



OPEN

A model for simulating emergent patterns of cities and roads on real-world landscapes

Takaaki Aoki^{1✉}, Naoya Fujiwara^{2,3,4,5,6}, Mark Fricker⁷ & Toshiyuki Nakagaki^{8,9}

Emergence of cities and road networks have characterised human activity and movement over millennia. However, this anthropogenic infrastructure does not develop in isolation, but is deeply embedded in the natural landscape, which strongly influences the resultant spatial patterns. Nevertheless, the precise impact that landscape has on the location, size and connectivity of cities is a long-standing, unresolved problem. To address this issue, we incorporate high-resolution topographic maps into a Turing-like pattern forming system, in which local reinforcement rules result in co-evolving centres of population and transport networks. Using Italy as a case study, we show that the model constrained solely by topography results in an emergent spatial pattern that is consistent with Zipf's Law and comparable to the census data. Thus, we infer the natural landscape may play a dominant role in establishing the baseline macro-scale population pattern, that is then modified by higher-level historical, socio-economic or cultural factors.

Across the surface of the earth throughout human history, civilisations have developed from small, scattered settlements into substantial agglomerations of towns and cities that are linked by increasingly sophisticated road networks. The theory about why populated places and transport networks arise at certain geographical locations in the natural environment, and what drives the massive variation in spatial infrastructure are long-standing problems dating from the seminal work of de la Blache¹. These fundamental questions still attract growing interest nearly a century later, because the resultant geospatial distribution of cities and road networks have major impacts on substantive issues in contemporary society such as urbanisation, traffic congestion, rural depopulation, land use, food supply, and ultimately migration.

In the 1930s and 40s, Central Place Theory emerged as an explanation for the observed hierarchical urban organisation from conceptual models advanced by Christaller² and Lösch³, who argued that cities are located hierarchically on a regular 2-D hexagonal lattice as a result of differing population thresholds for viability of particular goods or services, offset against maximum travel distances^{4,5}. In the 50s, Isard was instrumental in combining these ideas with earlier work on agricultural land use⁶, and industrial location⁷, in an attempt to construct general equilibrium models within a regional science framework⁸. Further location-theory models were subsequently developed within the New Economic Geography framework, which integrated spatial development into economic models^{9,10}, although these also have links back to von Thünen¹¹. The critical argument posited by these models is that locations with high or low populations emerge intrinsically without any external environmental differences, with behaviour reminiscent of Turing-type reaction-diffusion pattern formation¹². Such self-organising spatio-temporal pattern formation is common in both physical¹³ and biological¹⁴ systems, where structure can spontaneously emerge as a result of the collective behaviour of many agents following simple local rules, without centralized planning or control.

Nevertheless, in these theories, the transport network is typically considered to exist a priori rather than emerge as product of the developing human society. Thus, in most studies, the transport system is treated as an exogenous factor given by external datasets detailing existing road, highway, and railway infrastructure^{15–19}. However, in reality, the transport system develops in concert with shifts in human population^{20,21}, and iteratively

¹Faculty of Education, Kagawa University, Takamatsu 760-8521, Japan. ²Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan. ³PRESTO, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan. ⁴Center for Spatial Information Science, The University of Tokyo, Kashiwa 277-8568, Japan. ⁵Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan. ⁶Tough Cyberphysical AI Research Center, Tohoku University, Sendai 980-8579, Japan. ⁷Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK. ⁸Research Institute for Electronic Science, Hokkaido University, Sapporo 001-0020, Japan. ⁹Global Station for Soft Matter, Global Institution for Collaborative Research and Education, Hokkaido University, Sapporo, Japan. ✉email: takaaki.aoki.work@gmail.com

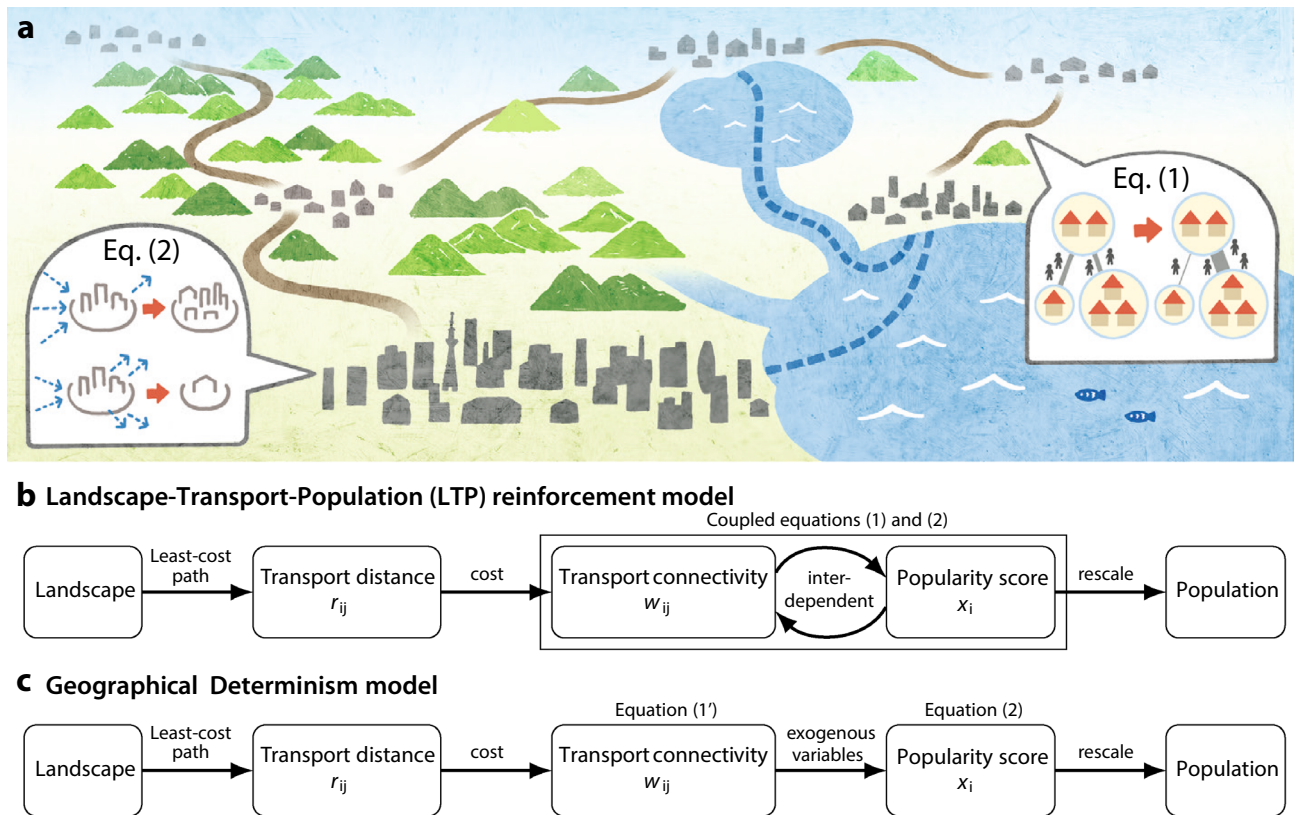


Figure 1. (a) Schematic illustration of dynamical pattern formation in the LTP-reinforcement model where populated places and their inter-connections emerge naturally across the landscape. Transport links following least-cost paths across the landscape are strengthened between locations with large populations, whilst well-connected centres in turn tend to grow to a greater extent. (b,c) Schematic diagram of the LPT-reinforcement model (b) and a counterfactual, 'Geographical Determinism' model with no reinforcement of the transport network (c).

feeds into the overall dynamics of spatial pattern formation, particularly over longer timescales. Furthermore, transport routes are very sensitive to the natural landscape, which is intuitively well recognised, but absent from most models.

