

Article

# An Empirical Agent-Based Model for Regional Knowledge Creation in Europe

Martina Neuländtner 

AIT Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria; martina.neulaendtner@ait.ac.at

Received: 26 May 2020; Accepted: 29 July 2020; Published: 30 July 2020



**Abstract:** Modelling the complex nature of regional knowledge creation is high on the research agenda. It deals with the identification of drivers for regional knowledge creation of different kinds, among them inter-regional networks and agglomeration factors, as well as their interplay; i.e., in which way they influence regional knowledge creation and accordingly, innovation capabilities—in the short- and long-term. Complementing a long line of tradition—establishing a link between regional knowledge input indicators and knowledge output in a regression framework—we propose an empirically founded agent-based simulation model that intends to approximate the complex nature of the multi-regional knowledge creation process for European regions. Specifically, we account for region-internal characteristics, and a specific embedding in the system of region-internal and region-external R&D collaboration linkages. With first exemplary applications, we demonstrate the potential of the model in terms of its robustness and empirical closeness. The model enables the replication of phenomena and current scientific issues of interest in the field of geography of innovation and hence, shows its potential to advance the scientific debate in this field in the future.

**Keywords:** regional knowledge creation; geography of innovation; collaboration networks; agent-based modelling; spatial simulation; Europe

## 1. Introduction

Understanding and explaining the complexities of regional knowledge creation constitutes an ongoing challenge for empirical scholars in regional science. Specifically, the literature has long been concerned with the spatial distribution of knowledge creation and innovation, concluding that these kinds of activities are not equally distributed in space but rather tend to be spatially clustered [1,2]. With knowledge being not easily accessible at every point in space, the location of knowledge creation, as well as the processes of knowledge diffusion, become a crucial point in understanding regional development and growth [3,4]. In this respect, attention has been shifted to the investigation and modelling of regional knowledge creation processes as an interplay between (i) geographically localised knowledge interactions within the region, and (ii) the embedding of the region in inter-regional R&D collaborations (see e.g., [5]), in particular by means of region-internal and region-external knowledge interactions in the form of R&D collaborations (see [6] for an overview).

Modelling regional knowledge creation follows a long line of research tradition, often applying the Knowledge Production Function (KPF) framework to model determinants of regional knowledge creation and innovation [7–9]. These studies typically attempt to establish a direct link between some kind of regional knowledge input, such as industrial and university R&D, and knowledge outputs measured in terms of patents, innovation or publication counts (see e.g., [7,10–12]) In this context, the role of knowledge spillovers [8,13,14], spatial proximity [15–17], and nonspatial forms of proximity [18–20] on regional knowledge creation and innovation are widely studied. However, all these studies are done at an aggregate regional level, and accordingly do not account for the regions' underlying micro-structure, for instance by assuming that each regional organisation benefits in the same way

from inter-regional knowledge spillovers. However, a better approximation and understanding of the real-world complexity of regional knowledge creation processes requires models accounting for the heterogeneity of the agent population, for the non-linearity of the interactions between agents, and for the complexity of the environment. Considering these aspects allows the observation of emergent phenomena such as specialisation and concentration tendencies in regional knowledge creation driven by the structure of R&D collaborations.

Recent contributions to the discussion on knowledge creation have been made by adding a dynamic perspective using computer simulation techniques, especially agent-based modelling (ABM; the abbreviation ABM is used for 'agent-based models' and 'agent-based modelling' consecutively), but are mainly implemented at an abstract, theoretical level: for instance, from a spatial perspective, theoretical contributions by Batty [21], and Crooks et al. [22] propose general spatial modelling frameworks and pose key challenges in geo-spatial modelling (see [23] for overview), whereas, Ausloos et al. [24] discuss simulating spatial interactions in ABM. From a conceptual viewpoint, Dawid [25], and Gilbert et al. [26] target innovation and technological change, and knowledge dynamics in innovation networks, respectively. Moreover, Vermeulen and Pyka [27] address the spatial distribution of knowledge in the setting of regional innovation policy. Whereas, theoretical models are built as tools for theory-building, very recently, a few empirical models of regional knowledge creation have aimed at analysing real-world scenarios; Wang et al. [28], use an agent-based model for analysing the diffusion of technologies across Chinese regions, while Beckenbach et al. [29] present an agent-based simulation of regional innovation dynamics including agents with explicit and implicit knowledge endowment, and Paier et al. [30] focus on the evolution of a single region's technological profile in a context of policy analysis (in the Viennese biotechnology sector). However, so far, the limited simulation studies following this research path, are either of a purely theoretical and conceptual nature, lacking empirical foundation and hence, real-world applicability, e.g., in a (regional) innovation policy context (e.g., [27,31]), have only a limited geographical and/or sectoral scope (e.g., [30,32]), neglecting the theoretical fundamentals of regional knowledge creation, and/or deal in a quite narrow way with network formation (e.g., [33,34]), as well as knowledge transfer and diffusion [28,35,36].

Hence, we propose an empirically founded agent-based simulation model for regional knowledge creation in Europe. It intends to better approximate the complex nature of the multi-faceted regional knowledge creation processes by specifically accounting for (i) region-internal characteristics, (ii) agent heterogeneity in the knowledge creation process, and (iii) a specific embedding in the system of region-internal and region-external interdependencies in the form of R&D collaboration linkages. By this, we particularly reflect the idea of geographical and relational aspects of the knowledge creation process, which is driven by the debate on 'local buzz' and 'global pipelines' as two forms of interactive knowledge creation [37]. This allows us to model local dynamics, such as learning and knowledge transfer, as well as structural evolution in the form of inter-regional network formation and transformation on a global level. The strong theoretical, and explicitly empirical, foundation enables us to apply the model to real-world contexts, such as simulation experiments referring to Research, Technology, and Innovation (RTI) policy measures at the European, the national as well as the regional levels (e.g., smart specialisation, mission-oriented public funding). In this study we present a comprehensive model overview, providing details on the model elements and processes, as well as technical specifications and robustness checks. The potential of the model is demonstrated by small example applications on currently debated research issues in the geography of innovation literature, namely regional concentration and specialisation patterns, as well as the role of networks as drivers for regional knowledge creation.

The remainder of this study is organised as follows. In Section 2, we shortly outline the agent-based modelling approach and give a detailed presentation of the proposed simulation model of multi-regional knowledge creation, subsuming a description of the model elements and processes, as well as the empirical foundations. In Section 3, we demonstrate the potential of the simulation model by small example applications to current scientific debates. In Section 4, we conclude with a discussion of the

model results and a critical assessment of the functionality of the model. Moreover, future development steps of the model are outlined and ideas for further fields of application are presented.

