Turing instability in an economic–demographic dynamical system may lead to pattern formation on a geographical scale

Anna Zincenko, Sergei Petrovskii, Vitaly Volpert, Malay Banerjee

INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Zincenko A, Petrovskii S, Volpert V, Banerjee M. 2021 Turing instability in an economic–demographic dynamical system may lead to pattern formation on a geographical scale. *J. R. Soc. Interface* **18**: 20210034.
https://doi.org/10.1098/rsif.2021.0034

Received: 12 January 2021
Accepted: 17 March 2021

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
biogeography, biomathematics, biometrics

Keywords:
population distribution, wealth distribution, population dynamics, long transients

Author for correspondence:
Sergei Petrovskii
e-mail: sp237@le.ac.uk

Turing instability in an economic–demographic dynamical system may lead to pattern formation on a geographical scale

Anna Zincenko[1,2], Sergei Petrovskii[1], Vitaly Volpert[3,4,5] and Malay Banerjee[6]

[1]School of Mathematics and Actuarial Science, University of Leicester, Leicester LE1 7RH, UK
[2]School of Mathematics, University of Birmingham, Birmingham B15 2TT, UK
[3]Institut Camille Jordan, UMR, 5208, CNRS, University Lyon 1, 69622 Villeurbanne, France
[4]INRIA Team Dracula, INRIA Lyon La Doua, 69603 Villeurbanne, France
[5]Peoples Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St, Moscow 117198, Russian Federation
[6]Department of Mathematics and Statistics, IIT Kanpur, Kanpur 208016, India

SP, 0000-0001-6259-2695

Spatial distribution of the human population is distinctly heterogeneous, e.g. showing significant difference in the population density between urban and rural areas. In the historical perspective, i.e. on the timescale of centuries, the emergence of densely populated areas at their present locations is widely believed to be linked to more favourable environmental and climatic conditions. In this paper, we challenge this point of view. We first identify a few areas at different parts of the world where the environmental conditions (quantified by the temperature, precipitation and elevation) show a relatively small variation in space on the scale of thousands of kilometres. We then examine the population distribution across those areas to show that, in spite of the approximate homogeneity of the environment, it exhibits a significant variation revealing a nearly periodic spatial pattern. Based on this apparent disagreement, we hypothesize that there may exist an inherent mechanism that may lead to pattern formation even in a uniform environment. We consider a mathematical model of the coupled demographic-economic dynamics and show that its spatially uniform, locally stable steady state can give rise to a periodic spatial pattern due to the Turing instability, the spatial scale of the emerging pattern being consistent with observations. Using numerical simulations, we show that, interestingly, the emergence of the Turing patterns may eventually lead to the system collapse.

1. Introduction

Fast growth of the global human population has long been regarded as a major challenge that faces mankind [1–4]. Presently, this challenge is becoming even more serious than before, in particular because many natural resources are estimated to deplete before the end of this century. The increasing population pressure on agriculture and ecosystems and the environment more generally is predicted to result in worldwide food and water shortages, pollution, lack of housing, poverty and social tension. The situation is exacerbated by global climate change as considerable areas of land are predicted to be flooded and hence taken out of human's use. It is widely believed that, unless alternative scenarios of sustainable population growth and social development are identified and implemented, mankind is likely to experience stagnation or even decline [5].

Population growth in time is complemented with the population dynamics in space. Population distribution over space is hugely heterogeneous for a variety of reasons, for example, climate, history and the economy, etc. The spatial heterogeneity may result in significant migration flows that in turn can have a significant

THE ROYAL SOCIETY PUBLISHING

feedback on the local demography and the population growth. On a smaller scale of individual countries and states, understanding the factors affecting the population distribution in space is needed to ensure adequate development of infrastructure, transport and energy network. Poorly informed decisions are likely to result in overcrowding and social problems in urban areas and/or lower quality of life in rural neighbourhoods.

Identification of scenarios of sustainable population growth and social development on various spatial and temporal scales requires good understanding of the relevant processes and mechanisms that affect both population growth and population distribution. Arguably, such understanding is unlikely to be achieved without a well-developed theory and the corresponding mathematical/modelling framework. Indeed, mathematical models of human population dynamics (e.g. [6,7]) have a long history dating back to the seventeenth century [8]. Over the last few decades, the need for an adequate and efficient mathematical theory of human population dynamics has been reflected by a steady growth in the number of studies where problems of demography along with related issues of economy were considered using mathematical models, tools and techniques, e.g. [9–12].

In this paper, we use mathematical modelling to address the phenomenon of heterogeneous spatial population distribution. Heterogeneity of geographical features (mountains, forests, rivers, etc.) and natural resources (e.g. coal, iron and copper ore) are commonly accepted as factors leading to the demographic and economic heterogeneity. Here we ask a question: is this natural heterogeneity the only underlying cause, or can there be another and perhaps more general principle responsible for emergence of heterogeneous population distribution?

In order to answer this question, we first revisit available data on population density over a few areas in different parts of the world to show that, in all cases considered, population distribution exhibits a clear nearly periodic spatial pattern in spite of the fact that the environmental conditions are relatively uniform. Being inspired by this finding, we then consider a novel model of coupled economic–demographic dynamics in space and time and endeavour to use it to simulate the spatial population distribution. The model consists of two coupled partial–differential equations of reaction–diffusion type. Following a similar modelling approach that was successfully used in ecology and biology [13–16], we then show that the emergence of spatial patterns in our model appears to be possible as a result of Turing instability. Although it is not our goal in this paper to provide any direct comparison between the real-world demographic patterns and the model properties, we regard the *qualitative* agreement between the model predictions and the data on the human population density as an indication that the heterogeneous population distribution observed across different countries in different continents may, at least in some cases, have been caused by endogenous rather than exogenous factors, i.e. may have appeared due to intrinsic Turing instability of the corresponding economic–demographic dynamical system.

2. Real-world examples

In many countries, the population distribution over space is distinctly heterogeneous, e.g. urbanized areas with a high population density alternate with rural areas with a low population density. Apparently, spatial variation in geographical and climatic factors can play a significant role in shaping the population distribution. Since our main hypothesis in this paper is the existence of a dynamical mechanism that may lead to the formation of heterogeneous population distribution regardless of the geographical heterogeneity, in our search for real-world examples we focus on the cases where the environment may be regarded, up to a certain spatial scale, as relatively uniform. The environmental properties that we consider here as proxies for environmental heterogeneity are elevation, annual mean temperature and annual mean precipitation. A brief overview of several relevant cases is given below.

2.1. Canadian southern region

Canada is a scarcely populated country and the majority of Canadian population live in the relatively narrow band along the USA border, except for Alberta and Saskatchewan where the densely populated areas expand somewhat further to the north (figure 1*a*). The distribution of the environmental properties across the country is highly heterogeneous, in particular in the south–north direction, ranging from temperate climate in the south to the rather extreme polar climate in the north. However, the magnitude of climatic variation in the east–west direction is significantly smaller (figure 1*c*), at least over the span between the Atlantic coast and the Rocky Mountains where the annual mean temperature normally varies within 6–8°C (contrary to more than 20°C in the south-north direction). The other environmental variables exhibit a significant change though (see figure 1*b*,*d*, respectively): the annual mean precipitation varies approximately between 250 and 1000 mm and the elevation varies between 50 and 150 m in the Canadian plains and 2000–3000 m in the Rockies.

We focus our analysis on the Canadian provinces that align along the USA–Canada border (creating a 'corridor' that is schematically shown by the black line in figure 1*a*) where the environmental conditions are relatively uniform compared to the north–south gradient (apart from the Rockies) but the population distribution is not. Figure 2*a* demonstrates how the population density varies in space along the border in the east-to-west direction. As the data on population density on a sub-provincial scale are not available in public domains, to represent the population density, we use data on the provincial scale. The circles in figure 2*a* show the average population density for the eight Canadian provinces located along the border, i.e. respectively, for Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan, Alberta and British Columbia. It is readily seen that the distribution exhibits three maxima with approximately equal spacing of 700 miles. We therefore regard it as a (nearly) periodic spatial distribution. Note that this pattern is persistent over time: a nearly periodic structure with the same properties is observed for different years (not shown here for the sake of brevity) starting from as early as late nineteenth century.

In order to reveal how strong the effect of environmental properties is on the population distribution, we perform the pairwise correlation analysis between the population density and each of the three environmental factors that we consider here. The results are shown in figure 2*b–d* and the corresponding values of the coefficient of determination R^2 [21] are given in table 1. We readily observe that the obtained values of

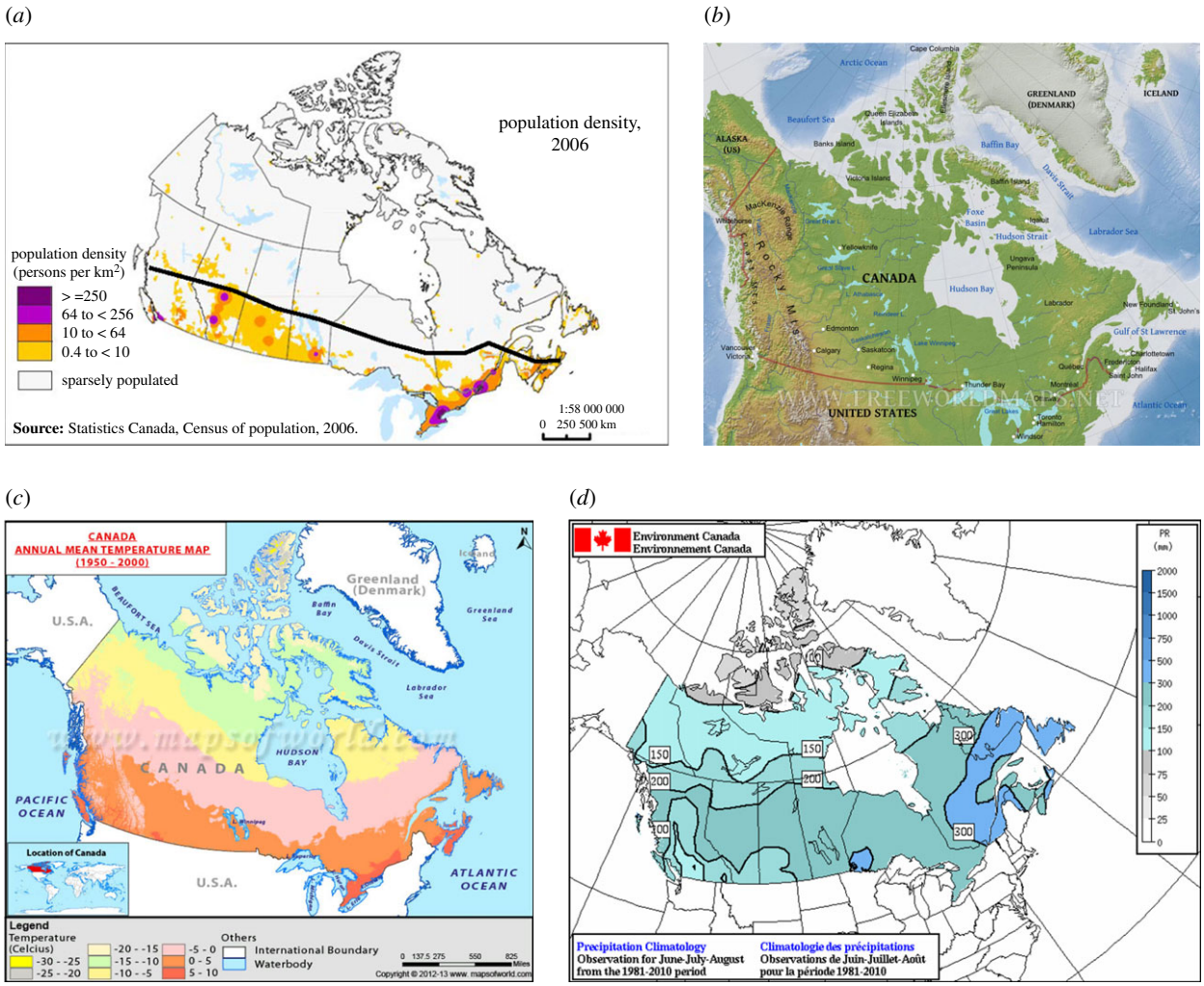

Figure 1. (*a*) Population distribution in Canada. Adapted from [17]. A strongly heterogeneous, 'patchy' structure is readily seen. The black line is drawn along the USA–Canada border where the environmental conditions are approximately homogeneous (except for the Rocky Mountains). (*b*) Geographical map of Canada showing the elevation [18]. One can see that, except for the Rocky Mountains, the elevation along the Canada–USA border is approximately uniform. (*c*) Canada annual mean temperature map, 1950–2000 [19]. It is readily observed that the temperature does not vary much along the southern border. (*d*) Precipitation map of Canada [20]. The amount of precipitation does not vary much along the border, except for the extreme west.

R^2 are quite small, hence only a small proportion of the variance in the population density can be explained by the environmental factors [21].

