Modelling the regional distribution of international travelers in Spain to estimate imported cases of dengue and malaria: statistical inference and validation study

David García-García, Beatriz Fernández-Martínez, Frederic Bartumeus, Diana Gómez-Barroso

Submitted to: JMIR Public Health and Surveillance
on: July 24, 2023

Table of Contents

Modelling the regional distribution of international travelers in Spain to estimate imported cases of dengue and malaria: statistical inference and validation study

David García-García[1]; Beatriz Fernández-Martínez[1]; Frederic Bartumeus[2]; Diana Gómez-Barroso[1]

[1]Epidemiology and Public Health Biomedical Network Research Consortium (CIBERESP) and National Centre of Epidemiology (CNE) Instituto de Salud Carlos III (ISCIII) Madrid ES
[2]Centre for Advanced Studies of Blanes (CEAB-CSIC) and Centre for Research on Ecology and Forestry Applications (CREAF) and Catalan Institute for Research and Advanced Studies (ICREA) Blanes ES

Corresponding Author:
David García-García
Epidemiology and Public Health Biomedical Network Research Consortium (CIBERESP) and National Centre of Epidemiology (CNE)
Instituto de Salud Carlos III (ISCIII)
Monforte de Lemos 5
Madrid
ES

Abstract

Background: Understanding the patterns of disease importation through international travel is paramount for effective public health interventions and global disease surveillance. While global airline network data has been utilized to assist in outbreak prevention and effective preparedness, accurately estimating how these imported cases disseminate locally in receiving countries remains challenging.

Objective: To describe and understand the regional distribution of imported cases of dengue and malaria upon their arrival in Spain via air travel.

Methods: We propose a method to describe these regional distributions, based on the computation of a “travelers’ index” from readily available socioeconomic data. We validate our estimates comparing them with the reported cases of malaria and dengue in Spain during 2015-2019.

Results: The estimates provided by the best fitted model showed high correlation with notified cases of malaria (0.94) and dengue (0.87), with economic motivations being the most relevant for imported cases of malaria and visits to friends and relatives for dengue imported cases.

Conclusions: Factual descriptions of the local movement of international travelers may improve substantially the design of cost-effective prevention policies and control strategies, and pose an essential contribution to decision-support systems. Our approach contributes in this direction by providing a reliable estimate of the risk of imported cases of non-endemic diseases, which could be generalized to other applications.

(JMIR Preprints 24/07/2023:51191)
DOI: https://doi.org/10.2196/preprints.51191

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Modelling the regional distribution of international travelers in Spain to estimate imported cases of dengue and malaria: statistical inference and validation study

David García-García[1][2]*, Beatriz Fernández-Martínez[1][2], Frederic Bartumeus[3][4][5], Diana Gómez-Barroso[1][2]

1 National Centre for Epidemiology, Instituto de Salud Carlos III (ISCIII), Madrid, Spain
2 Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain
3 Centre for Advanced Studies of Blanes (CEAB-CSIC), Blanes, Spain
4 Centre for Research on Ecology and Forestry Applications (CREAF), Cerdanyola del Vallès, Barcelona, Spain
5 Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain
* Corresponding author

Abstract

Background
Understanding the patterns of disease importation through international travel is paramount for effective public health interventions and global disease surveillance. While global airline network data has been utilized to assist in outbreak prevention and effective preparedness, accurately estimating how these imported cases disseminate locally in receiving countries remains challenging.

Objective
To describe and understand the regional distribution of imported cases of dengue and malaria upon their arrival in Spain via air travel.

Methods
We propose a method to describe these regional distributions, based on the computation of a “travelers’ index” from readily available socioeconomic data. We combine indicators representing the main drivers for international travel: tourism, economy and visits to friends and relatives, to measure the relative appeal of each region in the importing country for travelers. We validate the resulting estimates by comparing them with the reported cases of malaria and dengue in Spain during 2015-2019. We also assess which motivation provided more accurate estimates for imported cases of both diseases.

Results
The estimates provided by the best fitted model showed high correlation with notified cases of malaria (0.94) and dengue (0.87), with economic motivations being the most relevant for imported cases of malaria and visits to friends and relatives for dengue imported cases.

Conclusions
Factual descriptions of the local movement of international travelers may enhance substantially the design of cost-effective prevention policies and control strategies, and pose an essential contribution to decision-support systems. Our approach contributes in this direction by providing a reliable

estimate of the number of imported cases of non-endemic diseases, which could be generalized to other applications. Realistic risk assessments will be obtained by combining this regional predictor with the vectors' observed local distribution.

Keywords

Epidemiology, imported infections, modelling, surveillance system, vector-borne diseases

Introduction

Throughout history, human mobility has been a key determinant for the spread of infectious diseases. From the 14th century bubonic plague pandemic to the 1918 Spanish flu, and the more recent Ebola epidemic and COVID-19 pandemic, the way individuals travel across the globe has shaped the evolution and geographical dynamics of infectious diseases [1]–[3].

International mobility flows are especially relevant for the spread of vector-borne diseases (VBDs), which often receive less attention in routine epidemiological surveillance plans in countries where they are not endemic. In recent years, global warming and intensified urbanization processes have favored the establishment of previously foreign species around the globe, such as *Aedes, Anopheles* or *Culex* mosquitoes [4]–[7]. These are vectors for malaria, dengue, yellow fever, West Nile virus, Zika, and chikungunya, and thus pose a significant public health risk that needs adequate preparedness [8]–[10]. Air travel plays a central role in the diffusion of most of these diseases, allowing their spread through imported cases at non-endemic locations [11]–[13] with vector presence, and must be incorporated in decision-support systems to achieve operational preparedness and risk prediction [14]–[16]. Reliable descriptions and predictions of migration flows have proved to be a valuable tool for the design of more effective public health policies [17]–[20]. However, new developments are always in need as complex dynamics are expected to arise from the multi-step life cycle of VBDs [21]–[23].

A wide range of approaches have been used to model risks of imported cases of dengue [24]–[27] and malaria [28]–[30] from endemic to non-endemic regions. Several works [31]–[33] incorporate data on the global airline network to assist outbreak prevention and public health preparedness. However, accurate estimations of how such imported cases disseminate locally once arrived in the receiving countries are harder to devise. While fine-grained data on local mobility is available from cellular networks [34]–[36], it is not clear how the specific behavior of travelers can be differentiated from the local population dynamics. Moreover, travel-related data is usually obtained through coarse-grained spatial statistics, thus involving large territories. A reason for this is the inherent complexity and range of scales involved in human flows: while it is possible to record information at designated locations (for instance, at airports), it is much more difficult to reliably collect detailed movement data of targeted groups over larger geographical regions. In these cases, insights on the nature of human travelling behavior and motivations that uncover hidden patterns in these processes are crucial, as they may be used to sidestep the need for excessively detailed (and thus unreachable) data [37]–[39].

We aim to provide accurate descriptions of how infected travelers may distribute in a territory, as a valuable input for local authorities in the design of cost-effective VBDs prevention and control strategies. To do so, we approximate the local distribution of travelers arriving at a specific country (or any other territory) in terms of readily available indicators, rather than on travel information that is usually not quantified at a local scale. These statistics gauge the appeal of each region to foreign travelers, quantifying the number of imported cases each of them may receive. We calibrate our model with the number of imported cases of dengue and malaria at each province in Spain during 2015-2018, and then validate it comparing our model’s estimates for the number of imported cases in 2019 with official data.

Methods

We first develop a theoretical framework to estimate how travellers distribute throughout the territory after their arrival at the country. See figure 1 for a schematic description of this approach.

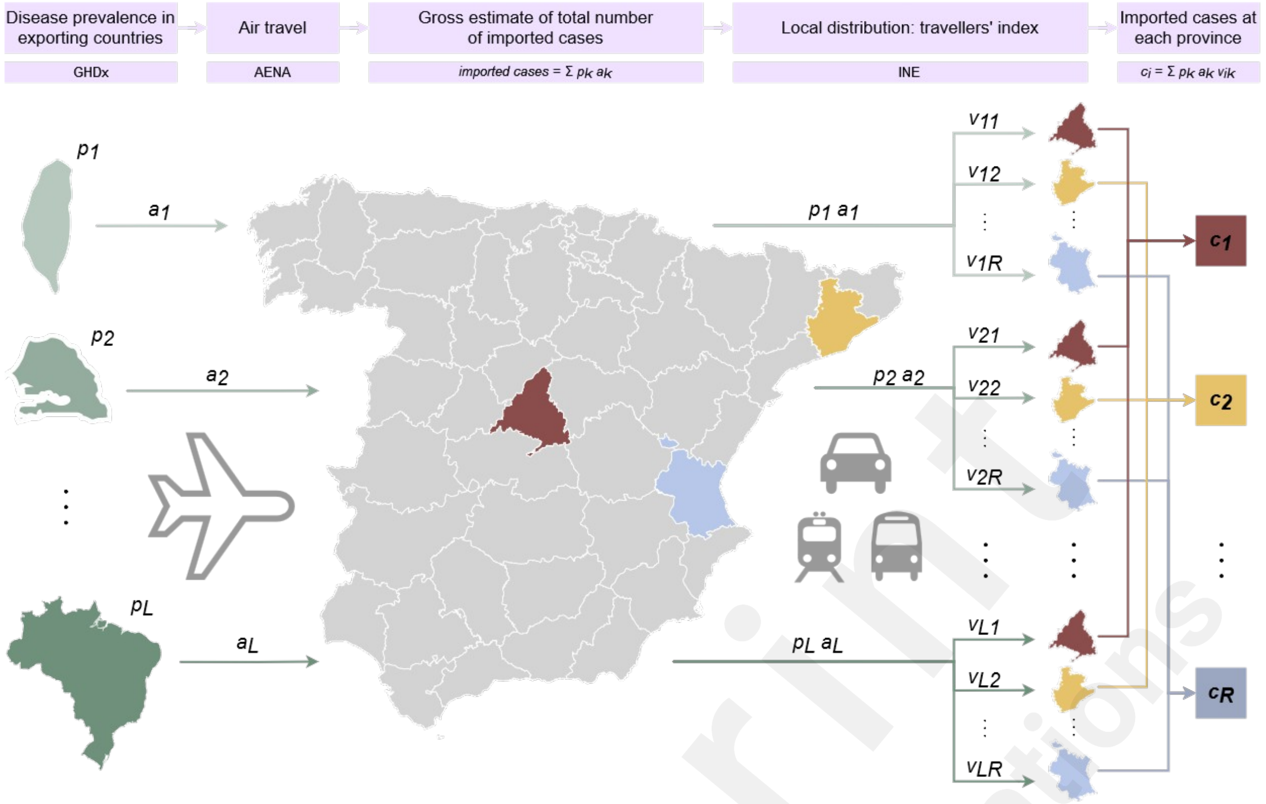


Figure 1. Summary of the rationale behind our approach. Infected travelers arrive to the importing country from L exporting countries over the world by means of air travel. For each country k, the prevalence of the disease at the country (p_k) and the number of travelers arriving from it to the importing country (a_k) are combined to obtain an estimate of the number of imported cases ($p_k a_k$). These spread across the R regions or spatial units of the importing country following a distribution that can be estimated by means of "traveler's indexes". These indexes are computed from local statistics concerning economic and touristic activity, and the number of foreign residents at each region. The traveler's index v_{ki} measures the proportion of imported cases from country k that moves to local region i upon arrival to the importing country. The total number of cases expected to arrive at the region (c_i) is then obtained as the sum of the estimated number of imported cases over all exporting countries: $c_i = \sum_k p_k a_k v_{ik}$.

Input data

We used the following yearly statistics during 2015-2019 to compute the relative appeal of each province in Spain to international travelers. All these are publicly available and curated by the Spanish National Statistics Institute INE [40]:

- **Touristic indicators**. For both hotels and touristic apartments, we used the following variables: total capacity, number of national travelers, number of foreign travelers, number of overnight stays by national travelers, and number of overnight stays by foreign travelers.
- **Economic indicators**. We considered each province's population, gross domestic product (GDP), GDP per capita, number of private limited companies (*Sociedades Limitadas*, SL) and number of public limited companies (*Sociedades Anónimas*, SA).

- **Indicators for visits to friends and relatives**. For each country and province, we used the number of foreign residents by nationality, the number of foreign residents by birthplace and the number of national residents by birthplace (other than Spain).

Three more data inputs are needed in our approach:

- **Arrivals data.** We computed the yearly number of travelers arriving to Spain during 2015-2019 from each of the 100 airports with the largest flows of incoming travelers aggregating publicly available monthly data provided by the public entity in charge of the Spanish Airports and Aerial Navigation (*Aeropuertos Españoles y Navegación Aérea*, *AENA*) [41].
- **Disease data.** We used yearly prevalence estimates for malaria and dengue during 2015-2019 provided by Global Health Data Exchange (GHDx) [42]. Data on malaria is supplied by the Malaria Atlas Project [43].
- **Cases in Spain**. The number of imported cases of malaria and dengue (including any type of infection by dengue virus) are reported by each province to the Spanish National Surveillance System (*Red Nacional de Vigilancia Epidemiológica*, RENAVE).

Travelers' index

Let us describe the rationale used to combine the above statistics in an informed indicator to estimate the international travelers' propensity to move to a specific region. We consider a country that receives travelers from any other countries of the world (these will be denoted as the *importing* and *exporting* countries respectively in the following). We assume that the importing country is divided in *regions*, which may represent geographical regions or administrative units, for instance.
Motivations for international travel are usually classified in three major categories: tourism, business, and visits to friends and relatives (VFR) [44,45]. Based on this principle, we compute the relative importance of each region in the importing country in terms of each of these drivers:

$$t_i = \text{relative importance of region } i \text{ in terms of tourism},$$
$$e_i = \text{relative importance of region } i \text{ in terms of economy},$$
$$r_{ik} = \text{relative importance of region } i \text{ in terms of visits to friends}$$
$$\text{and relatives for travelers arriving from country } k.$$

Note the last indicator depends on the exporting country k. For instance, when computing travelers' indexes with $k = Brazil$ in formula (1), only the number of Brazilian residents participates in the computation. The three indicators are computed from several available statistics concerning the regions of the country (see Section *Data*), as the relative contribution of each region to the country's total. For instance, when using GDP to measure economic status, the relative importance of region i in terms of economy would read

$$t_i = \frac{GDP\ of\ region\ i}{total\ GDP\ of\ the\ importing\ country}. \quad (1)$$

We then compute a *travelers' index* for each region i of the importing country and any exporting country k as the average of the three indicators:

$$v_{ik} = \frac{1}{3}(t_i + e_i + r_{ik}). \quad (2)$$

It follows from equations (2) and (1) that for each exporting country k, the sum of the travelers' indexes of this country over the regions i of the importing country equals

$$\sum_i v_{ik} = 1.$$

Therefore, given an exporting country k, the travelers' indexes v_{ik} may be understood as an estimate of the portion of travelers a_k arriving from this country to each province i. Hence, we can estimate the total number of travelers arriving at a given region i as

$$v_i = \sum_k a_k v_{ik}, \quad (3)$$

where the sum runs over all exporting countries k, and a_k is the total number of travelers arriving from any such country k. Accordingly, we estimate the number of imported cases of a disease at each province i as the sum of the imported cases from each of the exporting countries k

$$c_i = \sum_k p_k a_k v_{ik}, \quad (4)$$

where p_k denotes the prevalence of the disease at the exporting country k. Plainly speaking, we estimate the total number of cases arriving from country k as the product of the total number of travelers arriving from the country and the prevalence of the disease at the country. We then estimate how these cases disseminate across regions by means of the travelers' indexes, that assign to travelers a relative importance or preference for each of these regions. Adding these local distributions over all the exporting countries results in the total number of expected cases at each of the regions in the importing country. See Figure 1 for a visual representation of this reasoning.