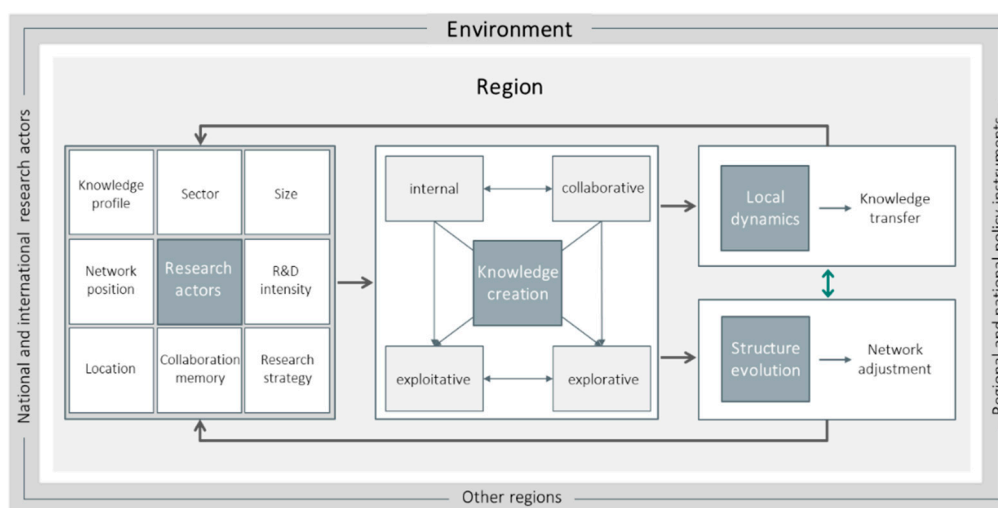
## 2. Materials and Methods

This section is dedicated to the description of the proposed multi-region agent-based model of knowledge creation and the specification of its empirical foundation, including the agent initialisation, calibration and output evaluation. The model description is deliberately kept brief; details on model elements and processes are given in Appendix A.

Generally, ABMs are developed to discover emergent properties from a bottom-up perspective and—in an attempt to replicate real-world concepts, actions, relations or mechanisms—are used to anticipate future developments and outcomes [38]. In this respect, ABM is particularly suited to examine the complex and adaptive nature of regional innovation systems, as it provides a framework to model and simulate the behaviour of heterogeneous agents, and to investigate the complex dynamics of system-wide interactions amongst them. Hence, the aim of this simulation model is to investigate inter-regional knowledge creation across European regions. In doing so, we adopt an empirically driven agent-based modelling (ABM) approach, utilising large-scale data sets on regional knowledge creation and research collaboration activities.

### 2.1. Model Description

The model conception closely follows state-of-the-art theoretical and conceptual contributions, as well as empirical findings in the fields of regional science, economic geography, and the geography of innovation literature (in particular [39–42]). Moreover, we integrate ABM and advanced methodological tools from social network analysis (SNA) and econometrics. The simulation model is implemented in Java, drawing on elements of the MASON (Multi-Agent Simulator Of Neighborhoods) multi-agent simulation environment. The proposed model comprises three key characteristics, (i) a set of interacting agents, their attributes and behaviours, (ii) a set of relationships and methods of interaction, situated within (iii) a model environment [43], that serve as cornerstones for the development of the model, visually illustrated by Figure 1.



**Figure 1.** Overview of model elements and processes—region perspective.

In this conception, agents are modelled as research actors characterised by organisation-level empirically-based attributes. Each agent is equipped with a *knowledge profile* representing the knowledge endowment of the agent, indicative of the technology classes the agent is active in, as well as the

expertise in the respective class. Hence, the knowledge endowment of agent  $i$  can be defined as a vector of length  $k$  (with  $k$  being the number of technology classes included in the model)

$$\kappa_i = \{\kappa_{i1}, \kappa_{i2}, \dots, \kappa_{ik}\} \quad (1)$$

where  $i = 1, 2, \dots, N$  with  $N$  being the total number of agents in the model, and the value of  $\kappa_{ik}$  determining the expertise in the respective technology class  $k$  (level of knowledge).

The agents' *location* is specified by the European NUTS-2 region [44] the agent is located in—in total, 283 regions in the EU-27 countries plus the United Kingdom and Norway are covered. Since there are no large-scale firm-level data on location, industry association, size and R&D intensity available in a systematic way for European regions, such data are constructed based on region and industry characteristics. Hence, in the process of empirically assigning the industry firm agents to regions, their *industry sector*, *size* (number of employees) and *R&D intensity* is specified as well. The primary objective for the agents in the model is to be representative for each region with respect to its characteristics (see Section 2.2 for details on agent initialisation). In addition to empirical attributes, agents are also characterised by model-inherent attributes: *research strategy*, *collaboration memory*, and *network position* indicating external networking capability.

The agents' relationships and interactions are defined within the *knowledge creation process* subsuming the total of agents' actions in creating new knowledge. This process is designed as a learning process along a specified research path, that each agent decides upon individually according to its *mode of knowledge creation*—exploitative or explorative (following the concept brought forward by March [45])—and with respect to its current knowledge and research target. Knowledge creation is based on the concept of *technology space*, which is defined as a network comprising a set of technology classes (TCs), with weighted links indicating the technological proximity between these classes (see Figure A3 in Appendix A for an illustrative example). Formally, the technology space  $S$  can be defined as symmetric matrix

$$S = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix}_{k \times k} \quad (2)$$

where  $p_{..}$  denote the Jaccard coefficients as a measure of proximity (empirically-based), and  $k$  is the number of technology classes considered in the model.

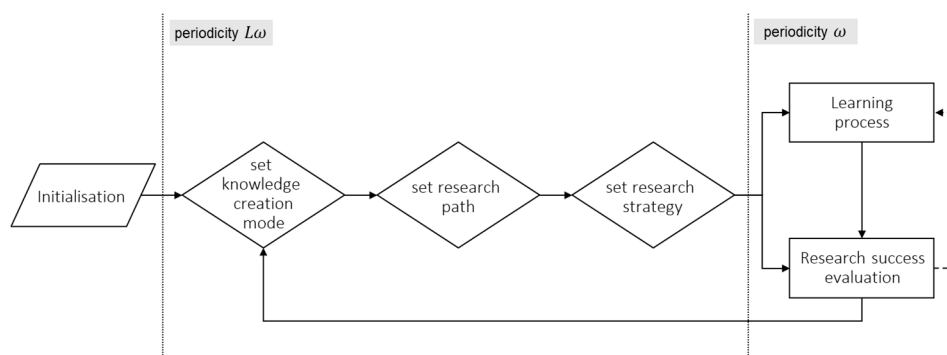
The technology space  $S$  serves as framework for the agents to gain new knowledge as they move along their *research paths*. Each research path comprises selected technology classes, indicating the way and direction of learning. Generally, a research path  $P$  is defined as a subset of all technology classes in the technology space

$$P = \{\tau_1, \tau_2, \dots, \tau_L\} \in T = \{\tau_1, \tau_2, \dots, \tau_k\} \quad (3)$$

where  $l = 1, 2, \dots, L$  with  $L$  being the length of the research path and  $T$  the set of all technology classes.