Thus, in this paper, we go back to the primary question of the impact of natural landscape on patterns of population, using a model where both cities and roads are allowed to emerge from dynamical Turing-like pattern formation, and the only a priori constraints are based on physical geography. In particular, we consider that mountains, hills, rivers, lakes, and coastlines are likely to have a substantial effect on both the population distribution and the transport network, but are rarely considered quantitatively^{19,22}. We capture these elements by inclusion of high-resolution topographic maps to sculpt the surface on which the model unfolds by progressively transforming the underlying space from the typical idealized isotropic 'flatland' to the real landscape. The resultant Landscape-Transport-Population (LTP) reinforcement model has three key elements: (i) Detailed landscape topography at 90 m resolution; (ii) The effective transport distance across the real-world landscape; and (iii) Turing-like pattern formation driven by flow-based reinforcement for population distributions. Our approach draws on our previous works^{23–27} where a biologically-inspired, current-reinforcement rule can construct realistic transport networks between food-supply points under physical and physiological constraints, and the principle of co-evolution of nodes and links that dynamically organize scale-free networks via a diffusion process^{28,29}.

Overview of the dynamical model

A schematic illustration of how the dynamical model evolves across a landscape is shown in Fig. 1a. Growing centres of population are connected via transport routes across land, rivers, lakes or sea. Preferential transport routes between locations are calculated by gradient-based least-cost path analysis on topographic GIS maps, taking into account the costs of moving over rough terrain or water bodies (see Methods for details). Given the effective transport distance of such a route between each pair of locations, i and j , we introduce a connectivity variable that weights the amount of traffic and trade. We model development of this transport connectivity using a mass-action formula, which originated from chemical reaction processes and has been applied to other population and social dynamics³⁰:

$$w_{ij}(t+1) - w_{ij}(t) = \epsilon [x_i(t) \cdot x_j(t) - f(r_{ij})w_{ij}(t)]. \quad (1)$$

Thus the change in transport connectivity is an increasing function of the popularity (normalised population) of the end-point locations given by variables x_i and x_j and offset by the increasing cost of transport with distance. The dependence on distance is given by $f(r) = \frac{r}{R} \exp(r/R)$, where R represents the typical length scale of transportation in the epoch under consideration. This specific form of $f(r)$ originates from the distance deterrence factor in spatial interaction models^{31–34}. The Eq. (1) leads a global equilibrium $w_{ij}^* = x_i x_j / f(r)$ when x_i is fixed. ϵ controls the time-scale of the evolution of this connectivity.

Next, we formulate the coupled hypothesis that centres of population will develop at important transport intersections that reflect the network centrality of each location. Here we use PageRank centrality³⁵ as a well-characterised estimator of the relative popularity of location i , according to Eq. (2):

$$x_i(t+1) - x_i(t) = d \sum_{j \in \mathcal{N}_i} [T_{ij}(t)x_j(t) - T_{ji}(t)x_i(t)] + (1-d) \left(\frac{1}{N} - x_i(t) \right), \quad (2)$$

where the transition probability $T_{ij}(t)$ from j to i is given by $\frac{w_{ij}(t)}{\sum_{k \in \mathcal{N}_j} w_{kj}(t)}$ and N is the number of locations. The first term sums the incoming and outgoing flows to location i from connected locations via the network, giving preference to locations with strong transport connectivity $w_{ij}(t)$. The second term models global dispersion driven by non-network factors, to ensure the agricultural hinterland around the cities remains populated. The balance between these two terms is controlled by parameter d . This popularity x_i accumulates the probability of visiting location i via a diffusion process on the weighted network, with all probabilities summing to unity. The actual population X_i is determined by scaling this normalised variable by the total population from the census data (see Supplementary Information 3.7 for details).

The initial state of the system variables x_i , w_{ij} is another important factor which determines the final outcome of the simulation. In the following simulation, we firstly do not provide any specific information on the initial state which is set to be nearly homogeneous with small random fluctuations (see Supplementary Information 3.4 for full details). Later, we set the initial state to be similar to the situation in the Roman era and discuss on the effect of the specific initial condition on the temporal evolution of the system.

Figure 1b summarises the coupled system of Eqs. (1) and (2) schematically. The natural landscape is incorporated into the model only through the impact on the least-cost transport distance r_{ij} . The resultant transport connectivity w_{ij} then interacts inter-dependently with the popularity at each location x_i . In other words, as populated places develop or decay, the connections among them are restructured to meet current needs. This network restructuring causes further expansion or contraction of populated places. Although the model has only three parameters: R , ϵ and d (see Supplementary Information 3 for full details), the reinforcement dynamics exhibit rich behaviour.

Results

Impact of varying landscape scenarios on population distribution. The impact of landscape factors on the emergent pattern of population distributions and transport networks was explored for four scenarios using Italy as an exemplar. The underlying space was changed from: (i) an idealised isotropic 'Flatland', with no geographical features and no boundaries (Fig. 2a); (ii) imposition of the Mediterranean Sea and Italian 'Coastline' (Fig. 2b); (iii) inclusion of terrestrial topography to give 'Elevation' (Fig. 2c); and (iv) addition of 'Water Bodies' such as rivers and lakes which augment the transport possibilities (Fig. 2d). In the isotropic 'Flatland' scenario the underlying space is uniform and symmetric with periodic boundary conditions. Nevertheless, a regularly-spaced, lattice-like pattern of populated places, with roughly equal population, still spontaneously emerges as a result of the reinforcement dynamics in the model (Fig. 2e). Each populated place also acts as a hub in an almost radially symmetric traffic network (Fig. 2i).

This phenomenon of pattern formation is widely seen in reaction-diffusion systems³⁶. Linear stability analysis of the isotropic state reveals that an Eigen mode with a finite wavelength initially becomes unstable as the control parameter d is varied above a critical threshold ($d_c = 0.8$), and the mode then develops to give the observed pattern (see Supplementary Information 7 for the details).

By adding landscape factors, the population and transport networks show more complex structure. In the 'Coastline' scenario, uniformity and symmetry are broken, with the sea providing a major barrier to movement, although the land is still flat. The emergent pattern now deviates from a perfect lattice, but still generates a reasonably regular array of similar sized cities (Fig. 2f), with some preferential strengthening of transport links to nearest neighbours (Fig. 2j). In the Elevation scenario, the least-cost paths for transport are affected by the local gradient of the landscape, and trigger a significant shift away from a regular lattice with substantial variation in population size and location (Fig. 2g). The transport network is also more complex with hubs emanating from the major cities on the northern plane, supplemented by coastal highways and occasional trans-Alpine routes (Fig. 2k). Finally, inclusion of 'Water bodies' permits transport across rivers, lakes, and seas, which previously acted as barriers in the other scenarios, and opens up the possibility of marine transport along the coast (Fig. 2l), that further modifies the population distribution (Fig. 2h).

Comparison of the model output with the 2011 census data. We then asked whether there are any similarities between the emergent patterns from the model and actual data on populations and transport. Figure 3 shows comparison with the 2011 population census data, and GPS-tracked traffic data at two administrative levels (see Methods for the details of these datasets). The aggregate population and traffic are shown at the *Regione* level in Fig. 3a, b, respectively. There is substantial variation in both the census population and traffic data between regions, but somewhat surprisingly, much of this heterogeneity is faithfully captured by the LTP