Model calibration and validation

In order to test the validity of our approach, we follow the pipeline depicted in Figure 2. The steps involved in this process are:

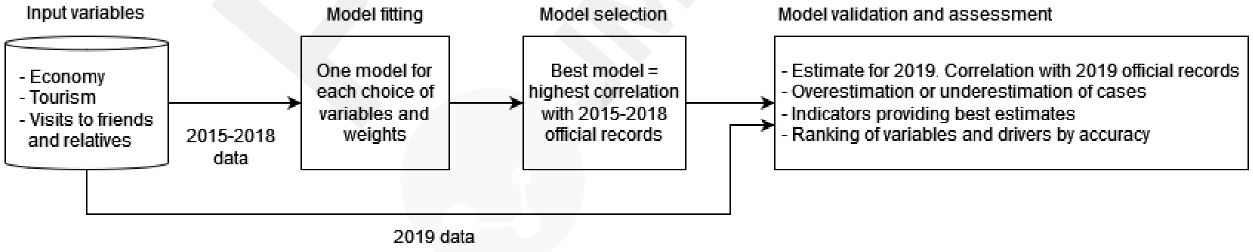


Figure 2. Summary of the model's building process. Key steps are the fitting, selection, validation and assessment of the model.

1. **Input variables.** We consider data from Spain (importing country) and its 52 provinces (regions over which the imported cases disseminate). We compute the relative importances t_i, e_i, r_{ik} using each of the statistics listed in Section *Data* for each of the three drivers (i.e., economy, tourism, friends and family).
2. **Model fitting.** We then use Eq. (2) to build travelers' indexes: one for each possible combination of the indicators. We combine these with arrivals and prevalence data to obtain estimates for the expected number of cases at each province for 2015-2018 (Eq. (4)),

resulting from simple averages of the indicators t_i, e_i, r_{ik}. We also consider a generalized version of the traveler's index, by replacing the average in Eq. (2) with a weighted average:

$$v_{ik}^{\alpha} = \alpha_1 t_i + \alpha_2 e_i + \alpha_3 r_{ik}, \qquad (5)$$

where $0 \leq \alpha_j \leq 1$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. The coefficients α_j measure the relative importance given to each of the three categories themselves (tourism, business, and VFR) in a particular model. For instance, an estimate obtained from a value of α_1 close to 1 and values of α_2 and α_3 close to 0 in equation (5) assumes that tourism is the most important driver for travelers from countries where the disease is endemic than business or VFR. Identifying the choice of weights α_j and indicators in formula (5) that provide a better estimate allows us to also understand which of the three drivers among t_i, e_i and r_{ik} plays a more significant role in the motivation of travelers carrying the disease.

3. **Model selection.** To identify which of the resulting travelers' indexes best approximate the number of imported cases per province, we compute the correlation between the estimates provided by each of our models (one for each combination of indicators) and the actually reported cases of malaria and dengue at each province for years 2015 to 2018. The model reporting the highest value for this correlation is then selected as the best model. We follow the same procedure for the case of the weighted averages; in this case, a different estimate is obtained for each choice of indicators and weights α_j.

4. **Model validation and assessment.** As a final test for accuracy, we compute the correlation between the best models' estimates for 2019 (data not used during the fitting and selection process) and the officially reported cases for this year, both for the simple and weighted averages. In case this correlation is high, we consider the models as validated and proceed to the next step.

   We assess three features of the resulting models. First, we fit a linear model explaining the estimated number of imported cases at each province in 2019 in terms of the officially reported ones. The coefficient of the linear model may be understood as the number of cases predicted by the model per officially reported case, thus informing of the overestimation or underestimation of the model's prediction of the actually reported cases. Note that this refers only to the raw number of cases, as the accuracy of their distribution is captured by the validated correlation. Second, we rank the contribution of each of the statistics considered in the model, by computing the average correlation with 2019 official data of those estimates obtained from models including each particular variable. We also compute the average loss of accuracy associated to each variable, as the difference in the average correlation of those models including and not including each statistic. This allows us to identify which statistics among the choices made for each indicator t_i, e_i, r_{ik} provide more reliable predictors of the disease. Third, we follow an analogous procedure for the assessment of the weighted averages, and compute the average correlation of those models built from each choice of indicator and weight α_j. Those indicators scoring higher for larger values of the corresponding weight are expected to inform of the motivations of the travelers carrying each disease, among business, tourism and VFR.

Human mobility model

Finally, we test the validity of our results against a well-established model for human mobility [46,47], assuming that the movement of travelers does not follow motivations different to those of resident populations. For this, we assume that, once arrived to their destination airport in Spain, a proportion q of the imported cases stays in that destination, and the rest moves to a different province following a well-known, generic, and random human mobility pattern [47,48]. The probability of these travelers to move to each province in Spain is assumed to follow a decaying power-law with exponent γ on the distance d between the origin and destination province centroids, $p(d) = d^{-\gamma}$.

We then group the total number of expected cases at each province as the sum of those that arrived to the province as their final flight destination, and those arriving from any other provinces by means of other transport modes reflected in a geographically bounded power law distribution. This model is then fitted to the data on the officially reported cases of dengue and malaria during 2015-2018 for values of q between 0 and 1 (the proportion of travelers that leave their destination province upon arrival) and values of γ between 1 and 5 (the exponent of the power law, with smaller values favoring longer-ranged movement, and higher values favoring shorter-ranged movement).

We follow an analogous philosophy for the model building and assessment process as in the travelers' index models. The parameters leading to a highest correlation with the reported cases for the period 2015-2018 are used to compute an estimate for 2019, and the correlation between this estimate and the 2019 official record is then computed to allow for comparison with the travelers' index model. A linear model between the human mobility model's estimate and the official 2019 records is also fitted to assess the underestimation or overestimation of the model.

Ethical Considerations

Our study used publicly available aggregated secondary data with no characteristics that allowed for individual identification. No relevant data protection and privacy issues to report.

Results

Input data

A preliminary analysis shows that among all statistics used, those concerning the same drivers are usually highly co-dependent, with some exceptions (for instance, GDP per capita, see Figures S1-S2). The relative importances appearing in the computation of the travelers' indexes $(\mathrm{Eq.}\,(2))$ showed little variation over the years (see Figure S3). This temporal stability has been observed before in the distribution of international [37] and national [49] human mobility flows across destinations, obtained from expressions analogous to Eq. (1).

The 100 airports with highest number of incoming travelers were located in 49 countries and accounted for a 99.75% of the total incoming travelers to Spain during years 2015 to 2019. Out of these, 10 countries were removed from our study since neither malaria nor dengue was present during the time span under study (according to prevalence data from GHDx), resulting in 39 exporting countries. Table 1 shows the number of incoming travelers from each of these countries, and the average prevalence of dengue and malaria for 2015-2019 as provided by GHDx.

Country	Incoming travelers	Malaria prevalence /100,000 population	Dengue prevalence /100,000 population
United States	6845337	0	0.54
Brazil	3472397	78.96	65.40
Colombia	3053635	197.47	54.97
Argentina	2972990	0	15.47
Peru	2101286	323.49	41.37
Mexico	2063577	7.26	37.29
Dominican Republic	1611336	1.50	49.11
Argelia	1562772	4.92	0
Venezuela	1241154	1065.28	48.17
Cuba	1062531	0	37.93
Cape Verde	1054106	135.17	36.41
Ecuador	1027244	86.97	37.84
Costa Rica	810214	0	67.74
Senegal	630192	2464.71	32.43
Panama	566765	115.70	56.47
Bolivia	556092	181.43	60.84
Gambia	522809	4237.62	33.78
Egypt	466705	0	10.88
Thailand	365318	68.10	58.10
Singapore	360619	0	68.44
Equatorial Guinea	360393	32981.86	39.18
China	359372	0.14	24.63
Pakistan	306321	536.27	41.90
Mauritania	300521	4095.70	24.57
El Salvador	293976	8.95	130.85
Republic of Korea	254457	15.80	0
Nigeria	200813	18792.47	38.73
Jordan	182431	0	12.45
Angola	176690	11182.79	27.97
Guatemala	174030	169.12	45.25
Saudi Arabia	166730	3.98	15.52
Ghana	149374	18512.96	40.65
Guinea	78545	30131.12	38.35
The Bahamas	49010	0	39.30
Gabon	38016	15756.90	44.25
Jamaica	14802	0	45.43
South Africa	12394	36.43	0
Cameroon	8994	19904.66	34.96
Mali	1165	16024.21	31.19

Table 1: Exporting countries. Total incoming travelers and average malaria and dengue prevalence (total cases per 100.000 inhabitants) for 2015-2019 for the 39 exporting countries considered in our study, ranked by number of travelers. Countries with no malaria and dengue prevalence have been removed from the list.

Estimates and model assessment

High correlation values were found for both malaria (0.94) and dengue (0.86) between the best models' estimates for 2019 and the notified cases. The models that provided the most accurate estimate included public limited companies, foreign travelers at hotels, and foreign residents by birthplace in their computations of the travelers' indexes. The same variables lead to the best estimate for both malaria and dengue. While considering weighted averages in the construction of the travelers' index did not improve the models' accuracy, different motivations were obtained for travelers carrying each of the diseases: economy seemed to best capture the appeal of each region for imported cases of malaria (a relative weight of 0.7, with GDP being the most accurate indicator), and VFR seemed to be the main motivation for travelers with dengue (relative weight of 0.9, assigned to the number of foreign residents in the province by birthplace). Different proportions of overestimation were found for each disease (99% for malaria, 86.5% for dengue). A summary of the relevant features of the models provided by the fitting and selection process is provided in Table 2.

Disease (model)	Economic indicator (weight)	Touristic indicator (weight)	VFR indicator (weight)	Pearson correlation of model's estimate with 2019 data	Overestimation
Malaria (simple)	Public limited companies (SA)	Foreign travelers at hotels	Foreign residents by birthplace	0.94	98.9%
Malaria (weighted)	GDP (0.7)	Foreign travelers at hotels (0.1)	Foreign residents by birthplace (0.2)	0.94	99%
Dengue (simple)	Public limited companies (SA)	Foreign travelers at hotels	Foreign residents by birthplace	0.86	86.5%
Dengue (weighted)	No contribution (0)	Foreign travelers at hotels (0.1)	Foreign residents by birthplace (0.9)	0.87	86.7%

Table 2: Summary of the models that most accurately approximated the 2015-2018 reported cases. Each row shows the statistics that provided the best estimate of imported cases of each disease), the correlation with the actually reported data in 2019, and the approximation for the proportion of overestimation as obtained from the linear models. For the models including weighted averages, the weight α_i of each indicator in included in parenthesis. If the weight of a given indicator is zero, this means that no contribution to the estimate is provided by the corresponding indicator.

Figure 3 shows the weighted models' fit and their estimates for 2019, together with the officially reported number of cases of each disease (malaria and dengue) at each province in Spain. Upon visual inspection, two provinces seemed to have a high influence in the models' fit. These correspond to Madrid and Barcelona, that host a much larger number of reported cases of both diseases. We excluded these provinces from the input dataset and repeated the analysis (Table S1, Figure S4). While a decrease in correlation was found overall (~0.12 over all models), the resulting estimates still showed high agreement with the official reports for 2019 (above 0.74 correlation with 2019 data, see Figure 3, Table S1 and Figure S4).

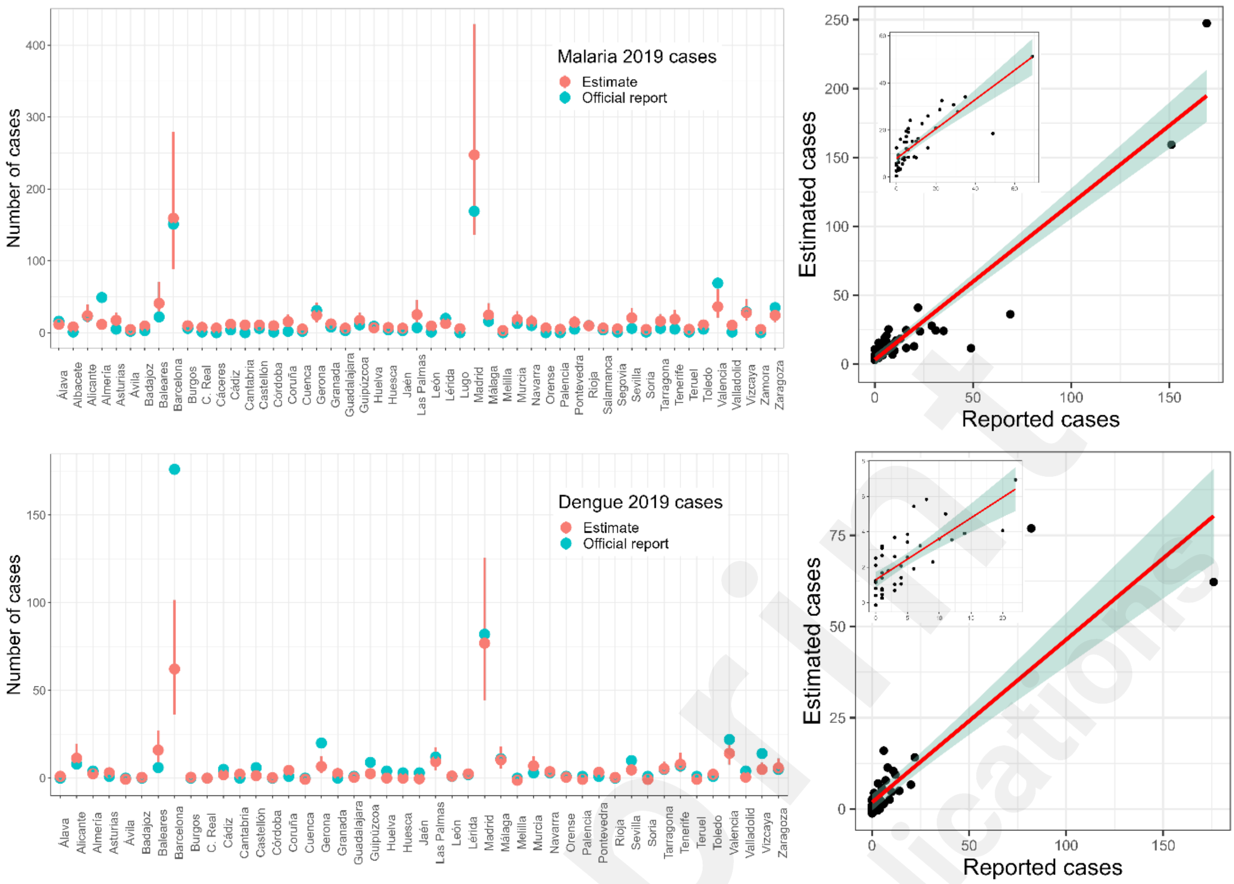


Figure 3. Summary of the best linear models for 2019 imported cases of malaria (top row) and dengue (bottom row). The left column shows the models' predictions (in red), together with the number of reported cases (in blue) for 2019 at each province in Spain. The right column shows the fit between the models' estimates and the official records (inset figures correspond to the fit after removing Madrid and Barcelona from the dataset, as explained in *Estimates and model assessment*).