The agent's individual knowledge creation process follows a predefined sequence of actions (see Figure 2 for simplified flow chart); however, it still implies many degrees of freedom, allowing for heterogenous interactions and processes that result in varying outcomes. In total, the agent's knowledge creation process includes selecting a *mode of knowledge creation*, setting a *research path*, selecting a *research strategy*, and a *learning* and *research success evaluation* process. The single processes follow two different periodicities: whereas, the subprocesses in the learning process and output evaluation are carried out at each model step (periodicity  $\omega$ ), setting a research agenda and research strategy occurs at every other step after completing the learning process (periodicity  $L\omega$ ); where  $\omega$  denotes one simulation step and  $L$  an integer value indicating the length of the agent's research path. To initiate the process at the beginning

of the simulation, a one-time-only initiation of a starting technology class for the research process takes place; i.e., a random technology with non-zero expertise of the agent's knowledge endowment.



**Figure 2.** The agent's knowledge creation process (simplified).

Setting the research path depends on the agent's *mode of knowledge creation*: *exploitative* or *explorative*; where exploitative knowledge creation reflects a direct and targeted, commercially oriented way of performing research, and exploration expresses a non-targeted and more indirect path selection for knowledge creation. In a next step, the agent decides on the *research strategy*—i.e., whether to follow this path by means of *internal research*, i.e., perform in-house research, or by looking for a suitable research partner to perform *collaborative research*. Hence, a core element of the collaborative research process is the agent's choice of a suitable collaboration partner. The partner choice relies on collaboration probabilities resulting from the estimation of a Spatial Interaction Model (SIM) considering the geographical distance between the regions, and variables indicating a neighbouring region and country. Spatial distance as well as country and region borders are generally acknowledged to be among the most important determinants to explain inter-regional R&D collaborations (e.g., [46,47]). Despite increasing globalisation and new information and communication technologies, spatial proximity is (still) a crucial factor in establishing and maintaining R&D network links. Especially, more complex knowledge requires the exchange of tacit knowledge elements via face-to-face interaction [48,49].

The probabilities resulting from the SIM are complemented by collaboration shares based on statistical data. A final criterion for a suitable partner is cognitive proximity, as in the presence of a certain overlap of knowledge endowments. In collaborative research, a further distinction is made between two *modes of collaboration*, (i) research-mode and (ii) service-mode. Whereas, the first mode is aimed at representing basic and applied research projects, primarily focused on the creation and deepening of knowledge, the second mode exemplifies commission projects, usually characterised by an efficient and straight-forward research agenda.

Knowledge creation is defined as a *learning process* along the specified research path, i.e., performing research along trajectories of technology classes. To determine whether a new step (i.e., new technology class) on the research path is reached, the *research success* is *evaluated*. This is a necessary intermediate step for every transition from one to another technology class. The research success is a scaled composite indicator (interpretable as success probability) depending on (i) agent-specific characteristics (overall expertise, R&D quota, internal and external capability) and (ii) technological proximity between the involved technology classes. In the case of successful research, the level of expertise is updated in the respective technology class. If the research is evaluated as not successful, the agent either chooses a new, but similar research path or stays on the original path. For collaborative research, the process of research success evaluation is dependent on the collaboration partners' expertise. Moreover, possible knowledge transfer in the collaborative knowledge creation process is based on the cognitive distance (knowledge distance) between the two collaborating agents. The actual knowledge gain depends on the absorptive capacity—representing the trade-off between novelty value and understandability of new knowledge. Specifically, it is assumed that the amount of knowledge gain

corresponds to an inverted u-shaped relationship, i.e., both low and high knowledge complementary results in low knowledge gains, indicating the presence of an ‘optimal distance’ entailing a trade-off between ‘learning something new’ and ‘mutual understanding’ (e.g., [50,51]).

The environment defines the space in which the agents operate; accordingly, it contains all the information external to the agents used in the decision-making processes and provides a structure or space for agent interaction [38]. From the perspective of a regional innovation system, we see national and international research actors, and other regions, as external elements of the specific region. Especially, studies in the vein of regional innovation systems (RIS) stress the importance of such external factors on the knowledge creation of individual research actors located within a certain region, e.g., universities and public research organisations that conduct basic and applied research, and regional policy institutions that implement regional innovation policies [41]. External to the whole system of regions and interrelated agents, national and European policy interventions may also affect the region-specific knowledge creation processes. Evidently, these external factors are by no means isolated from the region-internal processes and dynamics, but rather strongly interrelate with agent-specific capabilities. In an ABM, this fact is reflected in the relationship between agent behaviour and its environment comprising external factors—steered by the modeller by means of exogenous parameters.

## 2.2. Empirical Foundations

While theoretical models need to be less concerned with methods for initialising the simulation with empirical data, practical applications and policy analyses do require such methods [52]. The empirical foundation is one of the crucial aspects in which the proposed simulation model differs from purely theoretical and conceptual models of regional knowledge creation. A thorough empirical foundation is essential for the representation of real-world processes, practical applications, and policy analyses since it increases their integrative strength and liability. The empirical foundations of the model complement the conceptual model as presented in the previous section. In particular, in this model three central elements are driven by empirical data: agent initialisation, calibration of model parameters, and output evaluation. In addition, throughout the model, agents’ decision-making processes are empirically driven by means of statistical figures.

**Agent initialisation using spatial microsimulation.** The empirical agent initialisation focuses on the generation of a representative agent population for each region. Since detailed micro-level data on an organisational level is not available for European regions in a comprehensive way, model agents are created based on region-level empirical data. Agent-level data has been constructed in an elaborate process of drawing samples from empirical distributions of industry sectors, R&D intensity and number of employees (determined from the Eurostat Structural Business Statistics for an initialisation period corresponding to the years 2012 to 2014), while considering feasible combinations of characteristics for each agent based on the characteristic’s empirical correlations using Cholesky decomposition. Cholesky decomposition can be used to create correlations among random variables by decomposing the correlation matrix of empirically observed correlations between agent characteristics [53].