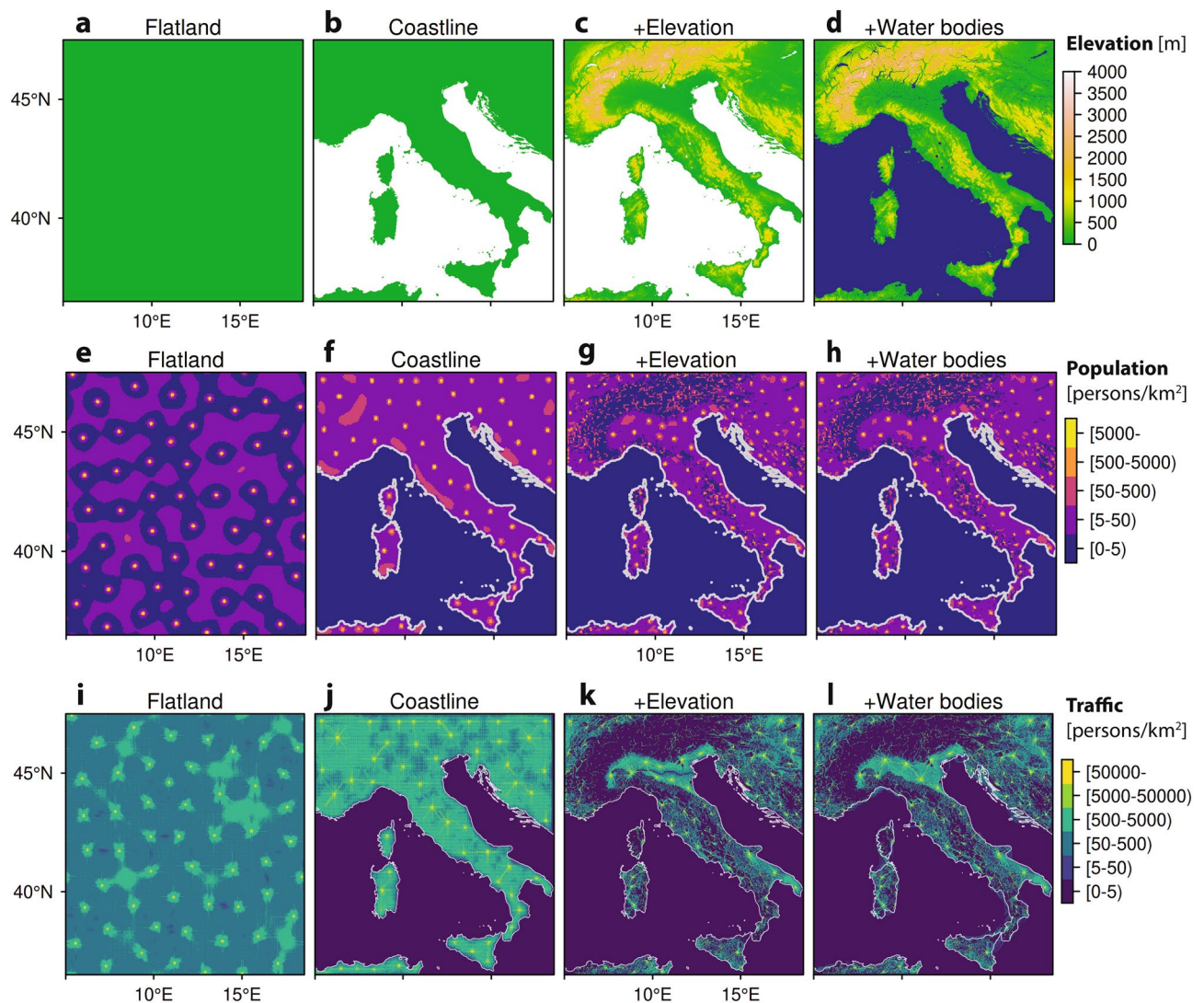


Figure 2. The effect of landscape factors on emergent geospatial patterns of both human population distribution and the interconnecting transport network. The underlying space is changed from an idealised flatland condition (a) to a more realistic one by sequentially adding coastlines (b), elevation (c), and water bodies (d), using Italy and its surrounding regions as a case study. By adding these factors, the emergent population pattern produced by the model shifts from a lattice-like Turing pattern with regular population sizes to a markedly heterogeneous distribution (e–h). In parallel, the transport network also emerges between centres of population, as depicted by the net traffic $\hat{W}(x)$ passing through each location (see Supplementary Information 3.8 for details) (i–l).

model, even though the model only has information of the landscape. Nevertheless, the match is not perfect, with the most obvious differences in *Lombardia* and *Campania*, which are under-estimated, and *Apulia* which is over-estimated.

At higher spatial resolution, the population at *Provincia* level is shown in Fig. 3c–f for the four different landscape scenarios. Error bars depict the standard deviation ($n = 64$) with different initial conditions (see Supplementary Information 3.4 for details). The Flatland scenario leads to substantially different populations in each *Provincia* at each realisation, apparent from the large error bars, and gives little overall match to the census data (Fig. 3c). Inclusion of more landscape features progressively reduces the variance at each location, stabilises the output, and improves the match between histogram profiles from simulation and census (Fig. 3d–f). Much of this improvement arises from inclusion of elevation (Fig. 3e). However, the additional transport routes available in the ‘Water Bodies’ scenario, allow increased growth of *Milan* (MI), for example (Fig. 3f), and the population of other major cities, such as *Turin* and *Palermo* match the census well. Nevertheless, populations of *Rome* and *Naples* are under-estimated, whilst those in *Alessandria*, *Modena*, *Pisa*, and *Foggia* are over-estimated. The comparison for traffic data at the *Provincia* level also showed a similar improvement with the addition of landscape factors (Figure S5 in Supplementary Information 5).

It is noted that changes in response to the underlying landscape reflect the response of the complete system of cities linked by adaptive connections, not just cities in isolation. For example, a population increase in *Naples*