We performed a residual analysis to check for normality and autocorrelation of the models' residuals. The malaria model showed close-to-normal residuals with no autocorrelation (statistically significant scores W=0.67 and DW=2.03 in Shapiro-Wilk and Durbin-Watson tests respectively). For the dengue model, a relevant deviation was caused by the estimate for Barcelona (see Figure 3). Exclusion of this outlier resulted in normally distributed and not autocorrelated residuals (statistically significant W=0.94 and DW=1.87). See Table S3 in the Supplementary Material for full details on the residual analysis.

Variable performance

The models constructed using simple averages provided a unanimous choice of indicators associated to economy, tourism and VFR. On the contrary, the best weighted models included different economic indicators: GDP provided the best estimate for imported malaria cases, while no influence of the economic indicator was considered in the best dengue model. In addition to this, different drivers were the most important ones for each disease, as shown by the much higher relative weight found for the economic motivations for malaria cases and for VFR for dengue cases respectively (Table 2).

When ranking the contribution of each of the variables to the accuracy of the models, similar results were found for both diseases, with some minor variations across variables (Table 3). Several statistics concerning tourism ranked the highest in this classification, although several others ranked in low position, indicating that appropriate choices of indicators may be important and need careful examination. All economic indicators provided an improvement (or absence of decrease) in correlation except for GDP per capita, which resulted in less accurate estimates (average decrease of ~0.08 in correlation with 2019 data). Indicators corresponding to VFR had mild average effects on the models’ outputs (largest variation in correlation with 2019 data was -0.02).

Variable	Malaria	Malaria (loss)	Dengue	Dengue (loss)
National travellers in hotels	0.83	0.08	0.74	0.05
Overnight stays by national travellers in hotels	0.83	0.07	0.73	0.05
Foreign travellers in hotels	0.82	0.07	0.76	0.08
Total hotels' capacity	0.81	0.05	0.74	0.05
Public limited companies (SA)	0.81	0.05	0.72	0.04
National travellers in touristic apartments	0.79	0.03	0.70	0.01
Overnight stays by national travellers in touristic apartments	0.79	0.03	0.70	0.02
GDP	0.78	0.02	0.71	0.02
Private limited companies (SL)	0.78	0.02	0.70	0.01
Foreign residents by country of nationality	0.77	0.01	0.70	0.01
Foreign residents by country of birth	0.77	0.01	0.70	0.01
Population	0.76	0.00	0.69	0.00
Overnight stays by foreign travellers in hotels	0.75	-0.01	0.69	0.00
National residents by country of birth	0.75	-0.02	0.68	-0.02
Total touristic apartments' capacity	0.71	-0.06	0.64	-0.06
GDP per capita	0.69	-0.09	0.63	-0.07
Foreign travellers in touristic apartments	0.67	-0.10	0.62	-0.08
Overnight stays by foreign travellers in touristic apartments	0.63	-0.15	0.58	-0.13

Table 3: Contribution of each variable to the models’ accuracy. For both diseases, the first column shows the average correlation of the estimates of the models including each variable in their fit with the officially reported 2019 data. The second column shows the average difference in correlation between models including each variable in their fit and the models not including each of the variables (variables ranked by the average correlation for malaria predictions).

A similar procedure was followed for the weighted models: the average correlation between those models including each variable and 2019 official data was computed, in this case stratified by the weight assigned to the variable (Figure 4). In addition to a ranking of variables, similar to Table 3, this also provided a measure of the variability of each variable’s contribution to the accuracy of the model in terms of the weight assigned to it. Smaller overall variations in models’ accuracy were found for variables measuring VFR, for instance, while much higher variability was recorded for some touristic indicators. This shows the higher potential loss in accuracy that would result from including these variables in the models compared to other ones.

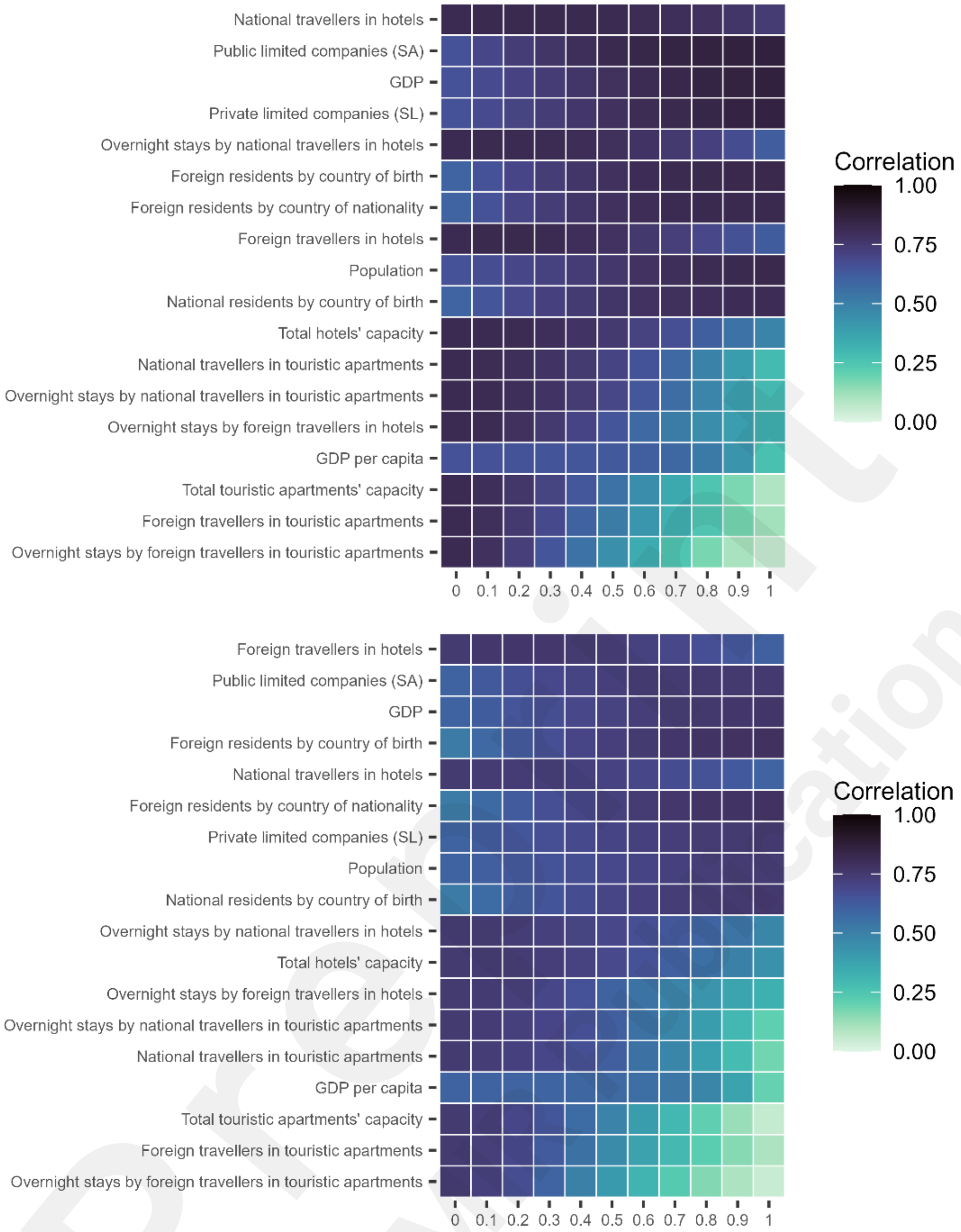


Figure 4. Summary of each input variable's performance on the estimates for malaria (top) and dengue (bottom). Each square in the figures is colored according to the average correlation between the official 2019 reports and the estimates provided by the weighted models including each of the variables, with the associated weight ranging from 0 (no contribution from the variable is assumed in the model) to 1 (the model only includes that variable). The variables are ranked from top to bottom according to the overall average correlation with 2019 data of the estimates of the models including each variable.

Comparison with a generic mobility model

For both dengue and malaria, the human mobility models ranked higher in terms of correlation with 2019 data for higher values of the assumed proportion of travelers that do not move from their destination province upon arrival (q, see Figure S5 and S6). Models also favored the choice of smaller values of the exponent of the power law distribution (highest average correlation with 2019 data for $\gamma = 1$), indicating that longer movements may take place if a displacement occurs after arrival. Much higher variability in the correlation with 2019 data was due to the choice of q than to the choice of γ, however (Figure S6).

In general, the estimates of the generic mobility model for the distribution of imported cases were less accurate when compared to actually reported cases than those resulting from the travelers' index models. This was the case for both malaria and dengue (0.59 and 0.66 correlation with 2019 data respectively, Table 4). In both cases, the best estimate was obtained assuming that imported cases are indeed reported at the region of arrival via air travel in Spain ($q = 1$), signaling that international travelers more often choose their final destination as the end of their trip, and rendering the choice of γ (the scaling exponent relating the length of the displacements) arbitrary. Visual examination of the resulting fit revealed the high influence of some provinces in these results, as in the traveler's index (Madrid, Barcelona and Las Palmas for malaria, Madrid and Barcelona for dengue). We performed the same analysis after removing these provinces from the input dataset, finding a strong decrease in the correlation of the models' estimates of the number of imported cases with the reported cases for 2019 (0.003 and 0.12 for malaria and dengue respectively). See Table S2 and Figures S5 and S6 for more detailed diagnostics on the generic human mobility model.

Model	Proportion of cases that do not move after arrival (q)	Exponent of power law distribution (γ)	Correlation with 2019 data	Overestimation
Malaria	1	Any	0.59	99.5%
Dengue	1	Any	0.66	95.2%

Table 4: Summary of the human mobility models that most accurately approximated the 2015-2018 reported cases (including all provinces). Each row shows the parameters of the model that provided the best estimate of imported cases of each disease, the correlation with the actually reported data in 2019, and the overestimation of the models as obtained from the linear fit with official records.

Discussion

We have computed estimates for the number of imported cases of malaria and dengue at each province in Spain based on simple methodological assumptions. Our approach makes use of readily available data and provides approximations of the actually declared number of cases of the disease. This advance may contribute to the adequate modelling and monitoring of VBDs, relevant for effective outbreak prevention strategies. More efficient resource allocation strategies for both vector control and disease prevention can be designed if reliable predictions of the geographical location of imported cases are available. By circumventing the need for detailed large-scaled data on human mobility or travelers' behavior, this methodology is accessible and suitable to be used in countries lacking more exhaustive data infrastructures, for instance [39,50]. The reasoning presented here could also be generalized to other choices of territories.

The high correlation found between our estimates and real data support the validity of our approach, based on an *a priori* theoretical conceptualization. This agreement in trend suggests that our estimates are reliable enough for the elaboration of scale-less risk indicators, for instance. On the other hand, our estimates of the raw number of imported cases were simplistic (products of yearly prevalences and total number of travelers), which resulted in substantial overestimation of the number of imported cases. For the case of malaria, this is coherent with the epidemiology of the disease, being more severe unless treatment is available and having a higher incidence in economically deprived populations [51]. These factors may prevent individuals with malaria from engaging in international displacements. For dengue, however, the found overestimation (8 predicted cases per notified case) lies relatively close to previously obtained estimates of the underreporting of cases in other contexts [52]. This suggests that our approach could also provide a valid method for assessing the sensibility of epidemiological systems. In any case, we remark that our focus was to assess the validity of the travelers' index, as a method to improve risk analysis, rather than developing a predictive model for imported cases of the diseases.

The proposed computation of the key indicators involved in our model (the travelers' index v_{ik}) has the advantage of being partially robust upon errors in declaration or incomplete data collection. Indeed, as these only involve the relative importance of each region in the country, correction factors are unnecessary in our approach, and incomplete data will yield equally valid estimates as long as the underreporting can be assumed to be comparable for all of the regions. Moreover, the little variation in time shown by these quantities (see Figure S3) could allow for reliable estimates even when only past statistics are available.

A key finding in this direction is that while the impact of each particular indicator in the quality of the estimate was similar for both diseases, the relevant drivers for case importation were different (economic motivations for malaria cases and visits to friends and relatives for travelers carrying dengue). This may be due to the different nature of the motivation for international travel across countries in the world: most malaria cases are imported from African countries, while dengue-bearing travelers usually arrive from America or Asia (see Table S1). Travelers arriving from these continents are expected to follow different motivations for international displacement. Actually, malaria cases imported to Spain in the pre-pandemic era were mainly due to VFR or migration in almost 75% of cases, which correspond to travelers following economic motivations [53]. On the other hand, dengue cases were imported mainly by tourist travelers or VFR [54].

Further evidence towards the appropriateness of our approach was given by comparison with the human mobility model. While the validity of this model has been established in many contexts and is widely acknowledged [47], it provided much less accurate estimates and was highly sensitive to data retrieval from the provinces with larger number of disease cases. This demonstrates the need to consider specific designs that take into account travelers' behavior and differentiates it from general resident population dynamics.

Future developments of our approach should cover the following improvements:

- **Coupling with post-importation dynamics**. Our framework could be integrated in more complicated models incorporating transmission dynamics, involving the life cycle of the disease within the vector and the host [55]–[58]. Well-developed approaches such as compartmental models could benefit from more precise estimates on the expected location of arrival of imported cases of diseases.
- **Refining the gross estimate of imported cases**. As mentioned above, we computed simple estimates for the total number of cases arriving to the importing country (products of yearly prevalences and total number of travelers). Our focus lied rather on how these cases distribute over the regions of the importing country. Considering more elaborate estimates on these quantities or the local distribution of the disease in the exporting countries would probably yield more precise final estimates.
- **Extending the scope of the model to other diseases**. A particular feature of the treated examples is that virtually all the incoming flows of travelers from regions where malaria and dengue are endemic into Spain that may result in transmission may be assumed to be via air travel. This may not be the case for other diseases and countries, for which detailed data on the total traveler flows or further developments of the proposed methodology could be necessary. Similarly, other importation phenomena that may depend on human behavior or allocation of resources could be analyzed under our assumptions, such as vectors' passive mobility by human means of transportation [59] or migratory flows [60,61].

It should be noted that our model is focused on countries with high dengue and malaria prevalence, hence, likely to export these diseases to Spain. However, this concept could be generalized to other type of risk-related importation scenarios, like the transport of new vector or exotic species (invasion biology), another crucial process in the spread of VBDs.

Several factors may be limiting the extent of our results. First of all, both malaria and dengue are diseases known to be subject to high underreporting [51]–[52]. Secondly, we validated our models with annual data, as data on the number of monthly reported cases was too noisy. In any case, model predictions could be generated at higher temporal resolution by incorporating monthly numbers of arriving travelers in the country, for instance. Thirdly, our model is designed to address the motivations of international travelers, while a significant number of imported cases may correspond to national travelers returning to the country or individuals from other non-endemic regions, especially for dengue. It would be desirable to devise an accurate method to differentiate between these two types of travelers and incorporate both motivations in the model. Finally, geographical borders are not always the best spatial human structure [62], and therefore the availability of data with a finer geographical resolution could result in significant improvements in our estimates. We also note that our model has been validated with data obtained prior to the COVID-19 pandemic; differences may arise in the post-pandemic era and therefore further validation with future data is desirable.