2.2. South-eastern Australia

As another relevant example of the heterogeneous population distribution in an approximately uniform environment, we consider south-eastern Australia. As well as Canada, Australia is a scarcely populated country, with most of the Australian population concentrated in three regions, i.e. southeast, east and southwest (figure 3*a*). The most densely populated area is the southeast. This area has the shape of a strip (approx. 400 km wide) along the coast of the Tasmanian sea and Bass strait. It appears that the climatic properties along this narrow strip such as precipitation and temperature are approximately uniform (see figure 3*b–d*), e.g. the variation in the annual mean temperature is just a few degrees (compared to more than 30°C over the continent as a whole). The stripe includes the Great Dividing Range and the Australian Alps, which therefore accounts for a significant variation in the elevation.

In spite of the relatively uniform environment (apart from the elevation, its effect being discussed below), the population distribution along the strip is strongly heterogeneous, with the population density varying more than two orders of magnitude between the more dense areas and the less dense ones. The origin of the data that we used is similar to the Canadian case: since the data on a small scale of Local Government Areas (Australian-specific small scale territorial unit roughly equivalent to a county) are not publically available, we used the data for larger regions (depicted by the lines in figure 4*a*, see also [26]). The results are shown in figure 4*a*. Interestingly, it exhibits a nearly periodic pattern where the three maxima are approximately equally spaced by about 700–850 km.

An immediate intuitive explanation of the heterogeneous population distribution can be sought in the heterogeneity of the environmental properties. Correspondingly, we look into the effect of the environmental factors more carefully by considering the correlation between each of the three factors chosen above and the population density. Figure 4*b*,*c* shows the scatterplots of the population density in Australia versus the mean annual temperature and the mean annual precipitation, respectively. In both cases, the straight line shows the best-fitting of the data to maximize R^2; the corresponding values of R^2 are shown in table 1. It is readily seen that in both cases R^2 is quite small. We therefore conclude that the climatic variation is unlikely to be the factor that defines the spatial distribution of the population.

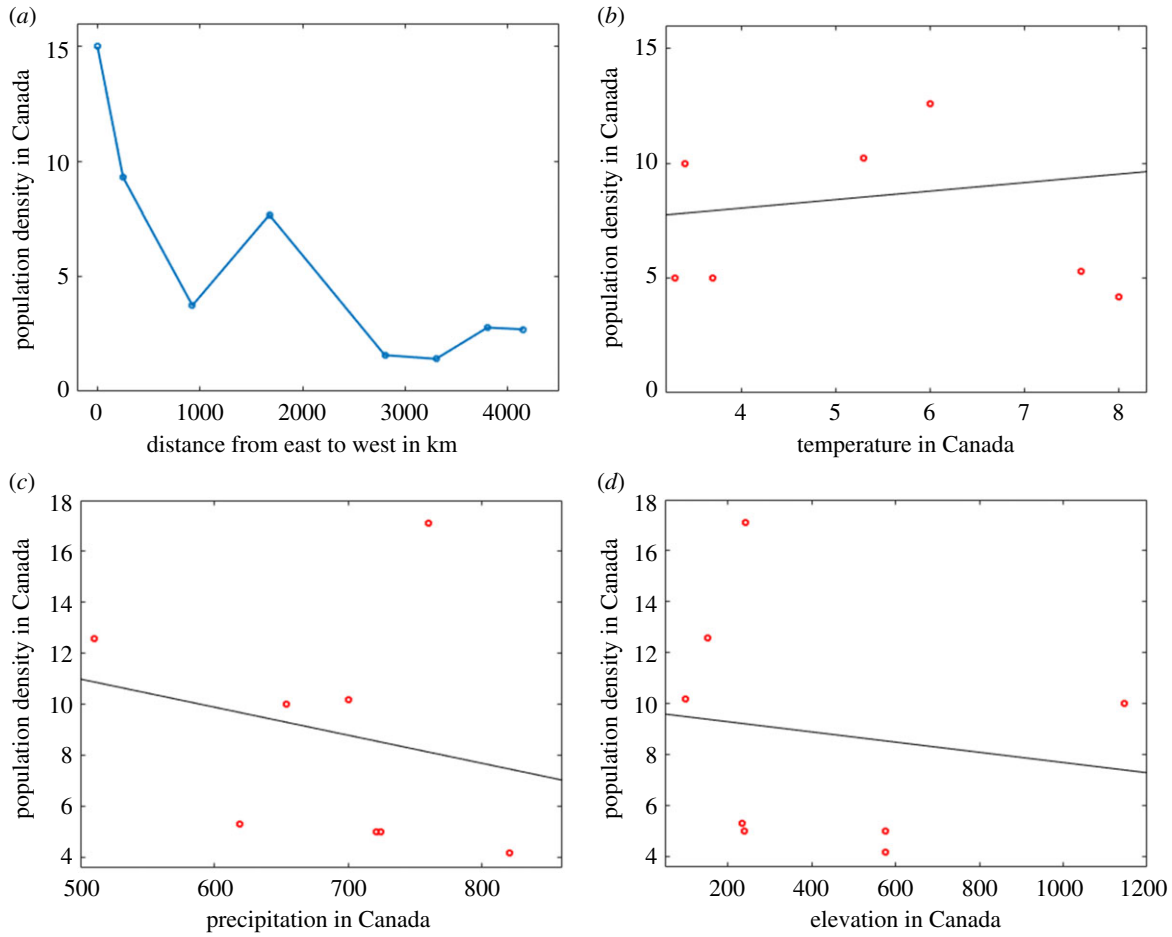

Figure 2. (*a*) Population density versus space along the USA–Canada border in 2010 in the east-to-west direction. The circles show the average population density for the eight Canadian provinces that align along the border. (*b*) Population density versus the annual mean temperature in Canada. (*c*) Population density versus the annual mean precipitation in Canada. (*d*) Population density versus the elevation. In all cases shown in panels (*b–d*), the best-fitting straight line is drawn by maximizing R^2; for details, see table 1.

Now we recall that the study area includes the mountain ranges and exhibits considerable variation in the elevation. The question hence arises as to whether that can be a relevant factor. However, we first notice that the vast majority of the Australian population lives at the elevation below 250 m (figure 4*d*). We then perform the correlation analysis by looking for the best-fitting straight line in the scatterplot of the population density versus the elevation. The corresponding value of R^2 (table 1) appears to be very small. We therefore rule out the elevation as a factor affecting the heterogeneous spatial population distribution along the southeast coast of Australia.

Table 1. Values of the coefficient of determination R^2 [21] for the best-fitting linear function between the population density in the three considered countries and the three environmental properties.

	elevation, m	annual mean temperature, °C	annual mean precipitation, mm
Canada	−0.14	−0.15	−0.09
Australia	−0.14	−0.04	−0.06
Mongolia	−0.08	−0.08	−0.04

2.3. Mongolian grassland

Mongolia, a Central-Asian country situated between China in the south and Russia in the north, has an elongated territory that extends from east to west for about 2400 km. It is the most sparsely populated country in the world. South of Mongolia is occupied by the Gobi Desert, which is barely populated at all due to the harsh climate and lack of resources. The majority of three million population of Mongolia live in grasslands, which are located in the north of the country. In order to reveal the features of the spatial population distribution as is needed in the context of this study, we focus on the densely populated 'corridor' located along the latitude at 47.7° north; see the black line in figure 5*a*. Similarly to the two cases considered above, for the population density we used the data on the provincial (aimag) scale, as smaller scale data are not available. Correspondingly, the circles in figure 6*a* show the average population density for the eight aimags located along the line. Interestingly, we readily observe that, as well as in the two previous cases, the population distribution in the east–west direction exhibits a periodic-like pattern (figure 6*a*). The three distinct peaks are separated by 700 and 900 km intervals.

Variation of the environmental properties (cf. figure 5*b–d*) along the latitude is considerably less than in the north–south direction. However, it appears to be larger than it is in the cases of Canada and Australia, e.g. the annual mean temperature varies over about 10°C and the annual mean

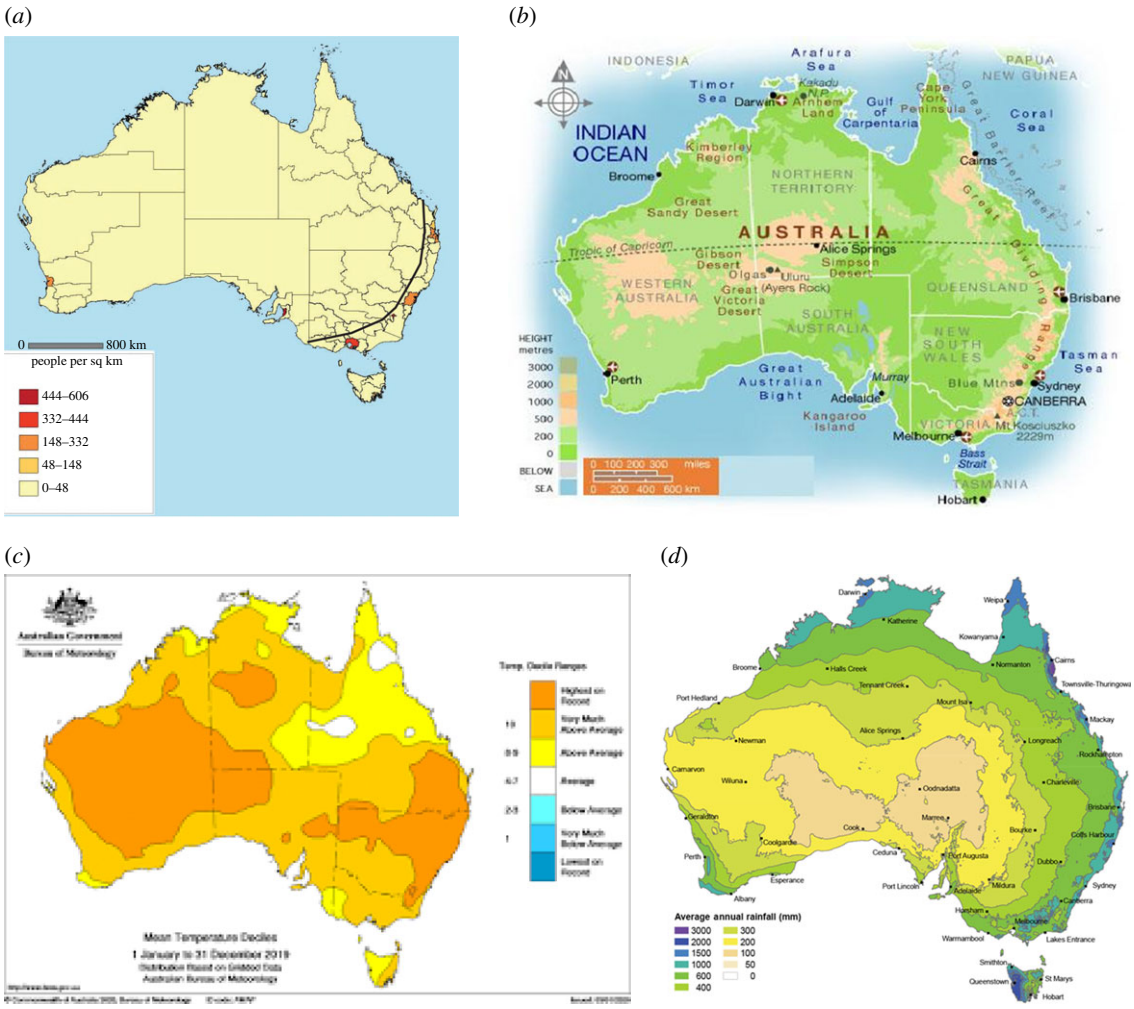

Figure 3. (*a*) Population distribution in Australia in 2011. It can be seen that the population is concentrated in the southeast; the thick black line indicates the location of the stripe where the majority of the population lives. Adapted from [22]. (*b*) Geographical map of Australia showing the elevation. Adapted from [23]. (*c*) Annual mean temperature map based on 30 years observations, 1961–1990. Adapted from [24]. (*d*) Annual precipitation map based on 30 years observations, 1961–1990. Adapted from [25].

precipitation from 50 to 350 mm m^{-2} year^{-1}). Also the elevation varies over about 1500 m, which is somewhat less than in Australia but larger than in Canada (where our analysis did not include the Rocky Mountains).

In order to reveal whether the variation of the environmental properties has any significant effect on the distribution of the population, we now perform the pairwise correlation analysis. The scatterplots of the population density versus the mean annual temperature, mean precipitation and the elevation are shown in figure 6*b*,*c* and *d*, respectively. The straight line is the best-fitting linear function; the corresponding values of R^2 are given in table 1. Apparently, the correlation between the population distribution and the environmental factors is very weak. We therefore conclude that the nearly periodic pattern clearly seen in the population distribution is unlikely to be caused by the environmental conditions.