To generate a representative agent population for each region, we employ spatial microsimulation techniques. Spatial microsimulation is a method to allocate individuals (organisations) to zones (regions), by combining individual (organisation-level) and geographically aggregated data [54]. Here, we opt for Iterative Proportional Fitting (IPF) as a statistical technique for combining individual and geographical data to allocate the primarily specified agents to European NUTS-2 regions using reweighting algorithms, resulting in maximum likelihood values for each zone-individual combination represented in a weight matrix (see e.g., [54] for details on spatial microsimulation and IPF). Since the overall aim of this model is the simulation of knowledge creation, a special focus lies on the initialisation of the agents’ knowledge profiles. Each agent is endowed with a unique set of technological fields—empirically represented by patent classes—representing their knowledge profile. The patenting records are extracted from the PATSTAT database, the Worldwide Patent Statistical Database by the

European Patent Office, which is the most important data source for scientific research on patent activities and patent data. We use patent classes on a three-digit subclass level (e.g., A61K) as specified by the International Patent Classification (IPC) that are assigned to the agents based on their industry sector (NACE classification) using the table of concordance proposed by [55]. In total, over 21,000 agents are included in the model, which is a fraction of 1000 of the actual number of local firm entities located in the NUTS-2 regions of interest (based on the Eurostat Structural Business Statistics).

**Calibration of model parameters.** The calibration process aims at finding values for the input parameters that make the model reproduce patterns observed in reality sufficiently well [56]. Parameter fitting must span the entire set of parameters, which rapidly increases the number of possible parameter combinations to be tested. To reduce the dimension of parameter combinations that have to be tested, we employ Latin hypercube sampling, which is a technique that considers the entire set of parameters to get the most representative subset of the space in a relatively efficient (and computationally saving) manner by means of uniform sampling of the scenario space given a certain parameter space and with a limit of a specified number of experiments [57].

The core of the empirical calibration is the fitting of model parameters in a way that the resulting output variables fit best; here, they lie within a range of the selected empirical measures. Thiele et al. [56] point out two different strategies for fitting model parameters to observational data: (i) best-fit and (ii) categorical calibration. Whereas, best-fit calibration aims at finding the parameter combination that best fits the observational data (i.e., there exists one exact value as a quality measure to evaluate the fit of the parameter values), using categorical calibration, not a single value is obtained, but a range of plausible values is defined for each calibration criterion. As proposed by Thiele et al. [56] a hybrid approach by transforming the categorical criteria to a best-fit criterion is followed here. This is done by means of conditional equations and the specification of a cost function, evaluating the cost for a parameter value of not being in the acceptable value range (which is defined externally)

$$criterion_r(x_r) = \begin{cases} 0 & \text{if } x_{min} \leq x_r \leq x_{max} \\ \left( \frac{mean(x_{min}, x_{max}) - x_r}{mean(x_{min}, x_{max})} \right)^2 & \text{else} \end{cases} \quad (4)$$

$$cost(x_r) = \sum_{r=1}^R criterion_r(x_r), r = 1, \dots, R \quad (5)$$

where  $x_r$  are corresponding simulation results of criterium  $r$ ,  $x_{min}$  and  $x_{max}$  denote the respective minimum and maximum value, and  $R$  is the total number of calibration criteria included. For each selected empirical measure, an acceptable value range is defined. If the simulated value lies within this interval, no costs incur. If this is not the case, a cost factor based on the squared relative deviation to the mean value of the acceptable range is assigned. The final cost function is the sum of the individual costs of each criterion. Finally, the parameter combination with the lowest cost is chosen as the one that best fits the real-world system. Applying this cost function approach enables combining multiple calibration criteria to one single decision criterion [56].

Empirically, four measures are chosen as criteria for the cost function: (i) the total number of patents in the agent population, (ii) the patenting profile across regions, (iii) the patenting profile across technological fields (as defined by Schmoch [58]), and (iv) the regions' degree centralities, i.e., number of collaboration partners in the collaboration network. The empirical reference datasets are the patent data on European regions, as well as—for the centrality measure—data on collaborative research projects in EU framework programmes that are widely used to proxy inter-regional R&D collaboration in Europe (see e.g., [59,60]). The empirical measures are calculated as aggregate values over the years 2014 to 2018; the calibration is performed with the simulated output after 60 time steps (12 time steps representing one year; see Appendix A for calibrated system parameters). The calibrated parameter set defines the so-called *baseline scenario*, as a reference for the simulations presented in the results section.

**Output evaluation.** Successful research efforts by agents result in knowledge gains (see Appendix A for details), which furthermore can result in patents. In the simulation model, we use the number of patents as a proxy variable to capture knowledge outputs and to establish a link between a rather generic knowledge gain—as a pure result of learning processes in the model—and patents. Patents are considered a suitable indicator to measure the ability to create commercially relevant new knowledge; specifically, as output of industrial innovation efforts in firms (see e.g., [61,62]). This allows for an interpretation closer to empirical observations. Whether or not a patent emerges from a knowledge gain follows an independent evaluation criterium. To that, we implement an empirical output filter, by means of econometrically estimated coefficients, that determines the patenting propensity of an individual agent based on a region-specific probability that itself depends on regional characteristics (human resources, GRP per capita, R&D per capita, degree centrality). Due to the lack of organisation-level data, the regression model is estimated on the regional level. We estimate a Poisson regression model to account for the true integer nature and the distributional assumptions of the number of patents as the dependent variable (see Appendix A for details).

### 3. Results

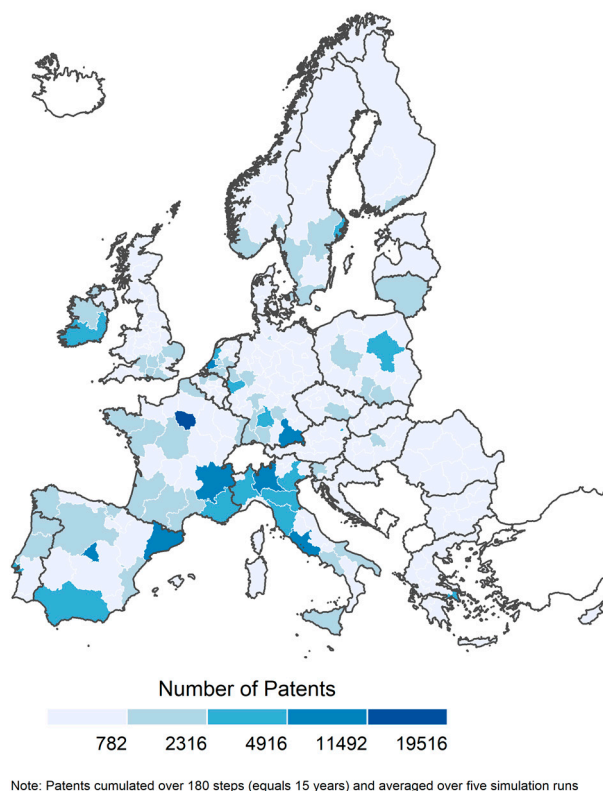
In this section, we demonstrate the potential of the simulation model by means of three small example applications derived from current scientific debates. We evaluate the knowledge creation exclusively from an aggregate regional perspective, since the model itself is designed to be representative on a regional level (e.g., representativity of agent population). Nevertheless, agent-level processes are reflected in the regional knowledge creation output via local and structural dynamics. On a local level, knowledge is transferred between agents through collaborative knowledge creation processes, which subsequently results in an update of the agents' knowledge profiles. On a network structural level, changes in the agents' knowledge status as well as knowledge creation performance affect their network embeddedness and the global network structure as a whole. To ensure empirical interpretability and allow for empirical calibration, a link between the model's knowledge gains, and patenting, as an empirical knowledge output, is established (see Section 2.2 on output evaluation). Note that the simulation results presented in this section are averages over five model runs to ensure robust findings and hence, limit the possibility of artefacts occurring by variability in the results (see Appendix B for robustness checks). The three small example applications for demonstrating the potential of the model are on regional concentration patterns, regional specialisation dynamics, and networks as drivers for regional knowledge creation, all of them intensively debated in the geography of innovation literature.