We have shown the validity of the travelers' index as a method to estimate the distribution of imported cases of malaria and dengue from endemic regions. This is an appropriate way to improve disease risk predictions on the basis of human mobility patterns. Our methodology adds value to available socioeconomic information relevant for public health. Nonetheless, human mobility is just one component of VBD risk models. The other key components that need to be added are vector (mosquito) distributions and suitability. Our work will be combined with multi-sourced presence/absence and suitability vector data in Spain, including both authoritative and citizen science data collections [63], and integrated in the Spanish National Surveillance System for VBDs. Pairing the risk for importation of cases and the risk of local transmission through the presence of vectors will provide a more comprehensive evaluation of the threats posed by VBDs to public health.

Acknowledgements

The project leading to these results has received funding from "la Caixa" Foundation (ID 100010434) under agreement HR-18-0036.

Data availability

Touristic, economic and VFR indicators are available from the Spanish National Institute for Statistics (INE, www.ine.es). Arrivals data are available from the Spanish Aviation and Airports Office (AENA, https://www.aena.es/es/estadisticas/inicio.html). Disease prevalences at exporting countries are available from Global Health Data Exchange (GHDx, https://ghdx.healthdata.org/). Reported cases of dengue and malaria in Spain are collected by the national surveillance system and managed by the Spanish National Center for Epidemiology and are available upon reasonable request at vigilancia.cne@isciii.es.

Conflicts of Interest

None to declare.

References

[1] D. Brockmann, «Human Mobility and Spatial Disease Dynamics», en *Reviews of Nonlinear Dynamics and Complexity*, John Wiley & Sons, Ltd, 2009, pp. 1-24. doi: 10.1002/9783527628001.ch1.

[2] S. A. Lee, C. I. Jarvis, W. J. Edmunds, T. Economou, y R. Lowe, «Spatial connectivity in mosquito-borne disease models: a systematic review of methods and assumptions», *Journal of The Royal Society Interface*, vol. 18, n.º 178, p. 20210096, doi: 10.1098/rsif.2021.0096.

[3] K. Linka, A. Goriely, y E. Kuhl, «Global and local mobility as a barometer for COVID-19 dynamics», *Biomech Model Mechanobiol*, vol. 20, n.º 2, pp. 651-669, abr. 2021, doi: 10.1007/s10237-020-01408-2.

[4] B. Wahid, A. Ali, S. Rafique, y M. Idrees, «Global expansion of chikungunya virus: mapping the 64-year history», *International Journal of Infectious Diseases*, vol. 58, pp. 69-76, may 2017, doi: 10.1016/j.ijid.2017.03.006.

[5] D. J. Weiss *et al.*, «Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study», *The Lancet*, vol. 394, n.º 10195, pp. 322-331, jul. 2019, doi: 10.1016/S0140-6736(19)31097-9.

[6] S. Brugueras *et al.*, «Environmental drivers, climate change and emergent diseases transmitted by mosquitoes and their vectors in southern Europe: A systematic review», *Environmental Research*, vol. 191, p. 110038, dic. 2020, doi: 10.1016/j.envres.2020.110038.

[7] A. Wilder-Smith, D. J. Gubler, S. C. Weaver, T. P. Monath, D. L. Heymann, y T. W. Scott, «Epidemic arboviral diseases: priorities for research and public health», *The Lancet Infectious Diseases*, vol. 17, n.º 3, pp. e101-e106, mar. 2017, doi: 10.1016/S1473-3099(16)30518-7.

[8] J. P. Messina *et al.*, «The current and future global distribution and population at risk of dengue», *Nat Microbiol*, vol. 4, n.º 9, Art. n.º 9, sep. 2019, doi: 10.1038/s41564-019-0476-8.

[9] Y. Tozan, O. L. H. Branch, y J. Rocklöv, «Vector-Borne Diseases in a Changing Climate and World», in *Climate Change and Global Public Health*, K. E. Pinkerton y W. N. Rom, Eds. Cham: Springer International Publishing, 2021, pp. 253-271. doi: 10.1007/978-3-030-54746-2_12.

[10] C. M. Gossner *et al.*, «Risks Related to Chikungunya Infections among European Union Travelers, 2012–2018 - Volume 26, Number 6—June 2020 - Emerging Infectious Diseases journal - CDC», doi: 10.3201/eid2606.190490.

[11] N. Li *et al.*, «COVID-19 Border Restriction Reduced Dengue Transmission». Rochester, NY, 3 de marzo de 2021. doi: 10.2139/ssrn.3796910.

[12] C. M. Gossner *et al.*, «Dengue virus infections among European travellers, 2015 to 2019», *Eurosurveillance*, vol. 27, n.º 2, p. 2001937, ene. 2022, doi: 10.2807/1560-7917.ES.2022.27.2.2001937.

[13] A. Wilder-Smith y D. J. Gubler, «Geographic Expansion of Dengue: The Impact of International Travel», *Medical Clinics of North America*, vol. 92, n.º 6, pp. 1377-1390, nov. 2008, doi: 10.1016/j.mcna.2008.07.002.

[14] M. M. I. Meslé, I. M. Hall, R. M. Christley, S. Leach, y J. M. Read, «The use and reporting of airline passenger data for infectious disease modelling: a systematic review», *Eurosurveillance*, vol. 24, n.º 31, p. 1800216, ago. 2019, doi: 10.2807/1560-7917.ES.2019.24.31.1800216.

[15] «Comparing sources of mobility for modelling the epidemic spread of Zika virus in Colombia | PLOS Neglected Tropical Diseases». https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0010565.

[16] A. Findlater y I. I. Bogoch, «Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel», *Trends in Parasitology*, vol. 34, n.º 9, pp. 772-783, sep. 2018, doi: 10.1016/j.pt.2018.07.004.

[17] M. Chinazzi *et al.*, «The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak», *Science*, vol. 368, n.º 6489, pp. 395-400, abr. 2020, doi: 10.1126/science.aba9757.

[18] H. Tegally, K. Khan, C. Huber, T. de Oliveira, y M. U. G. Kraemer, «Shifts in global mobility dictate the synchrony of SARS-CoV-2 epidemic waves», *Journal of Travel Medicine*, p. taac134, nov. 2022, doi: 10.1093/jtm/taac134.

[19] J. P. Messina *et al.*, «Global spread of dengue virus types: mapping the 70 year history», *Trends in Microbiology*, vol. 22, n.º 3, pp. 138-146, mar. 2014, doi: 10.1016/j.tim.2013.12.011.

[20] Gossner, C.M., Hallmaier-Wacker, L., Briet, O., Haussig, J.M., de Valk, H., Wijermans, A., Bakonyi, T., Madubuko, T., Frank, C., Noel, H. and Abdulaziz, M., 2023. Arthropod-borne diseases among travellers arriving in Europe from Africa, 2015 to 2019. *Eurosurveillance*, *28*(7), p.2200270.

[21] «The introduction of dengue follows transportation infrastructure changes in the state of Acre, Brazil: A network-based analysis | PLOS Neglected Tropical Diseases». https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0006070 (accedido 4 de enero de 2023).

[22] M. A. Johansson *et al.*, «An open challenge to advance probabilistic forecasting for dengue epidemics», *Proceedings of the National Academy of Sciences*, vol. 116, n.º 48, pp. 24268-24274, nov. 2019, doi: 10.1073/pnas.1909865116.

[23] D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, H. J. Martínez, S. Meloni, y J. Gómez-Gardeñes, «Vector-borne epidemics driven by human mobility», *Phys. Rev. Res.*, vol. 2, n.º 1, p. 013312, mar. 2020, doi: 10.1103/PhysRevResearch.2.013312.

[24] E. Massad *et al.*, «Estimating the probability of dengue virus introduction and secondary autochthonous cases in Europe», *Sci Rep*, vol. 8, n.º 1, Art. n.º 1, mar. 2018, doi: 10.1038/s41598-018-22590-5.

[25] J. Liebig, C. Jansen, D. Paini, L. Gardner, y R. Jurdak, «A global model for predicting the arrival of imported dengue infections», *PLOS ONE*, vol. 14, n.º 12, p. e0225193, dic 2019, doi: 10.1371/journal.pone.0225193.

[26] D. Salami, C. Capinha, M. do R. O. Martins, y C. A. Sousa, «Dengue importation into Europe: A network connectivity-based approach», *PLOS ONE*, vol. 15, n.º 3, p. e0230274, mar. 2020, doi: 10.1371/journal.pone.0230274.

[27] D. Salami, C. A. Sousa, M. do R. O. Martins, y C. Capinha, «Predicting dengue importation into Europe, using machine learning and model-agnostic methods», *Sci Rep*, vol. 10, n.º 1, Art. n.º 1, jun. 2020, doi: 10.1038/s41598-020-66650-1.

[28] A. J. Tatem *et al.*, «The geography of imported malaria to non-endemic countries: a meta-analysis of nationally reported statistics», *The Lancet Infectious Diseases*, vol. 17, n.º 1, pp. 98-107, ene. 2017, doi: 10.1016/S1473-3099(16)30326-7.

[29] L. Lei, J. S. Richards, Z.-H. Li, Y.-F. Gong, S.-Z. Zhang, y N. Xiao, «A framework for assessing local transmission risk of imported malaria cases», *Infectious Diseases of Poverty*, vol. 08, n.º 03, pp. 56-63, jun. 2019, doi: 10.1186/s40249-019-0552-6.

[30] C. A. Guerra, D. T. Citron, G. A. García, y D. L. Smith, «Characterising malaria connectivity using malaria indicator survey data», *Malaria Journal*, vol. 18, n.º 1, p. 440, dic. 2019, doi:

10.1186/s12936-019-3078-2.

[31] D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, y A. Vespignani, «Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model», *Journal of Computational Science*, vol. 1, n.º 3, pp. 132-145, ago. 2010, doi: 10.1016/j.jocs.2010.07.002.

[32] «epiflows: an R package for risk assessment of travel-related spread of disease - PMC». https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6738191/ (accedido 20 de diciembre de 2022).

[33] P. Siriyasatien, S. Chadsuthi, K. Jampachaisri, y K. Kesorn, «Dengue Epidemics Prediction: A Survey of the State-of-the-Art Based on Data Science Processes», *IEEE Access*, vol. 6, pp. 53757-53795, 2018, doi: 10.1109/ACCESS.2018.2871241.

[34] J. R. B. Palmer, T. J. Espenshade, F. Bartumeus, C. Y. Chung, N. E. Ozgencil, y K. Li, «New Approaches to Human Mobility: Using Mobile Phones for Demographic Research», *Demography*, vol. 50, n.º 3, pp. 1105-1128, nov. 2012, doi: 10.1007/s13524-012-0175-z.

[35] R. Becker *et al.*, «Human mobility characterization from cellular network data», *Commun. ACM*, vol. 56, n.º 1, pp. 74-82, ene. 2013, doi: 10.1145/2398356.2398375.

[36] C. Chen, J. Ma, Y. Susilo, Y. Liu, y M. Wang, «The promises of big data and small data for travel behavior (aka human mobility) analysis», *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 285-299, jul. 2016, doi: 10.1016/j.trc.2016.04.005.

[37] «The Hidden Geometry of Complex, Network-Driven Contagion Phenomena». https://www.science.org/doi/full/10.1126/science.1245200 (accedido 15 de marzo de 2022).

[38] C. E. Walters, M. M. I. Meslé, y I. M. Hall, «Modelling the global spread of diseases: A review of current practice and capability», *Epidemics*, vol. 25, pp. 1-8, dic. 2018, doi: 10.1016/j.epidem.2018.05.007.

[39] M. U. G. Kraemer *et al.*, «Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings», *Sci Rep*, vol. 9, n.º 1, Art. n.º 1, mar. 2019, doi: 10.1038/s41598-019-41192-3.

[40] «INE. Instituto Nacional de Estadística», *INE*. https://www.ine.es/ (accedido 15 de marzo de 2022).

[41] «Inicio». https://www.aena.es/es/estadisticas/inicio.html (accedido 15 de marzo de 2022).

[42] «Global Health Data Exchange | GHDx». https://ghdx.healthdata.org/ (accedido 15 de marzo de 2022).

[43] «The Malaria Atlas Project», *MAP*. https://www.tki-dev.malariaatlas.org/ (accedido 15 de marzo de 2022).

[44] «The Social Psychology of Tourist Behaviour - 1st Edition». https://www.elsevier.com/books/the-social-psychology-of-tourist-behaviour/pearce/978-0-08-025794-5 (accedido 22 de diciembre de 2022).

[45] S. Gössling, P. Hanna, J. Higham, S. Cohen, y D. Hopkins, «Can we fly less? Evaluating the 'necessity' of air travel», *Journal of Air Transport Management*, vol. 81, p. 101722, oct. 2019, doi: 10.1016/j.jairtraman.2019.101722.

[46] Brockmann, Dirk, Lars Hufnagel, and Theo Geisel. "The scaling laws of human travel." *Nature* 439.7075 (2006): 462-465.

[47] Clauset, A., Shalizi, C.R. and Newman, M.E., 2009. Power-law distributions in empirical data. *SIAM review*, *51*(4), pp.661-703.

[48] Barbosa, Hugo, Marc Barthelemy, Gourab Ghoshal, Charlotte R. James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J. Ramasco, Filippo Simini, and Marcello Tomasini.

"Human mobility: Models and applications." *Physics Reports* 734 (2018): 1-74.

[49] Del-Águila-Mejía, Javier, et al. "Epidemic Diffusion Network of Spain: A Mobility Model to Characterize the Transmission Routes of Disease." *International Journal of Environmental Research and Public Health* 20.5 (2023): 4356.

[50] L. Nelli *et al.*, «Distance sampling for epidemiology: an interactive tool for estimating under-reporting of cases from clinic data», *International Journal of Health Geographics*, vol. 19, n.º 1, p. 16, abr. 2020, doi: 10.1186/s12942-020-00209-1.

[51] R. E. Cibulskis, M. Aregawi, R. Williams, M. Otten, y C. Dye, «Worldwide Incidence of Malaria in 2009: Estimates, Time Trends, and a Critique of Methods», *PLOS Medicine*, vol. 8, n.º 12, p. e1001142, dic 2011, doi: 10.1371/journal.pmed.1001142.

[52] N. T. Toan, S. Rossi, G. Prisco, N. Nante, y S. Viviani, «Dengue epidemiology in selected endemic countries: factors influencing expansion factors as estimates of underreporting», *Tropical Medicine & International Health*, vol. 20, n.º 7, pp. 840-863, 2015, doi: 10.1111/tmi.12498.

[53] Herrador, Zaida, et al. "Imported cases of malaria in Spain: observational study using nationally reported statistics and surveillance data, 2002–2015." *Malaria journal* 18.1 (2019): 1-11.

[54] Informe epidemiológico sobre la situación de dengue en España. Años 2019, 2020 y 2021. Centro Nacional de Epidemiología. Instituto de Salud Carlos III. Madrid, 14 noviembre 2022. Available from https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/Resultados_Vigilancia_Dengue.aspx

[55] H. Tao *et al.*, «A comprehensive framework for studying diffusion patterns of imported dengue with individual-based movement data», *International Journal of Geographical Information Science*, vol. 34, n.º 3, pp. 604-624, mar. 2020, doi: 10.1080/13658816.2019.1684497.

[56] T.-H. Wen, C.-T. Tsai, y W.-C.-B. Chin, «Evaluating the role of disease importation in the spatiotemporal transmission of indigenous dengue outbreak», *Applied Geography*, vol. 76, pp. 137-146, nov. 2016, doi: 10.1016/j.apgeog.2016.09.020.