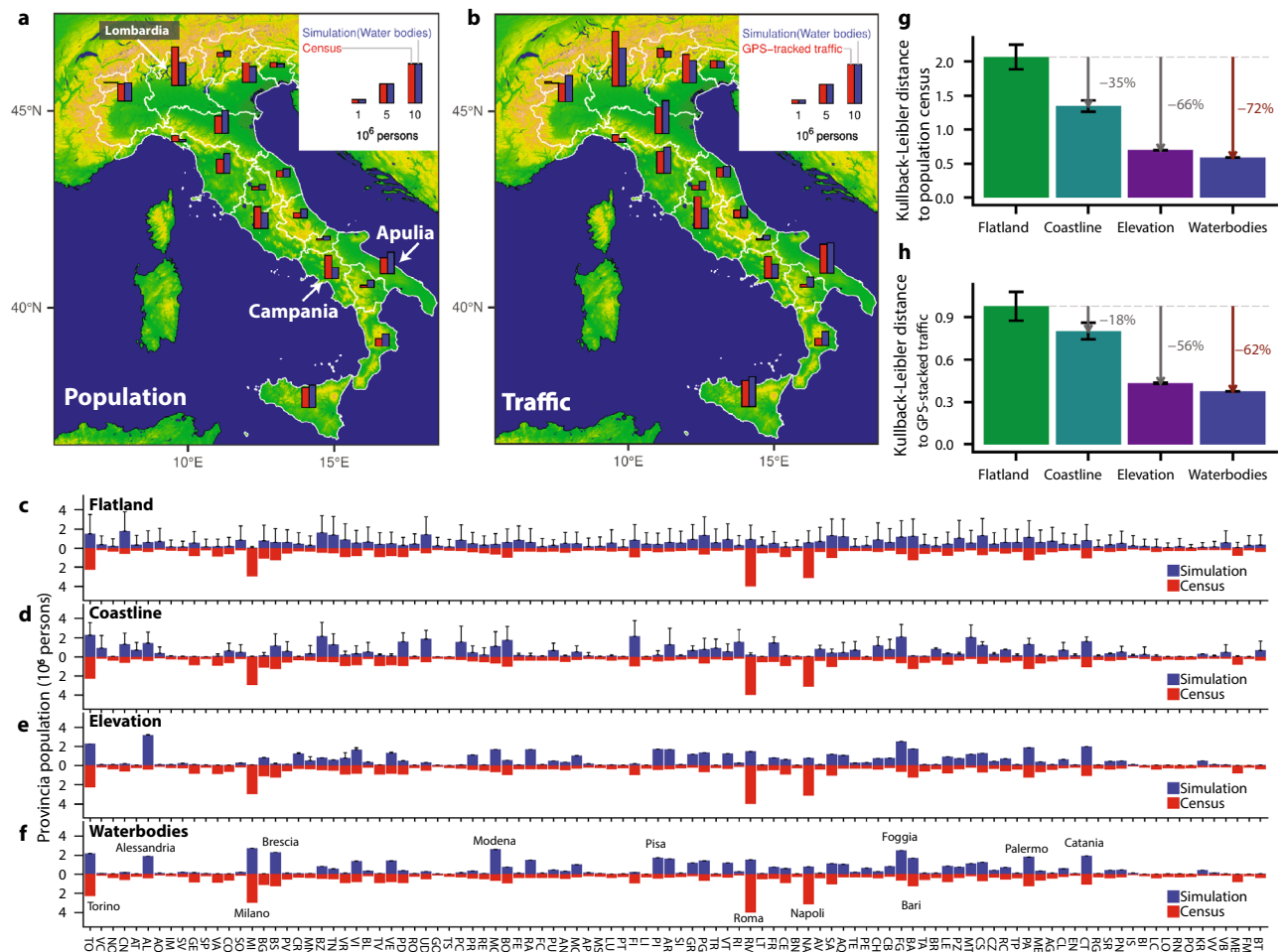


Figure 3. Quantitative comparison between the population and traffic datasets for Italy and the LTP simulation results for different landscape scenarios. (a) Population distribution from the 2011 census data for Italy (red) and the LTP simulation (blue) when aggregated at the *Regione* level for the 'Water bodies' scenario. (b) GPS-tracked traffic distribution from OpenStreetMap for Italy (red) and the LTP simulation (blue) when aggregated at the *Regione* level for the 'Water bodies' scenario. (c–f) Comparison in population distribution as in (a), but at *Provincia* level for each scenario. Error bars depict the standard deviation that comes from the 64 simulation runs for each scenario. (g,h) The Kullback–Leibler (KL) distance from the simulation data to the census population (g) and the GPS-tracked traffic datasets (h) for *Provincia*-level distributions. The error bars are the same as in (c–f).

induces not only population changes in neighboring places, but also the changes in the connections around it. These changes induce further changes at distant locations, and propagate over the entire network in subsequent time steps. Therefore, the effect of the landscape has to be interpreted through its impact on the entire system, not as separate local influence on individual cities.

To evaluate the match between these distributions quantitatively, we used the Kullback–Leibler (KL) distance³⁷ from the simulation distribution (q) to the census data (p), according to Eq. (3):

$$D_{KL}(p||q) = \sum_i p(i) \log \left(\frac{p(i)}{q(i)} \right), \quad (3)$$

where p, q are the normalized spatial distribution such that $\sum_i p(i) = \sum_i q(i) = 1$. The index i of p, q indicates the individual *Provincia* or *Regione*, not the size rank. The KL distance progressively decreases as the landscape factors are included for both the population estimates (Fig. 3g) and traffic predictions (Fig. 3h). The Water Bodies scenario gives a total improvement of 72% for population and 62% for traffic. This improvement is also confirmed by other metrics, such as Pearson's correlation and cosine similarity, which test slightly different aspects of the distributions (see Supplementary Information 6 for the details).

Counterfactual comparison without dynamic reinforcement. One explanation for these results might be that the population pattern simply reflects the external landscape alone—an idea reminiscent of geographical determinism^{38,39}, whereby the local population would essentially match the slope metric. However, we argue that the observed pattern could require the reinforcement dynamics linking population aggregation and

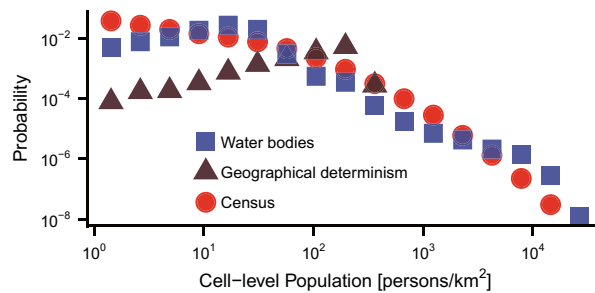


Figure 4. Population density distributions on a 6 km grid showing a typical Zipf's Law relationship for the census data (red), and the LTP model (blue), which shows a comparably broad distribution. In contrast, a counterfactual 'Geographical Determinism' scenario (brown), with the same landscape conditions as 'Water Bodies', but lacking the reinforcement dynamics of the model, does not fit the census population data.

transport connectivity. To clarify this argument, we introduce a counter-factual scenario termed Geographical Determinism (Fig. 1c). In this scenario, the underlying landscape conditions are the same as the Water Bodies scenario, but with no transport reinforcement dynamics (see Supplementary Information 4 for details). The Geographical Determinism model yields some variation in population size, but only within a narrow range (Fig. 4). In contrast, the LTP simulation show a strongly heterogeneous distribution, in agreement with the census data (Fig. 4). It is noted that this result does not reject geographical determinism in general. Instead, we argue that the reinforcement dynamics provides an alternative mechanism that could be a contributing factor to the heterogeneity.

The fat-tailed distribution of population is already well-known as Zipf's law (Fig. 4)⁴⁰, and its potential origin extensively investigated^{41–44}. The LTP model provides an alternative mechanism that could be a contributing factor to the heterogeneity: we already showed that roughly equal-sized populated places emerge in an idealized 'flatland' scenario but shifted to be strongly heterogeneous by inclusion of the underlying real-world landscape in the dynamical model, while the landscape on its own was insufficient. Thus, the combination of the heterogeneity of the underlying landscape and the intrinsic reinforcement dynamics in the model could be a potential driver of a fat-tailed population distribution.

Inclusion of historical population estimates improves model predictions. We then explore the effects of including two elements of "history" in the model. First, we use the population distribution of Roman cities^{45,46} to specify the initial conditions (Fig. 5a). Second, we vary the characteristic transport distance R in the cost function (Fig. 5d) to reflect changes in the transport system for each historical age (see Methods and Supplementary Information 1 for details). In the Roman era, the predominant transport modes on land were limited to walking, ox-carts or horse-drawn carriages. We assume that such a transport system has a shorter characteristic distance R ($= 8$ km) than that of a contemporary motorized society ($R = 40$ km), and that sailing is comparably more efficient than movement on land. We therefore follow the evolution of the population with the ramped increase in R and the resultant variation in least-cost path parameters (Fig. 5d). At an intermediate state with $R = 24$ km, the model predicts many medium-sized towns and the beginning of larger cities on the northern plain (Fig. 5b). In the final state with $R = 40$ km, some of these populated places merge into several larger metropolises, particularly in north Italy, at the expense of locations that were dominant in previous epochs, but have subsequently declined (Fig. 5c).

Figure 6 compares the intermediate predicted population distributions up to $R = 24$ km with the empirical population distribution estimates from 1300 to 1861⁴⁷ and the census data for 1911. These are aggregated at the *Provincia* level using boundaries existing in 1911 following re-unification, as these are available in digital format for the whole Italian Peninsula. As shown in Fig. 6, much of the spatial heterogeneity is faithfully captured by the model across these ages. The quantitative comparison shows that the KL distance remained around 0.2 during this period (Fig. 6i), with the notable exception of 1400 when city populations were drastically reduced by the plague pandemic.

The effect of including these historical elements was evaluated by the improvement of KL distance (Fig. 5e, f). Compared to initial near-isotropic starting conditions, the constrained historical evolution gave a further improvement of 16%, which provides a quantitative measure of the extent that integration of these historical pathways has had an impact on contemporary society.