#### (i) *Spatial distribution and concentration of knowledge creation*

The first application aims at the phenomenon of spatial concentration of knowledge creation. To a large extent, knowledge creation is driven by geographically localised knowledge flows, especially in the case of learning processes that are driven by tacit and region-specific knowledge elements. This highlights the facilitative role of spatial proximity for knowledge creation [3,63–65]; however, recent findings also suggest a decreasing effect of distance [59,66]. As described in the previous section, we use patents as an empirically-based measure of the model's knowledge output. Figure 3 illustrates the spatial distribution of the simulated patents as main model outputs resulting from the agents' individual knowledge creation processes, aggregated to a regional level.

It can be seen that some typical regions, such as e.g., Île-de-France (FR), Madrid (ES), Catalunya (ES), Oberbayern (DE), Rhône-Alpes (FR), and Northern regions of Italy such as Lombardia, clearly stand out in terms of their patent outcome, whereas, the majority of regions exhibit only a fair number of patents; we can, however, observe (almost) no distinct spatial clusters of multiple regions showing high patenting activity (except Northern Italy and South-East France). In terms of demonstrating the potential of the model, these results are quite promising. Clearly, the model and the implemented processes (see Sections 2 and 3) are able to approximate the empirically observed spatial distribution of knowledge (see [65]).





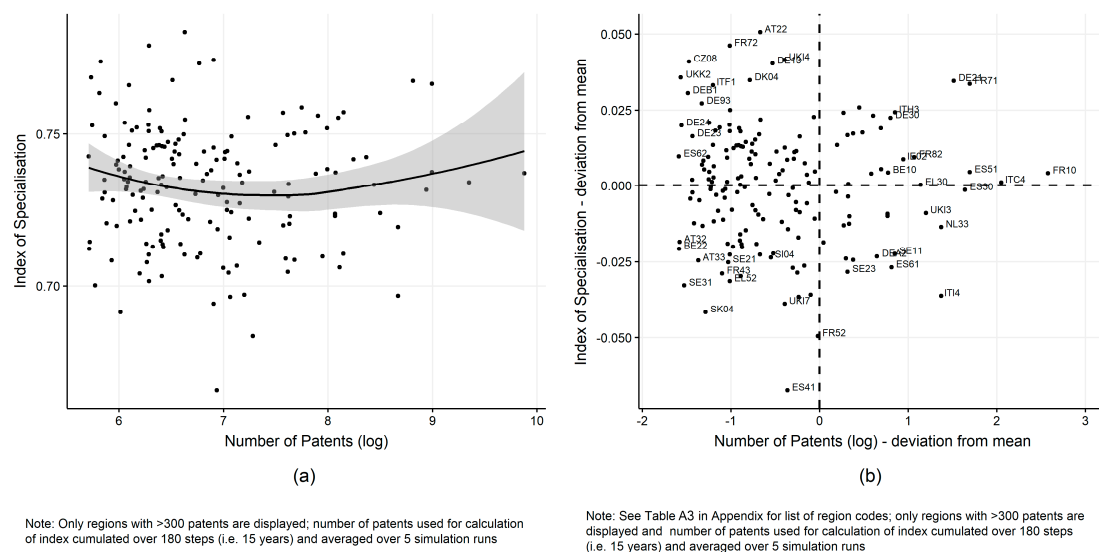
**Figure 3.** Spatial distribution of patents (natural breaks).

(ii) *Specialisation of regional knowledge creation*

The second application focuses on the debate of the relative importance of sectoral specialisation versus diversification for a region's knowledge created [67,68]. This dichotomy is rooted in the two concepts of localisation and diversity economies, as originally put forward by Marshall [69] and Jacobs [70], respectively. In this example, we use not only the total counts of simulated patents by region as in the previous example, but also their technological field, to calculate the degree of technological specialisation of regions based on the simulated patents. We use the *Index of Specialisation* to assess the degree of specialisation of each region relatively to the other regions (see Appendix C for the index definition). The spatial distribution of regional technological specialisation is given in Figure A8 (see Appendix C), while the relation between technological specialisation and simulated knowledge output is illustrated in Figure 4. The relationship between the degree of specialisation and number of patents—as shown in Figure 4a—shows no significant correlation. Hence, there is no direct link between the regions' sectoral specialisation and their respective knowledge output in the model.

On the one hand, according to the concept of localisation economies, a high degree of sectoral specialisation of regions points towards considerable advantages of these regions due to economies of scope when making use of local and specialised R&D infrastructure and local and dense R&D networks that facilitate the exchange of knowledge at relatively low costs. Looking at Figure 4b, which displays the centralised number of patents and specialisation indices (i.e., deviation from the respective mean values), this applies in the model simulation to the regions Rhône-Alpes (FR71) and Oberbayern (DE21) that exhibit relatively high technological specialisation and knowledge output. This may signal importance of sectoral specialisation to gain a higher output—a finding in the vein of Marshall [69,71]. On the other hand, the regions Lazio (IT14) and Andalucía (ES61), for example, suggest that a relatively low degree of specialisation (i.e., diversification) and knowledge output are positively related—supporting diversity economies as put forward by Jacobs [70]. Hence, although there is no clear relationship between knowledge output and degree of technological specialisation of regions,

we find specialist regions with industrial districts or sector-specific spillovers, and generalists that benefit from industry diversification, among the leading regions in terms of knowledge creation.