[57] T.-H. Wen, C.-S. Hsu, y M.-C. Hu, «Evaluating neighborhood structures for modeling intercity diffusion of large-scale dengue epidemics», *International Journal of Health Geographics*, vol. 17, n.º 1, p. 9, may 2018, doi: 10.1186/s12942-018-0131-2.

[58] K. Liu *et al.*, «Enhancing fine-grained intra-urban dengue forecasting by integrating spatial interactions of human movements between urban regions», *PLOS Neglected Tropical Diseases*, vol. 14, n.º 12, p. e0008924, dic. 2020, doi: 10.1371/journal.pntd.0008924.

[59] R. Eritja, J. R. B. Palmer, D. Roiz, I. Sanpera-Calbet, y F. Bartumeus, «Direct Evidence of Adult Aedes albopictus Dispersal by Car», *Sci Rep*, vol. 7, n.º 1, Art. n.º 1, oct. 2017, doi: 10.1038/s41598-017-12652-5.

[60] A. M. Mayda, «International migration: a panel data analysis of the determinants of bilateral flows», *J Popul Econ*, vol. 23, n.º 4, pp. 1249-1274, sep. 2010, doi: 10.1007/s00148-009-0251-x.

[61] G. J. Abel y J. E. Cohen, «Bilateral international migration flow estimates for 200 countries», *Sci Data*, vol. 6, n.º 1, Art. n.º 1, jun. 2019, doi: 10.1038/s41597-019-0089-3.

[62] S. Rinzivillo, S. Mainardi, F. Pezzoni, M. Coscia, D. Pedreschi, y F. Giannotti, «Discovering the Geographical Borders of Human Mobility», *Künstl Intell*, vol. 26, n.º 3, pp. 253-260, ago. 2012, doi: 10.1007/s13218-012-0181-8.

[63] Palmer, J.R.; Oltra, A.; Collantes, F.; Delgado, J.A.; Lucientes, J.; Delacour, S.; Bengoa, M.; Eritja, R.; Bartumeus, F. Citizen science provides a reliable and scalable tool to track disease-

carrying mosquitoes. Nat. Commun. 2017, 8, 916.

Supplementary Material for

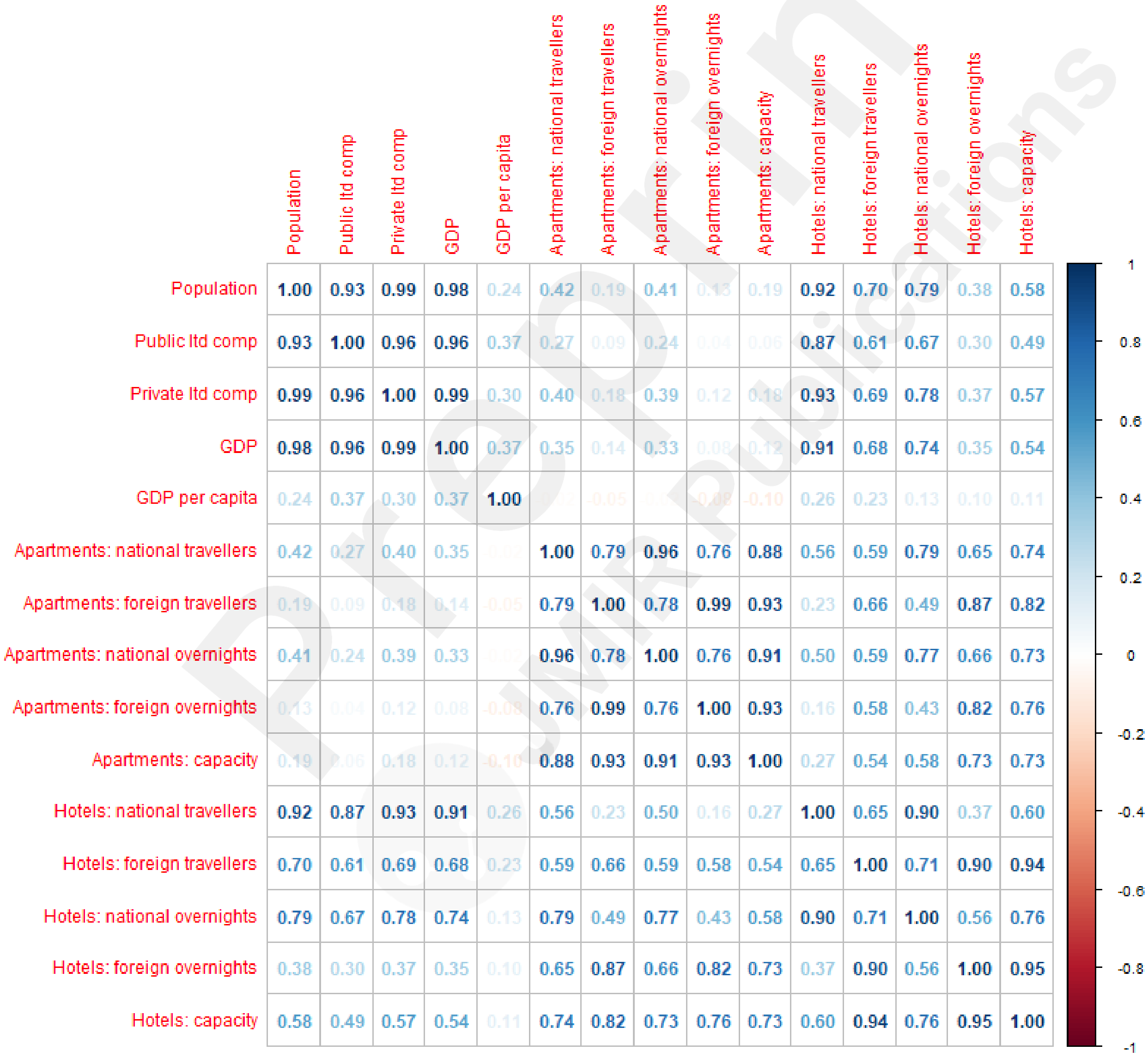
Modelling the regional distribution of international travelers in Spain to infer imported cases of dengue and malaria, 2015-2019

David García-García[12]*, Beatriz Fernández-Martínez[12], Frederic Bartumeus[345], Diana Gómez-Barroso[12]

1 National Centre for Epidemiology, Instituto de Salud Carlos III (ISCIII), Madrid, Spain
2 Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain
3 Centre for Advanced Studies of Blanes (CEAB-CSIC), Blanes, Spain
4 Centre for Research on Ecology and Forestry Applications (CREAF), Cerdanyola del Vallès, Barcelona, Spain
5 Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain
* Corresponding author

Local indicators

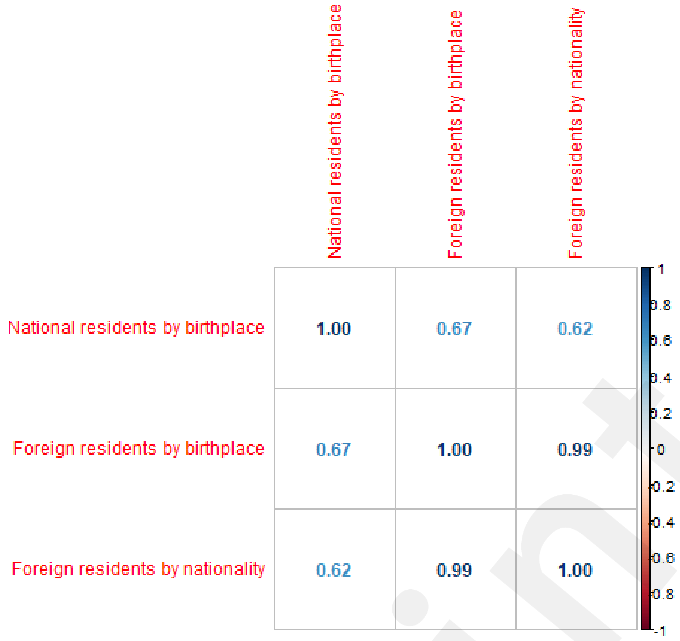
We found substantial correlations between most of the economic and touristic indicators used in the construction of the travelers’ index (Figure S1). Among the economic variables, GDP per capita displayed a much lower correlation with the remaining statistics, that shared very high pair-to-pair correlations. The variables quantifying the importance of tourism in the regions were also highly correlated, with those corresponding to hotels showing a higher variability, both among each other and with respect to those corresponding to apartments. Variables across the two different categories showed much smaller correlations in general, with the exception of the high correlations found between the economic variables (other than the GDP per capita) and some of the indicators measuring hotel activity.



	Population	Public ltd comp	Private ltd comp	GDP	GDP per capita	Apartments: national travellers	Apartments: foreign travellers	Apartments: national overnights	Apartments: foreign overnights	Apartments: capacity	Hotels: national travellers	Hotels: foreign travellers	Hotels: national overnights	Hotels: foreign overnights	Hotels: capacity
Population	1.00	0.93	0.99	0.98	0.24	0.42	0.19	0.41	0.13	0.19	0.92	0.70	0.79	0.38	0.58
Public ltd comp	0.93	1.00	0.96	0.96	0.37	0.27	0.09	0.24	0.04	0.06	0.87	0.61	0.67	0.30	0.49
Private ltd comp	0.99	0.96	1.00	0.99	0.30	0.40	0.18	0.39	0.12	0.18	0.93	0.69	0.78	0.37	0.57
GDP	0.98	0.96	0.99	1.00	0.37	0.35	0.14	0.33	0.08	0.12	0.91	0.68	0.74	0.35	0.54
GDP per capita	0.24	0.37	0.30	0.37	1.00	-0.02	-0.05	-0.02	-0.08	-0.10	0.26	0.23	0.13	0.10	0.11
Apartments: national travellers	0.42	0.27	0.40	0.35	-0.02	1.00	0.79	0.96	0.76	0.88	0.56	0.59	0.79	0.65	0.74
Apartments: foreign travellers	0.19	0.09	0.18	0.14	-0.05	0.79	1.00	0.78	0.99	0.93	0.23	0.66	0.49	0.87	0.82
Apartments: national overnights	0.41	0.24	0.39	0.33	-0.02	0.96	0.78	1.00	0.76	0.91	0.50	0.59	0.77	0.66	0.73
Apartments: foreign overnights	0.13	0.04	0.12	0.08	-0.08	0.76	0.99	0.76	1.00	0.93	0.16	0.58	0.43	0.82	0.76
Apartments: capacity	0.19	0.06	0.18	0.12	-0.10	0.88	0.93	0.91	0.93	1.00	0.27	0.54	0.58	0.73	0.73
Hotels: national travellers	0.92	0.87	0.93	0.91	0.26	0.56	0.23	0.50	0.16	0.27	1.00	0.65	0.90	0.37	0.60
Hotels: foreign travellers	0.70	0.61	0.69	0.68	0.23	0.59	0.66	0.59	0.58	0.54	0.65	1.00	0.71	0.90	0.94
Hotels: national overnights	0.79	0.67	0.78	0.74	0.13	0.79	0.49	0.77	0.43	0.58	0.90	0.71	1.00	0.56	0.76
Hotels: foreign overnights	0.38	0.30	0.37	0.35	0.10	0.65	0.87	0.66	0.82	0.73	0.37	0.90	0.56	1.00	0.95
Hotels: capacity	0.58	0.49	0.57	0.54	0.11	0.74	0.82	0.73	0.76	0.73	0.60	0.94	0.76	0.95	1.00

Figure S1. Pair-to-pair correlations between local touristic and economic indicators obtained from INE [39].

The numbers of foreign residents by birthplace showed a very high correlation with those by nationality, as expected (Figure S2). A significantly smaller, nevertheless large correlation was found for nationalized individuals born abroad.



	National residents by birthplace	Foreign residents by birthplace	Foreign residents by nationality
National residents by birthplace	1.00	0.67	0.62
Foreign residents by birthplace	0.67	1.00	0.99
Foreign residents by nationality	0.62	0.99	1.00

Figure S2. Pair-to-pair correlations between indicators concerning foreign and born-abroad population obtained from INE [39].

As stated in the main text, we found relatively small variations on the relative importances t_i, e_i, r_{ik} over the years for most provinces and choice of statistics (Figure S3). This suggests that the role of the provinces within the national context may be more stable than the global evolution displayed by the country as a whole, and that these indicators may provide reliable projections in the short-term.

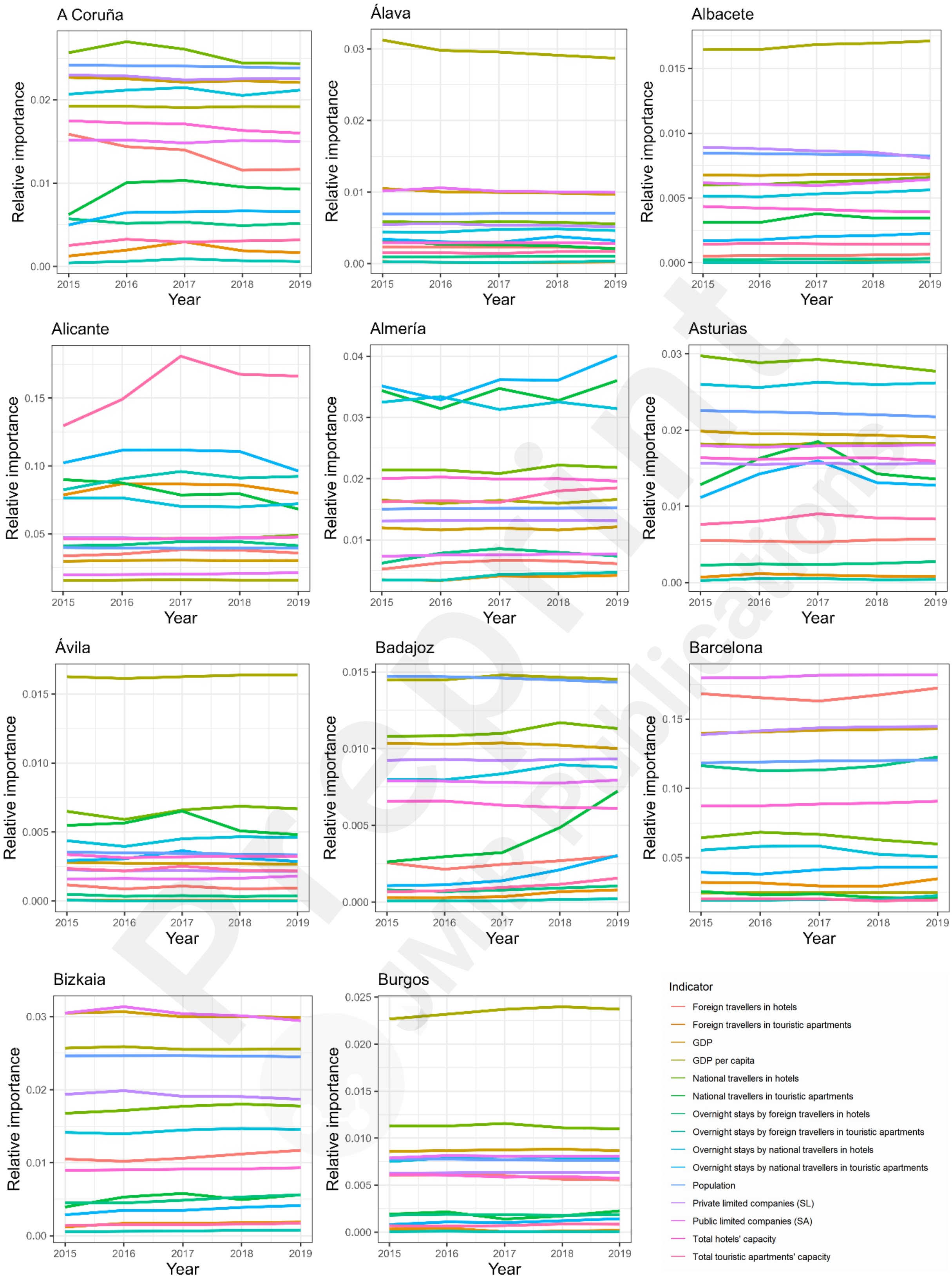


Figure S3.1: Relative importances of provinces in Spain in terms of economic and touristic indicators. Evolution of the indicators e_i and t_i during the years 2015-2019 for several provinces in Spain (see equation (2) in main text) for several choices of statistics.