Discussion

Large-scale regional pattern formation of cities and transport networks is a key theme of long-standing importance in the study of human civilization. Here we present a dynamical model in which populated places and their connections co-evolve inter-dependently, and we used this to examine how the emergent patterns are influenced by the detailed topographical landscape. The model incorporates spatial scales from tens of meters (from topographic maps) to thousands of km (the overall modelling domain) and is demonstrated for Italy as a test case. We show that in an idealized 'flatland' scenario, a regular pattern of cities with roughly equal population spontaneously emerge, consistent with the arguments set out in Central Place Theory by Lösch³, and observed

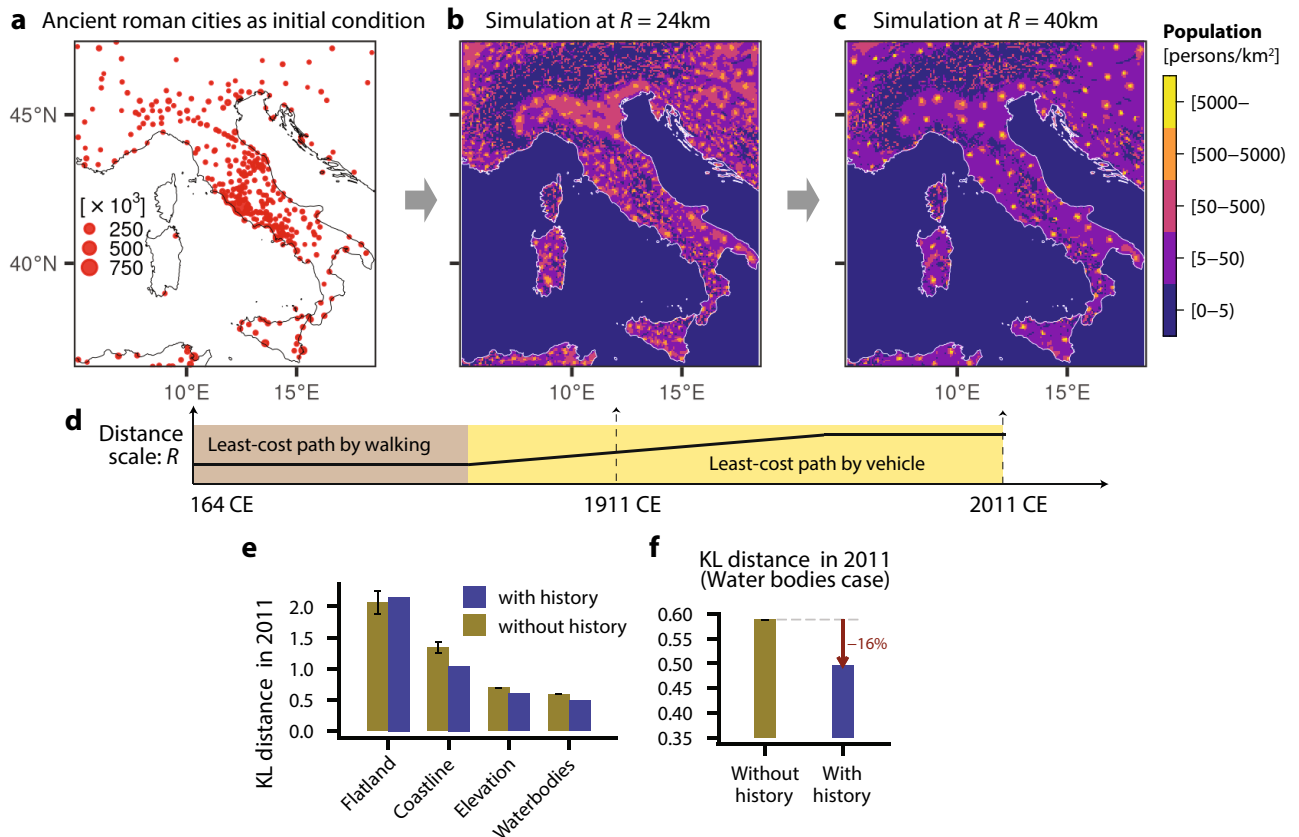


Figure 5. Historical dependency of the simulation on the past population distribution in Italy. (a–d) The initial state (a) is given by estimates of population distribution in ancient Roman cities. Starting from this initial state, the intermediate state of the simulation at $R = 24$ km (b) and the final state with $R = 40$ km (c) are depicted, as the distance scale, R , is increased ramped up from 8 to 40 km during the simulation (d). (e,f) Improvement in the KL distance from the 2011 census data at the *Provincia*-level by the inclusion of historical development is shown for all scenarios (e), and zoomed-in for the final Water Bodies scenario (f).

in the initial NEG models⁹. However, inclusion of the natural landscape adds considerably more richness (heterogeneity) to the emergent patterns of both population distribution and the transport network. The resultant patterns are broadly comparable to contemporary population census data and GPS-tracked traffic data, including emergence of a fat-tailed distribution of population density, as seen in the empirical Zipf's law. We consider it remarkable that such a minimal model can explain complex geospatial patterns not only of population, but also of transport networks. Thus, we would argue that a major contribution of this paper is that it reveals the importance of the natural landscape on human civilization with the phenomenological model of co-evolving dynamics of cities and roads, both qualitatively and quantitatively using information-theoretic measures referenced to the actual census data.

To set these results in context, we believe there are several important differences in this modelling approach to those adopted in recent urban economic models. First, our model has extremely few parameters, whose values are set completely independently from the actual population data. In comparison, existing models include detailed micro-economic processes, with the aim of providing as realistic an output as possible to inform policy, planning and investment. This results in elaborate systems with a huge number of location-specific fudge factors¹⁵, such as attractivity¹⁷, local amenity^{16,48}, or local productivity²², which are then tuned by calibration using the actual population data itself.

Second, our model only uses landscape datasets as a priori conditions to reproduce the spatial patterns of population. In contrast, in addition to calibration against the population data itself, alternative realistic models incorporate a wide range of external datasets, including *inter alia* employment levels, economic growth, birth and death rates, bilateral trade flows income and demographic characteristics^{15,16,49}. Nevertheless, the natural landscape itself is not included directly in most models, or plays a very limited role. For example, in a few studies landscape factors have been used to validate the calibrated local amenity⁴⁸, as a ruggedness parameter in counterfactual simulations²², or to evaluate the local suitability¹⁹ by elevation or slope in land-use allocation within cellular-automaton based models^{18,19}.

Third, in our model, the transport infrastructure is not given a priori, but develops inter-dependently with shifts in population. In existing models, contemporary rail, road and water transport networks are often supplied exogenously and the accessibility of different locations are imposed a priori and only vary to a limited extent, for example when modelling congestion in intra-urban models^{19,50,51}. The spatial pattern of these networks already provides a strong guide to where populated places are located. This is important when making short