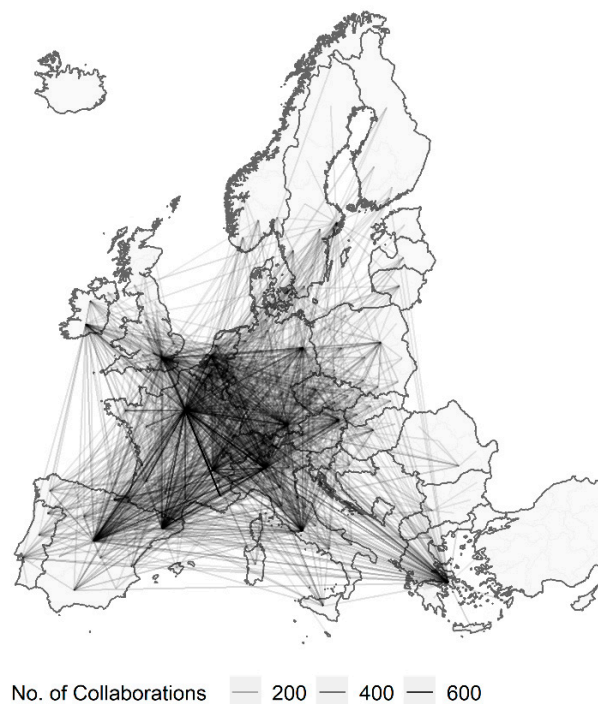
**Figure 4.** Relationship between specialisation and knowledge output (patents).

(iii) *Networks as drivers for regional knowledge creation*

In the third application, attention is shifted towards networks—viewed as inter-organisational arrangements in R&D—that are widely considered as essential for increasing a region’s knowledge creation capability (e.g., [8,9,37,72]). In this study, the network under consideration is defined as a regional knowledge network comprising a set of regions as nodes, inter-linked via edges that represent the knowledge flows resulting from collaborative R&D efforts [46,73]. Figure 5 displays the simulated regional knowledge network showing the collaboration links between the agents (when following the collaborative mode of knowledge creation) aggregated to a regional level. As for the spatial distribution in the first example, the model is able to re-create observed spatial network patterns in the literature using project collaboration with the European Framework Programmes (see. e.g., [74]). The European regions seem quite strongly engaged in research collaborations; however, only a few network hubs (in terms of their number of network partners) stand out: first, Île-de-France (FR), leading to the characteristic star-shaped network formation that is also known from empirical studies [46], followed by Oberbayern (DE), Madrid (ES), Lombardia (IT), and Rhône-Alpes (FR). These regions also exhibit the highest knowledge output in the model, suggesting a positive relationship between a region’s network connectivity and knowledge output (as measured by patents).

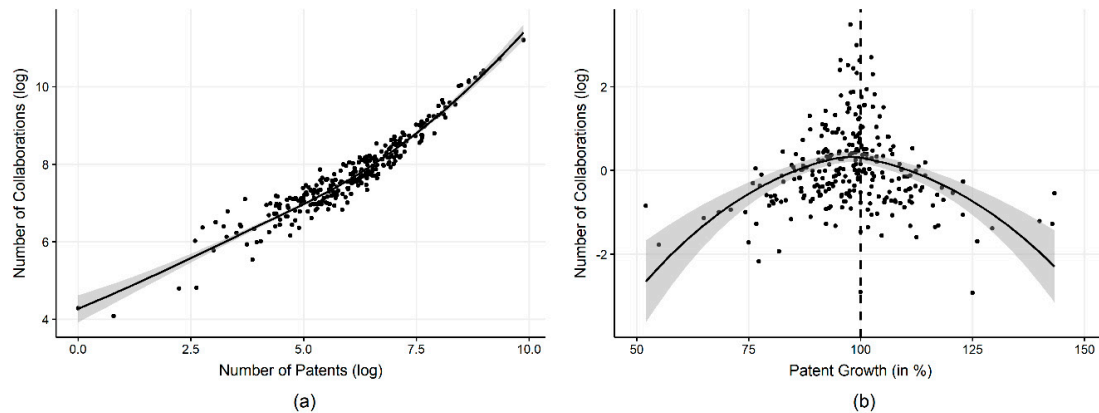
In Figure 6, we further reflect on the relation between the regions’ numbers of simulated collaborations and knowledge output. Figure 6a shows a positive and slightly exponential relationship between the number of collaborations and quantity of knowledge output (as measured by the number of patents), implying that the more collaborations the agents—located within the regions—have, the higher are the regions’ knowledge outputs. Note that these findings only reflect the quantity of network links (number of collaborations of regions), not their quality, which would identify certain regions as hubs with authoritative positions in the collaboration network.

However, looking at the relationship between the number of collaborations and patent growth in Figure 6b, we cannot observe that a high number of collaborations also coincides with high patent growth. Hence, a high number of inter-national collaborations is not a driving force for high patent growth (i.e., growth in knowledge output) in the model. Evidently, regions starting from a relatively high level of patent output exhibit lower rates of patent growth.



Notes: Collaborations cumulated over 180 steps (i.e. 15 years); only one exemplary simulation run, and only top 98% of links (w.r.t. number of collaborations) are displayed

Figure 5. European inter-regional R&D collaboration network.



Notes: Patents cumulated over 180 steps (i.e. 15 years) and averaged over five simulation runs; average number of collaborations over five simulation runs and 180 steps

Notes: Patent growth (in %) determined as  $(p_t - p_{t-1}) \cdot 100 / p_{t-1}$ , where  $p_t$  denotes cumulated patents over steps 36-108, and  $p_{t-1}$  denotes cumulated patents over steps 36-180; average number of collaborations over five simulation runs and 180 steps

Figure 6. Relationship between number of collaborations and knowledge output (patents).

#### 4. Discussion and Concluding Remarks

In this study, we introduce an empirical agent-based model of multi-regional knowledge creation and demonstrate its potential for applications to current research issues intensively debated in the geography of innovation literature. By employing an agent-based simulation approach, we intend to complement the prevailing research tradition of econometrically modelling regional knowledge creation, focusing on regional characteristics and determinants of knowledge creation and diffusion. ABM offers several benefits compared to these conventional modelling techniques, which allow for new perspectives and insights in the process of regional knowledge creation. In particular, agent heterogeneity, underlying micro-structures of the regions, and network dynamics as an interplay between region-internal and region-external interdependencies in the form of R&D collaboration

linkages, can be explicitly considered. Moreover, the possibility to conduct simulation scenarios allows the direct comparison of system behaviour within a controlled environment.

However, the ABM approach is, up to now, rarely used in the context of geography of innovation. In our understanding, this is to a large extent due to the lack of credibility and lack of empirical closeness and hence, lack of applicability to real-world questions. We react to this by drawing on large-scale data sets and applying state-of-the-art methods to empirically initialise and calibrate the simulation model. Moreover, we support the ABM by additionally employing well-established econometric tools and concepts of network science; and, by this, using 'the best of each world', which we believe changes the perception of methodological critics regarding simulation models being a black-box.

