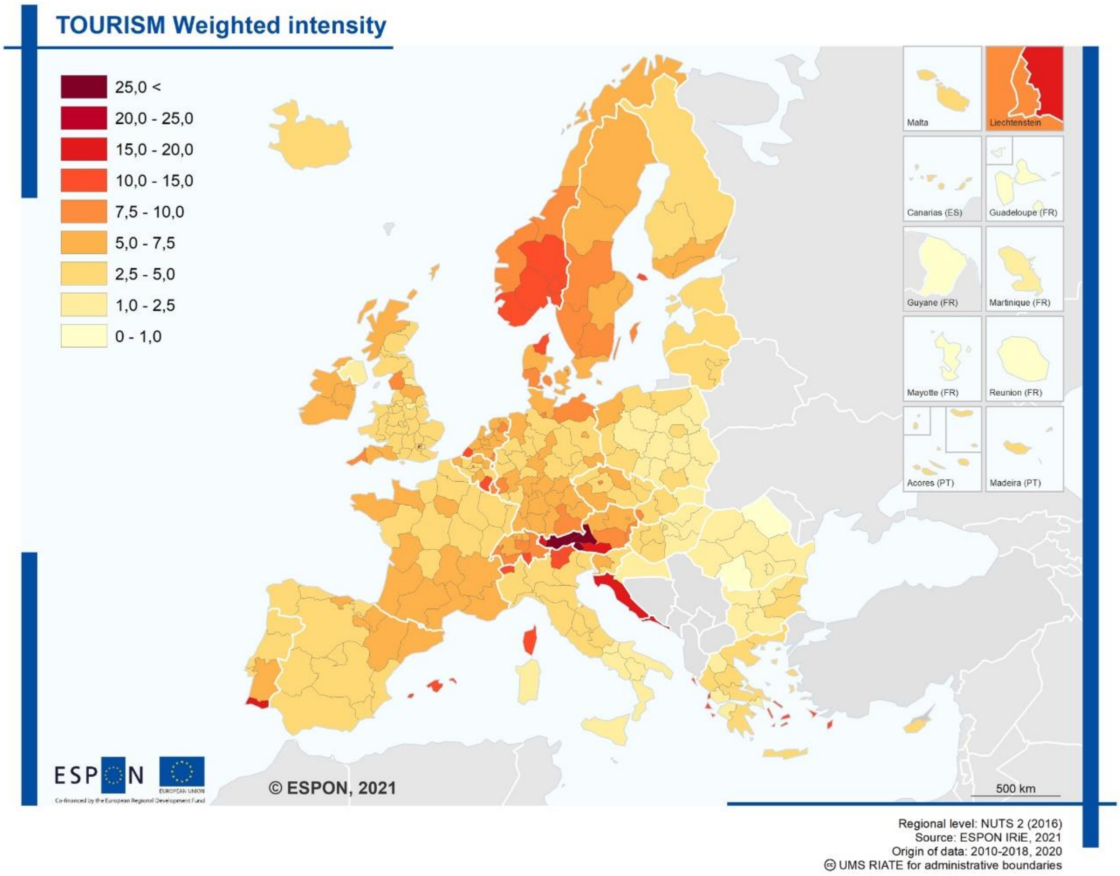


Sources: UNWTO & EUROSTAT databases / Authors' elaboration

One might find here a correlation between the tourism-flow intensity and a discontinuous corridor of urbanization spreading over Western and Central Europe, known as a Blue Banana (Brunet, 1989) — an area which stretches approximately from northwestern England through the English Midlands across Greater London to the European metropolis of Lille, the Benelux states and along the German Rhineland, southern Germany, Alsace-Moselle in France in the west and Switzerland (Basel and Zürich) to northern Italy (Milan and Turin) in the south. Behind this area, the regions of capital cities, regions situated near seashores, and regions with high mountains could also be described having great tourism intensity.

The following map (Map 8) represents the intensity of tourism flows in individual regions: the average annual number of all trips and tourist arrivals for the period 2010-2018 per inhabitant of the region. The *Tourism Intensity Rate* shows the share of tourists in the local population and is most often treated as a measure of tourism's impact on local society and culture (McElroy and Albuquerque, 1998). The weighted intensity of tourist flows in this way explains that tourism's impact on the researched area is greatest in the mountainous parts of Austria, Switzerland, and Italy (Alps, ski resorts), as well as in the NUTS 2 region of Adriatic Croatia, situated on the nearest part of the Mediterranean coast to the European continent (and markets of high tourism demand). Other areas with tourist intensity values of 10 to 20 were the southernmost region of Portugal, the Balearic Islands, Corsica, the Greek islands (in the Ionian and Aegean seas), southern Norway, northern Denmark, the Finnish autonomous region of the Åland Islands, and the southernmost NUTS 2 regions of the Netherlands and Belgium.

Map 8: Weighted intensity



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

The results for tourism flows weighted by intensity clearly show that regions of Eastern, Central, and Southeastern Europe — to be precise, huge parts of Finland, Estonia, Latvia, Lithuania, Poland, and Italy — have a lower weighted intensity of tourism flows than other parts of Europe, especially regions of the Scandinavian countries, Ireland, Scotland, southern England, the Alps, and the Mediterranean coast.

3.2.1.2 Connectivity

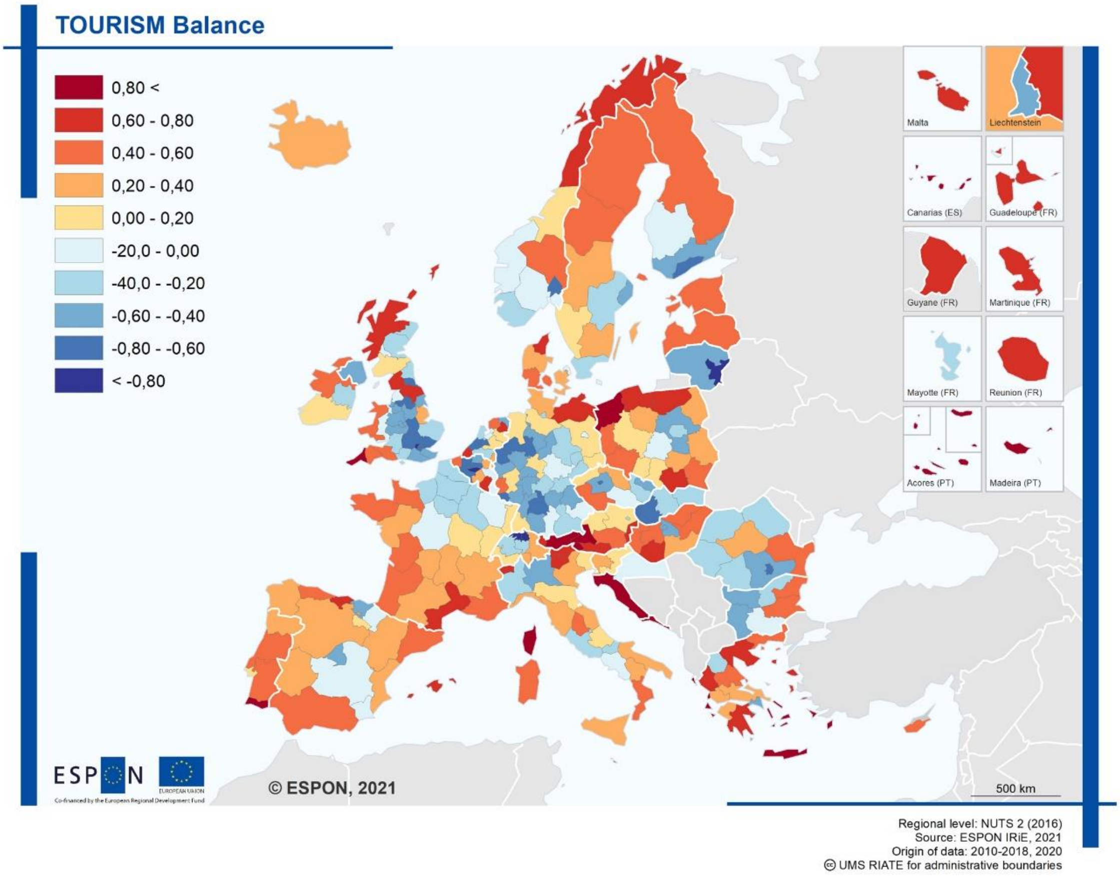
While the connectivity indicator might provide very interesting results for certain other flows (i.e. air passengers and Erasmus students), it does not for tourism flows. The connectivity indicator presents the percentage of regions with which a given region has experienced any flow. In the case of tourism flows, only 7 out of 297 regions have a region for which our calculation resulted in the departure of less than 0.5 tourists in 2010-2018 (i.e. after we rounded to the nearest zero).

3.2.1.3 Balance

The following map shows the balance of tourist flows (Map 9). Unlike the previous ones, it does not take into account the size of flows in a given region, but shows the dominant return: the relative value of the inflow and outflow imbalance. A positive balance denotes the advantage of arrivals over departures, while 0 represents equilibrium. If a region had only one flow direction, it would reach extreme theoretical values of 1 (only inflow) or -1 (only outflow).

The largest advantage of arrivals over departures (positive balance) occurred in mountainous regions (especially the Alps), on islands (e.g. the Balearic islands, Malta, Corsica, the Greek Ionian and Aegean islands, Canarias, Madeira, the Azores), and regions near seacoasts, which include almost all of the seas. In the lead were Mediterranean regions — especially in Greece, Croatia, Spain, southern France, and southern Italy — and southern Baltic regions in Denmark, northern Germany, and northern Poland. Many Scandinavian tourists, who by and large stayed within this macro-region, visited northern Scandinavia, a sparsely populated area. As a result, these regions had a great positive balance (regions of northern Norway, middle and northern Sweden, and Finland). There was an analogous situation in the northern UK – in Scotland (Map 9).

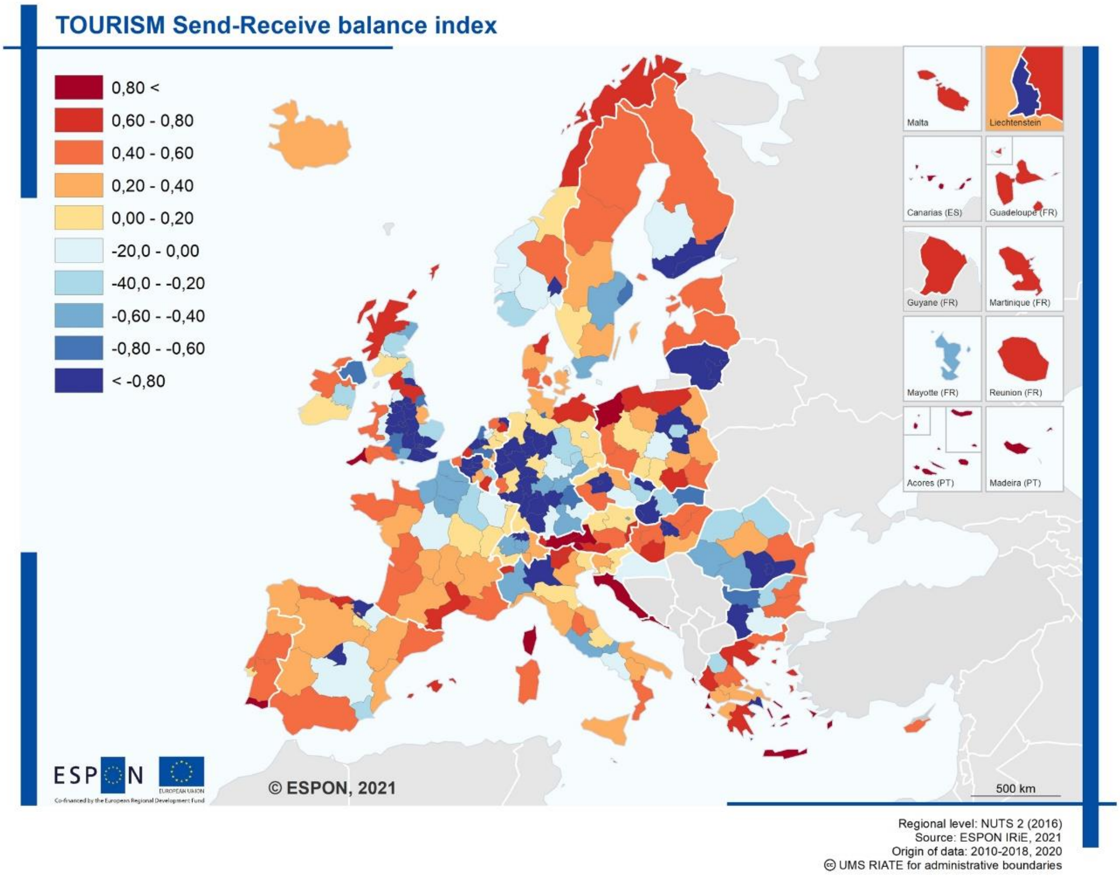
Map 9: Balance of tourist flows



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

The ESPON IRiE project analyses the balance of tourism flows also with the *send-receive balance index* (Map 10), an indicator similar to the one presented in the map above but without boundary values (it may exceed the range from -1 to 1).