2.4. Further statistical analysis

2.4.1. Multiple linear regression

In the above, we have shown that the spatial distribution of the population is unlikely to be affected, not to any considerable extent, by any single environmental property such as the mean annual temperature, mean precipitation or the elevation. However, generally speaking this does not rule out the possibility that a certain combination of those three factors may have a much stronger effect. In order to check this possibility, we applied the multiple regression:

$$y = a_0 + \sum_{i=1}^{3} a_i x_i, \tag{2.1}$$

where y is the population density, x_1, x_2 and x_3 are, respectively, the average annual temperature, the annual precipitation and the elevation. Model (2.1) was applied separately to the data for each of the three countries. The results are shown in table 2. We readily observe that the combined effect of the three environmental factors does not result in a stronger correlation with the population density. Small values of R^2 along with large p-values indicate that variation of population density is only to a small amount explained by the geographical and climatic properties.

Thus, we have examined three areas in three different countries chosen from three different continents to reveal that, in all three cases, the population distribution over an area with relatively uniform environmental conditions exhibits a clear spatial periodicity. Having considered the correlation between the population density and the main environmental properties, we have shown that the correlation is very weak and hence the nearly periodic pattern is unlikely caused by the effect of the environmental factors. Note that the

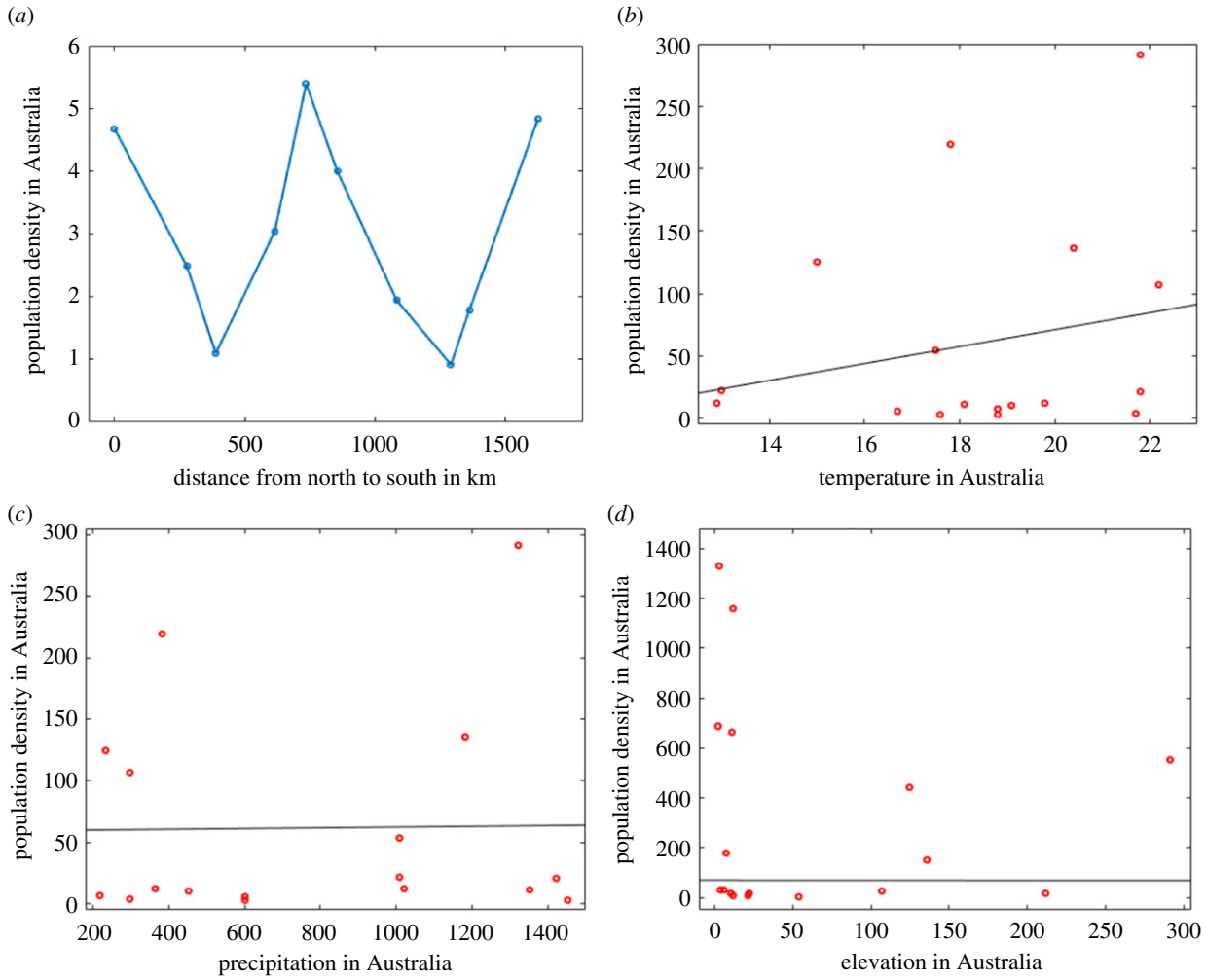

Figure 4. (*a*) Population density (on logarithmic scale) versus space inside the approximately 400 km wide stripe along the coast of Tasmanian Sea and Bass Strait in the southeast Australia. The circles show the average population density in regions along the coast. The horizontal axis shows the distance in kilometres from Brisbane along the coastline. (*b*) Population density versus the annual mean temperature in Australia. (*c*) Population density versus the annual mean precipitation. (*d*) Population density versus the elevation. In all cases shown in panels (*b–d*), the best-fitting straight line is drawn by maximizing R^2; for details, see table 1.

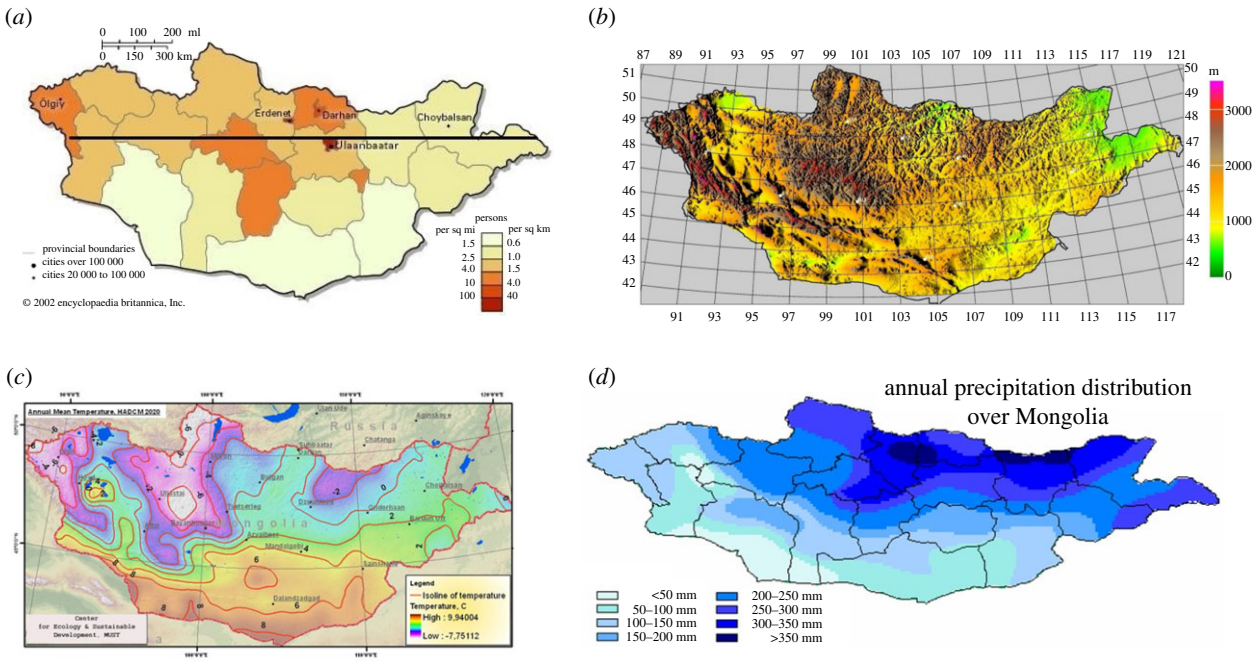

Figure 5. (*a*) Spatial population distribution in Mongolia. The horizontal line shows the transect (along the 37.7° latitude) across the most densely populated areas. Adapted from [27]. (*b*) Geographical map of Mongolia showing the elevation. Adapted from [28]. (*c*) Mean annual temperature. Adapted from [29]. (*d*) Mean annual precipitation in Mongolia for the period 1961–1990. Adapted from [29].

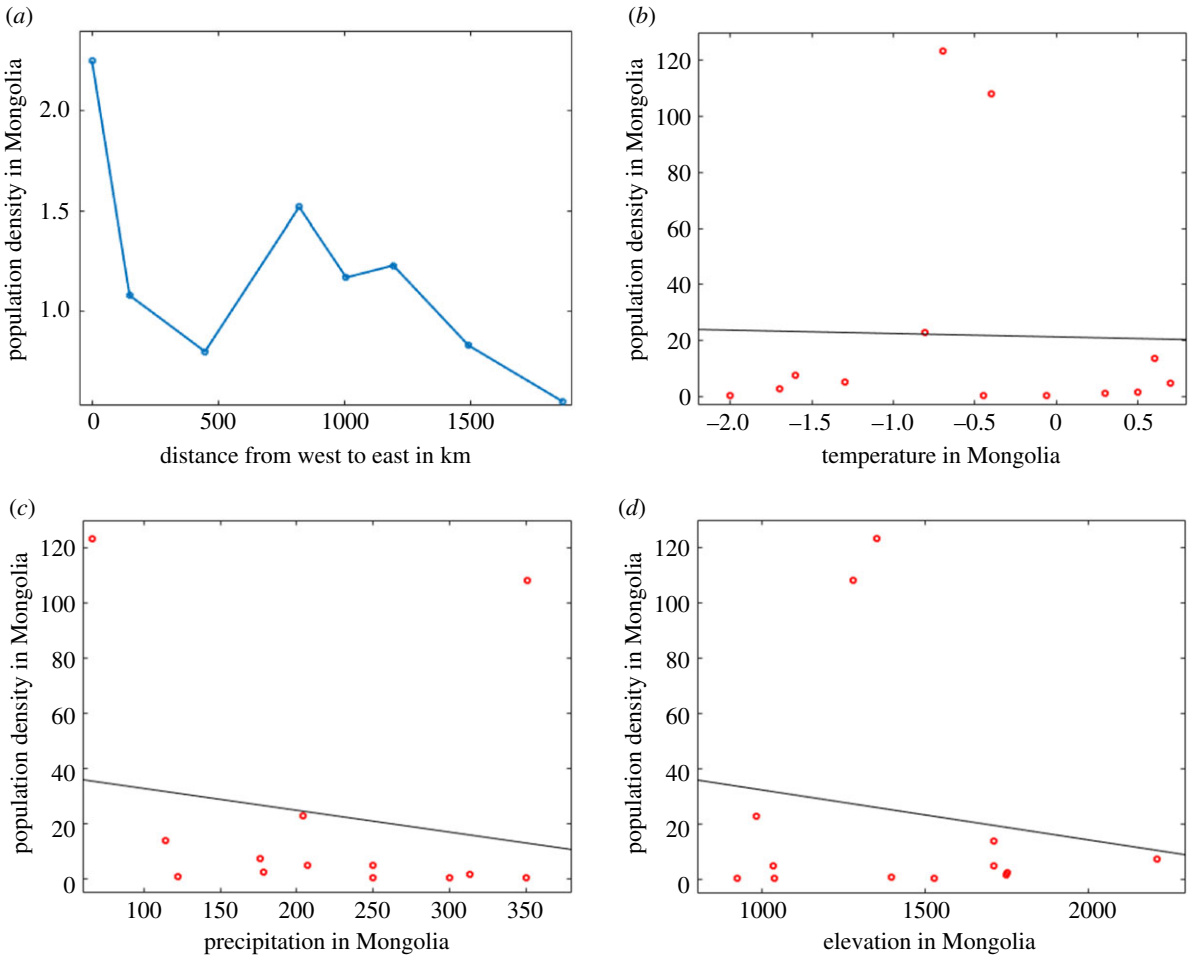

Figure 6. (*a*) Population density in Mongolia across the central areas of the country. (*b*) Population density versus the mean annual temperature in Mongolia. (*c*) Population density versus the mean annual precipitation in Mongolia. (*d*) Population density versus the elevation. In all cases shown in panels (*b–d*), the best-fitting straight line is drawn by maximizing R^2; for details, see table 1.

Table 2. Parameters of the linear model (2.1) and the *p*-values corresponding to the best fitting of the data obtained by maximizing R^2.

country	a_0	a_1	a_2	a_3	R^2	p
Canada	14.725	0.547	−0.013	−0.001	0.125	0.90
Australia	−74.207	7.656	0.003	−0.025	0.074	0.79
Mongolia	79.273	−5.273	−1.02	−0.025	0.070	0.71

three considered countries are vastly different in terms of their average climate, history and culture. This leads us to assume that there can be a generic mechanism resulting in the emergence of the observed spatial pattern. We further assume that this is a dynamical mechanism originated in the nonlinear interaction between the human demography and the distribution of resources or wealth. The corresponding mathematical model is considered in the next section.