The example applications of the model show quite promising results in terms of robustness and empirical approximation, speaking for the representativity of the model. We are able to show spatial concentration of knowledge creation, illustrate the mechanisms of the ambiguity of the effect of sectoral specialisation versus diversification on knowledge created, as well as to confirm the driving role of networks for regional knowledge creation. The replication of real-world phenomena, supported by empirical findings in related studies, is an essential step of the model validation. Hence, we conclude that the proposed simulation model indeed shows potential to advance the scientific debate in the field of geography of innovation in future applications making use of simulation experiments.

In addition to contributing to the scientific debate on regional knowledge creation, the proposed simulation model is also of high relevance in the field of research, technology and innovation (RTI) policy. Taking into account technological, institutional, as well as geographical aspects of knowledge creation in the model allows for simulation experiments referring to RTI policy measures at the European, the national as well as the regional levels. Such policy interventions may, for instance, refer to regional specialisation policies, the coordination of regional policies, increased incentives to engage in R&D collaborations or to mission-oriented public funding of specific thematic areas. One particular field of application, in this respect, is the use of the ABM for ex-ante impact assessment of policy interventions, such as public R&D programs. In particular, the evolutionary and forward-looking perspective of ABMs considers the openness of socio-technical development, and the micro-perspective on agent systems may help to understand the complexity of public policy interventions.

For both scientific research issues as well as policy aspects, the current model is sufficiently flexible to be easily tailored to new research issues of interest, while relying on the robustness of the core model elements and processes. Admittedly, the proposed model also has its limitations that the modeller has to be aware of: *First*, the knowledge creation process is tailored to Europe as a geographical entity. On the one hand, this is the case in terms of the data used for the empirical initialisation and calibration. On the other hand, this also applies implicitly with respect to the model elements and processes since the model conception is driven by the European spirit of performing R&D within the European Framework Programmes that connect regions all over Europe via collaborative, publicly funded research projects. It remains to be examined if the model is also suited for other regional, national and supra-national innovation systems, such as China or the US. *Second*, the aim of the model is to simulate regional knowledge creation in sufficient detail. We deliberately exclude any considerations on the valuation of the newly created knowledge, and its measurement. Hence, one has to be aware that respective statements cannot be made. Nevertheless, to have some kind of approximation, we include the distinction between simple knowledge gains, and patents that can result in a knowledge gain in the model (based on econometrically estimated probabilities using empirical information on region characteristics). *Third*, regarding the representativity of the model, it is explicitly adapted to a regional level. Although the agents are modelled at an organisational level and hence, also their knowledge creation and learning processes, the model's initialisation and calibration are targeted at model results that are representative for regions. This entails that we need to refrain from any analyses at the agent level, such as observing a single agent's behaviour, to ensure credibility of the results.

In this study, we demonstrated the potential of the proposed simulation model with first application examples. However, many possibilities and model features have not been exploited so

far (e.g., knowledge gain, learning processes), allowing for many future applications and simulation experiments. Specifically, characterising the influence of inter-regional R&D collaboration on regional knowledge creation, disentangling local effects from global network effects on regional knowledge creation, and the analysis of technological specialisation and geographical concentration tendencies, come to mind. In particular, scenario analyses referring to specific RTI policy measures are of interest to shed light on the mechanism of policy interventions at the European, the national and the regional level. Moreover, the application of the model to other geographical areas, such as China, is of great interest to gain an exceptional comparative perspective of regional knowledge creation in innovation systems showing different development paths, different overall socio-economic characteristics and conditions, different approaches in policy making and societal systems as a whole.

**Funding:** This research was funded by the Austrian Science Fund (FWF), grant numbers P28936-G27 and P32227-G27, and supported by RISIS2 (Research Infrastructure for and Innovation Policy Studies 2), funded by the European Union’s Horizon 2020 Research and innovation programme under the grant number n°824091.

**Acknowledgments:** The author particularly thanks two anonymous reviewers for their helpful suggestions, as well as Astrid Unger for technical support and Thomas Scherngell for valuable and thoughtful comments that helped to improve the quality of the paper. Open Access Funding by the Austrian Science Fund (FWF).

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A. Technical Appendix—Glossary of Model Elements and Processes

### In Alphabetical Order

**Collaboration memory.** The collaboration memory is specified as a vector of length  $s$  (steered by external model parameters) containing entry pairs of the last  $s$  former collaboration partners  $a_i$ , with respective probability value  $\gamma_s$  representing how successful the past collaboration has been.

$$\Gamma = \left\{ \left( \begin{array}{c} a_i \\ \gamma_1 \end{array} \right), \dots, \left( \begin{array}{c} a_i \\ \gamma_s \end{array} \right) \right\} \quad \text{for } i \in N \quad (\text{A1})$$

where  $N$  is the set of total agents. The collaboration memory vector is renewed, in a way that new collaboration partners are ranked first in the vector, while the partner ranked last (i.e., the one longest in the collaboration memory) drops out. To determine the degree of success of a collaboration, the share of the actual collaborative knowledge gain over the whole research path with respect to the maximum possible knowledge gain is evaluated; this share is interpreted as probability for a repeated collaboration. Within the process of *partner choice*, a random entry pair is selected, and the respective probability evaluated. Whereas, a positive return leads to a repeated collaboration between the two agents, a negative return initiates the remaining process of partner choice.

**Knowledge creation process.** The knowledge creation process subsumes the total of the agents’ actions in creating new knowledge, including the setting of the *knowledge creation mode*, setting of the *research path*, selection of the *research strategy*, *learning process*, and *research success evaluation* (see also Figure 2 in main text for simplified process diagram). In Figure A1, the agent’s knowledge creation process is illustrated in more detail.

**Knowledge endowment/profile.** The knowledge endowment of agent  $i$  represents the knowledge profile of the agent, i.e., indicates the technology classes the agent is active in, as well as the expertise in the respective class. Hence, the knowledge endowment can be defined as a vector of length  $k$  (with  $k$  being the number of technology classes in the *technology space*)

$$\kappa_i = \{\kappa_{i1}, \kappa_{i2}, \dots, \kappa_{ik}\} \quad (\text{A2})$$

where the value of  $\kappa_{ik}$  determines the expertise in the respective technology class  $k$  (level of knowledge). Combining all vectors of the agents’ knowledge endowments results in the *knowledge space*.