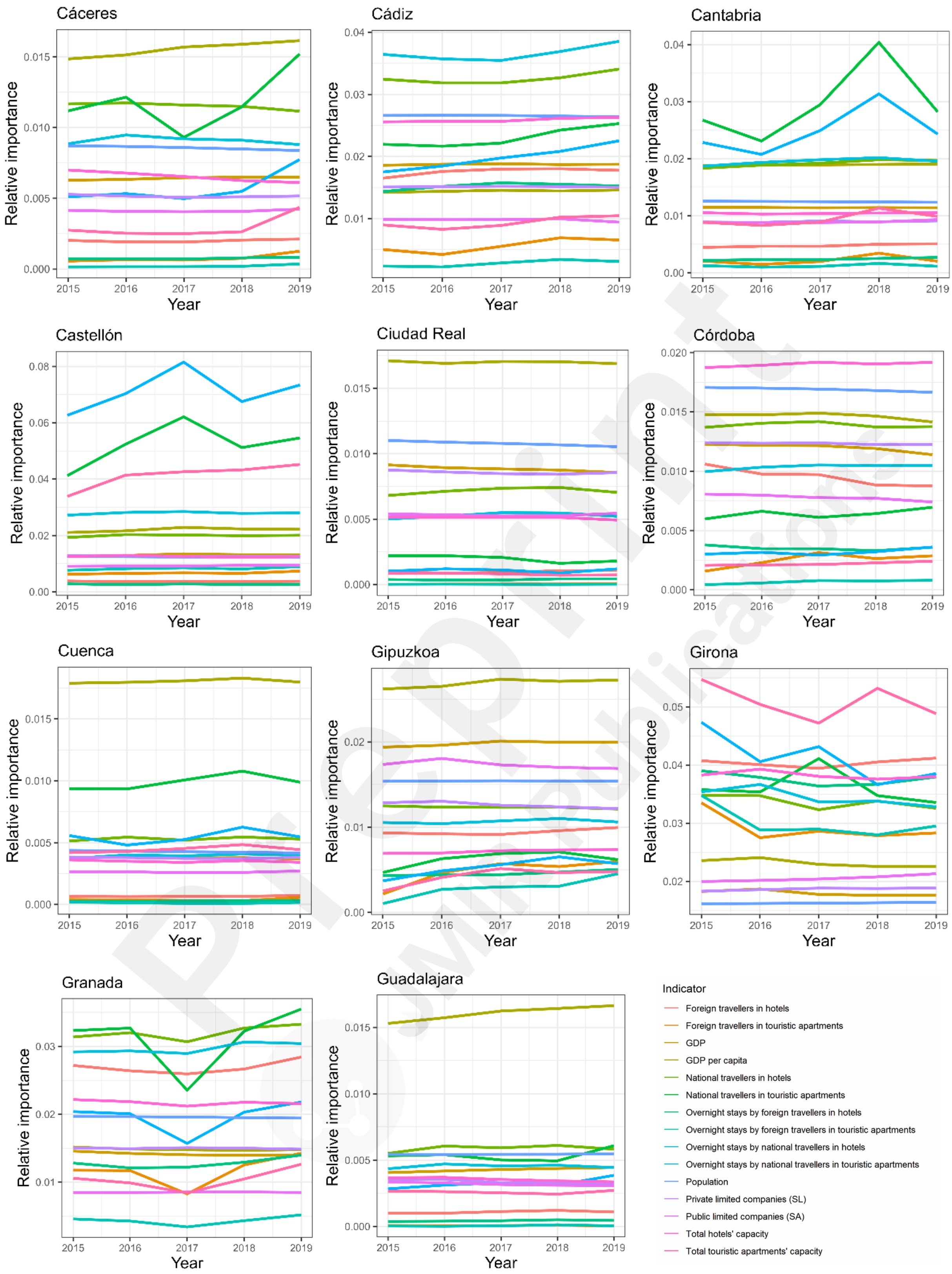


Figure S3.2: Relative importances of provinces in Spain in terms of economic and touristic indicators. Evolution of the indicators e_i and t_i during the years 2015-2019 for several provinces in Spain (see equation (2) in main text) for several choices of statistics.

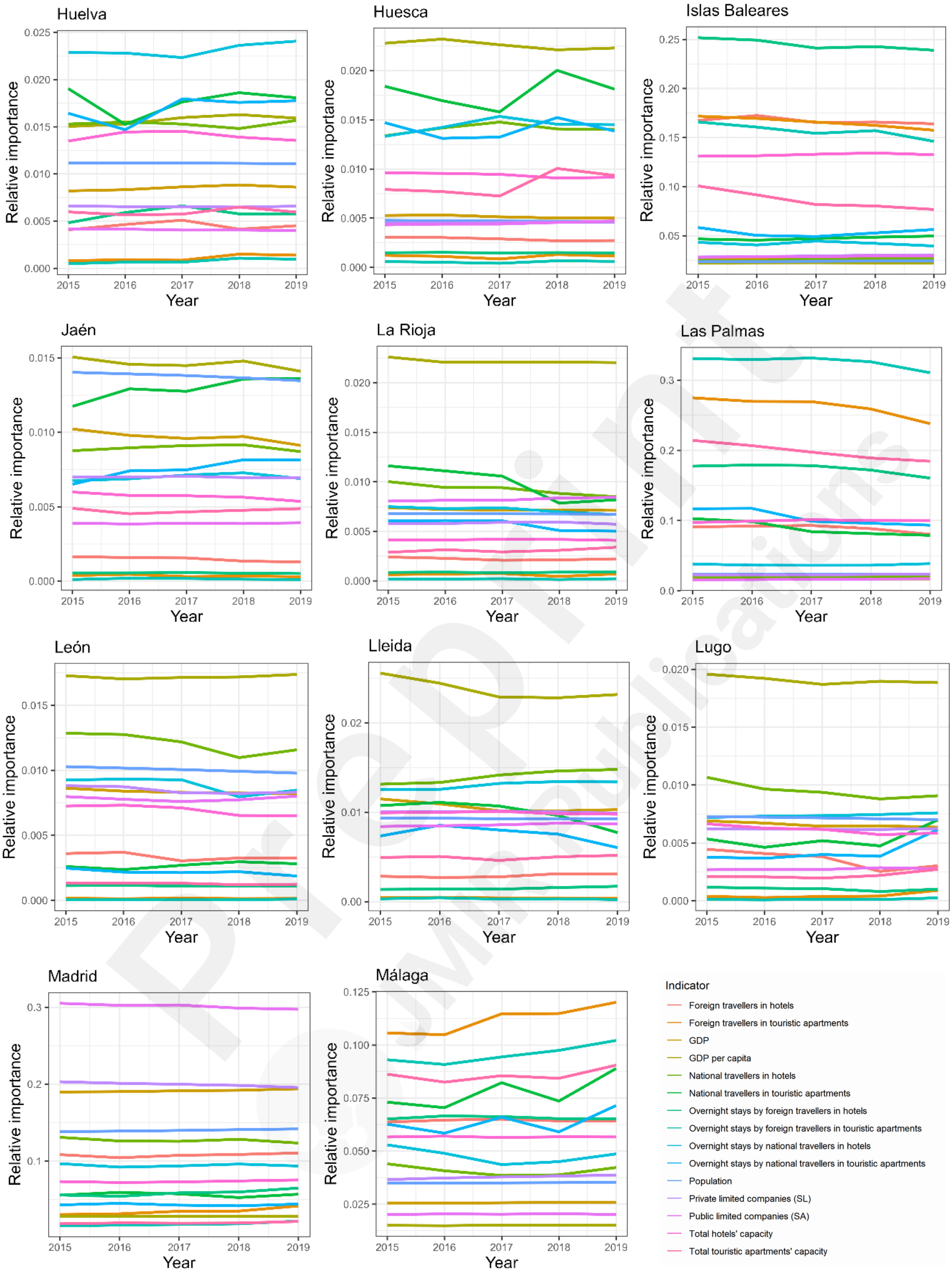


Figure S3.3: Relative importances of provinces in Spain in terms of economic and touristic indicators. Evolution of the indicators e_i and t_i during the years 2015-2019 for several provinces in Spain (see equation (2) in main text) for several choices of statistics.

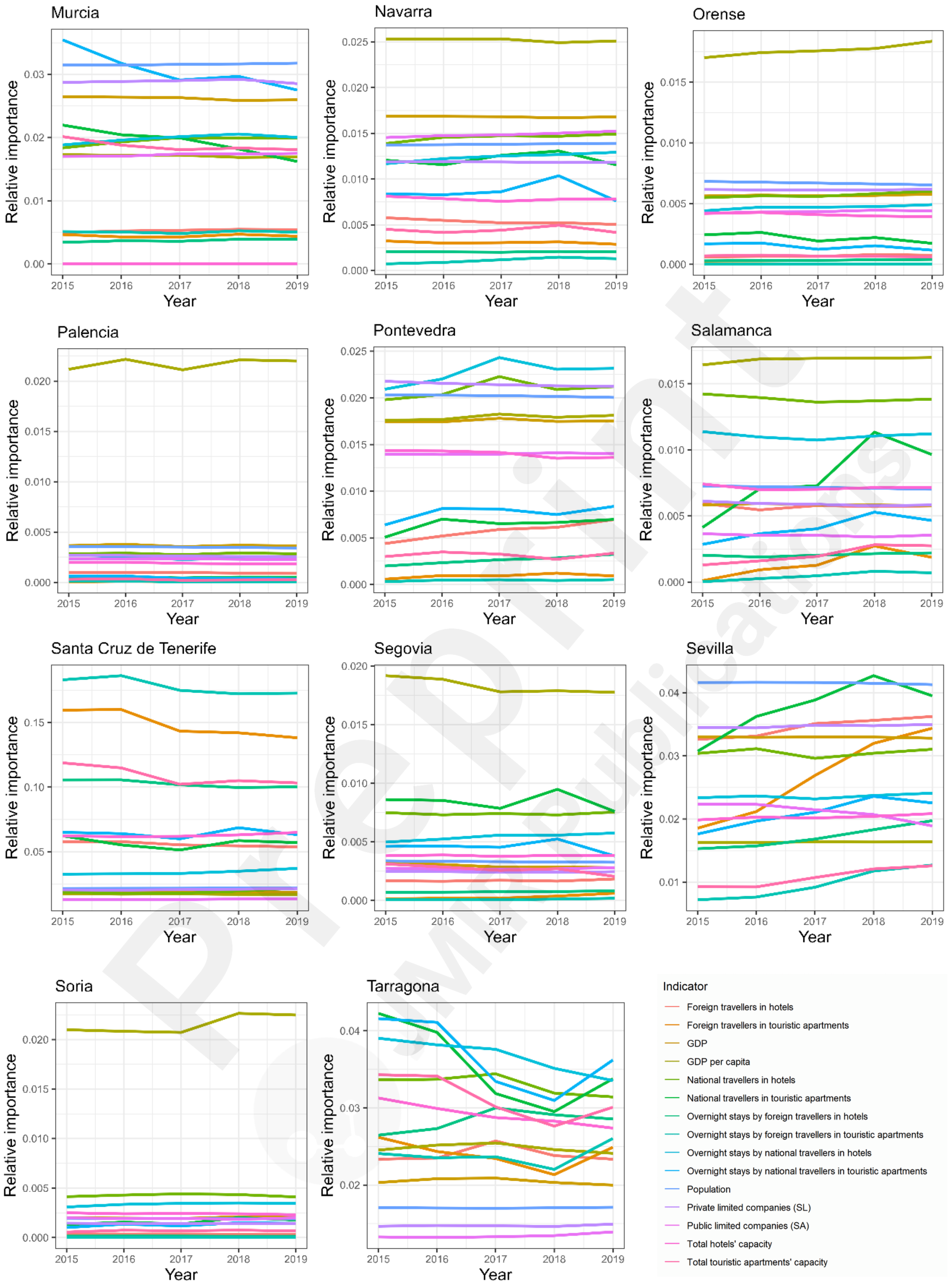


Figure S3.4: Relative importances of provinces in Spain in terms of economic and touristic indicators. Evolution of the indicators e_i and t_i during the years 2015-2019 for several provinces in Spain (see equation (2) in main text) for several choices of statistics.

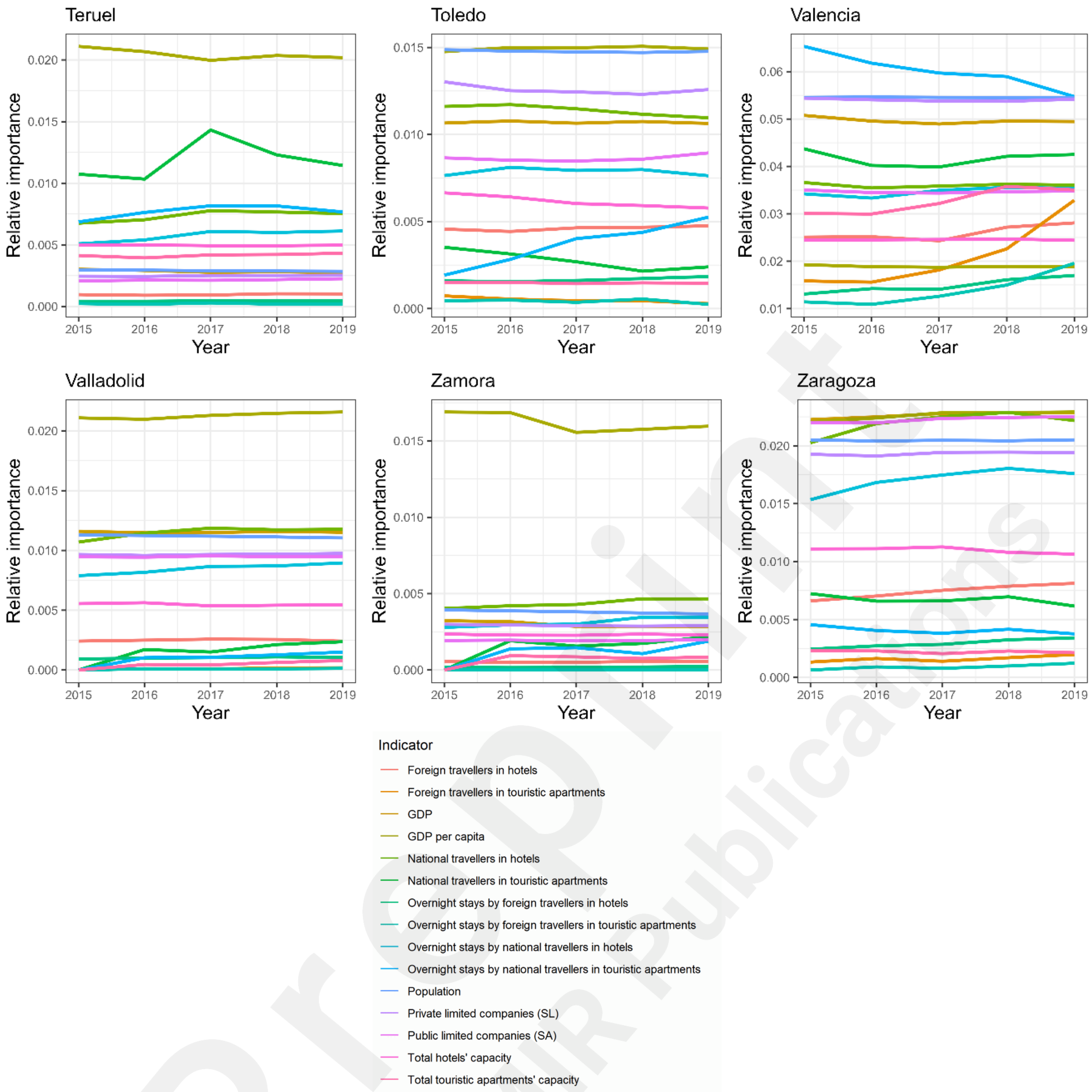


Figure S3.5: Relative importances of provinces in Spain in terms of economic and touristic indicators. Evolution of the indicators e_i and t_i during the years 2015-2019 for several provinces in Spain (see equation (2) in main text) for several choices of statistics.