2.4.2. Nonlinear regression

We now notice that the distribution of data on the population density against some of the selected environmental variables are peaked rather than distributed uniformly, e.g. see the scatter plot in figure 6*b*. This suggests that the linear regression may not be entirely appropriate for describing the data and a nonlinear regression should be used instead. The results of data fitting with a nonlinear function are shown in figure 7. For the data on Canada (figure 7*a*) and Mongolia (figure 7*c*), we use the Gaussian distributions (where the parameters of the distribution are obtained by maximizing the accuracy of the fit). The resulting negative values of R^2 indicates that the use of nonlinear regression is not justified as the mean of the data provides a better fit to the outcomes than does the fitted function values [30,31]. Thus, the apparently peaked correlation of population and temperature is not confirmed. Similar results are obtained for the correlation with the elevation and precipitation; we do not show the details for the sake of brevity.

For the data on Australia, the Gaussian fit appears to be effectively invalid, because of very large confidence intervals for the parameters. Therefore, the data are fitted with a third degree polynomial (figure 7*b*). Similarly to the above, it results in negative R^2, this indicating that the nonlinear regression in fact provides a worse description of the data than the linear one.

We mention here that using a nonlinear regression instead of the linear one brings an additional challenge as the choice of the fitting function becomes arbitrary. Correspondingly, our choice of the function(s) used to describe the data might not be the 'best one'. However, in the situation where the use of the nonlinear function instead of the linear one makes the quality of data fitting much worse, as it happens in our case, clearly indicates that the nonlinearity in the dependence of the population density on the environmental variables is not an important factor, at least in the considered range. Along with

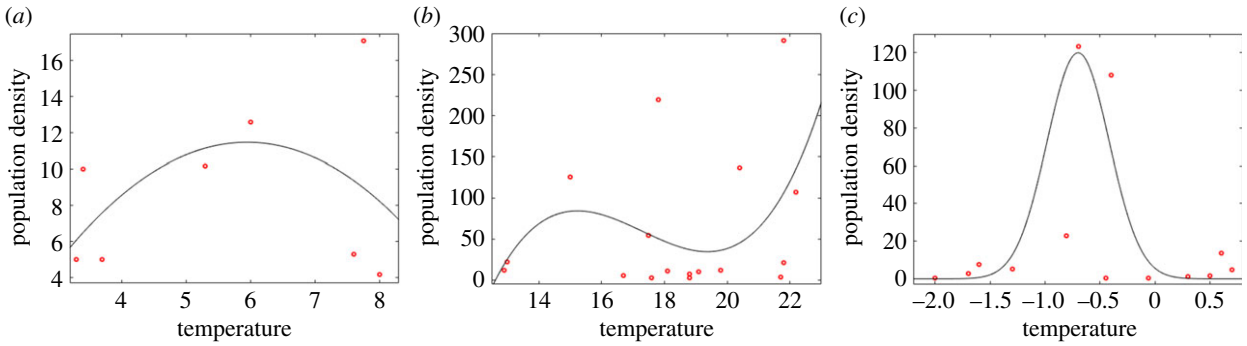

Figure 7. Population density versus temperature analysed by using the nonlinear regression for (*a*) Canada, (*b*) Australia and (*c*) Mongolia. The corresponding values of R^2 are −8.78, −1.72 and −2.7. The negative values of R^2 indicate that the nonlinearity in the dependence of the population density on the temperature is not important. See details in the text.

the results of the linear regression above, it indicates that the environment (quantified by the three selected variables) is not the factor that determines the spatial distribution of the population.

3. Mathematical model

In order to describe the dynamics of the human population, we use the simple, 'conceptual' economic–demographic model earlier developed in [32]. The model quantifies the state of the human society at a given location in space x at a given time t by two state variables, the population density $p(x, t)$ and the concentration of wealth $u(x, t)$. Note that, while due to its meaning $p \geq 0$, variable u must not necessarily be non-negative; negative values of wealth can be regarded as debt. In the baseline one-dimensional case (which is relevant in case of the population distribution in a narrow stripe, cf. the examples in the previous section), the model consists of two partial differential equations of reaction–diffusion type:

$$\frac{\partial u}{\partial t} = D_u \frac{\partial^2 u}{\partial x^2} + F(u, p) \tag{3.1}$$

and

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} + G(u, p), \tag{3.2}$$

(where we neglect, for the sake of simplicity, possible effects of cross-diffusion [32]). Here the first term in the right-hand side of equation (3.2) accounts for the population movement in space, that we assume can be considered, at least over certain spatial and temporal scales, as random [33–35] (for a detailed discussion of the 'bugbear of randomness' see [34]) and can be described mathematically as standard Fickian diffusion. The diffusion term in equation (3.1) describes local wealth redistribution due to the economic activities such as trade and investments, and/or taxes.

For the reaction term in equation (3.2), we consider the logistic population growth:

$$G(u, p) = [K(u) - \sigma(u)]p - \alpha p^2, \tag{3.3}$$

where the *per capita* growth rate (the expression in square brackets) is the difference between the birth rate K and the mortality rate σ, both of them being wealth-dependent, and α is a coefficient quantifying the competition (cf. [16,36]).

To parametrize $K(u)$, it is reasonable to assume that in the absence of any resources the birth rate is zero. Since the availability of resources (understood here in a broad sense) in our model is described by wealth, we therefore assume that $K(0) = 0$. For u being positive but not large, $K(u)$ is an increasing function. However, when the resource (wealth) becomes plentiful, $K(u)$ seizes to be monotonic. There is a certain difference—a 'cultural shift'—between the low-income and high-income society groups with respect to their wealth-dependent behaviour, implying that wealthier people tend to have less children [37]. Correspondingly, we consider the birth rate $K(u)$ to be an increasing function of wealth for small u but decreasing function for large u, tending to a small value (ultimately, to zero) as u tends to infinity. More specifically, we consider it in the following form:

$$K(u) = \frac{a_2 u}{u^2 + c_2^2}, \tag{3.4}$$

where a_2 and c_2 are positive parameters. Note that in our model we do not require $u \geq 0$ as a necessary sensible conditions; negative values of u are regarded as debt. For $u < 0$, the growth rate becomes negative as well, which seems to be consistent with the generic effect of very harsh conditions of the corresponding environment.

In order to parametrize the mortality rate $\sigma(u)$, we consider it to be a monotonously decreasing function of wealth. It takes into account the general observation that, on average, the mortality rate is lower for rich people, e.g. due to access to better health services and/or healthier lifestyle [38]. In particular, there is evidence that in the USA wealthier people tend to live longer [39]. More specifically, we consider the following generic Monod-type parametrization:

$$\sigma(u) = \sigma_0 - \frac{\sigma_1 u}{c_0 + u}, \tag{3.5}$$

where c_0, σ_0 and σ_1 are positive parameters, $\sigma_0 > \sigma_1$.

We mention here that, in the parametrization (3.5), the ratio σ_1/σ_0 can be related to the degree of inequality in the society: $\sigma_1/\sigma_0 \approx 1$ means that the basic health service is poor and the access to extra, potentially expensive services makes a big difference, while $\sigma_1/\sigma_0 \ll 1$ means that a good quality health service is widely accessible with no extra costs. In many contemporary societies, the health service is accessible to everyone and its quality is usually good [40–42]. With that argument in mind, in the simulations below we will consider $\sigma_1 \ll \sigma_0$. Note that, in this case, $\sigma(u) \approx \sigma_0$.

In order to specify the reaction term in equation (3.1), we first write it as follows:

$$F(u, p) = W(u, p) - S(u, p), \tag{3.6}$$

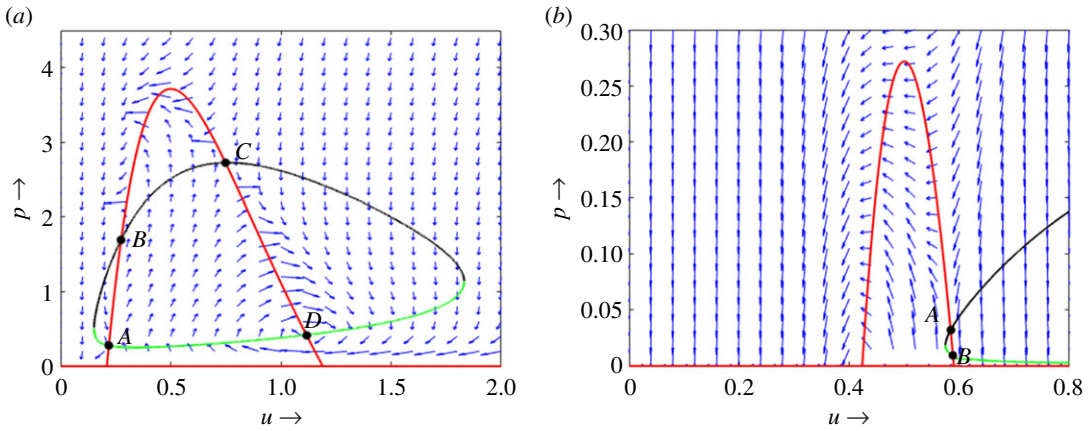

Figure 8. Isoclines of the dynamical system (4.1) shown for different parameter values, red colour for the *G*-isocline, black-and-green colour for the *F*-isocline. (*a*) The relative position of the isoclines allows for the existence of four positive steady states. Parameters are $c_0 = c_1 = c_2 = 1$, $a = 15$, $r = 7$, $s = 7.5$, $a_2 = 182$, $c_3 = 0.5$, $\sigma_1 = 0.066$, $\sigma_0 = 130$, $\alpha = 14$ and $a_1 = 150$. (*b*) Relative position of the isoclines allowing for the existence of only two positive steady states. Here *A* is a saddle point and *B* is a stable focus. Parameters are $c_1 = 5$, $c_0 = c_2 = 1$, $a = 0.01$, $r = 7.5$, $s = 16$, $a_2 = 192.6$, $c_3 = 0.5$, $\sigma_1 = 0.05$, $\sigma_0 = 190$, $\alpha = 9.63$ and $a_1 = 168$. The origin is always a stable node. Arrows show the direction of the phase flow.

where W and S are the rates of the wealth production and consumption, respectively. Production of wealth is often described by the Cobb–Douglas production function which in the simplest case can be written as [43,44]

$$W = bH^{\nu}Q^{\beta}M^{\gamma}, \tag{3.7}$$

where H is the labour, Q is the capital and M is the available natural resource, a positive coefficient b is a measure of technology, ν, β and γ are positive constants [43,44].

We assume that the natural resource is not a limiting factor, so that M can be kept as constant. This may look somewhat disputable, especially if the model predictions are projected to the future, as the natural resources are eventually becoming exhausted and this has been identified as a problem that already affects some modern economies [45]. However, since the purpose of our study is to contribute to the understanding of historical data, i.e. the spatial population distribution up to date, and not to predict the future, arguably the assumption of constant M remains relevant.

We further assume that capital Q is a function of wealth, $Q = f(u)$, and labour is a function of the population density, $H = g(p)$. Equation (3.7) then takes the following form:

$$W(u, p) = f(u)g(p). \tag{3.8}$$

Due to their meaning, it is reasonable to assume that $f(u)$ and $g(p)$ are increasing functions with saturation. Correspondingly, we choose them in the generic form as the Monod function:

$$f(u) = \frac{a_1 u}{u + c_1}, \quad g(p) = \frac{p}{p + c_2}, \tag{3.9}$$

where a_1, c_1 and c_2 are positive parameters.

For the wealth consumption S, we assume it to be the result of two processes, i.e. due to the depreciation (in particular in case of buildings, machinery, etc.) and consumption of goods and products by the people. For depreciation, we assume it to be a linear process with a constant rate a. The rate of the individual (*per capita*) consumption, say c, can be described by the Keynes linear consumption function, $c = r + sy$, where y is the *per capita* income and r and s are positive coefficients. Assuming additionally that average income is proportional to the wealth, we arrive at the following expression:

$$S(u, p) = au + (r + su)p. \tag{3.10}$$

From (3.3)–(3.10), we thus obtain the following expressions for the reaction terms:

$$F(u, p) = \frac{a_1 up}{(u + c_1)(p + c_2)} - [au + (r + su)p] \tag{3.11}$$

and

$$G(u, p) = \left[\frac{a_2 u}{u^2 + c_2^2} - \left(\sigma_0 - \frac{\sigma_1 u}{c_0 + u}\right)\right]p - \alpha p^2. \tag{3.12}$$

4. A glance at the non-spatial system

We begin with a brief look at the properties of the non-spatial counterpart of the reaction–diffusion system (3.1) and (3.2), which is given by the following equations:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = F(u, p) \quad \text{and} \quad \frac{\mathrm{d}p}{\mathrm{d}t} = G(u, p), \tag{4.1}$$

where functions G and F are given by equations (3.11) and (3.12). System (4.1) was studied in some detail in [46]. Here we only briefly revisit some of its properties, to the extent that is needed for the goals of this paper.