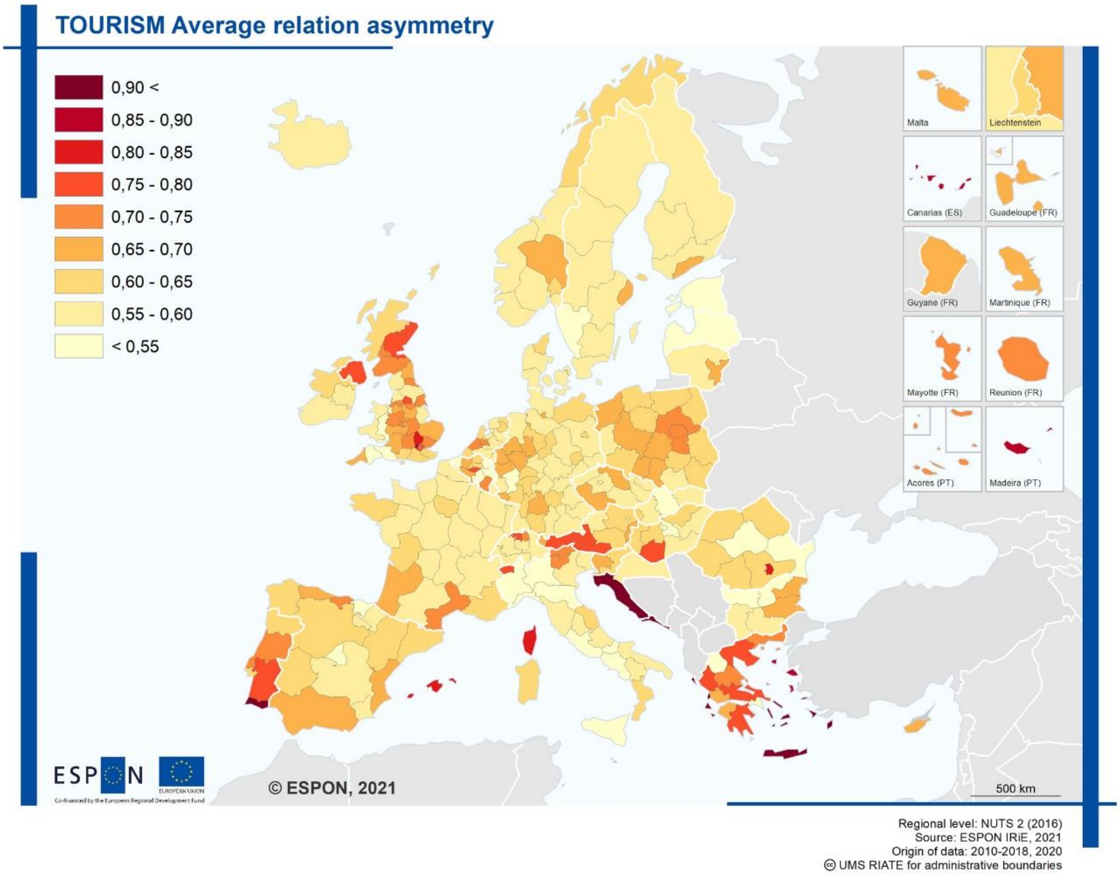
Map 10: Send-Receive balance index



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

Unlike the ratios above, average asymmetry does not apply to the general balance of a given region, but to the average imbalance of its individual relations. If there is a flow region, then the general inflow and outflow level out and the balance index is approximately 0, but the average asymmetry will be large, because relations with individual regions will be unbalanced (some positive and some negative). This indicator complements what the balance index expresses while deepening our knowledge of the situation. Therefore, the *average relation asymmetry* index (Map 11) shows which regions have unbalanced links, regardless of total balance. Suppose, for example, that a large number of tourists go to some European region but a similar number of tourists go only to another part of Europe. The Send-Receive balance index indicator may fail to detect this, because the total balance is close to 0. This is nevertheless important information for the balance sheet.

Map 11: Average relation asymmetry

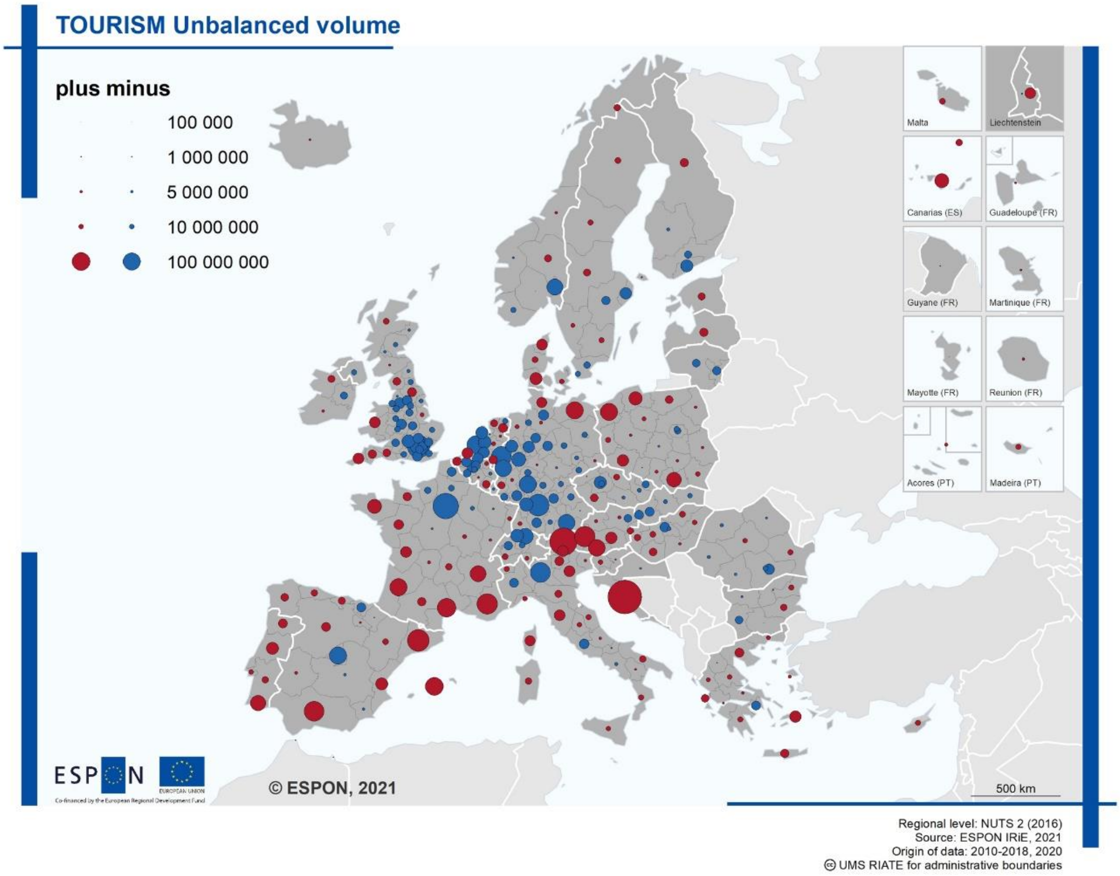


Sources: UNWTO & EUROSTAT databases / Authors' elaboration

The indicator for imbalanced volume of tourism flows shows how many tourist arrivals in a given region are more or less than balance. The balance index may have a large positive or strong negative value. While in less-populated regions this issue does not matter much, a slight imbalance in large agglomerations or in regions of mass tourism makes a difference.

When mapped, the unbalanced volume indicator clearly pinpoints the regions with the highest imbalance of tourism flows (Map 12). The three European spots of highly imbalanced volume of tourism where inflow of tourists far outstripped outflow were the Croatian Adriatic region (NUTS 2 code HR03), the Austrian region of Tirol (AT33), and the Spanish region of Cataluña (ES51). The Blue Banana figures in here as well, now as the continuous area with the greatest imbalance with dominant outflow (demand core of Europe). The greatest imbalance in tourism flows within regions of inflow dominance was observed in Alpine regions and along the Mediterranean and Baltic seas, more precisely in the coastal regions of Spain, France, and Croatia in the Mediterranean region, and in Denmark, northeastern Germany, and northwestern Poland in the Baltic region. An imbalance was present also in metropolitan areas and capital cities. Europe's largest cities and their respective surrounding areas were clearly the origin of tourists (see Map 12 for the negatives of the London metropolitan area, the Ruhr conurbation, Paris, Madrid, Stuttgart, Munich, Amsterdam, Oslo, Stockholm, Helsinki, Vilnius, Warsaw, Prague, Budapest, Bratislava, Zurich, Milan, Rome, Bucharest, Sofia, Athens). The majority of the mentioned cities were, however, important European tourism destinations.

Map 12: Imbalanced volume

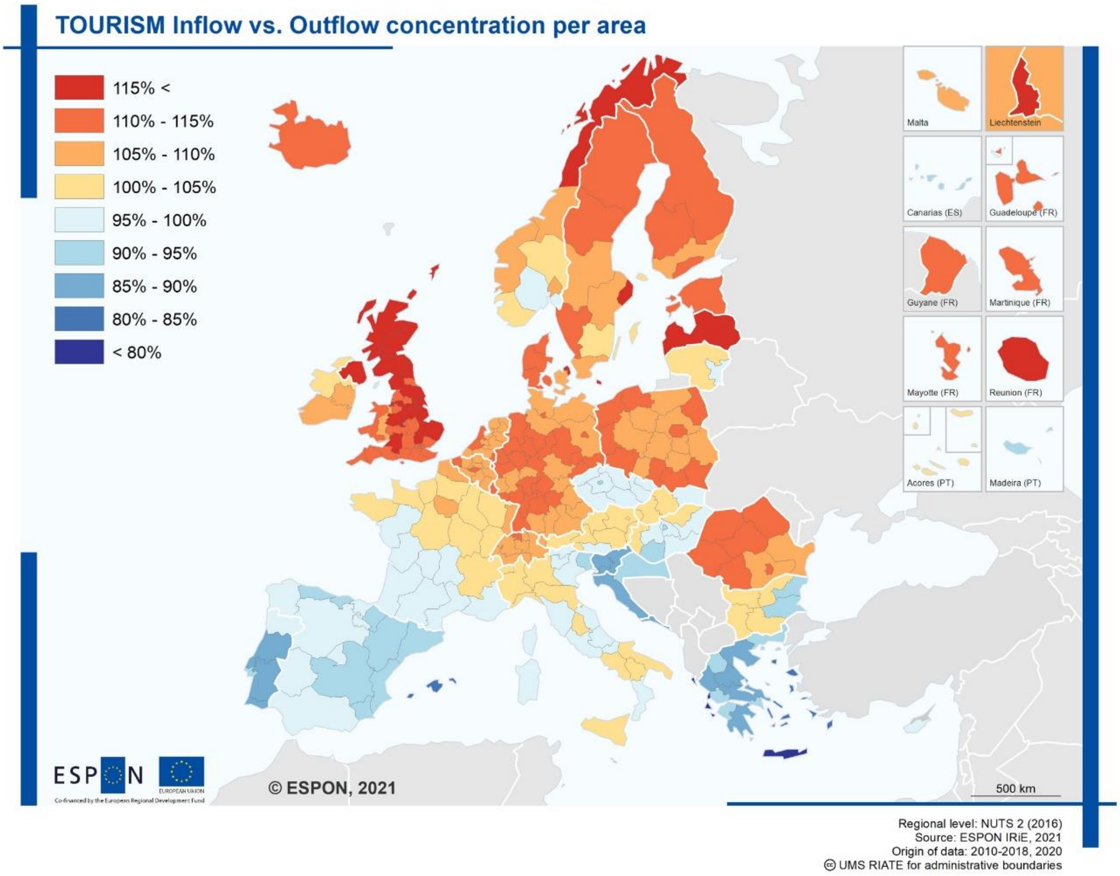


Sources: UNWTO & EUROSTAT databases / Authors' elaboration

3.2.1.4 Concentration

Analysis of the concentration of tourist inflow and outflow show that regions situated in northern and eastern Europe relied on tourists from a rather small area, mainly domestic regions, while inflow was dispersed to the tourism destinations in southern Europe. Outflow was the other way around: tourists from northern and eastern Europe visited different regions, while those from southern Europe did not move to dispersed areas outside the Mediterranean zone (Map 13).