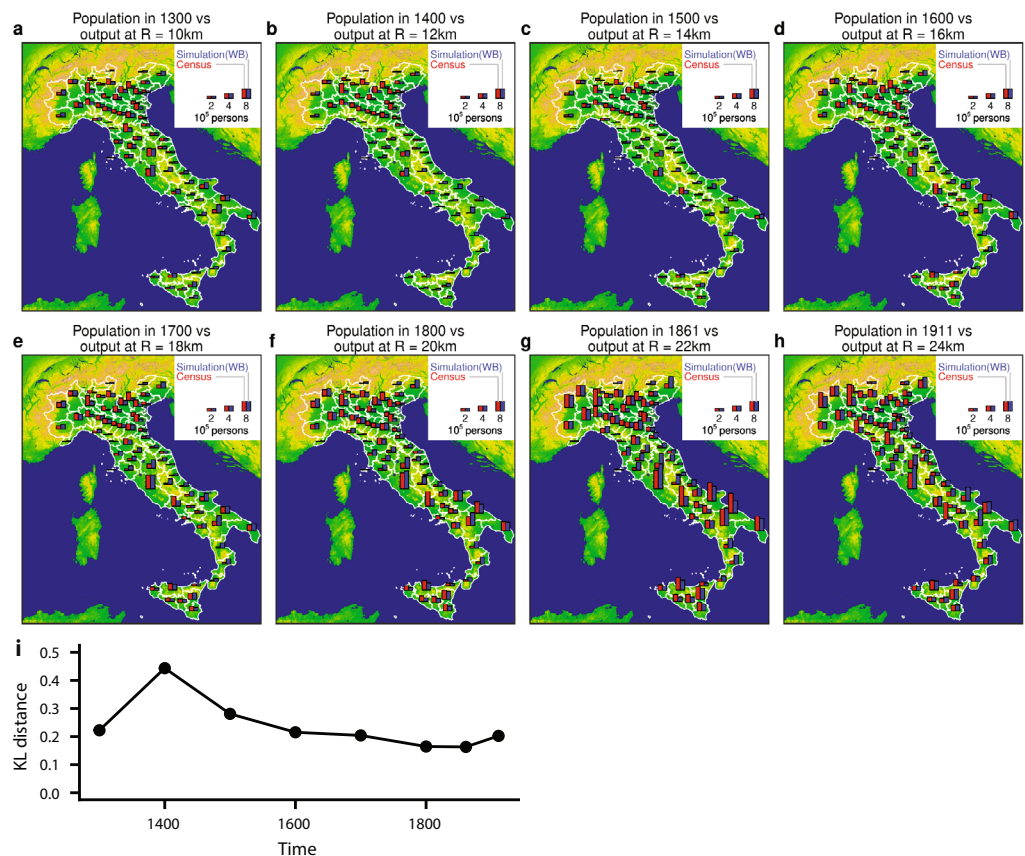


Figure 6. Comparison between the population distribution during the period 1300–1911 and the simulation output from intermediate levels of R . (a–g) As shown in Fig. 5d, the least-cost distance and characteristic transport distance R were varied during the simulation. The intermediate states of the model up to $R = 22$ km were compared with the estimated population distribution from city population data during the period 1300–1861⁴⁷ (see Methods for the details of the estimates). The population distributions from both the simulation output and the estimated data are aggregated by *Provincia* using the 1911 regions, as digital shape data is available at this time point. Red bars indicate the estimated populations and blue bars are the simulation output from the Water Bodies scenario. (h) Population distribution at 1911 for Italy (red) and the LTP simulation (blue) at $R = 24$ km when aggregated at *Provincia* level in 1911. (i) Quantitative comparison using the KL distance from the model outputs at intermediate values of R to empirical data shown in (a–h).

term predictions on local development relevant to policy and investment, but does not address the origin of these networks over longer time periods^{17,52}.

In summary, in contrast to existing models, which involve thousands of tunable parameters and detailed external datasets to fit population data as precisely as possible, here we adopted an opposite minimal approach. The system provides an abstract representation that captures the co-evolution rules governing emergence of populated places and their interconnections, when constrained solely by the topographic landscape without any other external datasets. The true population is only used to compare with the output of the model to evaluate its performance, not for calibrations. Thus, our modelling framework provides a straightforward explanation as to why some locations become so populated and others do not, by the direct influence of topographic landscape.

We next show that landscape on its own is not sufficient to explain the population distribution as a form of geographical determinism, but requires the inter-dependent dynamical feedback between population and the transport network emerging in parallel. Such co-evolution has been highlighted as a potential driver of self-organizing system for adaptive biological and social networks⁵³, including transportation networks in an idealized city⁵⁴ or region^{55,56}. In general, the reinforcement by such adaptive spatial connections represents a macroscopic, ‘rich-gets-richer’ effect in social dynamics. Several positive feedback effects, such as economy of scale, or agglomeration¹⁰ are well established in spatial economics, whilst other feedback mechanisms have also been incorporated in spatial interaction models^{57–59}. In particular, the recursive iteration feedback employed by Wilkinson et al.⁶⁰, embodies remarkably similar principles to the proposed LTP reinforcement model for transport dynamics.

It is notable that de la Blache¹ also considered the significance of historical factors on the subsequent development of the population distribution, in addition to the central role of the natural landscape. We have begun to explore these possibilities in the model, by incorporating aspects of historical evolution, such as the estimated

populations in ancient Roman cities as the starting condition. In the context of dynamical systems theory, the inclusion of historical elements is related to the issue of ‘multistability’: the model has multiple possible equilibrium states as the result of its evolution, whilst there is only one actual realisation in history. In the model, the initial condition determines which one of the possible states is realised in the simulation. In other words, the past condition can affect the future state by selection from multiple possible outcomes. Compared to the average from multiple runs using noisy near-isotropic initial conditions, starting with the “historical” condition from the ancient Roman period selects a better final outcome to the contemporary census dataset in a single model run (Fig. 5).

There are several limitations of the model that need to be considered carefully. The model appears to work well for Italy, which has natural boundaries set by the Mediterranean Sea and the Alps to the North. However, we would expect the population outside the simulation region to influence locations inside, but these are not captured in the scope of the model. Scaling the model up to a continental or global level to avoid these edge artefacts is technically plausible, but currently computationally prohibitive unless one adopts greater constraints on the initial conditions to limit the model runs, or sacrifices some of the spatial resolution to reduce the grid size.

There are also some limitations in the topography dataset, with implications for the predicted impact on transport. Although the GIS data is high-resolution, it does not capture small rivers or narrow passes in mountainous areas that might have a disproportionate impact on transport routes. Furthermore, whilst the coefficients for calculation of the least-cost paths across the landscape are externally derived from other studies, the values are only approximate for each period under study. There are also several coefficients that require refinement, but at the expense of generality. For example, inclusion of larger rivers in the Water Bodies scenario has the positive benefit of permitting river freight and preferential transport between linked river locations. However, as currently implemented, it also uniformly removes the river as a barrier to crossing along its length. In reality, fords and bridges are relatively sparse and have a strong impact on local centres of population, but their location cannot be predicted with the resolution currently available. In addition, the underlying landscapes are assumed to be unchanged in the model, whilst in practice the landscape has been altered by engineering of transport routes to explicitly circumvent barriers, using tunnels, bridges, or elevated roads. Thus, the situation assumed in the model would be more relevant to epochs before the introduction of highly structured and planned transport networks in the modern era, such as motorways, railways and air transport.

Another limitation arising from the assumptions in the simplified model, is that various resources needed to support populations are uniformly distributed. In reality, resources are spatially-varying, and the model can be partly reformatted to include a spatially-varying population capacity, C_i , in Eq. (2) (see Supplementary Information 3). Moreover, as resources are consumable, shortages may arise from large populations, curtailing local growth. The negative feedback by consumable resources will be important to consider for the sustainability of cities and the transport connections between them. In addition, the model is mathematically described as a dynamical system. It assumes deterministic and gradual changes of the system, while rare, but large, interurban migratory shocks might also have a significant effect on long-term growth of cities⁴⁴.

Overall, we argue that the LTP model should probably be best considered as providing a sophisticated null model to facilitate thought experiments about the spatial patterns of cities and roads, and the factors that might influence them. In this framework, the LTP model provides a baseline reference tool to predict the expected population distribution when constrained solely by topography, in the absence of higher-level socio-economic or cultural factors. In this sense, the deviation of the real-world data from the model at each location is perhaps the most interesting feature, as it quantitatively indicates the influence of (untested) spatially-varying environmental factors, such as climate, water, soil, and crop productivity, on the one hand, or local economic, social, cultural, historical and political factors, on the other. For example, the predicted population of Rome is about a third (35.9%) of the actual number, suggesting that its status as both a capital city and a religious centre has stimulated three-fold more growth than expected. To test such factors empirically, the model can be reformatted to include a spatially-varying population capacity or other factors with relevant datasets for future works. This approach provides a systematic way to test the impact of various local or regional factors individually or in combination and how their impact might propagate across the network. We argue that such a bottom-up, building-block approach starting with known external factors drawn from physical geography offers a new direction to deconstruct the complex phenomena of human civilization involving many natural and social factors.

Methods

Calculation of transport distances across landscapes. *Topographic datasets.* We used the Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global elevation dataset for the digital elevation model, and the SRTM Water Body, and Global River Classification (GloRic) datasets for ocean, lakes and rivers. These datasets are publicly available, distributed by the United States Geological Survey (USGS)⁶¹ and World Wildlife Fund (WWF)^{62,63}. In this study, we used the region ranging from 5° to 18.55° longitude, to 36.5° to 47.5° latitude that contains Italy and the surrounding Mediterranean Sea. The SRTM datasets contain the information on the type of cell, such as land, ocean, lake, or river, and the elevation in meters for each cell type. The data were downsized to 3 arcsecond spacing by median resampling to reduce the effects of outliers and reduce the computational complexity. To supplement the river network, rivers in the GloRic dataset greater than hydrologic class 1 were included.

Least-cost path analysis. We used standard GIS least cost path analysis⁶⁴ to calculate the route for any detour, taking into account the topography. Thus, for each cell in the topographic dataset, the transportation cost to move to its eight adjacent neighbours was evaluated. The cost depends on both the slope, and cell type of the origin and destination cells. The least cost path between any pair of cells was then calculated on the grid graph

to minimize the accumulated cost along the overall length of the route, r . See Supplementary Information 1 for full details of the calculation and parameters.