The phase plane of system (4.1) is shown in figure 8. It is readily seen that the origin is a steady state; a closer look reveals that it is a stable node. Inside the first quarter of the phase plane, i.e. for $u > 0$ and $p > 0$, the *F*-isocline is a convex closed curve (loop) and the *G*-isocline is an upward-convex, dome-shaped curve. Depending on the relative position of the isoclines (and hence on the parameter values, see [46] for details), the number of positive steady states can be anywhere from 0 to 4. Therefore, in a general case system (4.1) can exhibit a rich, multi-stable dynamics and a complicated bifurcation structure where positive states can emerge or disappear. A typical case corresponding to four positive steady states is shown in figure 8*a*.

A case where the relative position of the isoclines allows for only two positive steady states is shown in figure 8*b*. For these parameters, A is a saddle point and B is a stable focus. Interestingly, a closer look reveals that even in this case the phase plane has a complicated structure; figure 9 shows a magnified part of the phase plane containing steady states A and B only. There are two attractors: the

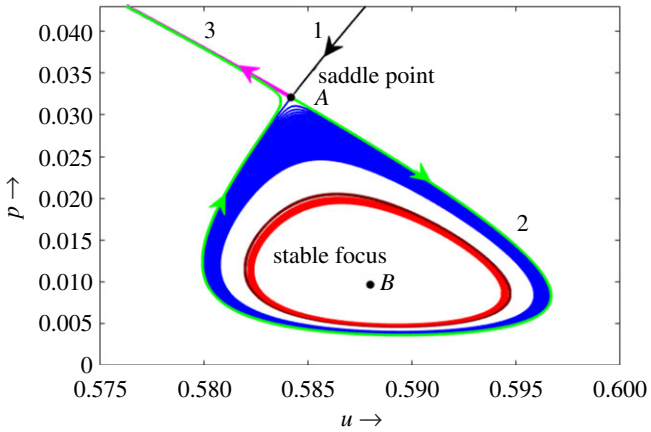

Figure 9. Fine structure of the phase plane near saddle point *A* and stable focus *B* (a magnified view of the corresponding part of the phase plane shown in figure 8). Black curve 1 and the blue curve show the stable manifolds of the saddle. Green curve 2 and magenta line 3 show the unstable manifolds. The brown curve shows the unstable limit cycle. Note that the green curve leaves the vicinity of steady states *A* and *B* to eventually go to the stable node (0, 0) which is the attractor for the rest of the phase plane (except for the part of the plane inside the limit cycle). The arrows show the direction of the phase flow. The red curve shows a trajectory that starts close to the limit cycle and eventually converges to the stable focus.

stable node (0, 0) and the stable focus *B*, so that the system is bistable. The attraction basin of stable focus *B* is bounded by an unstable limit cycle (shown by brown colour). Trajectories that start close to the limit cycle from inside will approach the stable focus over time; an example is shown by the red curve. We mention here that the eigenvalues of the system linearized in the vicinity of stable focus *B* have a very small real part (for the parameters of figure 9, $\lambda_{1,2} = -0.0000895 \pm 0.460312i$) so that the trajectory approaches the steady state at a very low rate. Trajectories that start outside of the limit cycle eventually approach the stable node (0, 0) except for the separatrix (the blue curve), which is part of the stable manifold of saddle point *A*; an example is shown in figure 10.

5. Turing instability conditions

We now consider the properties of the spatially explicit system (3.1)–(3.2) with the reaction terms given by (3.11)–(3.12):

$$\frac{\partial u}{\partial t} = D_u \frac{\partial^2 u}{\partial x^2} + \frac{a_1 up}{(u + c_1)(p + c_2)} - [au + (r + su)p] \tag{5.1}$$

and

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} + \left[\frac{a_2 u}{u^2 + c_2^2} - \left(\sigma_0 - \frac{\sigma_1 u}{c_0 + u}\right)\right] p - \alpha p^2. \tag{5.2}$$

Equations (5.1) and (5.2) are complemented with the Neumann 'zero-flux' boundary conditions:

$$\frac{\partial u}{\partial x}(0, t) = 0, \quad \frac{\partial u}{\partial x}(L, t) = 0, \quad \frac{\partial p}{\partial x}(0, t) = 0, \quad \frac{\partial p}{\partial x}(L, t) = 0, \tag{5.3}$$

and by initial conditions that we will discuss below.

Since our study is motivated by the existence of periodic spatial patterns, see §2, we are particularly interested in the possibility of Turing instability and the corresponding pattern formation. Turing instability is the property of a nonlinear reaction–diffusion system where a steady state that is stable in the corresponding non-spatial system can, under certain parameter constraints, become unstable in the spatial system with respect to a heterogeneous perturbation containing a certain wavelength [47].

Let $(\bar{u}, \bar{p})$ is a steady state of the non-spatial system and J is the Jacobian matrix evaluated at this steady state:

$$J = \begin{pmatrix} F_u & F_p \\ G_u & G_p \end{pmatrix}, \tag{5.4}$$

where the subscript denotes the corresponding partial derivative, for instance $F_u = \partial F(u, p)/\partial u$. We require that the steady state is stable, so that the following conditions hold:

$$\text{(a)} \quad \text{tr}(J) < 0, \quad \text{and} \quad \text{(b)} \quad \det(J) > 0, \tag{5.5}$$

e.g. see [36]. In the spatial system, for stability of the corresponding uniform steady state $p(x, t) \equiv \bar{p}$ and $u(x, t) \equiv \bar{u}$ with respect to a periodic perturbation with the wavenumber k, conditions (5.5) change to

$$\text{(a)} \quad \text{tr}(J_k) < 0, \quad \text{and} \quad \text{(b)} \quad \det(J_k) > 0, \tag{5.6}$$

where

$$J_k = J - k^2 \begin{pmatrix} D_u & 0 \\ 0 & D_p \end{pmatrix}. \tag{5.7}$$

The Turing instability occurs if one of the conditions (5.6) is broken. It is readily seen that condition (5.6*a*) holds for any k. Therefore, the instability can only occur if there is a range of values of k that satisfy the following inequality [36,48]:

$$\min(R(k^2)) < 0, \tag{5.8}$$

where the characteristic function $R(k^2) = \det(J_k)$. Taking into account (5.7), $R(z)$ appears to be a quadratic polynomial, so that inequality (5.8) is equivalent to [36,47]:

$$DF_u + G_p > 2\sqrt{D \det(J)}, \tag{5.9}$$

where parameter $D = D_p/D_u$ is the ratio of the diffusion coefficients. In its turn, it appears that a necessary condition for (5.9) is that F_u and G_p must be of a different sign. Consider $F_u > 0$ and $G_p < 0$; in this case, u is called the 'activator' and p the 'inhibitor' [36]. Then another necessary condition for (5.9) is $D > D_{\text{cr}} > 1$ where D_{cr} is a certain critical value that depends on the parameters in the reaction terms [36,47].

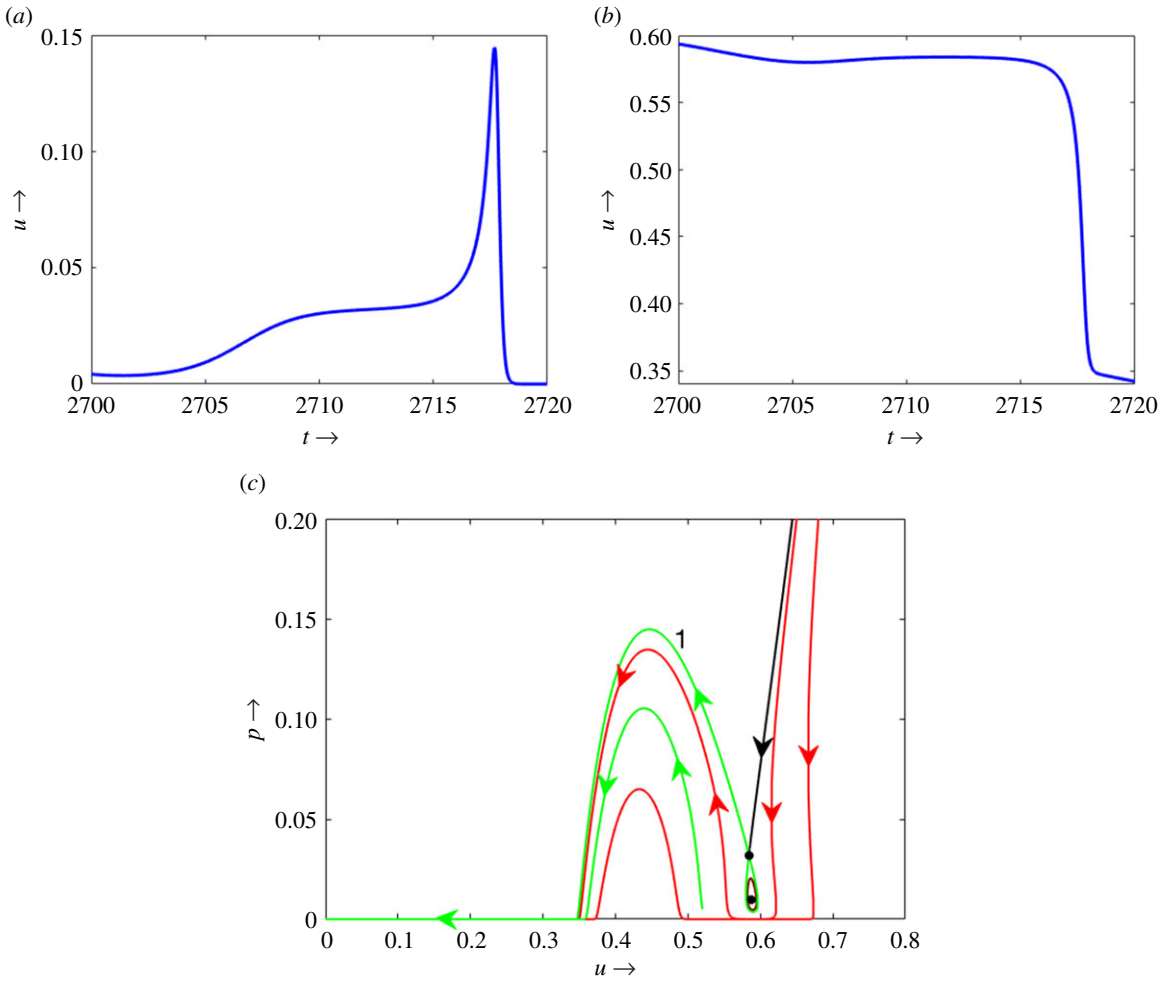

Figure 10. (*a*,*b*) Population and wealth, respectively, versus time in case the initial conditions correspond to a point in the phase plane outside of the unstable limit cycle (just below the unstable manifold of the saddle point, see curve 2 in figure 9). (*c*) Typical system's trajectories passing through the vicinity of steady states *A* and *B*, arrows show the direction of the flow. The solution shown in (*a*,*b*) corresponds to green curve 1.

Now we consider how the generic relation (5.9) between the system's feedbacks works in the case of our model (5.1)–(5.2). Given the complexity of the bifurcation structure of non-spatial system (3.12)–(4.1), see §4 and [46] for more details, a comprehensive study addressing the Turing instability of all (stable) steady states over the entire parameter range does not seem possible. We therefore concentrate on the specific yet instructive case where there are two positive steady states, a saddle and a stable focus (cf. figures 8*b* and 9), in particular to investigate whether the Turing instability may occur for stable steady state $B = (\bar{u}, \bar{p})$.

As realistic parameters are not presently available, we use some hypothetical values instead. To partly compensate for the lack of realistic values, below we will consider how the pattern formation may be affected by a variation in (some of) the parameters values. As a starting point, we consider the following set: $a_1 = 170$, $c_1 = 5$, $c_0 = c_2 = 1$, $a = 0.01$, $r = 7.5$, $s = 7.5$, $a_2 = 200$, $c_3 = 0.5$, $\sigma_1 = 0.05$, $\sigma_0 = 190$, $\alpha = 10$; the corresponding steady-state values are $\bar{u} = 0.69$ and $\bar{p} = 0.00087$. For these parameters, the critical value of the diffusivity ratio is readily obtained as $D_{cr} = 59746$. Figure 11 shows the function $R(z)$ for a subcritical case $D < D_{cr}$ where the steady state is stable (as $R(z) > 0$ for any z and condition (5.6*b*) holds for any k) and a supercritical cases $D > D_{cr}$ where the steady state is unstable with respect to perturbation with the wavelength from the interval where $R(z) < 0$ and hence condition (5.6*b*) is broken.