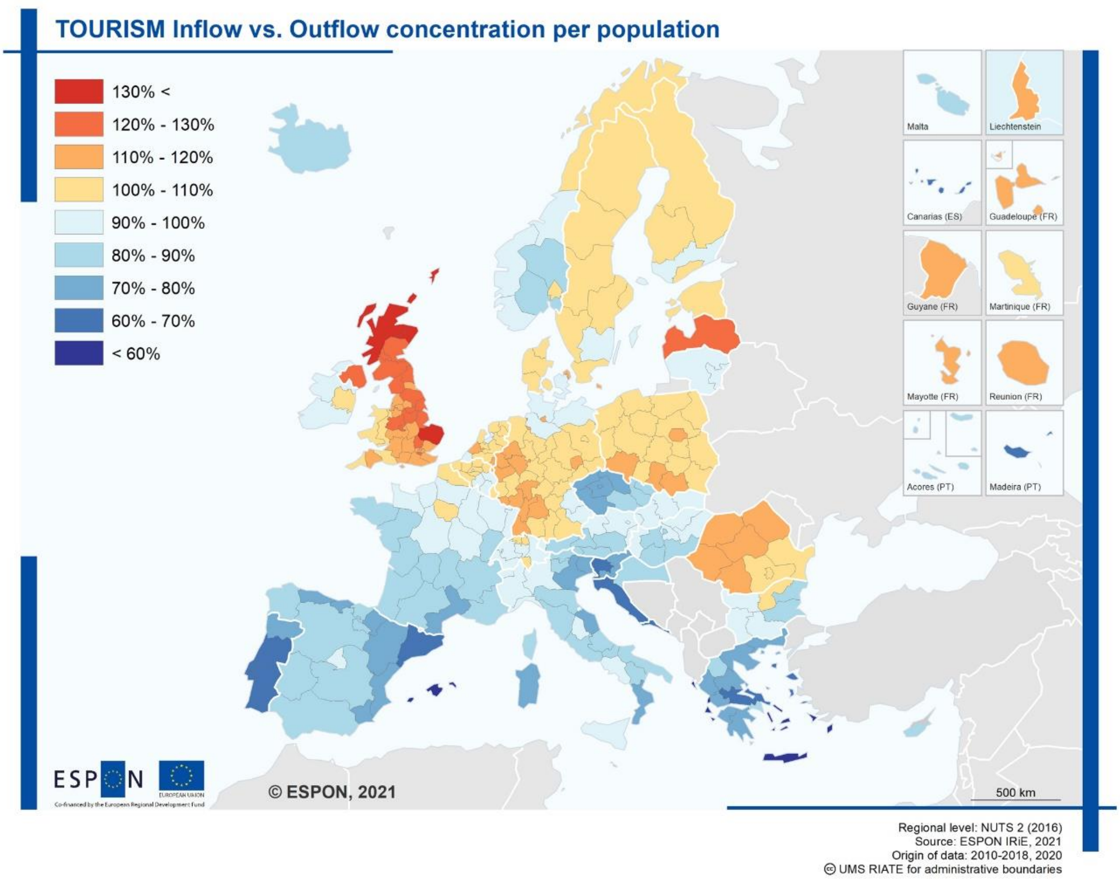
Map 13: Tourism inflow vs outflow concentration per area



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

Scandinavia was special when it comes to outflow per population, because, with visiting tourism destinations mainly in its cultural circle, it had a high concentration of tourists per population, while the indicator was much lower for outflow per area. In outflow generally most regions had a greater concentration per population than per area, because the destinations are peripheries, mountains, and such — sparsely populated areas. Hence the large proportion of tourists in a small population. In inflow most regions had a greater concentration per area than per population, because much of the inflow in tourist regions was made up of the inhabitants of small urbanized areas. Tourism inflow vs outflow concentration per population is presented on Map 14.

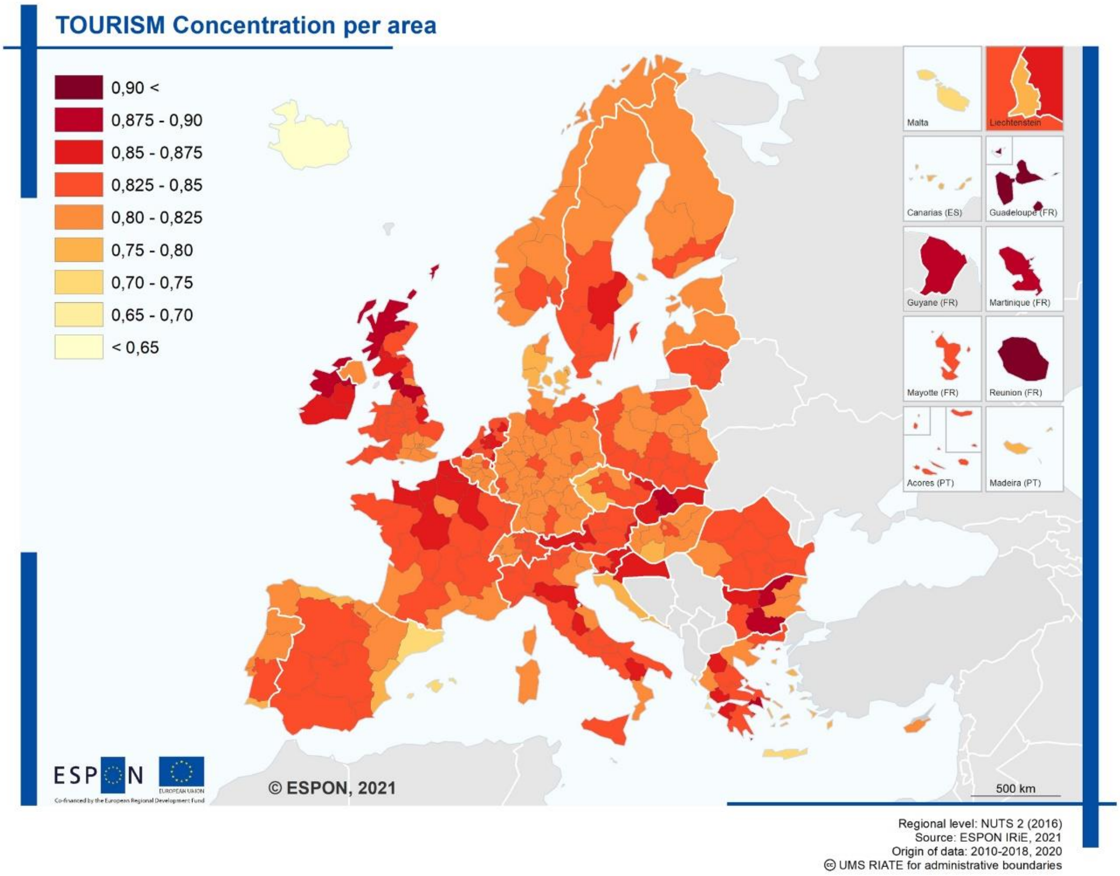
Map 14: Tourism inflow vs outflow concentration per population



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

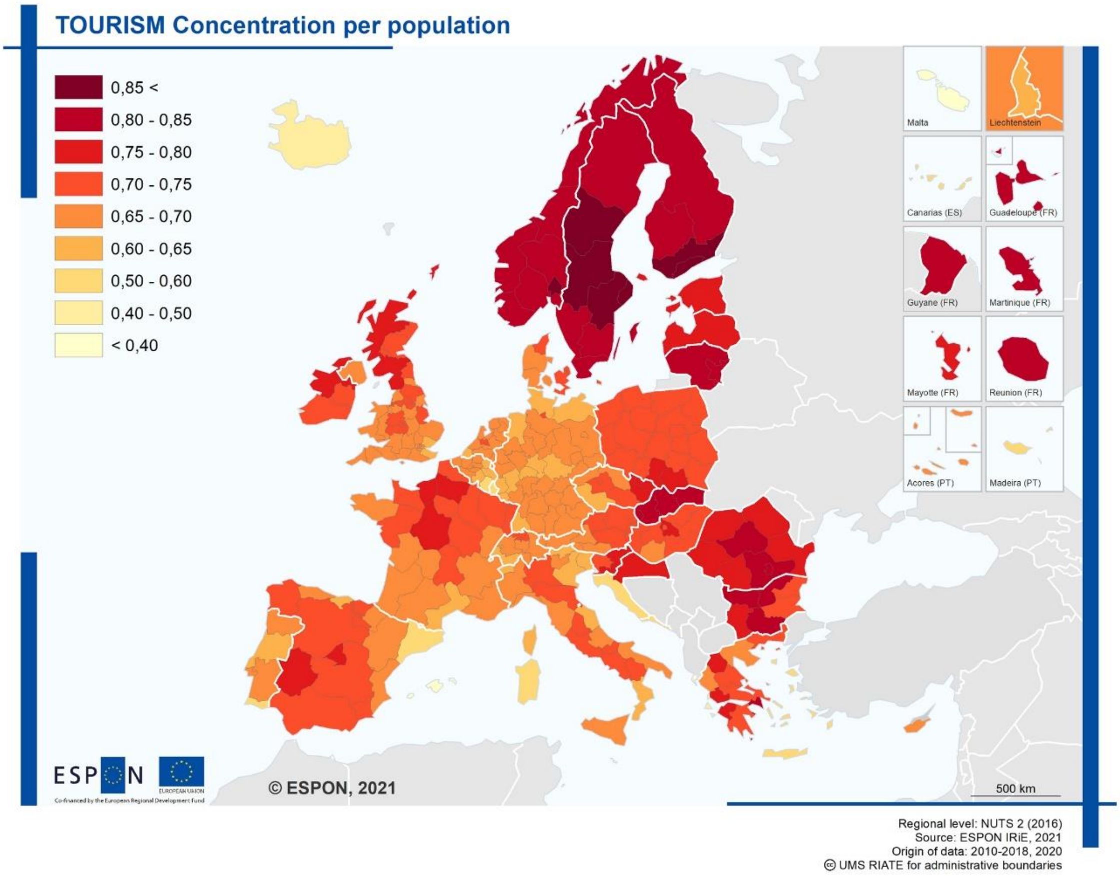
The lowest concentration was in total two-way tourism flows (Maps 15 and 16), because the areas of tourism inflow and outflow were complementary, so the overall volume of flows was dispersed. This applied especially to the per population term (Map 16), because much of the flows still concerned densely populated areas (see intensity: Maps 7 and 8), so it was concentrated per area.

Map 15: Tourism concentration per area



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

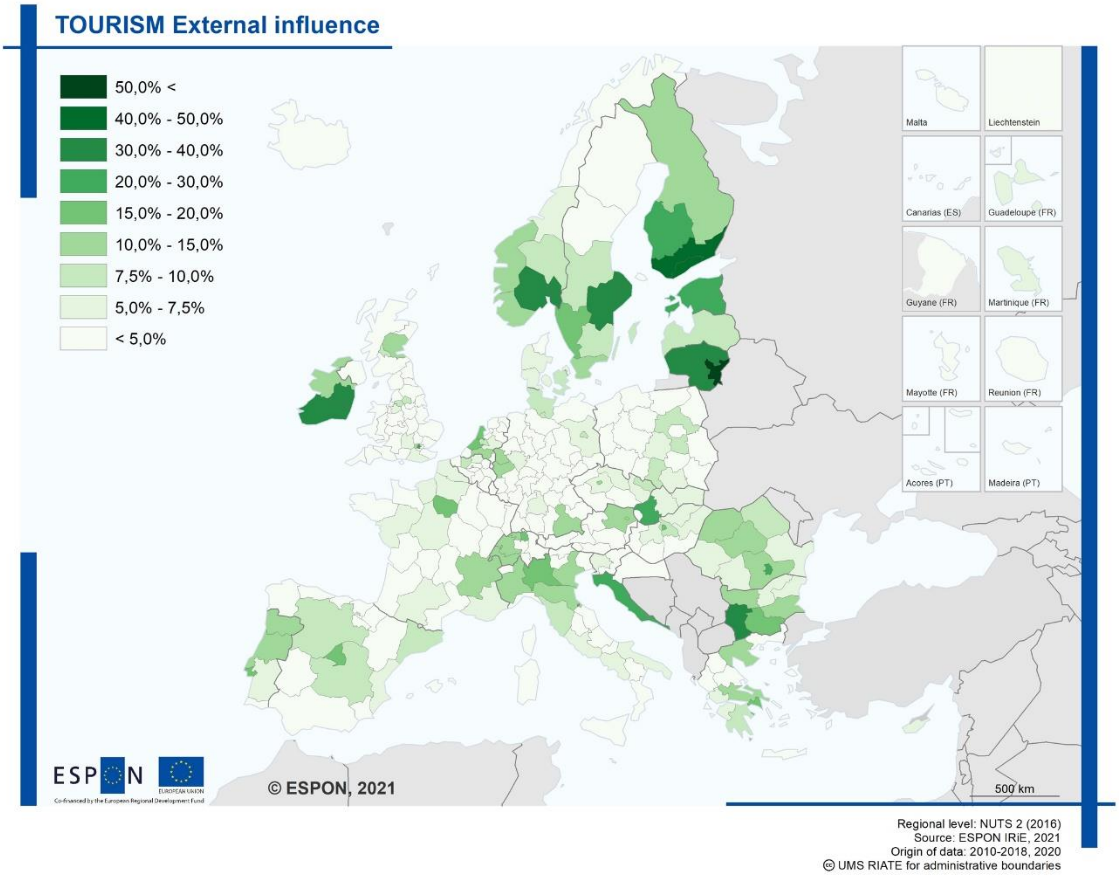
Map 16: Tourism concentration per population



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

Another measure of the outflow concentration is *external influence*, presented on Map 17. For each NUTS 2 region, it shows what percentage of its greatest outflow made up the entire inflow of the target region. For example, the Lubelskie region in Poland scored 3.21%, because its greatest outflow was to the region of Warsaw, and the amount of this outflow was 3.21% of the total inflow to Warsaw. The largest external influence was observed in the regions of Scandinavia and the Baltic States, Ireland, and southeastern Bulgaria.