Excluding outliers from the analysis

Table S1 shows the relevant features of the best models after excluding Madrid and Barcelona from the analysis. This was done in order to test the robustness of our approach, checking to what degree the accuracy of our models was due only to the extreme values of these provinces, both in terms of reported cases and travelers' index (see Figure S3).

Figure S4 shows the resulting models' fit and estimates for 2019 data, compared with the officially reported number of cases at each of the remaining provinces.

Disease (model)	Economic indicator (weight)	Touristic indicator (weight)	VFR indicator (weight)	Correlation with 2019 data	Overestimation
Malaria (simple)	Public limited companies (SA)	National travelers at hotels	Foreign residents by nationality	0.82	99.3%
Malaria (weighted)	No contribution (0)	No contribution (0)	Foreign residents by birthplace (1)	0.89	99.5%
Dengue (simple)	Public limited companies (SA)	National travelers at hotels	Foreign residents by birthplace	0.74	95.7%
Dengue (weighted)	Public limited companies (SA)	Overnight stays by national travelers at touristic apartments (0.1)	No contribution (0)	0.78	95.7%

Table S1: Summary of the models that most accurately approximated the 2015-2018 reported cases after excluding Madrid and Barcelona from the analysis. Each row shows the statistics that provided the best estimate of imported cases of each disease), the correlation with the actually reported data in 2019, and the proportion of overestimation as obtained from the linear models. For the models including weighted averages, the weight α_i of each indicator in included in parenthesis. If the weight of a given indicator is zero, this means that no contribution to the estimate is provided by the corresponding indicator.

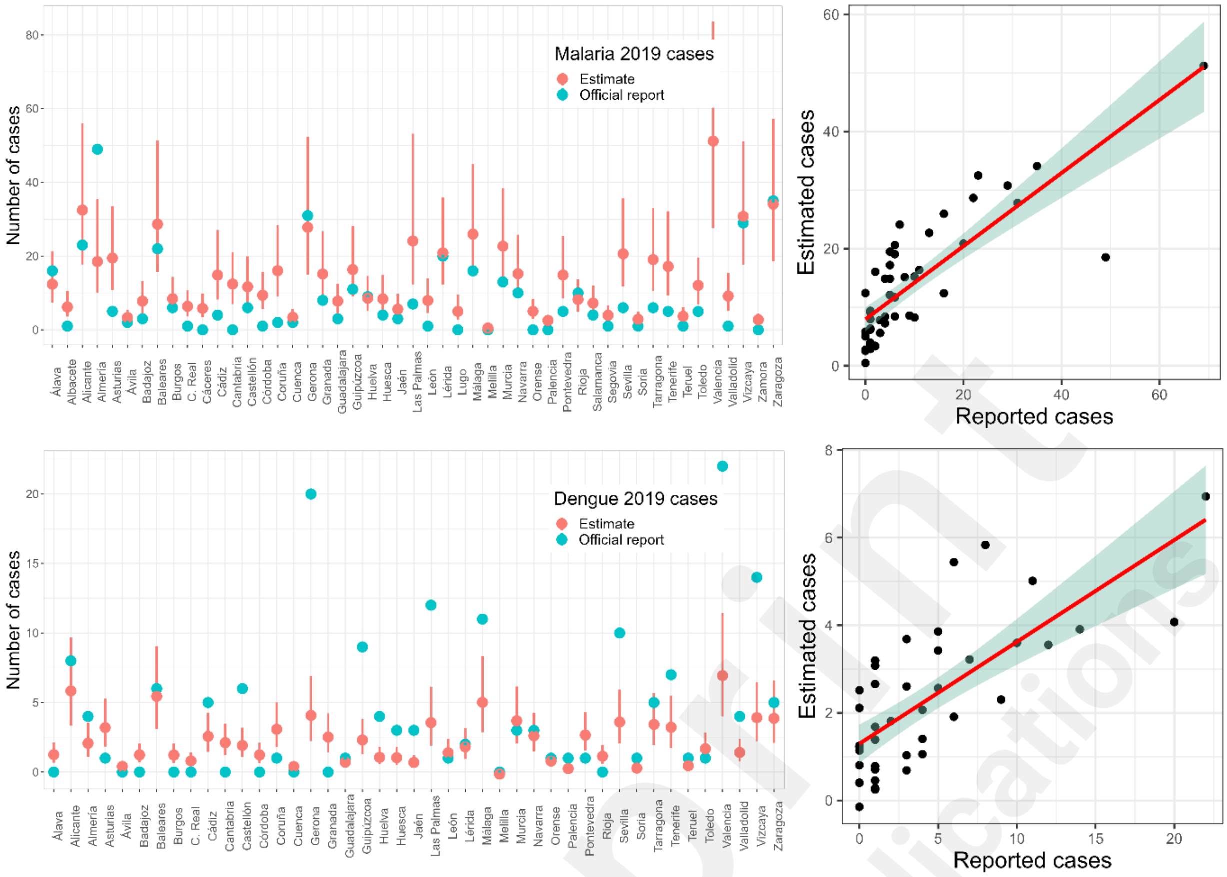


Figure S4. Summary of the best simple models for imported cases of malaria (top row) and dengue (bottom row), after excluding Madrid and Barcelona from the analysis. The left column shows the shows the models' predictions (in red), together with the actually reported cases (in blue) for 2019 for each province in Spain. The right column shows the fit between the models' estimates and the official records.

Human mobility models

Despite the relatively good score in terms of correlation with 2019 data found for the most accurate human mobility models (Table 3 in the main text), visual inspection of the models' fit showed that these results could be due to the dominance of some provinces from the dataset (Figure S5). After excluding these provinces from the analysis, the models' accuracy decreased significantly (Table S2). Average correlations with 2019 data in terms of models' parameters are provided in Figure S6 for general diagnostics.

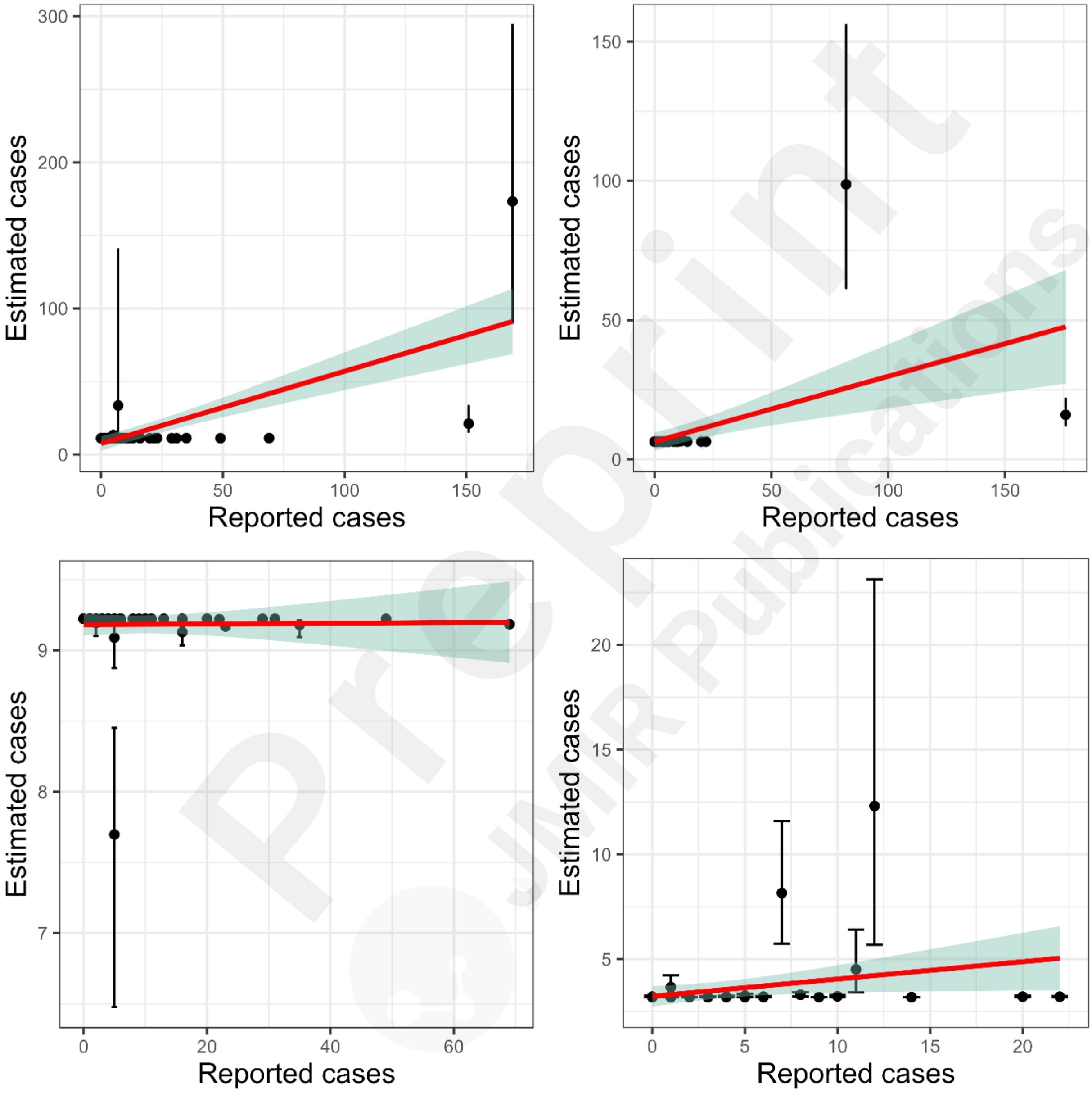


Figure S5. Models' fit for the human mobility models for malaria (left column) and dengue (right column). Top row shows the fit for the complete dataset of 52 provinces in Spain; bottom row shows the fit after removing the outliers from the analysis. Madrid, Barcelona and Las Palmas were removed from the malaria dataset, and Madrid and Barcelona were removed from the dengue dataset.

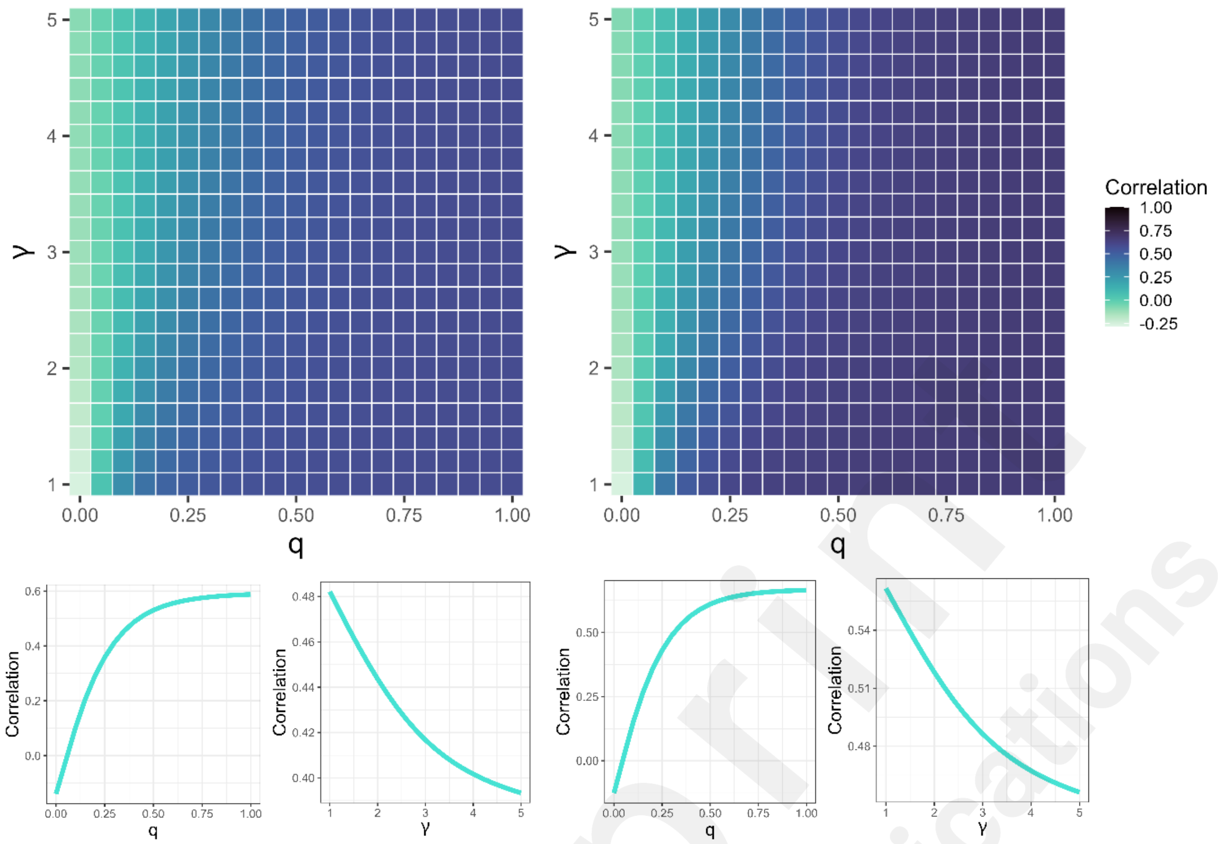


Figure S6. Accuracy of the human mobility model in terms of the models' parameters for malaria (left column) and dengue (right column). In each column, the top row shows the average correlation of the mobility models' estimates of imported cases with the 2019 official data in terms of the proportion of travelers that do not move after their arrival (q) and the exponent of the power law distribution (γ). Bottom row shows the average correlation of the models in terms of q and γ separately (left plot for q, right plot for γ).