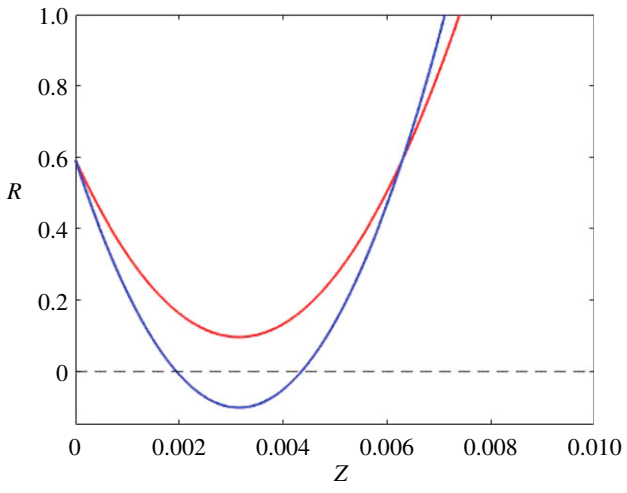

Figure 11. Examples of the characteristic function $R(z)$ for $D = 50\,000$ (blue curve) and $D = 70\,000$ (orange curve). Other parameters are given in the text; the critical value is $D_{cr} = 59\,746$.

For the above parameter set, the critical ratio of the diffusion coefficients is very large, which may rise doubts whether it is at all realistic in terms of the real-world dynamics. Therefore, now we are going to consider how the critical relation responds to changes in the parameter values and whether it can be diminished. Indeed, it appears that D_{cr} is rather

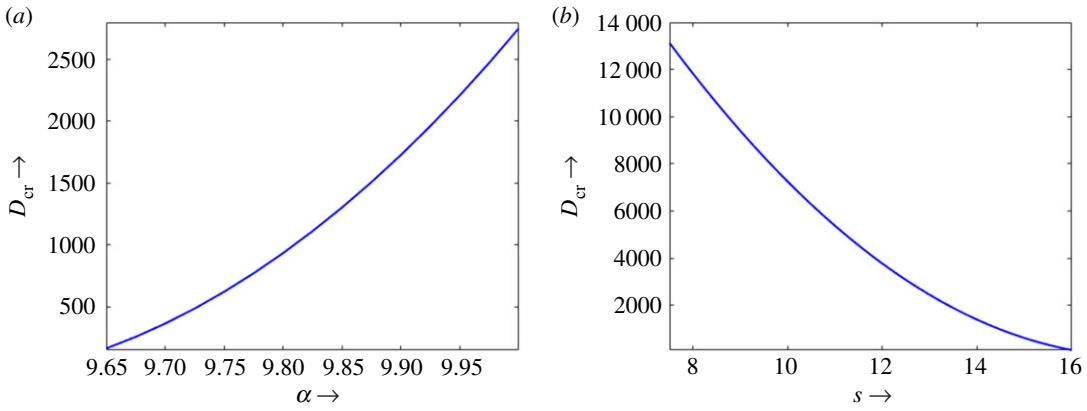

Figure 12. Dependence of the critical diffusivity ratio on model parameters: (*a*) D_{cr} versus α, (*b*) D_{cr} versus s. Other parameters as in figure 11.

sensitive with respect to the variation of some of the model parameters; examples are shown in figure 12. We have found that by varying α, a_1 and s, the critical ratio can be made as small as $D_{cr} = 98.5$ (obtained for parameter values $c_1 = 5$, $c_0 = c_2 = 1$, $a = 0.01$, $r = 7.5$, $s = 16$, $a_2 = 192.6$, $c_3 = 0.5$, $\sigma_1 = 0.05$, $\sigma_0 = 190$, $\alpha = 9.63$, $a_1 = 168$. The corresponding steady state values are $\bar{u} = 0.59$ and $\bar{p} = 0.01$), and the Jacobian is

$$J = \begin{pmatrix} 0.0937 & 0.434 \\ -0.508 & -0.0939 \end{pmatrix}. \tag{5.10}$$

Therefore, at this steady-state wealth acts as the activator and population as the inhibitor: the relation similar to the classical resource–consumer system. We mention here that a further reduction of D_{cr} does not appear to be possible: for instance, a further decrease in α (as in figure 12*a*) or a further increase in s (as in figure 12*b*) make the steady-state unstable.

6. Spatio-temporal dynamics: numerical results

In this section, we consider the spatiotemporal dynamics of system (5.1)–(5.2) that arises as a result of the Turing instability. Note that the fact that the steady state becomes, in a certain parameter range, unstable with regard to spatially heterogeneous perturbations is established analytically (see §5) and hence, as such, do not require any confirmation (e.g. by simulations). However, the mathematical analysis of the instability is based on linearization of the system in the vicinity of the steady state and thus is limited to the time interval when the deviation of the solution from the steady state is small. That raises the question: what are the dynamics at the later time, after the deviation from the steady state becomes large enough to be affected by the nonlinearity of the system? Turing instability is known to often lead to the formation of a stationary spatially periodic pattern [14], however more complicated dynamics can occur too [49,50].

In order to elucidate the above question, the reaction–diffusion system (5.1)–(5.2) is solved numerically by finite-differences using the following initial conditions:

$$u(x, 0) = \bar{u}, \quad p(x, 0) = \bar{p}\left(1 + 0.01 \sin\left(\frac{\pi x}{L}\right)\right), \tag{6.1}$$

with the size of the spatial domain $L = 120$. The diffusion coefficients are chosen as $D_u = 1$ and $D_p = 100$. For the reaction parameters, we use the values that are given at the end of the previous section as they were shown to (almost) minimize the diffusivity ratio. At the boundaries of the domain, the zero-flux conditions (5.3) are used.

For these parameter values, $D_p/D_u > D_{cr} = 98.5$ so that we expect that a small initial perturbation of the steady state leads to the emergence of a spatial pattern. This is indeed seen in the numerical simulations (figure 14). At an early stage of the dynamics, the initial conditions (6.1) fast evolve (over $t \sim 10^2$, not shown here) to a nearly stationary periodic spatial pattern, which then remains almost unchanged over a considerable time (until $t \approx 7000$, see figure 14*a*). The spatial distribution then starts evolving fast to develop a double-frequency spatial mode, so that the spatial distribution contains eight peaks instead of four (figure 14*c–e*). The double-frequency mode appears to be transient; at a slightly later time, it disappears and the spatial distribution returns to four distinct peaks but now at different locations, so that the peaks and the troughs exchange places (cf. figure 14*a*,*g*).

Further dynamics lead to a decrease in the amplitude of the peaks and eventually to the decay in both system's components and eventually to the population extinction. Figure 13 shows (for the same parameter values) the population and wealth densities versus time obtained at two fixed locations in space, i.e. at the boundary of the domain $x = 0$ and at $x = (3/8)L$, see figure 13*a*,*c*, respectively. It is readily seen that the change of the nearly stationary dynamics (until approx. $t = 7000$) to much faster dynamics occurs when the local evolution of the system's variables takes them away from the vicinity of the saddle point (figure 9).

7. Scales and units

It is readily seen from our numerical simulations that the wavelength of the emerging pattern is approximately $l = 30$. This is in dimensionless units; the choice $D_u = 1$ that we used in the simulations clearly indicates that, as the dimensional diffusion coefficient is measured in length$^2 \cdot$ time^{-1}. The question therefore arises here as to what is the corresponding dimensional spatial scale of the pattern, i.e. whether $l = 30$ corresponds to metres, kilometres, hundreds of kilometres, etc.? Recall that in our simulations we used a hypothetical parameter set. A rigorous, quantitative answer to the above question requires the knowledge of realistic parameter values, which are currently unknown (and there are difficult to find). However, a preliminary, semi-qualitative insight into this issue appears to be possible.

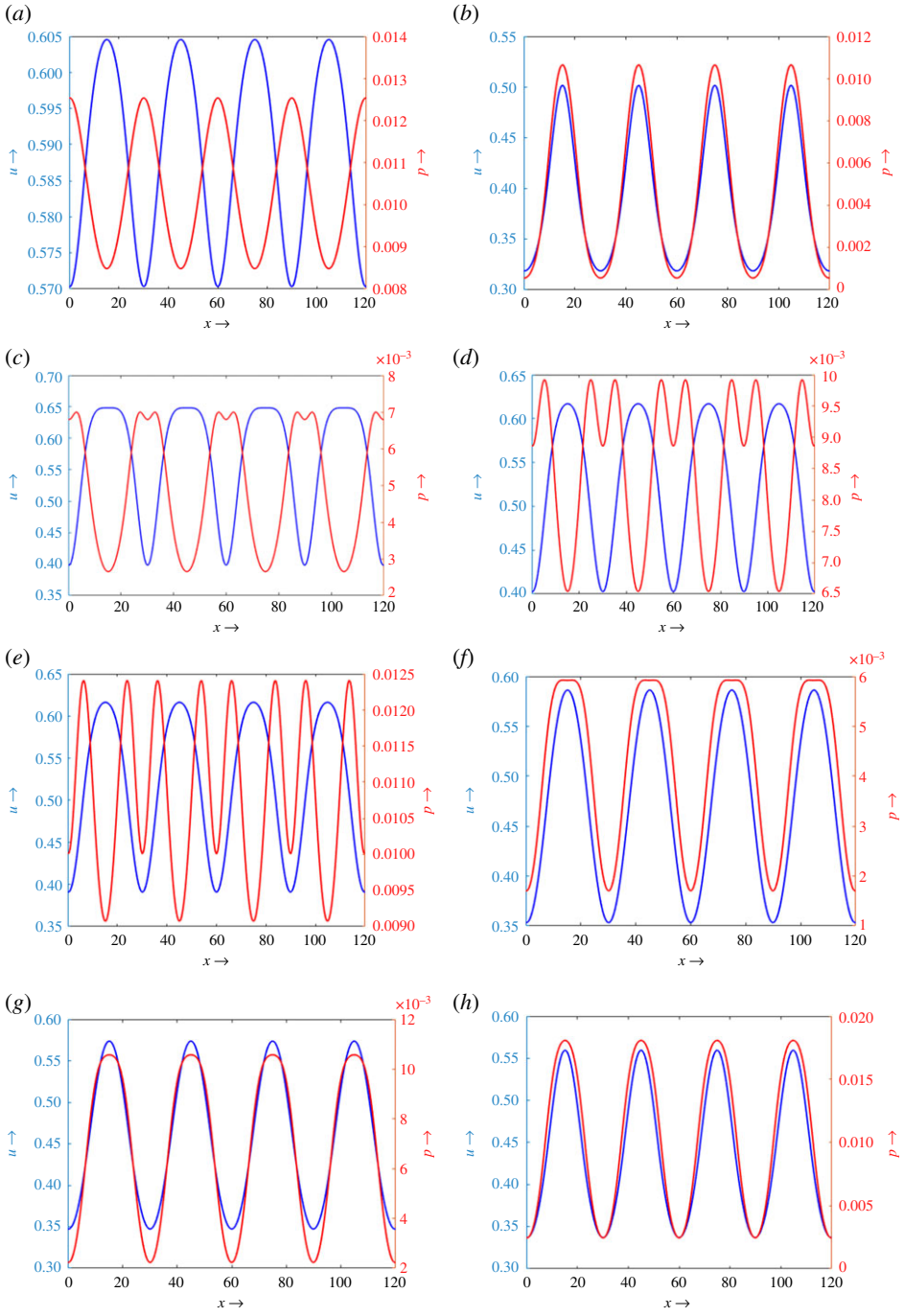

Figure 13. Snapshots of the spatial distribution of population (red) and wealth (blue) shown at different moments of time: (*a*) $t = 7000$, (*b*) $t = 7004$, (*c*) $t = 7030$, (*d*) $t = 7038$, (*e*) $t = 7040$, (*f*) $t = 7044$, (*g*) $t = 7046$ and (*h*) $t = 7048$.

The critical wavenumber, i.e. the wavenumber where characteristic function $R(z)$ reaches its minimum, cf. equation (5.8), is given by the following expression:

$$k_{\mathrm{cr}}^2 = \frac{D_u a_{22} + D_p a_{11}}{2 D_u D_p}, \tag{7.1}$$

e.g. see [48]. The corresponding wavelength can be written as

$$\begin{aligned} l_{\mathrm{cr}} &= 2\pi \left(\frac{2 D_u D_p}{D_u a_{22} + D_p a_{11}} \right)^{1/2} \\ &= 2\pi \sqrt{D_p/|a_{22}|} \left(\frac{2}{(D_p/D_u)(a_{11}/|a_{22}|) - 1} \right)^{1/2}, \end{aligned} \tag{7.2}$$

(taking into account that $a_{22} < 0$). Note that the expression in the brackets is dimensionless as it contains only the ratios of the parameters; for the parameter values used in simulations it gives $\sqrt{2/(100 \cdot 1 - 1)} \approx 0.14$.

Since the Jacobian elements have the dimension of inverse time, the square root in the right-hand side of (7.2) has the dimension of length and hence, once estimates of D_p and a_{22} are available, is the factor that determines the spatial scale. Estimates of the diffusivity of the human movement are available in the literature; here we use the value $D_p = 1000\ \mathrm{km}^2\ \mathrm{yr}^{-1}$ [51].