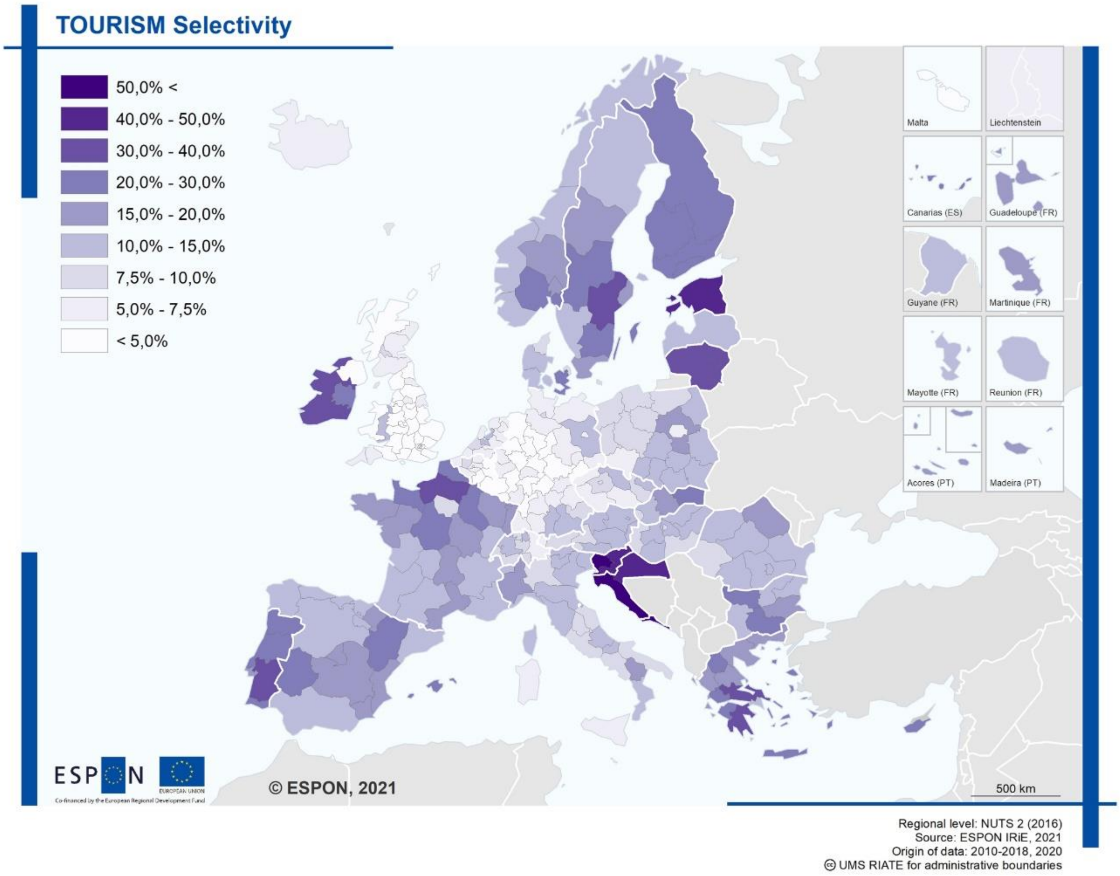
Map 17: External influence



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

The map below (Map 18) shows the indicator *selectivity*, which is symmetrical to the previous one. This time it's a question of what percentage of the inflow to a given region is the region from which this inflow is the greatest. In other words, the indicator therefore measures inflow concentration, but again on the basis of the share of only one region: the one providing the highest inflow. Therefore, in comparison with the Tourism External Influence (map 17), it considers the greatest inflow, not the greatest outflow. Our analysis shows that Estonia, Lithuania, East Middle Sweden (Östra Mellansverige), the Irish regions, and the regions of southern Greece, Slovenia, Croatia, Portugal, and northern France had the highest selectivity indicators.

Map 18: Selectivity

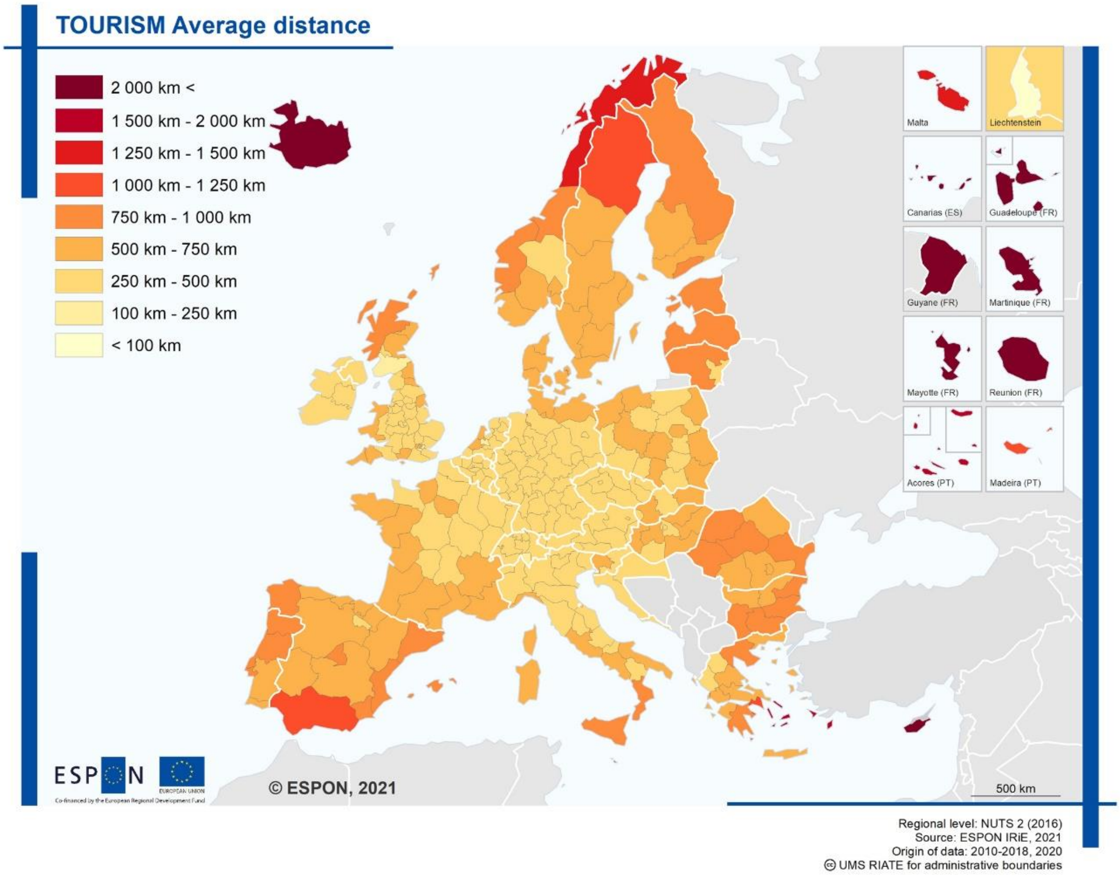


Sources: UNWTO & EUROSTAT databases / Authors' elaboration

3.2.1.5 Distance impact

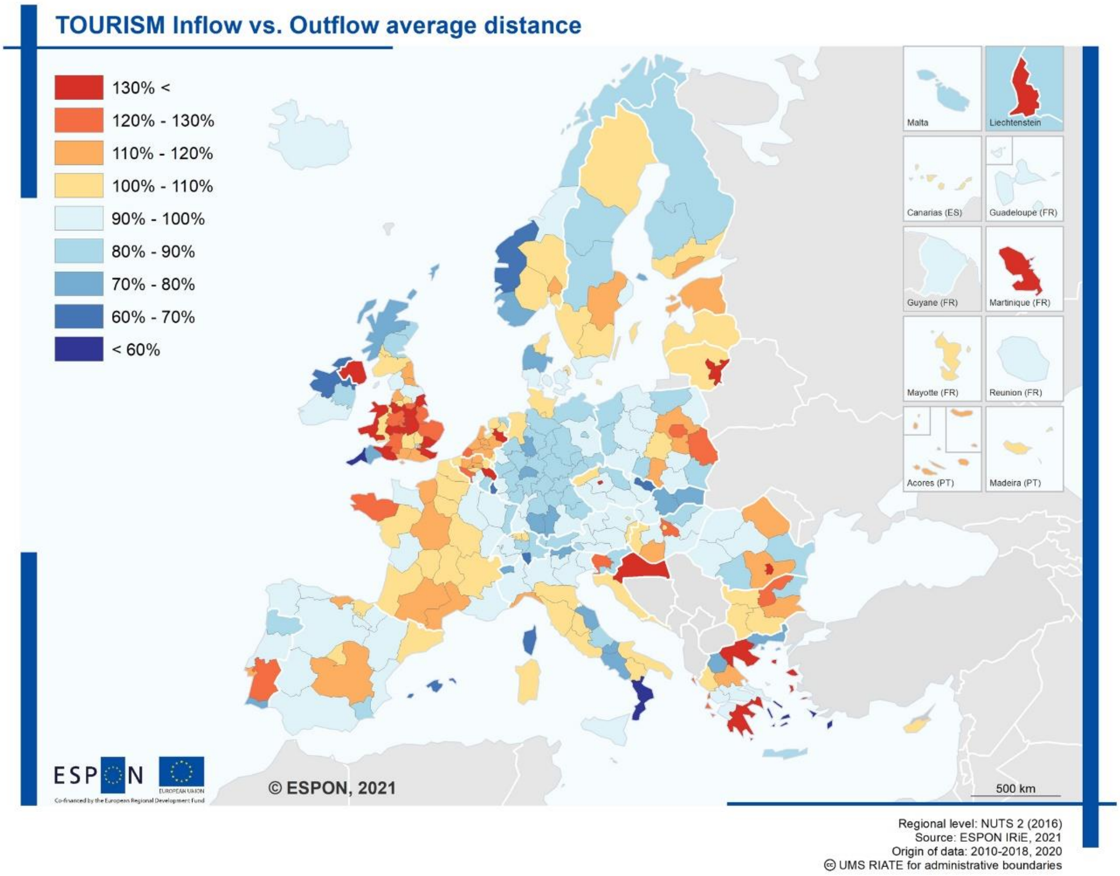
Distance is understood as the average distance in kilometers for departures, arrivals, and travel to or from a region. Our analysis shows that the longest distances were travelled toward regions situated on the outskirts of the researched area: northern Scandinavia, Iceland, Portugal, the southern Spanish regions, southern Italy, Malta, Greece, Cyprus, the Bulgarian regions on the Black Sea coast, and Atlantic, Mediterranean, and other ocean islands belonging to European countries (Madeira, Azores, Canarias, Reunion, etc.). Outflow followed the aforementioned rule, with some exceptions: tourists from Mediterranean regions generally travelled to tourist destinations within their region, not far away. On the other hand, tourists from northern European regions travelled more kilometers to tourism destinations than the tourists who came to visit *them*. In other words, tourism in northern Europe remained largely within the region, with few tourists travelling from faraway places to visit (e.g. regions of northern Scandinavia, the northern UK, Estonia, or Latvia). On the other hand, inhabitants of the aforementioned regions travelled longer distances to reach their tourism destination. Total average distance in kilometers is presented on Map 19, while Map 20 shows inflow vs. outflow average distance.

Map 19: Average distance total



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

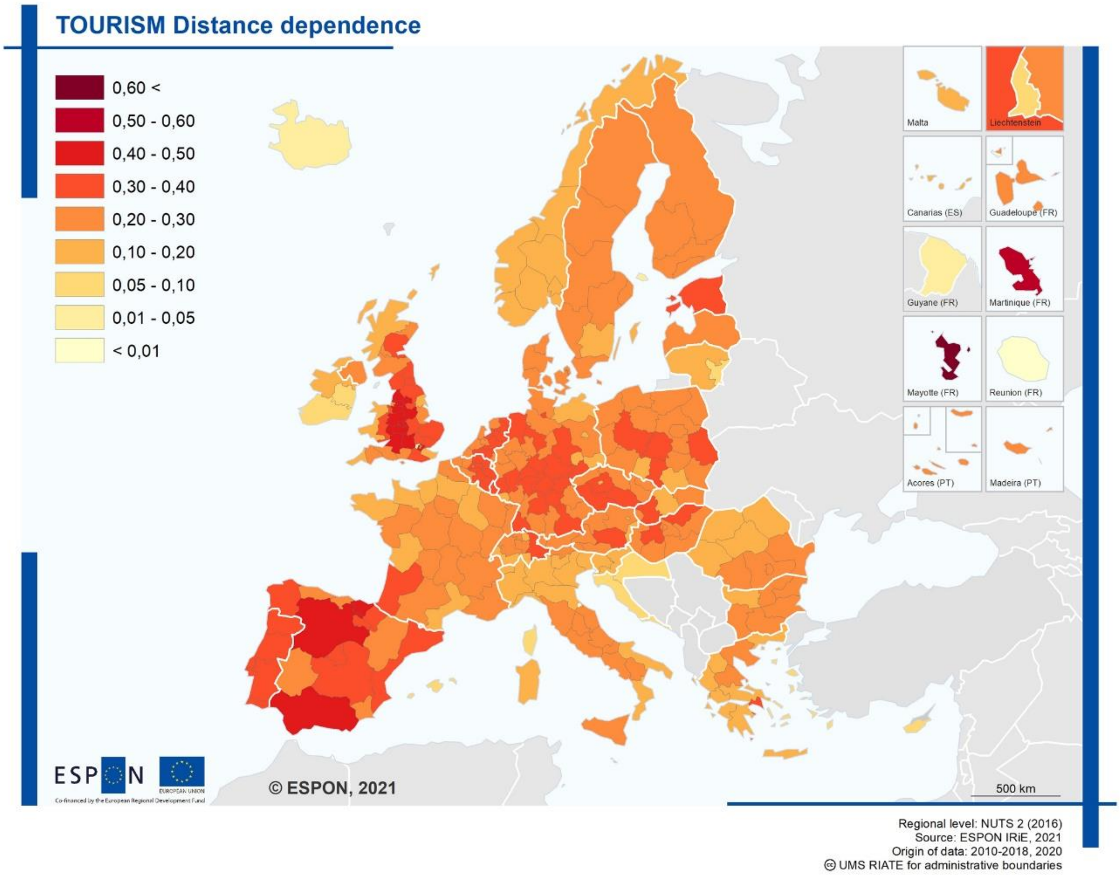
Map 20: Inflow vs outflow average distance