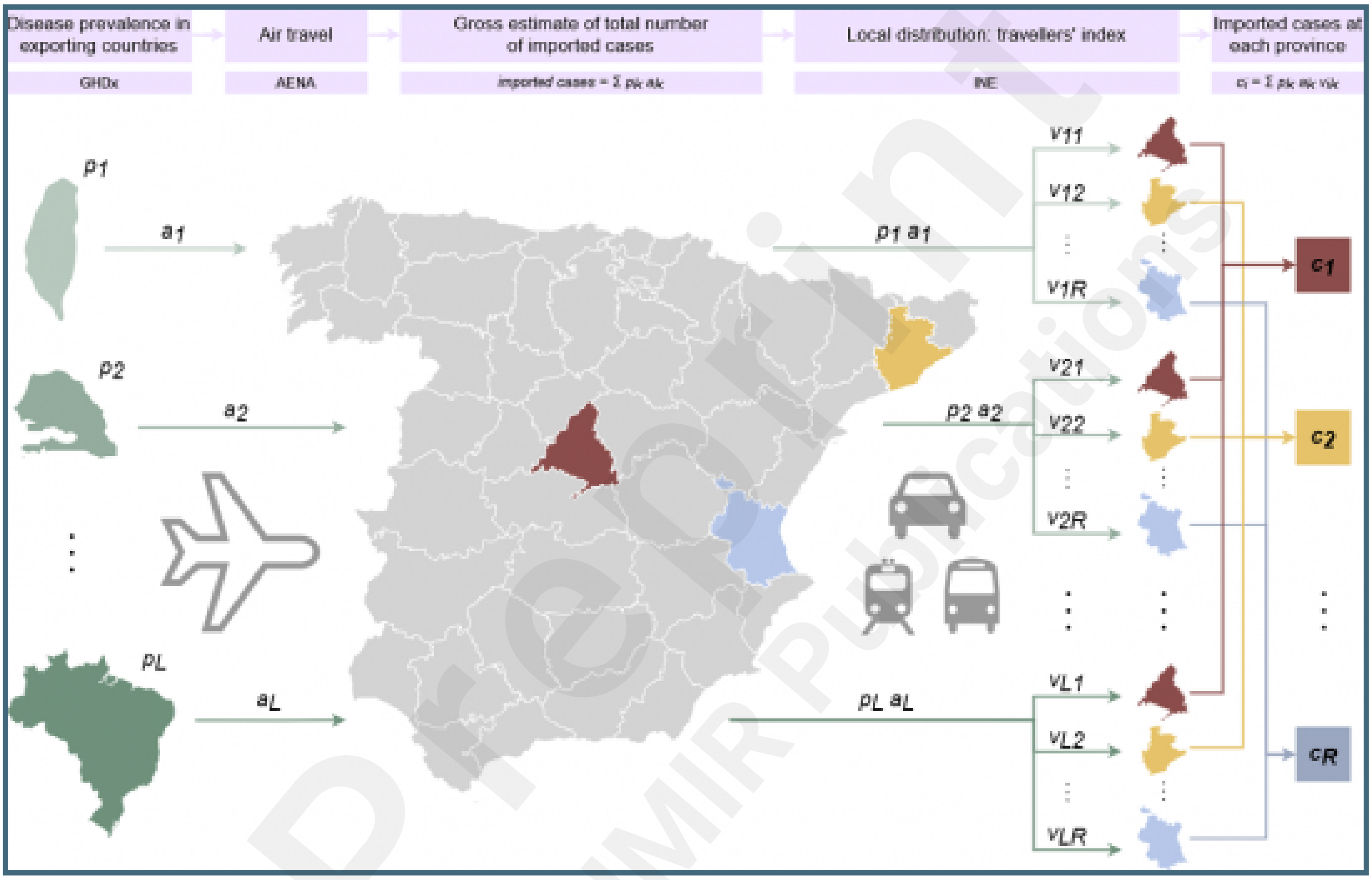
Model	Proportion of cases that do not move after arrival (q)	Exponent of power law distribution (γ)	Correlation with 2019 data	Overestimation/ underestimation
Malaria	1	Any	0.003	97.7%
Dengue	1	Any	0.12	17%

Table S2: Summary of the human mobility models that most accurately approximated the 2019 reported cases after excluding outliers from the analysis. Each row shows the parameters of the model that provided the best estimate of imported cases of each disease, the correlation with the actually reported data in 2019, and the overestimation/underestimation of cases as obtained from the linear models. Madrid, Barcelona and Las Palmas were removed from the malaria dataset, and Madrid and Barcelona were removed from the dengue dataset.

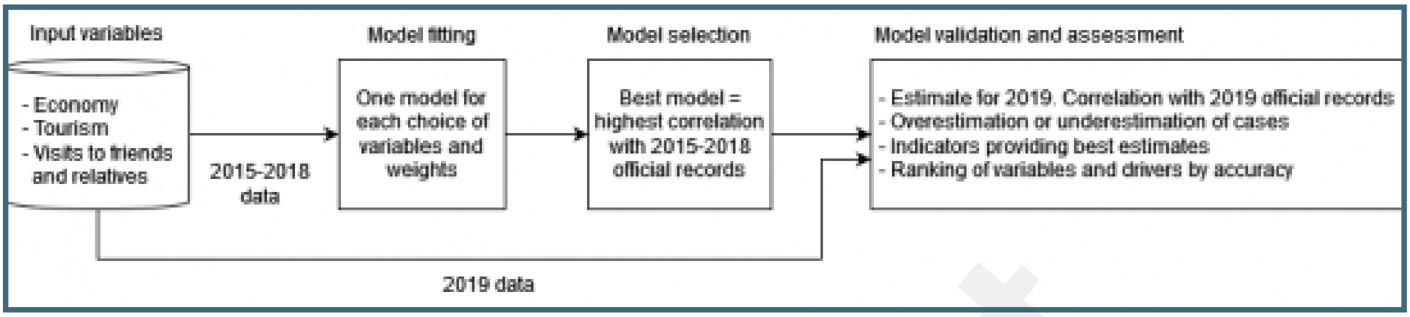
Supplementary Files

Figures

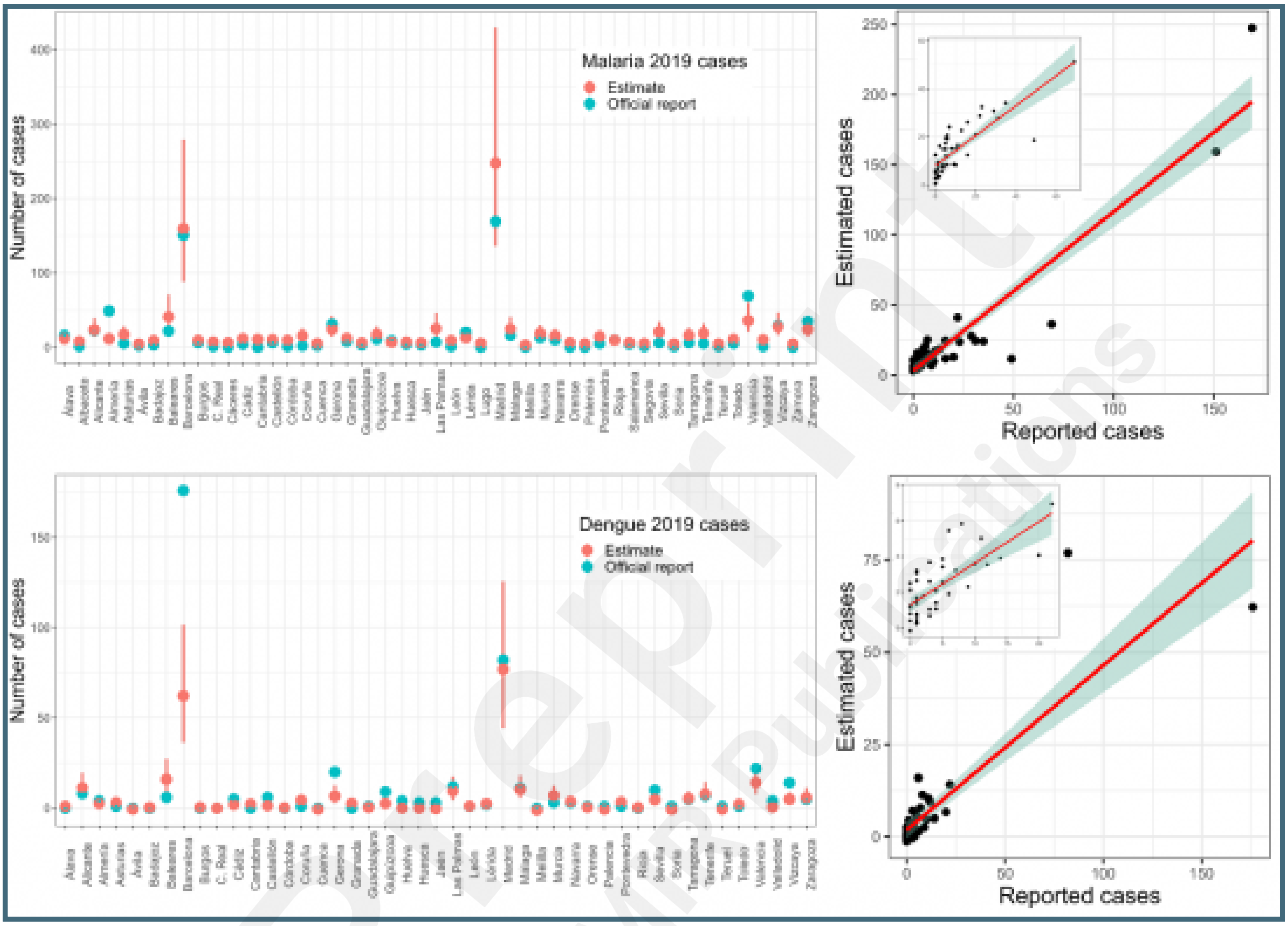
Summary of the rationale behind our approach. Infected travelers arrive to the importing country from L exporting countries over the world by means of air travel. For each country k, the prevalence of the disease at the country (p_k) and the number of travelers arriving from it to the importing country (a_k) are combined to obtain an estimate of the number of imported cases (p_k a_k). These spread across the R regions or spatial units of the importing country following a distribution that can be estimated by means of “traveler’s indexes”. These indexes are computed from local statistics concerning economic and touristic activity, and the number of foreign residents at each region. The traveler’s index v_ki measures the proportion of imported cases from country k that moves to local region i upon arrival to the importing country. The total number of cases expected to arrive at the region (c_i) is then obtained as the sum of the estimated number of imported cases over all exporting countries: c_i=? p_k a_k v_ik.



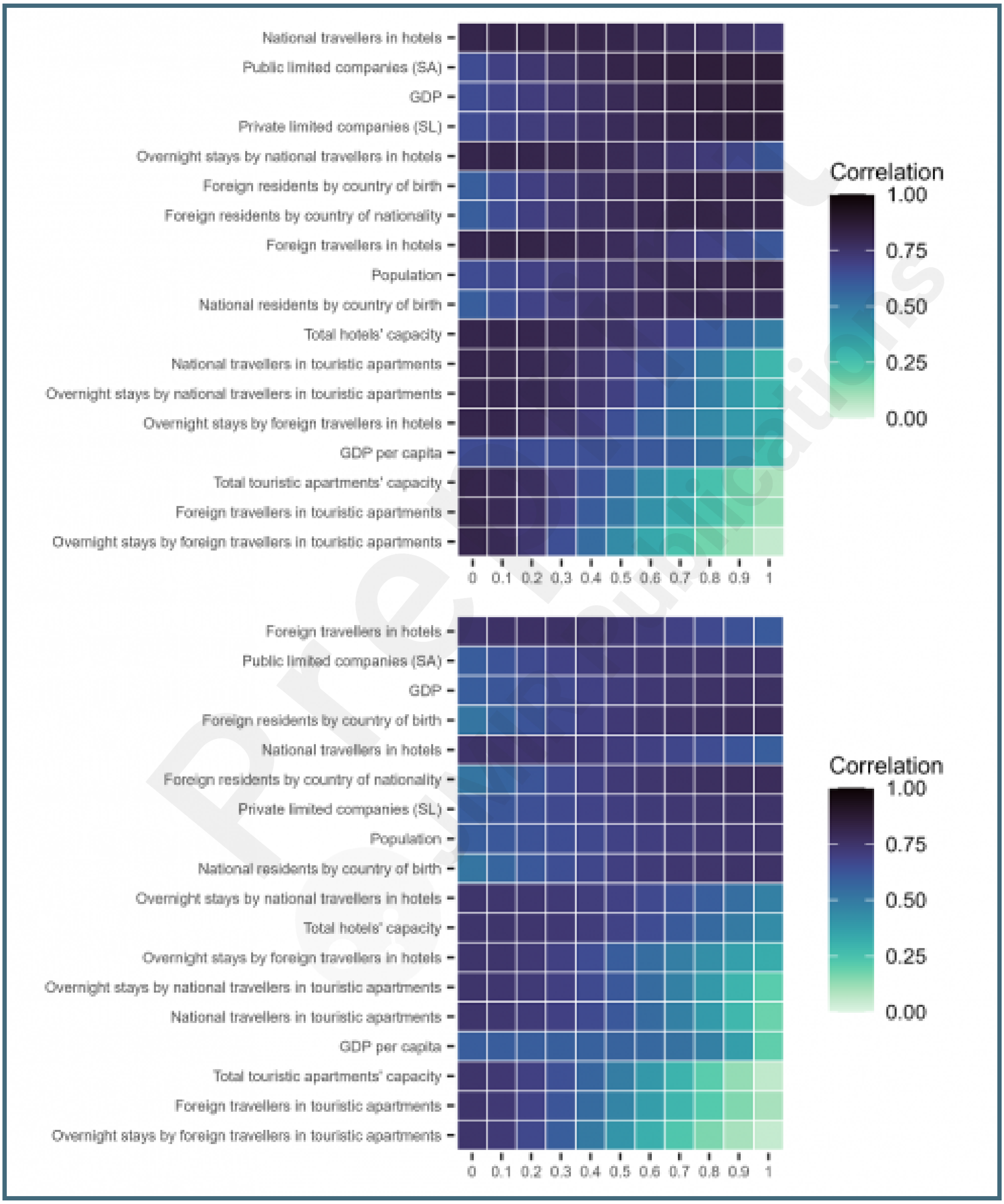
Summary of the model’s building process. Key steps are the fitting, selection, validation and assessment of the model.



Summary of the best linear models for 2019 imported cases of malaria (top row) and dengue (bottom row). The left column shows the models’ predictions (in red), together with the number of reported cases (in blue) for 2019 at each province in Spain. The right column shows the fit between the models’ estimates and the official records (inset figures correspond to the fit after removing Madrid and Barcelona from the dataset, as explained in "Estimates and model assessment").



Summary of each input variable’s performance on the estimates for malaria (top) and dengue (bottom). Each square in the figures is colored according to the average correlation between the official 2019 reports and the estimates provided by the weighted models including each of the variables, with the associated weight ranging from 0 (no contribution from the variable is assumed in the model) to 1 (the model only includes that variable). The variables are ranked from top to bottom according to the overall average correlation with 2019 data of the estimates of the models including each variable.



Multimedia Appendixes

Supplementary material for main manuscript.
URL: http://asset.jmir.pub/assets/bb5c64b4b646666d86c4930509eae8c7.pdf

TOC/Feature image for homepages

The travellers' index: regional distribution of imported cases of dengue and malaria can be described using local socioeconomic statistics.