Datasets for comparison with simulation results. *Census in 2011.* The census in 2011 was taken from the GEOSTAT 2011 grid dataset, distributed from Eurostat, EFGS⁶⁵. In Fig. 3, this census data was aggregated by Italian administrative units (*Regione* and *Provincia*). The boundary information of these administrative units in 2011 was obtained from the Italian National Institute of Statistics⁶⁶.

Census in 1911. The *Provincia* population in 1911 (Fig. 6) was taken from a data repository⁶⁷, originally from the Almanach de Gotha⁶⁸. The boundary information of *Provincia* in 1911 was obtained from the Italian National Institute of Statistics⁶⁶.

Estimated city population sizes in the ancient Roman period (164CE). The population of the ancient Roman cities shown in Fig. 5a were obtained from the open-access database of the Oxford Roman Economy project^{45,69}. The cities that existed at 164CE in the target region were selected from the database. The populations of the cities were estimated using a scaling law between city area and population, as proposed by Hanson and Ortman⁴⁶. The sum of city populations over the cities shown in Fig. 5 was 2,800,832, which is less than the estimated population in the entire area of about 12 million⁷⁰. The urbanisation rate rarely exceeded 20% of total population at this time⁷¹, thus the difference between these numbers represents the rural population outside the cities in total. The spatial distribution of the rural population is less clear because of methodological difficulties, but it is probably higher close to urban sites in highly urbanized areas, whereas it may be distributed further away from urban sites in low-populated regions⁷². We therefore modelled the spatial distribution of the rural population by a biased distribution to the locations close to populated cities, which is proportional to a distance-discounted score $s(i)$ at each location i :

$$s(i) = \sum_{u \in \text{ancient roman cities}} [r_{ju}/R]^{-1} \times \text{Population}(u), \quad (4)$$

where j_u denotes the location of a city u and r_{ju} is the least-cost distance specified for walking with $R = 8$ km.

Estimated population distribution during 1300–1861. During 1300–1861, the populations in Italian cities with greater than 5000 residents were taken from the Italian Urban Population database^{47,73}. The locations of the cities were determined by matching their names with the records in the GEONAMES database⁷⁴. The rural population in total outside the cities was calculated from the urbanisation ratio during the ages^{47,75}, and its spatial distribution is determined in the same way to that in 164CE—a distribution proportional to the distance-discounted score to populated cities, given by the Eq. (4) with $R = 8$ km.

In Fig. 6, the population distributions during 1300–1861 were aggregated using the *Provincia* boundaries in 1911 following re-unification. This is because (i) a common agglomeration is required to compare the value of KL distance across the middle ages. (ii) High-resolution shape files for *Provincia* in 1911 are available in digital format from the Italian National Institute of Statistics⁶⁶. (iii) *Provincia* in 1911 covers the whole Italian Peninsula, including Roma, Verona, and Vicenza that were excluded from the *Provincia* dataset in 1861.

GPS-tracked traffic. We use the traffic data which is publicly available from OpenStreetMap⁷⁶, as a proxy of transportation volume at each location. We used a set of GPS tracks in Italy, collected up to 2013. A track is an ordered list of GPS points describing a path. We excluded uncorrected data which only comprised a single point, and selected the tracks which had a minimum 1 km trip from a location in Italy. The resultant dataset has 63999 tracks with 90600639 points. The number of these points were counted in cells $45 \text{ arcsecond} \times 30 \text{ arcsecond}$ (about 1 km), to give the traffic data describing the number of people passing through each cell. It is noted that anonymous contributors who uploaded the GPS tracks may not necessarily be representative samples from Italy as a whole. Thus, to correct the sample, counts were weighted by the populations in provinces where the tracks originated.

Data availability

The topographic datasets and the census data that support the findings of this study are publicly available as noted in Methods.

Code availability

The simulation code of the model is available in a GitHub repository at <https://github.com/TakaakiAokiWork/geodynamics/>.

Received: 10 December 2021; Accepted: 27 May 2022

Published online: 16 June 2022

References

- de la Blache, P. V. *Principles of Human Geography*. Translated by M. T. Bingham (Lonstable and Co., Ltd., 1926).
- Christaller, W. *Central places in southern Germany* (Prentice Hall, Hoboken, 1966).
- Lösch, A. *The economics of location* (Yale University Press, New Haven, 1954).