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

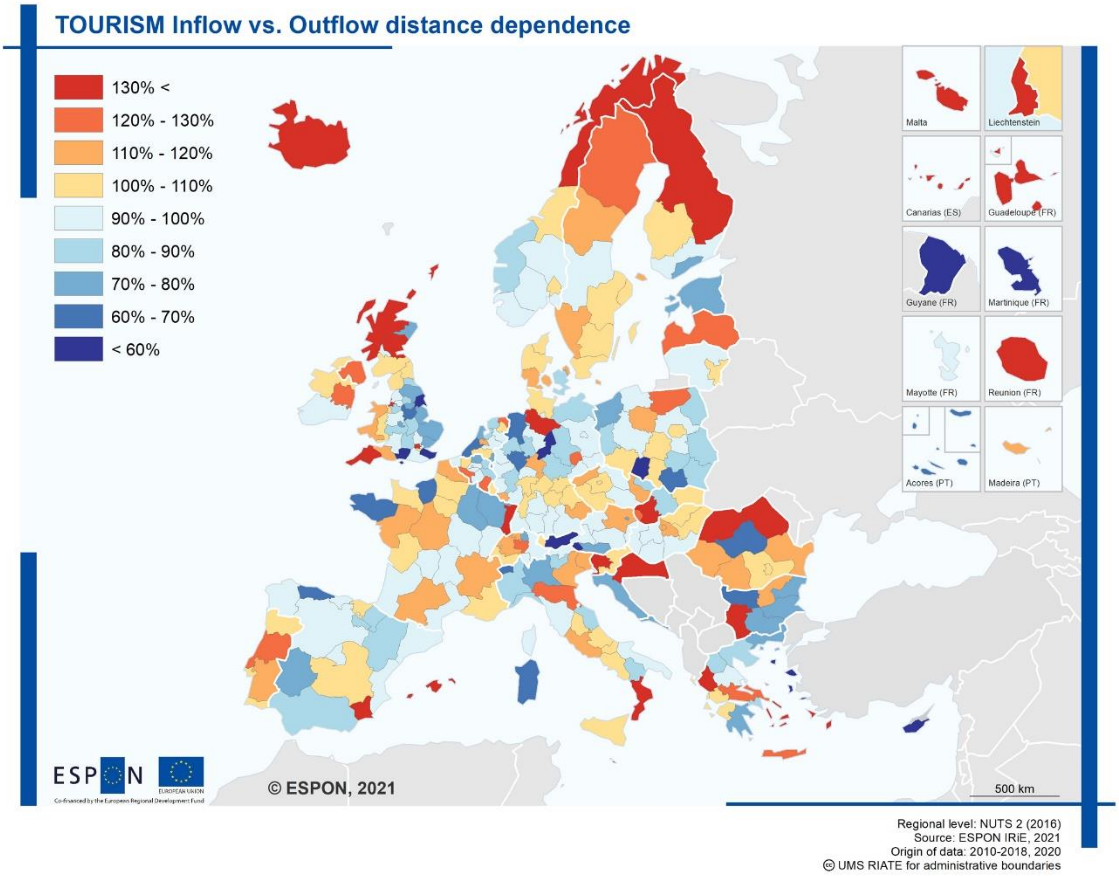
Map 21 shows the total distance dependency between NUTS 2 regions in the researched area. The distance dependency indicator shows the extent to which distance determined the magnitude of a flow, regardless of the dependency's specifics. A value of 0 (theoretical minimum) denotes that no relationship was found between distance and variation in flow rate. A value of 1 (theoretical maximum) denotes the perfect dependence of a given region's flow rate on distance. Map 22 presents inflow vs. outflow distance dependence.

Map 21: Distance dependence total



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

Map 22: Inflow vs outflow distance dependence



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

There is a regularity within the distance-impact analysis: regardless of the direction of the inflow/ outflow, the longest journeys are in regions of the northern and southern edges of Europe. The respective maps are generally similar, but when proportions have been observed, tourists from the northern regions travel much farther away in their tourist trip than those from southern regions. In other words, southern regions attract tourists from an even greater distance.

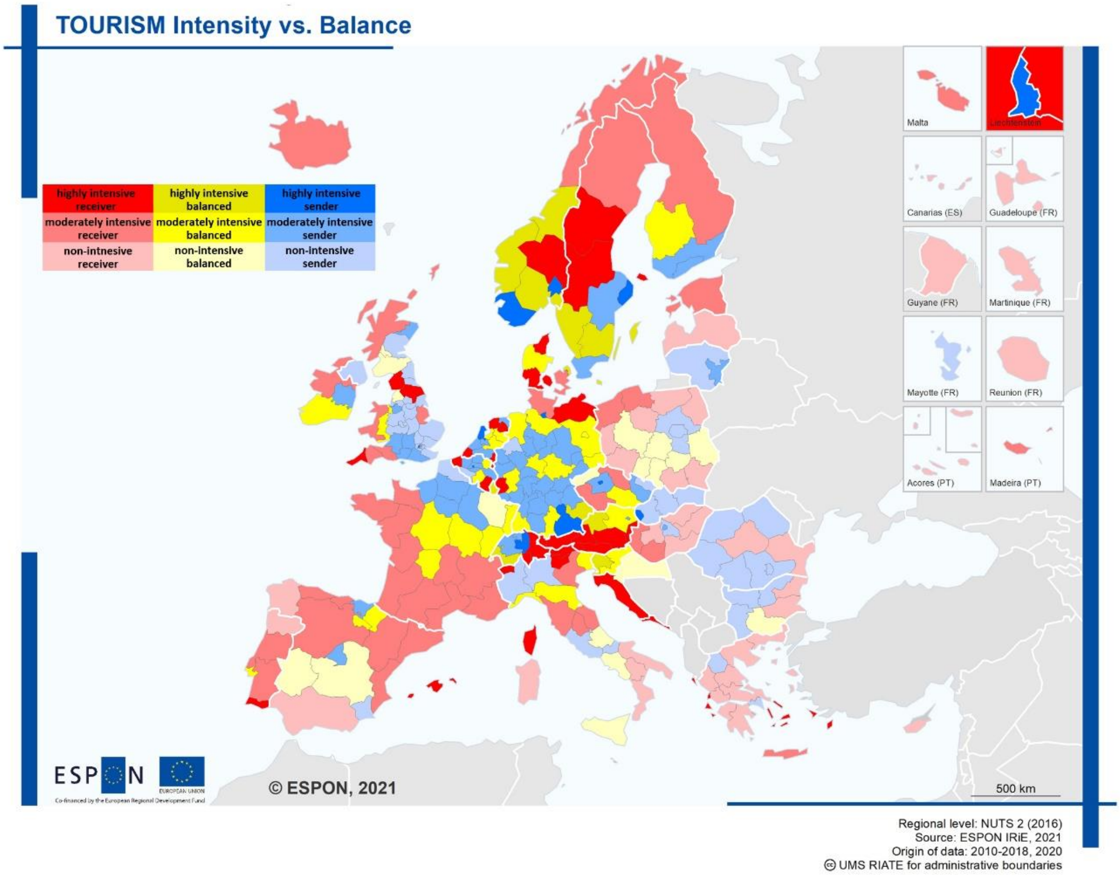
3.2.2 Typologies

In this part, we have confronted the indicator for intensity of tourism flows with the indicators for balance, concentration, and distance of tourism flows.

The intensity of tourism flows in a given region is the sum of the volume of all of its tourism flow relations. When combined with the balance of tourism flows indicator, which explains the relative value of the inflow and outflow imbalance, three ranges of each of the two indicators could be distinguished. Therefore, each of the NUTS 2 regions could be classified into one of 9 groups, ranging from *highly intensive receivers* to *non-intensive sender* (Map 23).

Out of 297 NUTS 2 regions in the researched area, there were only 39 *highly intensive receivers* (13.1%). These were situated in different parts of Europe, but generally close to coasts (Baltic, Mediterranean, North seas), mountains (Alps, Scandinavia), or small regions in the vicinity of regions with high population density (i.e. parts of the UK, the Netherlands, the Ruhr area). The sum of *highly intensive receivers* and *moderately intensive receivers* our research covers almost half of Europe in area. With this subsequent group, almost the entire Mediterranean region, European regions on the Atlantic coast, north Scandinavia, and the northern part of the Baltic states (Estonia and Latvia) are covered. By comparison with intense countries, balanced countries cover mostly regions in Central Europe (regions in northern and central Germany, northeastern France, central Spain, southern Ireland, central and eastern Poland, southern Czechia, Slovenia, northern and central Italy, and Sicily and southern Bulgaria). *Senders* were most often regions with high urbanization and population density, and relatively unattractive to tourists in comparison with regions with extraordinary natural and cultural heritage resources for the development of tourism products (Map 23).

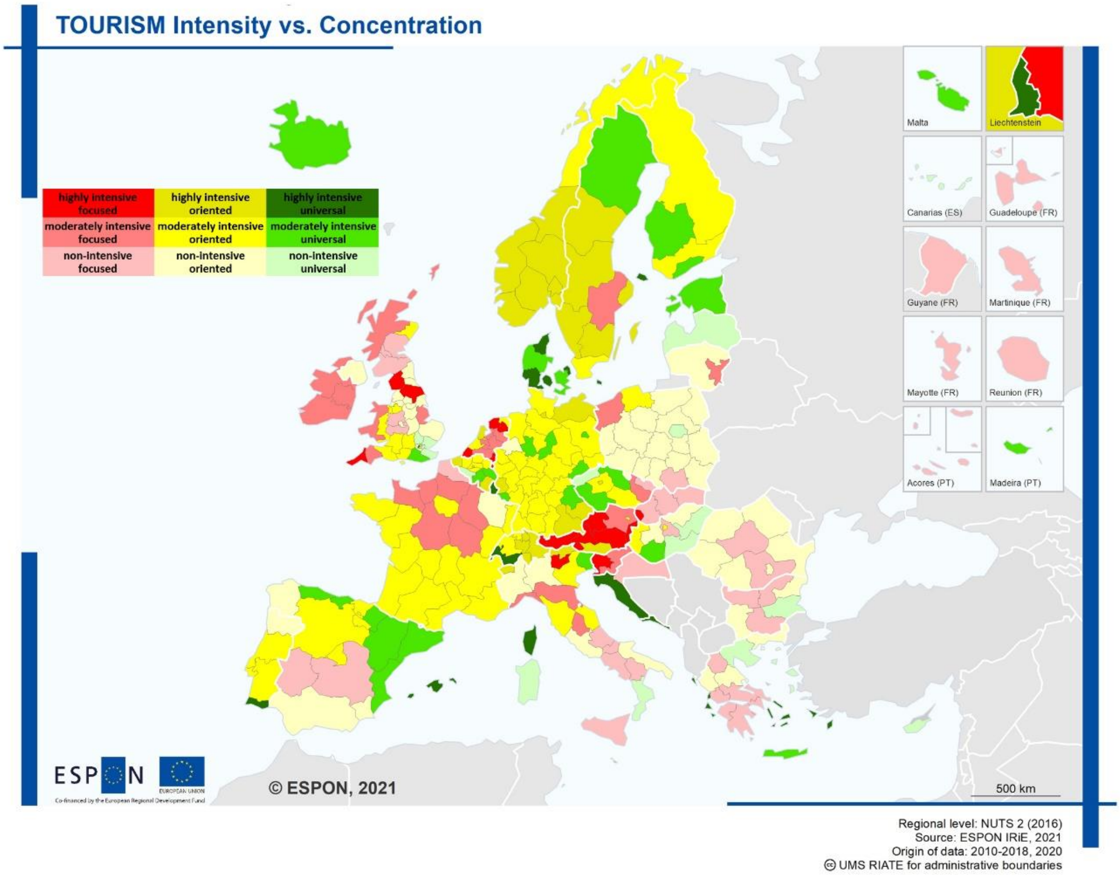
Map 23: Intensity vs. Balance



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

When we combine intensity with concentration of tourism flows, we find these two dimensions could cover a group of regions ranging from the *highly intensive focused* to the *non-intensive universal* (Map 24).

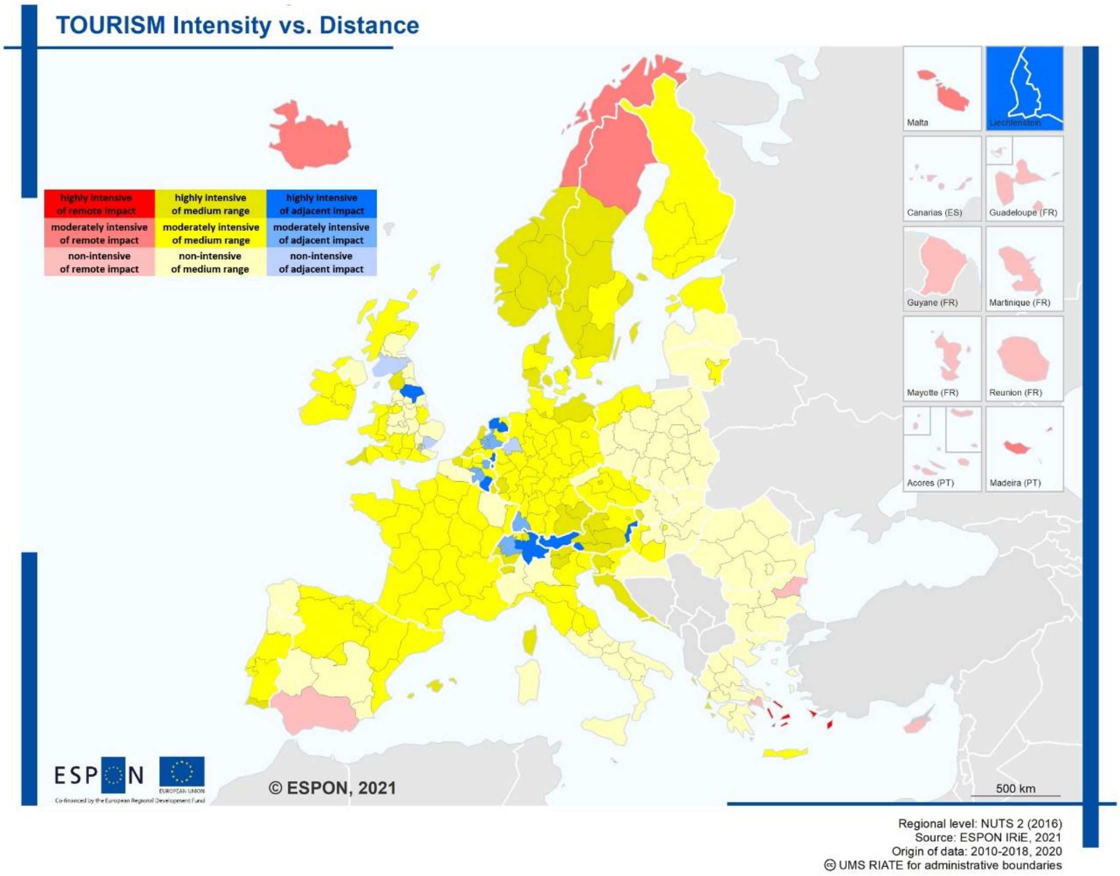
Map 24: Intensity vs. Concentration



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

A comparison of interrelations between intensity of tourism flows and distance between regions brought out nine groups of regions: *highly, moderately* and *non-intensive of remote impact, medium range* and *adjacent impact* (Map 25). While large majority of observed NUTS 2 regions were assigned to the medium distance range, the highest intensity within the remote distance regions was observed for the Greek Ionian islands, followed by regions situated in northern Scandinavia, in Iceland, in southern Spain, on islands (i.e. Balearic, Malta, Cyprus etc.), and in the extra-continental territories of European countries (i.e. Canarias, Madeira, Reunion). *High intensity of adjacent impact* regions could be found primarily in Alpine areas, as well as in urbanized areas of the Western Europe: several regions in western Germany, Belgium, the Netherlands, and the UK.