In order to estimate the unit of time, we consider equation (5.2) in a special case where $u = 0$ and p is small. In the non-spatial case, neglecting the quadratic term, its solution is given as $p(t) = \mathrm{const} \cdot \exp(-\sigma_0 t)$, so that over time $t = 1/\sigma_0$ the population size decreases by the factor of e, i.e. approximately threefold. Recalling that the value of σ_0 used in simulations was 190, we obtain that the threefold decrease

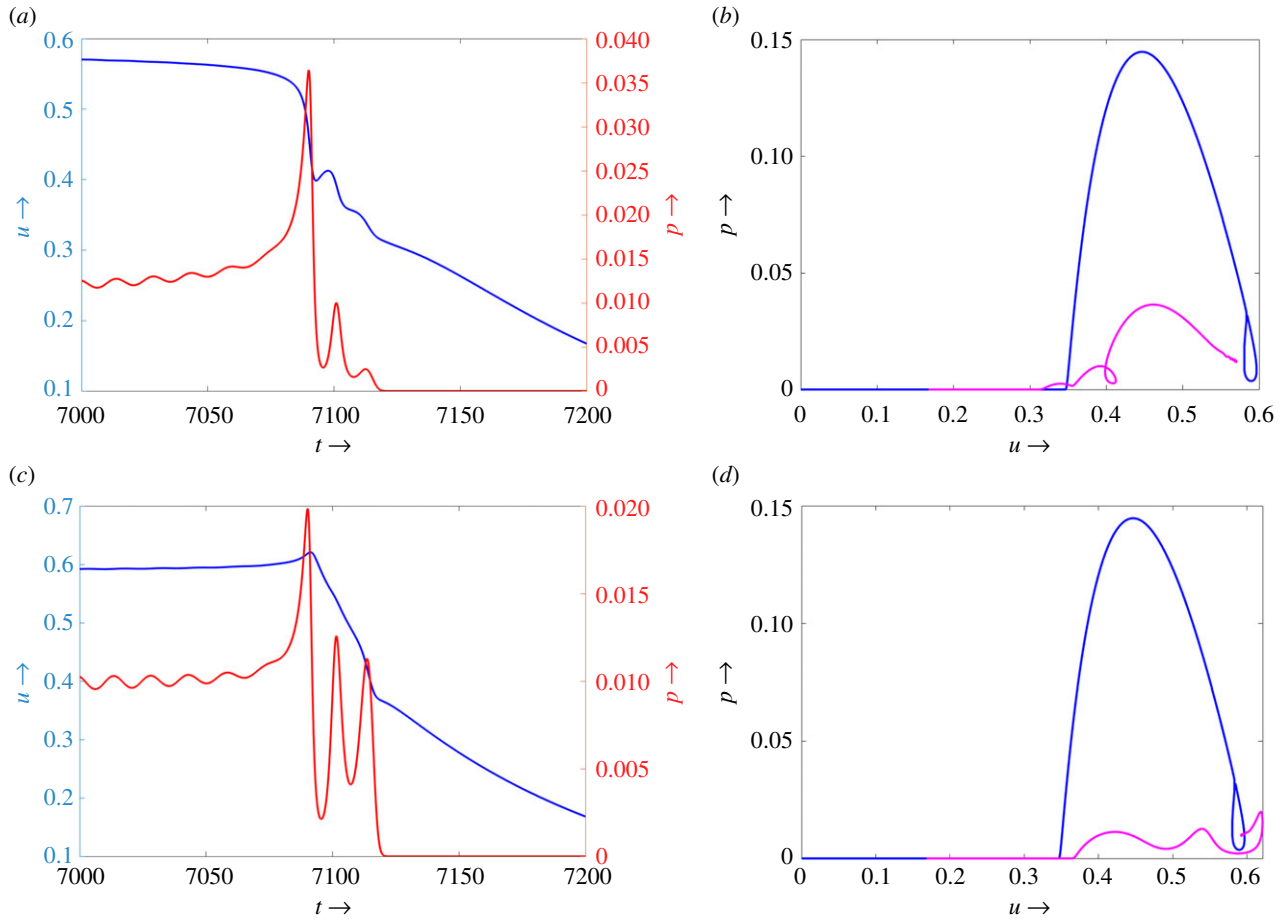

Figure 14. (*a*,*c*) The evolution of population and wealth densities at different spatial points, (*a*) for $x = 0$ and (*c*) for $x = 3L/8$. (*b*,*d*) The corresponding system's trajectories in the phase plane, the blue curve shows the unstable separatrix of the saddle point in the corresponding non-spatial model.

occurs by $t \approx 0.005$, in dimensionless units. In order to estimate what can be this value in dimensional units, e.g. in years, we note that the special case $u = 0$ corresponds to some extreme conditions where there are no resources whatsoever. Under such conditions, the population would decrease very fast (e.g. dying from hunger), the timescale of the decrease being defined by the physiological traits of the human body. Considering the corresponding time is 0.1 year, we readily obtain that the dimensionless $t = 1$ corresponds to approximately 20 years. Therefore, in dimensional units,

$$|a_{22}| \approx 0.1 \cdot (\text{time unit})^{-1} = 0.1/(20\,\text{years}) = 0.005\ \text{yr}^{-1}. \tag{7.3}$$

Correspondingly, we obtain that the scaling factor is

$$\sqrt{D_p|a_{22}|} = \sqrt{1000 \cdot 200}\ (\text{km}) = 450\ (\text{km}), \tag{7.4}$$

so that, from equation (7.2) we obtain:

$$l_{\text{cr}} = 2\pi \cdot 0.14 \cdot 450\ (\text{km}) \approx 400\ (\text{km}). \tag{7.5}$$

We notice that, although this number is somewhat less than that observed in the data, it is of the same order of magnitude.

8. Discussion and concluding remarks

In this paper, we have revisited factors and mechanisms affecting the spatial distribution of the human population. Such distributions often exhibit remarkable heterogeneity so that the population density in some areas (e.g. urban) can be much larger than in others (e.g. rural). In modern times, there are many factors that contribute to this difference [9,12,52]. However, how the heterogeneity in the population distribution has developed over a longer, historical timescale is not always clear. A textbook explanation relates the emergence of densely populated areas to the heterogeneity of the environment. Centuries ago, humankind was much more exposed to the forces of nature than it is nowadays. Areas with milder climatic conditions would have more likely been selected to establish a settlement. Recall also that agriculture was the main driver of the economy, its efficiency to a large extent being dependant on the properties of the natural environment.

Convincing as it may sound, in this paper we endeavoured to challenge the above explanation. We first identified a few areas (selected from different parts of the world) where the environmental properties such as the average annual temperature, the average annual precipitation and the elevation do not show much variability in space over stretches of thousands of kilometres. We have revealed that, in spite of this approximate spatial homogeneity of the environment, the population density distribution over those areas is clearly heterogeneous—in fact, in all cases exhibiting a nearly periodic spatial pattern (although the wavelength is somewhat different in different countries). We mention here that, at least in the case of Canada where the population density data are known starting from early twentieth century, this pattern does not show any considerable variation with time, with the location of peaks and troughs being approximately the same.

We then ask the following question: can there exist another mechanism potentially leading to the formation of a heterogeneous spatial distribution, even in an approximately homogeneous environment? We hypothesize that one such mechanism can result from the nonlinear interplay between the human population and the resources that support its growth, in the manner of resource–consumer interaction [16,53]. Indeed, it is well known that, in a system of two or more interacting components, a locally stable steady state can become unstable with respect to a spatially heterogeneous perturbation with a certain wavelength: the phenomenon called the Turing instability [47]. As a result, a spatially periodic pattern can arise [14,16].

Pattern formation due to the Turing instability is well known for chemical and biological systems [14,16,36]. More recently, the Turing instability has also been suggested as a possible relevant mechanism in spatial economy [9], albeit using a quite different modelling approach. For demographic systems, however, to the best of our knowledge, the Turing instability has never been considered. In this work, we have investigated this phenomenon within the framework of the demographic–economic model (5.1)–(5.2) proposed in our earlier work [32,46]. Conditions of the Turing instability in terms of the model parameters are found analytically (see §5). The spatio-temporal dynamics of the system resulting from the instability, including the development of the periodic spatial pattern, was considered in numerical simulations (§6). As realistic parameter values are not presently available, simulations were performed for hypothetical parameter values.

We mention here that, on one hand, the Turing instability is a generic property of a certain class of reaction–diffusion systems, hence its discovery in the model (5.1)–(5.2) is perhaps not surprising. On the other hand, since understanding of spatial economic-demographic systems is at its infancy, the formation of Turing patterns in a relevant mathematical model is thought-provoking, especially when considered in conjunction with the nearly periodic population distribution occurring in some parts of the world (cf. §2). Although we do not claim that the patterns emerging in the model are necessarily the same patterns that are observed in the real-world population distribution—indeed, without careful parameterization and realistic parameter values, such a claim could hardly be justified—the qualitative agreement (see §7) between the data and the simulations is encouraging.

Interestingly, a factor quantifying the spatial distribution of wealth in the northern part of USA (along the USA–Canada border) also exhibits a nearly periodic distribution with a comparable wavelength of 800–1000 km [54]. This observation may be regarded as indirect evidence in favour of our hypothesis.

We mention here that the Turing instability and the corresponding pattern formation in model (5.1)–(5.2) occur not just for a particular parameter set but in a certain parameter range (see figure 12) and hence are robust with respect to some variation in the parameters. This also follows from a more general argument that the solution of a reaction–diffusion system depends continuously on the variation of the form of the reaction term for an appropriately chosen norm of the functions, e.g. see [55], so that a small change (that does not change the existence or stability of the system steady states) in the form of the functions can only lead to a small change in the solution. Thus, the parameter set chosen for our numerical simulations (§6) is representative and by no means unique. We also checked the robustness of the results to the choice of the initial conditions by repeating the simulations for a few other types including random. The results (not shown here for the sake of brevity) are essentially the same, showing only a minor difference from those presented in figures 13 and 14.

Our simulations reveal that the emerging periodic pattern is not the large-time asymptotics of the system, as it usually happens in the case of Turing patterns, but the long-term transient dynamics (cf. [56]). After the pattern emerges (at $t \sim 10^2$), it remains almost unchanged over a long period of time (for parameters of figure 14, until $t \approx 7000$). Eventually, this quasi-stationary regime turns into a fast spatiotemporal dynamics where the emergence of a higher-frequency spatial mode is followed by the system collapse.

Recall, however, that our simulations were performed for hypothetical parameter values. Correspondingly, in the context of real-world demographic–economic systems, the system collapse as a result of pattern formation is hypothetical too; by no means it can be regarded as a forecast. Moreover, given our simulation parameters and the corresponding time units, it is not going to happen until a very remote future. However, we believe that, by adding this 'worst case' to the range of possible scenarios, it contributes to the understanding of ways in which the society may develop. Besides, it remains to be checked whether the collapse may happen at an earlier time for other parameter values.

Note that, although the corresponding non-spatial system exhibits multistability, without an external perturbation (which is not a part of the model (4.1)), the system forever remains in the vicinity (basin of attraction) of the positive stable steady state. Also, for the parameter values outside of the Turing instability range, the spatial system can persist at its positive steady state indefinitely. The Turing instability and the subsequent pattern formation are therefore the factors that push the system out of the basin of attraction of the positive steady state (so that the system goes to extinction). Interestingly, although the Turing pattern formation in bistable and multistable reaction–diffusion systems has been a focus of interest since 1990s [57–59], it has been only recently that this phenomenon—i.e. the formation of quasi-stable patterns eventually converging to a spatially uniform steady state—was observed and studied [60,61].

Our study leaves a few open questions. Firstly, we mention that, although our theoretical findings, such as the formation of Turing's patterns on a relevant spatial scale are in a qualitative agreement with the real-world data (see §7), a direct comparison between theory and data is hardly possible at this stage of research. Such a comparison would required a sufficiently accurate estimate of the value of all model's parameters. This is a challenging and tedious task and will become the focus of a separate study.

Secondly, our model is quite schematic. One question that arises here is what can be the effect of different history and/or ethnicity on the population dynamics, although the latter is difficult to quantify [62]. Different ethnicity can mean a different culture [63] and that can affect the ways in which the resources are consumed and the wealth is generated and distributed. Although we do not expect that it will change the system's properties completely—the existence of the nearly periodic spatial pattern in countries as different as Canada, Australia and Mongolia points out at the universality and robustness of this phenomenon—yet this issue

should be addressed more carefully, for instance by using more complicated models.

Another factor that was completely disregarded in our present study is the way in which the given country's population has originally emerged. For instance, Canada's population mostly emerged as a result of migration from Europe, eventually propagating across Canada from east to west. In Australia, Europeans first settled in and around bays and then diffused over the continent. By contrast, Mongolia has a significantly different history [64] and, albeit not ethnically uniform, its present population mainly consists of the same nations that have lived there at least since Middle Ages. In mathematical terms, the different ways in which different countries came to be populated may correspond to different initial conditions, which, in its turn, may affect the selection of the emerging spatial mode.