4. Mulligan, G. F., Partridge, M. D. & Carruthers, J. I. Central place theory and its reemergence in regional science. *Ann. Reg. Sci.* **48**, 405–431 (2012).
5. Parr, J. B. Central place theory: an evaluation. *Rev. Urban Reg. Devel. Stud.* **11**, 214–256 (2017).
6. von Thünen, J. H. *Isolated state: an English edition of der isolierte staat* (1826) (Pergamon Press, Oxford, 1966).
7. Weber, A. *Theory of the location of industries* (The University of Chicago Press, Chicago, 1929).
8. Isard, W. *Location and space-economy: a general theory relating to industrial location, market areas, land use, trade, and urban structure* (MIT Press, New York, 1956).
9. Krugman, P. Increasing returns and economic geography. *J. Polit. Econ.* **99**, 483–499 (1991).
10. Fujita, M., Krugman, P. R. & Venables, A. J. *The spatial economy: cities, regions, and international trade* (MIT Press, New York, 2001).
11. Fujita, M. Thünen and the new economic geography. *Reg. Sci. Urban Econ.* **42**, 907–912 (2012).
12. Turing, A. M. The chemical basis of morphogenesis. *Philos. Trans. R. Soc. B* **237**, 37–72 (1952).
13. Nicolis, G. & Prigogine, I. *Self-organization in nonequilibrium systems: from dissipative structures to order through fluctuations* (Wiley, Hoboken, 1977).
14. Camazine, S. *et al.* *Self-organization in biological systems* (Princeton University Press, Princeton, 2003).
15. Allen, P. M. *Cities and regions as self-organizing systems: models of complexity* (Gordon and Breach Science Publishers, London, 1997).
16. Allen, T. & Arkolakis, C. Trade and the topography of the spatial economy. *Q. J. Econ.* **129**, 1085–1139 (2014).
17. Pumain, D. An evolutionary theory of urban systems. In *International and transnational perspectives on urban systems* (eds Rozenblat, C. *et al.*) 3–18 (Springer, New York, 2018).
18. Portugali, J. *Self-organization and the city* (Springer, New York, 2000).
19. White, R., Engelen, G. & Uljee, I. *Modeling cities and regions as complex systems: from theory to planning applications* (MIT Press, Cambridge, 2015).
20. Rodrigue, J.-P., Comtois, C. & Slack, B. *The geography of transport systems* 3rd edn. (Routledge, Milton Park, 2013).
21. Cronon, W. *Natures metropolis: Chicago and the great west* (W.W. Norton & Company, New York, 1992).
22. Fajgelbaum, P. D. & Schaal, E. Optimal transport networks in spatial equilibrium. *Econometrica* **88**, 1411–1452 (2020).
23. Tero, A. *et al.* Rules for biologically inspired adaptive network design. *Science* **327**, 439–442 (2010).
24. Ma, Q., Johansson, A., Tero, A., Nakagaki, T. & Sumpter, D. J. Current-reinforced random walks for constructing transport networks. *J. R. Soc. Interf.* **10**, 20120864 (2013).
25. Watanabe, S., Tero, A., Takamatsu, A. & Nakagaki, T. Traffic optimization in railroad networks using an algorithm mimicking an amoeba-like organism. *Physarum Plasmod. BioSyst.* **105**, 225–232 (2011).
26. Latty, T. *et al.* Structure and formation of ant transportation networks. *J. R. Soc. Interf.* **8**, 1298–1306 (2011).
27. Kunita, I. *et al.* Adaptive path-finding and transport network formation by the amoeba-like organism physarum. In *Natural computing and beyond* (eds Suzuki, Y. & Nakagaki, T.) 14–29 (Springer, New York, 2013).
28. Aoki, T. & Aoyagi, T. Scale-free structures emerging from co-evolution of a network and the distribution of a diffusive resource on it. *Phys. Rev. Lett.* **109**, 208702 (2012).
29. Aoki, T., Yawata, K. & Aoyagi, T. Self-organization of complex networks as a dynamical system. *Phys. Rev. E* **91**, 012908 (2015).
30. James, C. F. *Introduction to population modeling* (UMAP Monograph, 1979).
31. Zipf, G. K. The P1 P2/D hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* **11**, 677–686 (1946).
32. Ullman, E. L. The role of transportation and the bases for interaction. In *Mans role changing face earth* (ed. Thomas, W. L.) 862–880 (University of Chicago Press, Chicago, 1956).
33. Wilson, A. G. *Urban and regional models in geography and planning* (Wiley, London; New York, 1974).
34. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
35. Page, L., Brin, S., Motwani, R. & Winograd, T. The PageRank citation ranking: Bringing order to the web. Technical Report for 1999–1966, Stanford InfoLab (1999).
36. Cross, M. & Greenside, H. *Pattern formation and dynamics in nonequilibrium systems* (Cambridge University Press, Cambridge, 2009).
37. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).
38. Febvre, L., Bataillon, L., Mountford, E. G. & Paxton, J. H. *A geographical introduction to history* (K. Paul Trench, Trubner & Co., Ltd.; A.A. Knopf, New York, 1925).
39. Ratzel, F. *Anthropogeographie* 2nd edn. (J. Engelhorn, German, 1899).
40. Gabaix, X. Zipf's law and the growth of cities. *Am. Econ. Rev.* **89**, 129–132 (1999).
41. Decker, E. H., Kerkhoff, A. J. & Moses, M. E. Global patterns of city size distributions and their fundamental drivers. *PLoS ONE* **2**, 2–7 (2007).
42. Mansury, Y. & Gulyás, L. The emergence of Zipf's Law in a system of cities: an agent-based simulation approach. *J. Econ. Dyn. Control* **31**, 2438–2460 (2007).
43. Lee, S. & Li, Q. Uneven landscapes and city size distributions. *J. Urban Econ.* **78**, 19–29 (2013).
44. Verbavatz, V. & Barthélemy, M. The growth equation of cities. *Nature* **587**, 397–401 (2020).
45. Hanson, J. W. *An urban geography of the Roman world, 100 BC to AD 300* (Archaeopress, Oxford, 2016).
46. Hanson, J. W. & Ortman, S. G. A systematic method for estimating the populations of Greek and Roman settlements. *J. Roman Archaeol.* **30**, 299–322 (2017).
47. Malanima, P. Italian cities 1300–1800. A quantitative approach. *Rivista di Storia Economica* **14**, 91–125 (1998).
48. Desmet, K., Nagy, D. K. & Rossi-Hansberg, E. The geography of development. *J. Polit. Econ.* **126**, 903–983 (2018).
49. Delventhal, M. The globe as a network: geography and the origins of the world income distribution. In *Proceedings of the 2019 Meeting Papers No 840, Society for Economic Dynamics* (2019).
50. Solé-Ribalta, A., Gómez, S. & Arenas, A. A model to identify urban traffic congestion hotspots in complex networks. *R. Soc. Open Sci.* **3**, 160098 (2016).
51. Brinkman, J. C. Congestion, agglomeration, and the structure of cities. *J. Urban Econ.* **94**, 13–31 (2016).
52. Bretagnolle, A. & Pumain, D. Simulating urban networks through multiscale space-time dynamics: Europe and the united states, 17th–20th centuries. *Urban Stud.* **47**, 2819–2839 (2010).
53. Gross, T. & Blasius, B. Adaptive coevolutionary networks: a review. *J. R. Soc. Interf.* **5**, 259–271 (2008).
54. Barthélemy, M. & Flammini, A. Co-evolution of density and topology in a simple model of city formation. *Netw. Spat. Econ.* **9**, 401–425 (2009).
55. Raimbault, J. Modeling the co-evolution of cities and networks. <https://arxiv.org/abs/1804.09430> (2018).
56. Raimbault, J. Hierarchy and co-evolution processes in urban systems. <https://arxiv.org/abs/2001.11989> (2020).
57. Harris, B. & Wilson, A. G. Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environ. Plan A* **10**, 371–388 (1978).
58. Wilson, A. Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems. *J. R. Soc. Interf.* **5**, 865–871 (2008).
59. Ellam, L., Girolami, M., Pavliotis, G. A. & Wilson, A. Stochastic modelling of urban structure. *Proc. Roy. Soc. A* **474**, 20170700 (2018).

60. Wilkinson, J., Emms, T. & Evans, T. S. Dynamical analysis of spatial interaction models. <https://arxiv.org/abs/1911.10123> (2019).
61. USGS. Digital elevation: shuttle radar topography mission (srtm) 1 arc-second global. <https://doi.org/10.5066/F7PR7TFT> (2015).
62. Dallaire, C. O. & Lehner, B. Global river classification gloric version 1.0. <https://www.hydrosheds.org/page/gloric> (2018).
63. Dallaire, C. O., Lehner, B., Sayre, R. & Thieme, M. A multidisciplinary framework to derive global river reach classifications at high spatial resolution. *Environ. Res. Lett.* **14**, 024003 (2019).
64. Herzog, I. A review of case studies in archaeological least-cost analysis. *Archeol. e Calcolatori* **25**, 223–239 (2014).
65. Eurostat. Geostat 2011 grid dataset v2.0.1. <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography>.
66. The Italian National Institute of Statistics. Boundaries of administrative units for statistical purposes. <https://www.istat.it/it/archivio/222527> (2020).
67. Lahmeyer, J. Italy: historical demographical data of the administrative division. <http://www.populstat.info/Europe/italyp.htm> (2019).
68. Perthes, J. *Almanach de Gotha* (Gotha, Germany, 1838–1944).
69. Hanson, J. W. Cities database (oxrep databases), version 1.0. <https://doi.org/10.5287/bodleian:eqapevAn8> (2016).
70. Scheidel, W. Demography. In *The Cambridge economic history of the Greco-Roman World* (eds Scheidel, W. et al.) 38–86 (Cambridge University Press, Cambridge, 2007).
71. de Ligti, L. *Peasants, citizens and soldiers: studies in the demographic history of Roman Italy 225 BC–AD 100* (Cambridge University Press, Cambridge, 2012).
72. de Haas, T. The geography of Roman Italy and its implications for the development and integration of rural economies. In *The economic integration of Roman Italy* (eds de Haas, T. C. A. & Tol, G.) 51–82 (Brill, Leiden, 2017).
73. Malanima, P. Italian urban population 1300–1861 (database). <https://www.cnr.it/en/institutes-databases/database/324/italian-urban-population-1300-1861> (2005).
74. Geonames. Geonames. <https://www.geonames.org/> (2020).
75. Bosker, M., Brakman, S., Garretsen, H., de Jong, H. & Schramm, M. The development of cities in Italy: 1300–1861. *CESifo* **1893**, 1–43 (2007).
76. OpenStreetMap. Openstreetmap gps data. <https://planet.openstreetmap.org/gps/> (2020).

Acknowledgements

We thank S. Murayama and M. Yamada for discussions. This work was supported by the Research Institute for Mathematical Sciences, a joint research centre at Kyoto University (TA); the joint research with CSIS, the University of Tokyo (No. 699) (TA); JSPS KAKENHI Grant Number 21H04357 (TA); the Human Frontier Science Program (RGP0053/2012, MDF); the Leverhulme Trust (RPG-2015-437, MDF); JSPS Fellowship to MDF; “Dynamic Alliance for Open Innovation Bridging Human, Environment and Materials” from the Ministry of Education, Culture, Sports, Science and Technology of Japan (TN); JST, PRESTO Grant Number JPMJPR21RA, Japan (NF); JSPS KAKENHI Grant Number JP18K11462 (NF).

Author contributions

All authors designed the study and wrote the manuscript. TA implemented the model and performed the numerical analysis. NF performed the theoretical analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13758-1>.

Correspondence and requests for materials should be addressed to T.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022