Map 25: Intensity vs Distance



Sources: UNWTO & EUROSTAT databases / Authors' elaboration

3.2.3 Analysis of results -> Explanatory factors: drivers and barriers

It is as important to explain the diversification of interregional tourism flows in Europe as it is to identify them. To make our inferences (study of a covariate, coexistence) we used the methods of macroeconomic analysis described in Chapter 2.4. Every dependent variable consisted of identified tourism flows between all regions in the studied European countries. On the other hand, the independent variables were adopted a priori. In other words, the model's mathematical form resulted from our substantive knowledge of the issues and an attempt to simplify our approach to the relationship between the variables.

In selecting the variables, we took both their statistical correctness and their substantive significance into account. The development of all types of flows, including the flow of tourists, is a derivative of the existing economic, human, institutional, ecological, infrastructural and financial conditions. The spatial structures and transformational dynamics of each of these conditions undergo significant change in the process of socio-economic development and the formation of interregional equations. The measures and indicators used to explain the flows are also changing. An important determinant of the given attributes is therefore not only availability, but above all usefulness in describing and explaining the analyzed issues. These are difficult assumptions to make in dynamic research, as certain measures lose their usefulness during the analyzed period. Moreover, the spatial scale proved to be extremely important in our analyses — in particular the spatial adequacy of our indicators. For example, we did not use interregional flows on a European scale as an explanatory variable to measure local development.

Correct measurement and results depend on the correct selection of explanatory variables. However, it should be remembered that the nature of social differentiation is qualitative, and a quantitative approach to the problem will always be a simplification. In the case of tourism flows, it is difficult to take fashion or tourists' perception of a given destination and translate it into a variable. The selection of features is therefore necessarily arbitrary and is a compromise between what is important in the study and what is available. There are no universally recognized solutions in this regard. When choosing the measures, however, one should take into account the spatial, temporal, and substantive scope of the analysis — and above all its purpose (Strahl, 2006; Runge, 2006). The measures used to explain the obtained tourist traffic dispersion should be based on three criteria: expertise (reflecting the essence of the described features as much as possible), formality (data should be collected with uniform methodology and should be reliable and measurable), and statistical use (high variability, lack of mutual correlation).

About 50 independent variables were used for the study. For our final analysis, however, we narrowed the field to the 20 described in Chapter 2.4. This resulted from our selection of the collected material on substantive criteria (low relationship with the phenomenon to be explained, inability to draw an appropriate hypothesis for the dependence of tourist flows on the analyzed phenomenon, or the occurrence of a better indicator from the set of variables that matched the others thematically), for formal reasons (above all, above-average data gaps in a large group of regions), and on statistical grounds (mutual correlation between features). We introduced these variables into our analyses gradually, maintaining standards of macroeconomic modeling. We initially produced eight models, but careful analysis enabled us to reduce the list to the four presented in this report.

The first model is the simplest and assumes the impact on the dependent variable of only three factors: GDP of the region sending tourists, the number of people living in this region, and the distance between the region of origin and region of destination. The other three models include additional variables. In the second model, these are the variables relating to GDP per capita, the significance of domestic tourism, the population density of origin and destination regions, atmospheric factors (such as precipitation and temperature), regional geographic features (such as location on islands, near a border, in countries that joined the EU European Union after 2004, in metropolitan regions, and regional demographic features (linguistic differences between regions)). In the third model, the GDP-per-capita variable is replaced by disposable income of the inhabitants of origin regions. The fourth has additional variables to define the level of education of origin residents and total regional employment. All these variables show an a priori substantive relationship with the tourism flow variable, and each can be treated as an independent single variable to explain the analyzed phenomenon. In other words, for each it is possible to formulate a hypothesis on the mutual dependence of the explained and the explanatory variable. Thus, the substantive condition can be considered fulfilled. All 20 of the variables selected for our analyses are also not mutually correlated and are characterized by significant spatial dispersion (fulfillment of the statistical condition). Also, the observations are available for the entire research period (2010-2018), and there is little lack of data (fulfillment of the formal condition).

In order to verify that the identified explanatory factors are the same for all types of regions, we resorted to econometric modeling for different groups of regions. Our divisions included the following groups of relationships:

- destinations are mountain regions and non-mountain regions;
- destinations are coastal regions and non-maritime regions;
- origins are coastal regions and non-coastal regions;
- destinations are regions with metropolitan areas (METRO 1 or 2 or 3 or 4) and regions without metropolitan areas;
- origins are regions with metropolitan areas (METRO 1 or 2 or 3 or 4) and regions without metropolitan areas.

All estimation results for each group are presented separately (Tables 4–14), below.

Already at the beginning of the analysis of the results, we should point out that the explanatory value achieved for each of the complex models is much higher than that of individual determination coefficients obtained in simple regression models: i.e. analyses of a variable of tourist flows with a single explanatory variable. The modeling results for all regions and the two-way flows will be described in detail below, and the differences between tourist flows for different types of regions will be outlined.

Results for main model T1 (for all observations in both directions of tourism flows) are the main results of the gravity estimation. The GDP and population variables showed a positive coefficient of 1.64 for the origin GDP and a negative coefficient of -0.43 for the origin population. Also, distance to destination, showing the importance of spatial aspects in determining tourism flows, showed a negative and significant coefficient (-1.18) consistent with the literature (between -0.5 and -1.5). Since all four variables in the model are given in logarithm, a 1% increase in the GDP value of the origin region will result in an increase of 1.64% in the number of tourists, and an increase of 1% in destination distance will result in a decrease of 1.18% in the number of potential tourists. The first model T1 had the largest number of observations (the smallest loss of information related to a lack of data), reaching about 90% of all possible variables in terms of the R2R matrix and the nine-year period 2010-2018. This model is therefore the most universal, not only because of its simplicity, but also because of it has the greatest spatial coverage. The explanatory value of the T1 model is $r^2 = 0.53$.

The introduction of further variables led to an increase in the explanatory level to approximately r^2=0.64 - r^2=0.66. Importantly, analysis of the obtained models suggests a significant relationship between total flows and domestic flows of tourists. A significant positive relationship was also shown by the researched variable with GDP per capita and the interchangeably used level of disposable income. All other variables showed statistical significance, but it is worth mentioning only the most important of them in detailed analysis, because the ß indexes of such variables as population density were close to zero and thus had little influence on the shape of the dependent variable. There was a positive impact of such variables as the share of naturally valuable areas, the difference in rainfall between origin and destination regions, the lack of linguistic differences between the two regions, and the choice of islands as a travel destination. Additionally, it should be pointed out that the number of tourists was correlated with the level of education of the inhabitants — namely, the participation of people with the highest level of education — and, to a lesser extent, with the general level of employment. In conclusion, it should be stated that the addition of a dozen or so explanatory variables did not lead to a significant increase in the level of explanation, but allowed us to significantly reduce the pure influence of GDP, population size, and distance by emphasizing the importance of other factors. The direction of the impact of these variables on the flow of tourism was fully in line with expectations. Analysis of their simultaneous impact showed the complexity and totality of the variables determining the movement of people for tourism purposes, although, as indicated earlier, it did not fully reflect all possible variables in this regard.

For a more complete analysis of explanatory factors, we executed models of tourist flows "to" and "from" specific areas (coastal, mountain, and metropolitan) and for equilibrium without these features (non-coastal, non-mountain and non-metropolitan). In summary, these analyses showed no significant differences between different types of areas, and the results generally coincided with those of the general model.

Table 4. Estimation results of the econometric model for the explanatory factors of tourist flows between European regions in 2010-2018 – entire dataset

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(T1)	(T2)	(T3)	(T4)
l_gdp_o	1.639***	-0.151***	1.045***	0.829***
l_dist	-1.184***	-0.882***	-0.914***	-1.036***
l_popul_o	-0.427***	1.333***	0.167***	0.096***
domestic		1.116***	1.151***	1.079***
l_gdp_pc_o		2.177***		
l_disp_inc_o			0.932***	0.720***
l_pop_den_o		-0.027***	-0.010***	-0.004
l_pop_den_d		0.031***	0.033***	0.025***
new_eu_o		0.089*	0.271***	0.187***
non_euro_rel		-0.122***	-0.083***	-0.083***
l_natura_d		0.214***	0.219***	0.258***
temp_rel		0.052***	0.049***	0.042***
precip_rel		0.410***	0.501***	0.416***
island_d		0.561***	0.599***	0.652***
bord_reg_d		0.218***	0.192***	0.162***
language		0.428***	0.349***	0.337***
metro_o		-0.433***	-0.277***	-0.210***
metro_d		0.534***	0.532***	0.545***
l_h_edu_o				0.548***
l_emp_tot_o				0.134***
Constant	4.859***	-25.780***	-8.613***	-4.950***
Observations	724,608	457,196	453,990	452,622
Number of id	85,248	53,815	53,815	53,815
pseudo R^2	0.525	0.656	0.639	0.656

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1, the Hausman-Taylor estimation resuls.
Source: Own elaboration.

Table 5. Estimation results of the econometric model of tourist inflows to mountain regions in 2010-2018 - mountain destinations

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(M1)	(M2)	(M3)	(M4)
l_gdp_o	1.740***	0.254***	1.141***	0.873***

l_dist	-1.474***	-1.134***	-1.193***	-1.317***
l_popul_o	-0.549***	0.930***	0.089***	-0.423***
domestic		1.071***	1.088***	1.009***
l_gdp_pc_o		1.879***		
l_disp_inc_o			1.056***	0.828***
l_pop_den_o		-0.043***	-0.028***	-0.016**
l_pop_den_d		0.026	-0.008	-0.004
new_eu_o		-0.240***	-0.148*	-0.218**
non_euro_rel		-0.076***	0.014	-0.002
l_natura_d		-0.014	-0.020	0.025
temp_rel		0.058***	0.057***	0.052***
precip_rel		0.449***	0.521***	0.449***
island_d		0.984***	1.036***	1.054***
bord_reg_d		0.482***	0.431***	0.390***
language		0.403***	0.339***	0.361***
metro_o		-0.456***	-0.318***	-0.231***
metro_d		0.757***	0.809***	0.817***
l_h_edu_o				0.441***
l_emp_tot_o				0.635***
Constant	7.894***	-19.902***	-7.964***	-0.279
Observations	209,876	127,764	126,868	126,484
Number of id	24,691	15,040	15,040	15,040
pseudo R2	0.534	0.616	0.591	0.618

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1;
the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 6. Estimation results of the econometric model of tourist inflows to non-mountain regions in 2010-2018 – non-mountain destinations

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(nM1)	(nM2)	(nM3)	(nM4)
l_gdp_o	1.597***	-0.261***	1.022***	0.833***
l_dist	-1.144***	-0.813***	-0.834***	-0.961***
l_popul_o	-0.395***	1.447***	0.191***	0.292***
domestic		1.100***	1.143***	1.087***
l_gdp_pc_o		2.249***		
l_disp_inc_o			0.880***	0.675***
l_pop_den_o		-0.016***	0.001	0.004
l_pop_den_d		0.032***	0.034***	0.026***
new_eu_o		0.195***	0.401***	0.317***
non_euro_rel		-0.129***	-0.104***	-0.096***
l_natura_d		0.243***	0.252***	0.282***
temp_rel		0.054***	0.051***	0.043***
precip_rel		0.322***	0.447***	0.336***
island_d		0.315***	0.331***	0.409***
bord_reg_d		0.104***	0.082***	0.061***
language		0.444***	0.358***	0.339***
metro_o		-0.429***	-0.270***	-0.210***
metro_d		0.487***	0.481***	0.492***
l_h_edu_o				0.588***

l_emp_tot_o				-0.071***
Constant	4.447***	-27.323***	-8.641***	-6.549***
Observations	514,732	329,432	327,122	326,138
Number of id	60,557	38,775	38,775	38,775
pseudo R2	0.524	0.657	0.646	0.659