We also mention here that a more advanced model should include more pathways for the effect of wealth on spatio-temporal human population dynamics, e.g. by accounting for cross-diffusion (cf. [32]). Indeed, in our current study, low or negative wealth (i.e. debt) only affects the population by increasing the mortality rate. However, intuitively, it can also increase the emigration rate. It is well known that cross-diffusion can lead to Turing instability and pattern formation in population systems of ecological origin (e.g. [65]); whether it can occur in a demographic–economic system is yet to be seen.

Finally, the effect of the system's spatial dimensions remains an open question. Recall that our spatially explicit model only include one spatial dimension. Arguably, it was appropriate for considering the population dynamics along a narrow stripe, as was the case identified in our real-world examples. Yet a mode general study on pattern formation in the demographic–economic system should consider it in the more realistic case of two spatial dimensions. That may reveal more complicated patterns and more complicated dynamics. That should be a focus of future research.

Data accessibility. This research did not produce any new data. All data mentioned in the text were taken from public domains.

Authors' contributions. S.P. conceived the idea and designed the methodology; S.P. and V.V. developed the mathematical model; A.Z. collected the data, conducted statistical analysis and performed numerical simulations (as part of her PhD programme under the supervision of S.P.). All authors contributed to interpretation of findings. S.P. produced the first draft of the manuscript with contribution from A.Z. All authors contributed critically to its development and completion.

Competing interests. We declare we have no competing interests.

Funding. V.V. was supported by the RUDN University through Strategic Academic Leadership Program.

Acknowledgements. Prof. Ehsan Gharaie (Melbourne) advised the authors on the choice of the territorial units in Australia, his help is much appreciated. The third author (V.V.) has been supported by the RUDN University Strategic Academic Leadership Program.

References

1. Ehrlich P. 1968 *The population bomb*. New York, NY: Ballantine Books.
2. Malthus TR, Flew A. 1982 *An essay on the principle of population and a summary view of the principle of population*. Harmondsworth, UK: Penguin.
3. United Nations. 2002 *Policy responses to population decline and ageing*. New York, NY: UN.
4. United Nations. 2017 World population prospects. key findings and advance tables. New York, NY: UN.
5. Meadows DH, Meadows DL, Randers J, Behrens III WW. 1972 *The limits to growth*. New York, NY: Universe Books.
6. Coale AJ. 1972 *Growth and structure of human populations: a mathematical investigation*. Princeton, NJ: Princeton University Press.
7. Impagliazzo J. 1985 *Deterministic aspects of mathematical demography*. Berlin, Germany: Springer.
8. Graunt J. 1662 Natural and political observations made upon the bills of mortality. London. Reprinted 1973 in *The earliest classics: pioneers of demography* (ed. P Laslett). Farnborough, UK: Gregg International.
9. Fujita M, Krugman PR, Venables A. 1999 *The spatial economy: cities, regions, and international trade*. Cambridge, MA: MIT Press.
10. Ikeda K, Murota K. 2014 *Bifurcation theory for hexagonal agglomeration in economic geography*. Tokyo, Japan: Springer.
11. Ikeda K, Aizawa H, Kogure Y, Takayama Y. 2018 Stability of bifurcating patterns of spatial economy models on a hexagonal lattice. *Int. J. Bifurcation and Chaos* **28**, 1850138. (doi:10.1142/S0218127418501389)
12. Krugman P. 1996 Confronting the mystery of urban hierarchy. *J. Jpn Int. Econ.* **10**, 399–418. (doi:10.1006/jjie.1996.0023)
13. Cantrell RS, Cosner C. 2004 *Spatial ecology via reaction–diffusion equations*. London, UK: John Wiley & Son.
14. Meinhardt H. 1982 *Models of biological pattern formation*. New York, NY: Academic Press.
15. Meron E. 2015 *Nonlinear physics of ecosystems*. Boca Raton, FL: CRC Press.
16. Murray JD. 1989 *Mathematical biology*. Berlin, Germany: Springer.
17. Statistics Canada. https://www150.statcan.gc.ca (accessed 6 January 2021).
18. Reddit. 2019 *Topographic map of Canada*. https://www.reddit.com/r/MapPorn/comments/8vy788/topographic-map-of-canada-1400-1211/ (accessed 21 June 2020).
19. Mapsofworld.com. *Canada annual mean temperature map*. https://www.mapsofworld.com/canada/thematic-maps/canada-annual-mean-temperature-map.html (accessed 7 January 2021).
20. Government of Canada. *Environment Canada: temperature and precipitation*. https://weather.gc.ca/saisons/ (accessed 5 January 2021).
21. Glantz SA, Slinker BK. 1990 *Primer of applied regression and analysis of variance*. New York, NY: McGraw-Hill.
22. Australian Bureau of Statistics. 1270.0.55.007-Australian Population Grid, 2011. Canberra, Australia: Australian Bureau of Statistics. See https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.007.
23. Australian Government. *Geoscience Australia: elevations*. https://www.ga.gov.au/scientific-topics/national-location-information/landforms/elevations (accessed 21 June 2020).
24. Australian Government. *Geoscience Australia: climatic extremes*. http://www.ga.gov.au/scientific-topics/national-location-information/dimensions/climatic-extremes (accessed 21 June 2020).
25. Australian Government. *Bureau of Meteorology*. http://www.bom.gov.au/jsp/ncc/climate-averages/rainfall/index.jsp (accessed June 20, 2020).
26. New South Wales Government, Office of Sport. *How to find what's on in your local area.* https://www.sport.nsw.gov.au/how-to-find-whats-on-in-your-local-area (accessed 5 January 2021).
27. Encyclopedia Britannica (2002) Mongolia.
28. Wikipedia. *Geography of Mongolia*. https://en.wikipedia.org/wiki/Geography-of-Mongolia (accessed 5 January 2021).
29. Ministry of the Environment (Japan). Climate change in Mongolia. In 40th Session of the Int. Panel on Climate Change (IPCC), October 2014. See https://www.env.go.jp/earth/ondanka/pamph_gcm/gcm_mongolia_en.pdf.
30. Barten A. 1987 The coeffecient of determination for regression without a constant term. In *The practice of econometrics* (eds R Heijmans, H Neudecker), pp. 181–189. Dordrecht, The Netherlands: Kluwer.
31. Cameron AC, Windmeijer FAG. 1997 An R-squared measure of goodness of fit for some common

nonlinear regression models. *J. Econom.* **77**, 329–342. (doi:10.1016/S0304-4076(96)01818-0)

32. Volpert V, Petrovskii S, Zincenko A. 2017 Interaction of human migration and wealth distribution. *Nonlinear Anal.* **159**, 408–423. (doi:10.1016/j.na.2017.02.024)
33. Codling EA, Plank MJ, Benhamou S. 2008 Random walk models in biology. *J. R. Soc. Interface* **5**, 813–834. (doi:10.1098/rsif.2008.0014)
34. Turchin P. 1998 *Quantitative analysis of movement*. Sunderland, MA: Sinauer.
35. Viswanathan GM, Raposo EP, Stanley HE. 2011 *The physics of foraging: an introduction to random searches and biological encounters*. Cambridge, UK: Cambridge University Press.
36. Edelstein-Keshet L. 2005 *Mathematical models in biology*. Classics in Applied Mathematics, vol. 46. Philadelphia, PA: SIAM.
37. Small ML, Harding DJ, Lamont M. 2010 Reconsidering culture and poverty. *Ann. Amer. Acad. Polit. Soc. Sci.* **629**, 6–27. (doi:10.1177/0002716210362077)
38. Deaton A. 2013 *The great escape—health, wealth, and the origins of inequality*. Princeton, NJ: Princeton University Press.
39. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, Bergeron A, Cutler D. 2016 The association between income and life expectancy in the United States, 2001–2014. *J. Am. Med. Assoc.* **315**, 1750–1766. (doi:10.1001/jama.2016.4226)
40. Falchetta G, Ahmed T, Hammad AT, Shayegh S. 2020 Planning universal accessibility to public health care in sub-Saharan Africa. *Proc. Natl Acad. Sci. USA* **117**, 31 760–31 769. (doi:10.1073/pnas.2009172117)
41. Hierink F, Okiro EA, Flahault A, Ray N. 2021 The winding road to health: a systematic scoping review on the effect of geographical accessibility to health care on infectious diseases in low- and middle-income countries. *PLoS ONE* **16**, e0244921. (doi:10.1371/journal.pone.0244921)
42. Wang D, Cao X, Huang X. 2021 Equity of accessibility to health care services and identification of underserved areas. *Chin. Geogra. Sci.* **31**, 167–180. (doi:10.1007/s11769-021-1181-0)
43. Groth C. 2016 Lecture notes in macroeconomics. Department of Economics, University of Copenhagen, Copenhagen, Sweden. See https://web2.econ.ku.dk/okocg/MAT-OEK/Mak%C3%98k2/Mak%C3%98k2-2015/Forel%C3%A6sninger/Pensum-M2-i-een-fil-2015.pdf.
44. The Digital Economist. http://www.digitaleconomist.org/dmacro.html (accessed 17 June 2020).
45. Diamond J. 2004 *Collapse. How societies choose to fail or succeed*. New York, NY: Viking.
46. Zincenko A, Petrovskii SV, Volpert V. 2018 An economic–demographic dynamical system. *Math. Mod. Natur. Phenom.* **13**, 27. (doi:10.1051/mmnp/2018035)
47. Turing AM. 1952 On the chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B* **237**, 37–72. (doi:10.1098/rstb.1952.0012)
48. Malchow H, Petrovskii SV, Venturino E. 2008 *Spatiotemporal patterns in ecology and epidemiology: theory, models, simulations*. Boca Raton, FL: Chapman & Hall/CRC.
49. Dangelmayr G, Oprea I (eds). 2004 *Dynamics and bifurcation of patterns in dissipative systems*. Nonlinear Science Series B, vol. 12. Singapore: World Scientific.
50. Liehr AW. 2013 *Dissipative solitons in reaction diffusion systems. mechanism, dynamics, interaction*. Springer Series in Synergetics, vol. 70. Berlin, Germany: Springer.
51. Young DA, Bettinger RL. 1995 Simulating the global human expansion in the late Pleistocene. *J. Arch. Sci.* **22**, 89–92. (doi:10.1016/S0305-4403(95)80165-0)
52. Glaeser EL, Matthew E, Kahn ME, Rappaport J. 2000 Why do the poor live in cities? Harvard Institute of Economic Research. Discussion Paper No. 1891. See https://scholar.harvard.edu/files/glaeser/files/why_do_the_poor_live_in_cities.pdf.
53. Murdoch WW, Briggs CJ, Nisbet RM. 2013 *Consumer-resource dynamics*. Princeton, NJ: Princeton University Press.
54. Chenevert R, Gottschalck A, Klee M, Zhang X. 2017 Where the wealth is: the geographic distribution of wealth in the United States. US Census Bureau. https://www.census.gov/content/dam/Census/newsroom/press-kits/2017/assa-geographic-distr-wealth.pdf (accessed 15 December 2020).
55. Volpert AI, Khudyaev SI. 1985 *Analysis in classes of discontinuous functions and equations of mathematical physics*. Dordrecht, The Netherlands: Nijoff.
56. Morozov A *et al.* 2020 Long transients in ecology: theory and applications. *Phys. Life Rev.* **32**, 1–40. (doi:10.1016/j.plrev.2019.09.004)
57. Leppänen T, Karttunen M, Barrio RA, Kaski K. 2004 Morphological transitions and bistability in Turing systems. *Phys. Rev. E* **70**, 066202. (doi:10.1103/PhysRevE.70.066202)
58. Mazin W, Rasmussen KE, Mosekilde E, Borckmans P, Dewel G. 1996 Pattern formation in the bistable Gray–Scott model. *Math. Comp. Simul.* **40**, 371–396. (doi:10.1016/0378-4754(95)00044-5)
59. Metens S, Dewel G, Borckmans P, Engelhardt R. 1997 Pattern selection in bistable systems. *Europhys. Lett.* **37**, 109–114. (doi:10.1209/epl/i1997-00119-4)
60. Rodrigues VW, Mistro DC, Rodrigues LAD. 2020 Pattern formation and bistability in a generalist predator–prey model. *Mathematics* **8**, 20. (doi:10.3390/math8010020)
61. Smith S, Dalchau N. 2018 Beyond activator–inhibitor networks: the generalised Turing mechanism. arXiv:1803.07886v1. See https://arxiv.org/abs/1803.07886v1.
62. Connelly R, Gayle V, Lambert PS. 2016 Ethnicity and ethnic group measures in social survey research. *Methodol. Innovations* **9**, 1–10.
63. Nagel J. 1994 Constructing ethnicity: creating and recreating ethnic identity and culture. *Soc. Problems* **41**, 152–176. (doi:10.2307/3096847)
64. Allsen TT. 2001 *Culture and conquest in Mongol Eurasia*. Cambridge, UK: Cambridge University Press.
65. Peng Y, Zhang T. 2016 Turing instability and pattern induced by cross-diffusion in a predator–prey system with Allee effect. *Appl. Math. Comput.* **275**, 1–12.