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 7. Estimation results of the econometric model of tourist inflows to coastal regions in 2010-2018 – coastal destinations

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(C1)	(C2)	(C3)	(C4)
l_gdp_o	1.715***	0.117**	1.190***	0.931***
l_dist	-1.440***	-0.846***	-0.853***	-1.005***
l_popul_o	-0.780***	1.105***	0.057**	-0.201***
domestic		1.453***	1.511***	1.393***
l_gdp_pc_o		2.038***		
l_disp_inc_o			0.947***	0.734***
l_pop_den_o		-0.046***	-0.025***	-0.018***
l_pop_den_d		0.139***	0.167***	0.111***
new_eu_o		-0.127	0.028	-0.029
non_euro_rel		-0.082***	-0.062***	-0.051***
l_natura_d		0.379***	0.381***	0.433***
temp_rel		0.043***	0.038***	0.034***
precip_rel		0.089***	0.154***	0.066***
island_d		0.228***	0.257***	0.274***
bord_reg_d		0.462***	0.454***	0.405***
language		0.326***	0.231***	0.232***
metro_o		-0.437***	-0.287***	-0.204***
metro_d		0.270***	0.239***	0.289***
l_h_edu_o				0.506***
l_emp_tot_o				0.357***
Constant	11.109***	-23.591***	-8.646***	-2.696***
Observations	358,628	199,655	198,255	197,655
Number of id	42,190	23,500	23,500	23,500
pseudo R2	0.529	0.62	0.593	0.619

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 8. Estimation results of the econometric model of tourist inflows to non-coastal regions in 2010-2018 – non-coastal destinations

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(nC1)	(nC2)	(nC3)	(nC4)
l_gdp_o	1.564***	-0.329***	0.952***	0.771***
l_dist	-1.022***	-1.081***	-1.133***	-1.249***
l_popul_o	-0.073***	1.440***	0.202***	0.272***
domestic		0.774***	0.794***	0.754***
l_gdp_pc_o		2.254***		
l_disp_inc_o			0.901***	0.687***

l_pop_den_o		-0.025***	-0.012***	-0.010***
l_pop_den_d		0.028***	0.026***	0.028***
new_eu_o		0.256***	0.436***	0.349***
non_euro_rel		-0.114***	-0.059***	-0.063***
l_natura_d		0.070***	0.077***	0.107***
temp_rel		0.020***	0.016***	0.007***
precip_rel		0.536***	0.647***	0.567***
bord_reg_d		0.181***	0.150***	0.138***
language		0.466***	0.393***	0.380***
metro_o		-0.376***	-0.226***	-0.171***
metro_d		0.514***	0.512***	0.518***
l_h_edu_o				0.587***
l_emp_tot_o				-0.035**
Constant	-0.664	-25.539***	-7.041***	-4.862***
Observations	365,980	257,541	255,735	254,967
Number of id	43,058	30,315	30,315	30,315
pseudo R2	0.507	0.633	0.627	0.635

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1;
the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 9. Estimation results of the econometric model of tourist outflows from coastal regions in 2010-2018 – coastal origins

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(C5)	(C6)	(C7)	(C8)
l_gdp_o	1.459***	-0.511***	1.005***	0.828***
l_dist	-1.007***	-0.988***	-0.935***	-1.067***
l_popul_o	-0.114***	1.537***	0.026	-0.115***
domestic		1.316***	1.455***	1.408***
l_gdp_pc_o		2.447***		
l_disp_inc_o			0.921***	0.487***
l_pop_den_o		-0.030***	0.005	0.007
l_pop_den_d		0.017***	0.028***	0.024***
new_eu_o		0.004	0.179**	0.096
non_euro_rel		-0.157***	-0.121***	-0.109***
l_natura_d		0.134***	0.148***	0.205***
temp_rel		0.060***	0.044***	0.023***
precip_rel		0.288***	0.333***	0.233***
island_d		0.513***	0.579***	0.684***
bord_reg_d		0.189***	0.166***	0.131***
language		0.424***	0.367***	0.355***
metro_o		-0.181***	-0.070***	-0.142***
metro_d		0.501***	0.492***	0.502***
l_h_edu_o				0.855***
l_emp_tot_o				0.260***
Constant	1.037*	-26.794***	-6.027***	-1.289***
Observations	363,488	203,652	200,904	200,904
Number of id	43,216	24,174	24,174	24,174
pseudo R2	0.516	0.661	0.636	0.651

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1;

the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 10. Estimation results of the econometric model of tourist outflows from coastal regions in 2010-2018 – non-coastal origins

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(nC5)	(nC6)	(nC7)	(nC8)
l_gdp_o	1.784***	0.198***	1.118***	0.903***
l_dist	-1.242***	-0.870***	-0.886***	-0.983***
l_popul_o	-0.632***	1.124***	0.248***	0.190***
domestic		0.962***	0.966***	0.918***
l_gdp_pc_o		1.867***		
l_disp_inc_o			0.887***	0.839***
l_pop_den_o		-0.018***	-0.011***	-0.007**
l_pop_den_d		0.038***	0.037***	0.029***
new_eu_o		0.185***	0.320***	0.251***
non_euro_rel		-0.083***	-0.044***	-0.046***
l_natura_d		0.265***	0.263***	0.292***
temp_rel		0.058***	0.060***	0.058***
precip_rel		0.709***	0.832***	0.735***
island_d		0.575***	0.580***	0.609***
bord_reg_d		0.226***	0.201***	0.179***
language		0.425***	0.369***	0.354***
metro_o		-0.719***	-0.491***	-0.358***
metro_d		0.560***	0.557***	0.569***
l_h_edu_o				0.316***
l_emp_tot_o				0.098***
Constant	6.702***	-23.709***	-10.543***	-7.849***
Observations	361,120	253,544	253,086	251,718
Number of id	42,032	29,641	29,641	29,641
pseudo R2	0.519	0.641	0.627	0.646

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1;
the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 11. Estimation results of the econometric model of tourist inflows to metropolitan regions in 2010-2018 – metropolitan destinations

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(MET1)	(MET2)	(MET3)	(MET4)
l_gdp_o	1.692***	-0.144***	1.071***	0.906***
l_dist	-0.794***	-0.734***	-0.717***	-0.948***
l_popul_o	-0.213***	1.472***	0.331***	0.294***
domestic		1.686***	1.740***	1.652***
l_gdp_pc_o		2.154***		
l_disp_inc_o			0.872***	0.691***
l_pop_den_o		-0.006	0.003	0.007
l_pop_den_d		0.375***	0.512***	0.254***
new_eu_o		-0.004	0.112	0.064
non_euro_rel		-0.140***	-0.142***	-0.114***

l_natura_d		0.285***	0.303***	0.326***
temp_rel		0.070***	0.061***	0.064***
precip_rel		0.275***	0.347***	0.318***
island_d		0.818***	0.976***	0.767***
bord_reg_d		0.489***	0.602***	0.309***
language		0.356***	0.273***	0.242***
metro_o		-0.545***	-0.423***	-0.284***
l_h_edu_o				0.551***
l_emp_tot_o				-0.051
Constant	-1.073	-29.951***	-14.040***	-8.286***
Observations	192,745	121,791	120,937	120,574
Number of id	22,676	14,335	14,335	14,335
pseudo R2	0.467	0.618	0.584	0.624

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 12. Estimation results of the econometric model of tourist inflows to non-metropolitan regions in 2010-2018 – non-metropolitan destinations

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(nMET1)	(nMET2)	(nMET3)	(nMET4)
l_gdp_o	1.619***	-0.101***	1.049***	0.829***
l_dist	-1.286***	-0.892***	-0.930***	-1.048***
l_popul_o	-0.497***	1.258***	0.136***	0.023
domestic		0.962***	0.992***	0.920***
l_gdp_pc_o		2.121***		
l_disp_inc_o			0.937***	0.720***
l_pop_den_o		-0.031***	-0.013***	-0.007**
l_pop_den_d		0.013***	0.011***	0.010***
new_eu_o		0.106**	0.278***	0.192***
non_euro_rel		-0.104***	-0.040***	-0.054***
l_natura_d		0.177***	0.179***	0.222***
temp_rel		0.041***	0.038***	0.030***
precip_rel		0.442***	0.533***	0.444***
island_d		0.691***	0.740***	0.793***
bord_reg_d		0.252***	0.222***	0.191***
language		0.445***	0.368***	0.355***
metro_o		-0.415***	-0.262***	-0.195***
l_h_edu_o				0.539***
l_emp_tot_o				0.183***
Constant	6.648***	-24.646***	-8.192***	-4.177***
Observations	531,863	335,405	333,053	332,048
Number of id	62,572	39,480	39,480	39,480
pseudo R2	0.534	0.637	0.62	0.637

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 13. Estimation results of the econometric model of tourist outflows from metropolitan regions in 2010-2018 – metropolitan origins

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(MET5)	(MET6)	(MET7)	(MET8)
l_gdp_o	1.560***	0.340***	1.190***	0.962***
l_dist	-0.958***	-0.821***	-0.787***	-0.958***
l_popul_o	0.233***	1.155***	0.334***	0.172***
domestic		1.125***	1.122***	1.169***
l_gdp_pc_o		1.538***		
l_disp_inc_o			0.886***	0.570***
l_pop_den_o		-0.080***	-0.009	-0.053***
l_pop_den_d		0.015***	0.018***	0.014***
non_euro_rel		-0.100***	-0.015	-0.037***
l_natura_d		0.136***	0.147***	0.230***
temp_rel		0.084***	0.078***	0.046***
precip_rel		0.563***	0.607***	0.384***
island_d		0.495***	0.518***	0.678***
bord_reg_d		0.240***	0.221***	0.175***
new_eu_o		0.588***	1.198***	0.798***
language		0.541***	0.566***	0.472***
metro_d		0.557***	0.552***	0.570***
l_h_edu_o				0.733***
l_emp_tot_o				0.099***
Constant	-5.759***	-23.082***	-14.027***	-7.234***
Observations	191,512	120,856	119,940	119,256
Number of id	22,496	14,366	14,366	14,366
pseudo R2	0.484	0.642	0.634	0.664

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; the Hausman-Taylor estimation results.
Source: Own elaboration.

Table 14. Estimation results of the econometric model of tourist outflows from non-metropolitan regions in 2010-2018 – non-metropolitan origins

	dependent variable: $log(Tourists\_flow_{ijt})$			
VARIABLES	(nMET5)	(nMET6)	(nMET7)	(nMET8)
l_gdp_o	1.648***	-0.269***	0.936***	0.729***
l_dist	-1.329***	-0.914***	-0.936***	-1.052***
l_popul_o	-0.608***	1.398***	0.221***	0.171***
domestic		1.051***	1.069***	0.997***
l_gdp_pc_o		2.363***		
l_disp_inc_o			1.013***	0.833***
l_pop_den_o		-0.016***	-0.012***	-0.003
l_pop_den_d		0.036***	0.038***	0.029***
new_eu_o		0.064	0.233***	0.143***
non_euro_rel		-0.144***	-0.137***	-0.123***
l_natura_d		0.242***	0.246***	0.273***
temp_rel		0.043***	0.039***	0.038***
precip_rel		0.382***	0.442***	0.403***
island_d		0.578***	0.607***	0.643***
bord_reg_d		0.206***	0.188***	0.160***

language		0.411***	0.346***	0.333***
metro_d		0.527***	0.526***	0.537***
l_h_edu_o				0.508***
l_emp_tot_o				0.125***
Constant	8.390***	-27.092***	-8.733***	-5.714***
Observations	533,096	336,340	334,050	333,366
Number of id	62,752	39,449	39,449	39,449
pseudo R2	0.533	0.652	0.642	0.656

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1;
the Hausman-Taylor estimation results.
Source: Own elaboration.

4 Key questions for subsequent tasks

4.1 Task 2. Pan-European systemic analysis

- Tourism flows should be contrasted with migration, commuting, and student flows, i.e. other types of flows analysed in the ESPON IRiE project. Such analyses would show the significance of differences between tourism flows and other types of inhabitant mobility, and thus allow us to construct a typology of tourism regions based on the balance of flows.
- To construct the typology, it was necessary to distinguish between tourism outflow and inflow. Our analyses indicated that the balance in individual countries varied greatly and that treating tourism flows together — as the sum of tourism flows in both directions — could result in erroneous conclusions.
- The typologies identified in our analyses are interesting and informative. However, by narrowing the list of indicators to a common list comparable with other flows, we found that some aspects could be extended.

4.2 Task 3. Scenarios

Every aspect of this report is subject to variation in the context of any of the big scenarios considered in Task 3 of the IRIE project. BREXIT, the new trends in the globalization, the long-term effects of the COVID-19 pandemic and the EU New Green Deal have direct and indirect effects on the intensity and shapes of the regional relationships described here as structural. Though we cannot exhaust all potential effects, we might consider the following:

- At the end of May 2021 there was a workshop with national and regional stakeholders. It concluded that tourism flows might be modified the most by the Covid-19 scenario, and in principle might not be affected by the Green Deal. Covid-19 has certainly already had an impact on tourism flows and will continue to have one in the short term, but the key question remains what its impact will be in the long term. On the other hand, the impact of the Green Deal seems to be underestimated. Reducing greenhouse gas emissions may cause fundamental changes in the structure of transport, which may, however, have a significant impact on the means of transport used, and thus on tourist destinations. Everything related to transport services, tourism, and the local mobility of people (commuting) is expected to experience the most severe adjustments if the EU and the whole planet really seeks to fulfil the national commitments of the Paris Agreement.
- It would seem that, since the UK was not part of the Schengen area, Brexit will not significantly affect tourism flows. However, as the ongoing processes after UK left the EU have shown, the relationship is not so obvious and the issue will require more detailed analysis. In the case of the New Globalization, it is difficult to see a direct impact on tourism flows. However, it should be remembered that the rise in public indebtedness will affect the economic growth of certain locations in the coming years. Therefore, the impact of scenarios on tourism flows cannot be considered only in the form of actual flows (e.g. restrictions in air traffic). An equally important element is the economic and geopolitical situation of the areas sending and receiving tourists. This approach seems to be absent in the current discourse and should find a special place in the analyses of Task 3.

4.3 Task 4. Policy implications

In this report we have described the methodology used to obtain consistent estimates for intranational and international trade tourism flows at the NUTS 2 level for the entire researched area (EU27+UK+EFTA). To the best of our knowledge, the ESPON IRIE project offers unique datasets of region-to-region tourism flows, with a common methodology and a basis in the available national statistics.

The dataset fills an important gap in the current European statistical system, providing a new understanding of the regional interlinkage for tourism, which accounts for a large part of the economy but had never been researched in this manner.

The analyses conducted here suggest that access even to country-to-country tourism flows was rare and region-to-region really unique. Until now, most EU countries have not enjoyed a solid and comparable estimate of intra-national and inter-national tourism flows, as it has not been possible to study the relevant matters, such as the current state of internal/external tourism integration.

From a methodological viewpoint, the long process followed here helps to illustrate how far we are from the automatic generation of official region-to-region flows, like the country-to-country flows we are accustomed to having at our disposal. All in all, we expect our approach to open a line of work to make this complex exercise less exceptional. We are convinced that by standardizing these procedures, and their corresponding improvements, we will be able to generate the solid figures required to follow up on the fulfilment of European tourism flows, not just at the country level, but also at regional or even lower spatial levels.

There is a high inertia in time in the observed flows. This means that the dominant directions of the tourism trips of the residents of a given country did not change substantially. This is very important for the stakeholders and policymakers of tourism development. Moreover, we can indicate countries with universal attractiveness for tourism flows (mainly Mediterranean and Alpine countries), where the number of tourist arrivals does not depend on the distance from places of residence. More importantly, we can indicate a whole range of country attractiveness, where a given country is a tourist destination for people from neighbouring countries. The presented results indicate potential directions of tourism marketing for individual countries — either to strengthen the dominant destinations or to open up to new destinations. These unique results, showing actual flows of tourists to and from a given region, provide great opportunities for the agencies and institutions responsible for tourism development to develop their strategies and policies.

The conducted research has unambiguously proved the great importance of internal tourism flows. Usually in tourism research the focus is on international tourism, in terms of marketing and the creation of the tourist offer. It is also often said that domestic tourism is neglected. However, the results of the ESPON IRiE project show the great importance of interregional tourism flows within individual countries — especially large ones such as Germany, France, Italy, Spain, and Poland. This observation also carries great weight for stakeholders creating national strategies for tourism development. For most countries foreign tourism is important but usually complementary to domestic tourism flows. The results presented should shed new light on tourism development activities. In addition, in the last two years, because of the Covid-19 pandemic, these processes have clearly deepened and further highlighted the need for greater attention.

Finally, we should mention how important it is to compare the results for tourism flows with strategic documents (regional strategies, sectoral strategies) on the European and national level. The project's results and recommendations should affect the long-term visions of the Member States and Brussels (at present the importance and dynamics of the interregional flows are difficult to perceive).

5 Recommendations for data providers to improve data quality

Data providers are understood in this report as the National Statistical Offices of the EU, the UK and the EFTA countries, which gather data on tourism flows from border crossings or various service providers, mainly of accommodations. National Statistical Offices then analyse these data and publishing raw data and/or results.

For tourism flows, we have several recommendations for data providers :

- Publish existing raw data: The ESPON IRiE's researchers observed that many countries would release raw data only on appeal to the respective statistical offices;
- Gather more information from the sources: Tourism data gathered in accommodation units usually do not detail the arriving tourist's origin beyond the name of the country. Data on the region of residence would be helpful for all interested parties in the tourism sector;
- EUROSTAT's international tourism database should provide series of data for both the national and the regional/interregional level (between NUTS 2 in Europe): If these data are unavailable (from national statistical offices), EUROSTAT should urge member countries to start gathering and providing them.

The definition and collection of tourism statistics remains a major methodological challenge. The results obtained in this report show the heterogeneity of approaches among countries, particularly for domestic tourism flows. Note that the dataset delivery in D4 and this report should be considered provisional. Although all our estimation processes have been developed with quality standards, certain aspects require further checks: the results of the estimations need to be double-checked, as individual cases with errors became apparent (as indicated in the methodology chapter, some of the data were estimated). However, the results presented in this report represent the best output we could get from the available data on tourist flows, and our methodology could be useful for further estimations of R2R tourist flows.

6 Conclusions

This report describes the methodology we used to generate the most updated series of C2C and R2R tourism flows in Europe. The authors are not aware of any other similar methodology of estimation. One of the goals of the ESPON IRiE project is to generate new data and relevant evidence for the main interregional tourism flows in the EU27, the UK and the EFTA countries.

The accessible data on tourism flows are incoherent in their definitional basis and in the methodology used in their gathering by different statistical agencies, but two complex sources (EUROSTAT and UNWTO) are incomplete even at the C2C level. The best data for R2R matrix estimation came from EUROSTAT and concerned stocks of yearly domestic arrivals to NUTS 2.

In order to create the data — complex information on spatial and temporal allocation of tourism flows within the researched area — we used two procedures: estimation of the gaps for the C2C matrix and disaggregation data for the R2R matrix. We developed a set of indexes for tourism-flow description with four dimensions (intensity, balance, concentration, and distance impact) as well as a synthesis of three simple bi-dimensional typologies.

The observed tourism flows between all researched countries in the period 2010-2018 have a high inertia in time. This means that the dominant directions of the tourism trips of the residents of a given country did not change substantially. It is possible to indicate a number of countries that are attractive for residents of at least several countries: Austria, Germany, Greece, Spain, France, Croatia, Italy, and Portugal. These are located on the Mediterranean (summer tourism) and in the Alps (winter tourism).

For R2R tourism flows, we observed a large number of interregional flows within the largest countries of the Mediterranean region (such as France, Spain, and Italy). We observed less intense but still large domestic interregional movements of tourists within the Scandinavian countries (Norway, Sweden, and Finland), Ireland, the Netherlands, Portugal, Poland, Switzerland, Germany, Romania, Bulgaria, and Greece. For international interregional tourism flows, the north-south axis dominated. Here we saw tourism flows between the regions of Central Europe and Adriatic Croatia; between the regions of Finland, Estonia, Latvia, and Lithuania; and between Germany and Austria, France and Spain, France and Italy, and the regions of the UK with NUTS 2 regions in France and Spain.

Tourist-intensity analysis showed the greatest impact of tourism in the mountainous parts of Austria, Switzerland, and Italy (Alps, ski resorts) and the NUTS 2 region of Adriatic Croatia, situated along the nearest Mediterranean coast to the European continent (and tourism demand markets). Other areas with tourist-intensity values of 10 to 20 cover southernmost Portugal, the Balearic islands, Corsica, the Greek islands in the Ionian and Aegean seas, southern Norway, northern Denmark, the Finnish autonomous region of the Åland Islands, and the southernmost NUTS 2 regions of the Netherlands and Belgium.

The greatest imbalance in tourism flows within the regions where inflow dominated was observed along the Mediterranean and Baltic seas, more precisely in the coastal regions of Spain, France, and Croatia in the Mediterranean region, and in Denmark, northeastern Germany, and northwestern Poland in the Baltic Region.

Inflow to regions in northern and eastern Europe came from a rather small area, and were mainly domestic, while inflow to tourism destinations in southern Europe was dispersed. Outflow was the other way around: tourists from northern and eastern Europe visited different regions, while those from southern Europe did not move to dispersed areas outside the Mediterranean zone.

As for distance of tourism flows, tourism in northern Europe remained largely within the region, with not many travelling from faraway places to the regions of northern Scandinavia, the northern UK, Estonia, or Latvia. From the other side, inhabitants of mentioned regions would travel larger distances to reach tourism destination.

The explanatory analysis showed a positive correlation tourism flows with the share of naturally valuable areas, the difference in rainfall between origin and destination regions, the lack of linguistic differences between regions, and the choice of island destinations. Additionally, it should be pointed out that the number of tourists correlated with the level of education of the inhabitants, namely with the participation of people with the highest level of education and, to a lesser extent, with the general level of employment. However, the models we developed for explanatory analysis of origin and destination grouped into specific areas (coastal, mountain, and metropolitan areas, and, for equilibrium without these features, non-coastal, non-mountain, and non-metropolitan areas) suggested no significant differences between the different types of areas. In general, our results coincided with those of the general model which does not take into the account different specific areas.

7 References

Baltagi, B. H., Khanti-Akom, S. (1990). On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. Journal of Applied Econometrics 5: 401–406.

Baltagi, B., Egger, P., Pfaffermayr, M. (2003). A Generalized Design for Bilateral Trade Flow Models. Economics Letters. 80: 391-397. 10.1016/S0165-1765(03)00115-0.

Basher, B, Fachin, S. (2008). The long-term decline of internal migration in Canada: the case of Ontario. Letters in Spatial and Resource Sciences 1. 171–81.

Beine, M., Bourgeon, P., Bricongne, J.C. (2019). Aggregate fluctuations and international migration. Scandinavian Journal of Economics. 121(1): 117–152. 10.1111/sjoe.12258

Beine, M., Parsons, C. (2015). Climatic factors as determinants of international migration. Scandinavian Journal of Economics. 117. 723–67.

Brunet, R. (1989). *Les villes europeénnes: Rapport pour la DATAR*. Montpellier: RECLUS

Egger, P., Pfaffermayr M. (2003). The Proper Panel Econometric Specification of the Gravity Equation: A Three Way Model with Bilateral Interaction Effects. Empirical Economics. 28: 571-580. 10.1007/s001810200146.

EUROSTAT, 2021. Tourism trips of Europeans https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism_trips_of_Europeans (obtained 02/06/2021).

Hausman, J, Taylor, W. (1981). Panel data and unobservable individual effects. Econometrica. 49: 1377–1398.

Khan, M.Z., Hossain, M. (2010). Model of Bilateral Trade Balance: Extensions and Empirical Tests. Economic Analysis and Policy, 40(3), 377-391. 10.1016/S0313-5926(10)50037-7

Leibenstein, H., Tinbergen, J. (1966). Shaping the World Economy: Suggestions for an International Economic Policy. The Economic Journal. 76(301): 92-95. 10.2307/2229041.

Mayda, A.M. (2010). International migration: a panel data analysis of the determinants of bilateral flows. Journal of Population Economics 23(4): 1249–1274. 10.1007/s00148-009-0251-x

McElroy J. L., Albuquerque, K. (1998). Tourism penetration Index in small Caribbean Islands. Annals of Tourism Research 25 (1). pp. 145-168.

Molloy, R., Smith, C., Wozniak, A. (2011). Internal migration in the United States. Journal of Economic Perspectives. 25(3): 173–196. 10.1257/jep.25.3.173

Runge J., 2006, Metody badań w geografii społeczno-ekonomicznej: elementy metodologii, wybrane narzędzia badawcze, Podręczniki i Skrypty Uniwersytetu Śląskiego, 59, Wydawnictwa Uniwersytetu Śląskiego, Katowice.

Santos-Silva, J.M.C., Tenreyro, S. (2006), The log of gravity, The Review of Economics and Statistics. 88(4): 641-658.

Serlenga, L., Shin, Y. (2021). Gravity models of interprovincial migration flows in Canada with hierarchical multifactor structure. Empirical Economics. 60, 365–390.

Shao, Y., Huang, S., Wang, Y., Li, Z., Luo, M., 2020. Evolution of international tourist flows from 1995 to 2018: A network analysis perspective. Tourism Management Perspectives, Volume 36, 100752, https://doi.org/10.1016/j.tmp.2020.100752.

Stock, J. H., Wright, J. H., Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20: 518–529.

Strahl D. (red.), 2006, Metody oceny rozwoju regionalnego, Wydawnictwa Akademii Ekonomicznej, Wrocław.

UNWTO, 2017. UNWTO Annual Report 2017. Madrid: World Tourism Organization.

UNWTO, 2018. European Union Tourism Trends. Madrid: World Tourism Organization.

UNWTO, 2021. Tourism Flows – Source markets and Destinations - Project powered by Telefónica. https://www.unwto.org/tourism-flows-source-markets-and-destinations (obtained 02/06/2021).

van Bergeijk, P., Brakman, S. (2010). The Gravity Model in International Trade. Cambridge University Press, Cambridge. 10.1017/CBO9780511762109.

Viegas, J., Silva, E.A., Martinez L., 2009, Effects of the Modifiable Areal Unit Problem on the Delineation of Traffic Analysis Zones, Environment and Planning B – Planning and Design, 36(4): 625–643.

Co-financed by the European Regional Development Fund

Inspire Policy Making with Territorial Evidence

espon.eu

ESPON 2021

ESPON EGTC
4 rue Erasme, L-1468 Luxembourg
Grand Duchy of Luxembourg
Phone: +352 20 600 280
Email: info@espon.eu
www.espon.eu

The ESPON EGTC is the Single Beneficiary of the ESPON 2020 Cooperation Programme. The Single Operation within the programme is implemented by the ESPON EGTC and co-financed by the European Regional Development Fund, the EU Member States and the Partner States, Iceland, Liechtenstein, Norway and Switzerland.