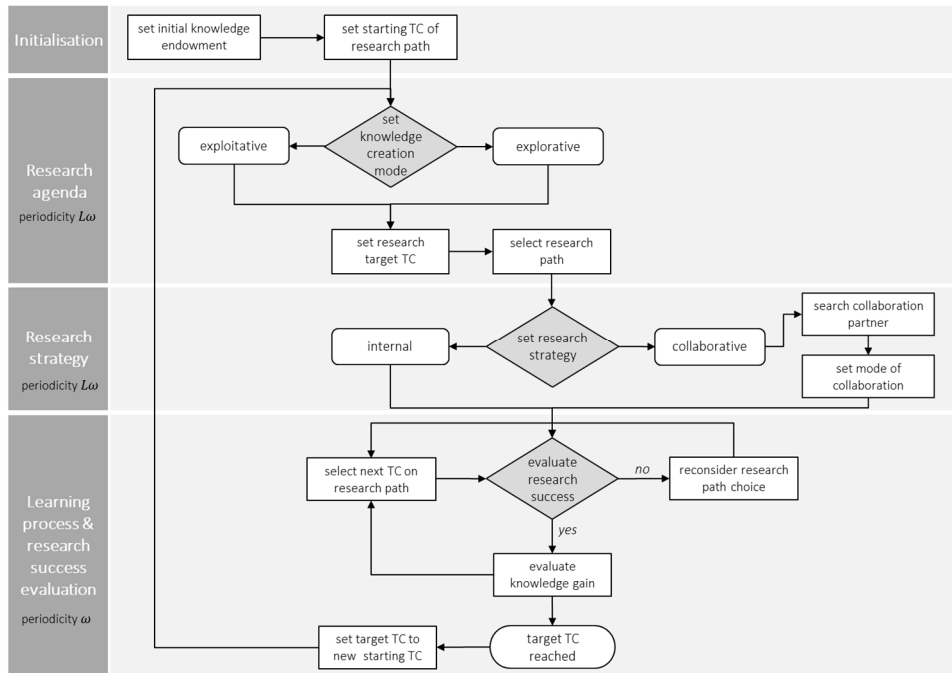


Figure A1. Detailed illustration of agent’s knowledge creation process.

**Knowledge gain.** In the case of successful research, the agent’s knowledge gain is evaluated, i.e., the level of expertise in the subsequent technology class (TC) is updated. In the case of *internal research*, the learning outcome  $\psi_t$  in time step  $t$  can be written as

$$\psi_t = \kappa_{il,t} * (1 - p_{\tau_l, \tau_{l+1}}) \tag{A3}$$

where  $p_{\tau_l, \tau_{l+1}}$  denotes the proximity between two consecutive TCs on the research path in the technology space (indicating the similarity of the two TCs); this results in an increase in the expertise level  $\kappa_{ik+1}$  of

$$\kappa_{il+1,t+1} = \kappa_{il+1,t} + \psi_t \tag{A4}$$

An example of an update in the agent’s expertise level is given in Figure A2. Departing from technology class A with a current expertise of 1.5, the knowledge gain with a transition to TC B on the learning path—bridging a distance of 0.8 ( $1 - p_{A,B}$ ) between the two TCs—amounts to 1.2.

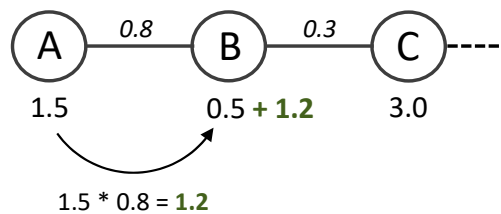


Figure A2. Example of expertise level update.

In the case of *collaborative research*, there is an additional *knowledge transfer* between the collaborating agents. Knowledge transfer in the collaborative knowledge creation process is based on the cognitive distance (knowledge distance)  $d_{ijk}$  between the two collaborating agents

$$d_{ij} = \sqrt{\sum (\kappa_{ik} - \kappa_{jk})^2} \tag{A5}$$

To determine the increase in expertise for the technology class  $k$  of interest, the distance of the levels between the collaborating agents is determined; however, the agent only gains from the partner's knowledge if the partner's expertise is higher than its own. This is represented by

$$d_{ijk} = \max[(\kappa_{jk} - \kappa_{ik}), 0] \quad (\text{A6})$$

The actual amount learnt depends on the absorptive capacity—representing the trade-off between novelty value and understandability of new knowledge. Hence, it is assumed that the amount of knowledge gain corresponds to an inverted u-shaped relationship. Specifically, this relationship is used to scale  $d_{ijk}$

$$\epsilon_k = -\frac{1}{\delta}(d_{ijk} - \delta)^2 + \delta \quad (\text{A7})$$

where  $\delta$  denotes the optimal learning distance (which is specified by an external model parameter). As in the case of knowledge gain in internal research, this scaled expertise level distance is set in relation to the technological distance that is overcome from one to the next technology class on the research path

$$\varphi_t = \epsilon_k * (1 - p_{\tau_l, \tau_{l+1}}) \quad (\text{A8})$$

In total, the knowledge gain in each technology class when performing collaborative research is specified as an additive function comprising knowledge gains from the *internal research*  $\psi_t$  the agent is performing regardless of the collaboration partner, and *collaborative research*. This can be written as

$$\Delta E_t = \psi_t + \varphi_t \quad (\text{A9})$$

**Knowledge space.** Combining all vectors of the agents' knowledge endowments results in the knowledge space. With  $N$  being the total number of agents in the model, that can be interpreted as the total portfolio and amount of knowledge available in the model

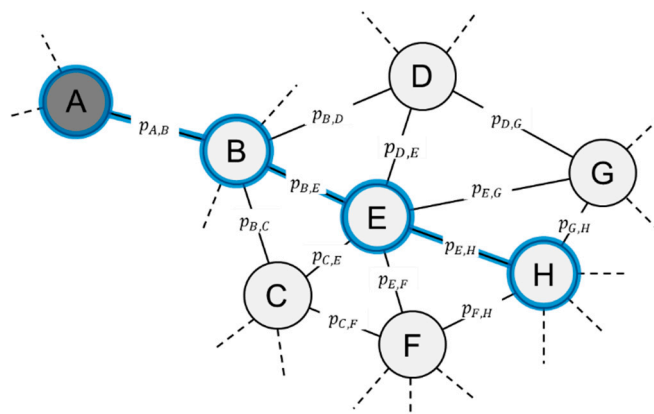
$$K = \begin{pmatrix} \vec{\kappa}_1 \\ \vdots \\ \vec{\kappa}_N \end{pmatrix} \quad (\text{A10})$$

**Learning process.** The core of the knowledge creation process is the learning mechanism. It is designed as a sequenced process along a specified *research path*, that each agent decides upon individually according to its *mode of knowledge creation* (exploitative or explorative) and with respect to its current knowledge and research target. The basis for knowledge creation along a research path is the concept of *technology space*. It serves as a framework for the agents to gain new knowledge as they move along their research paths comprising *technology classes*. Each research path  $P$  is a subset of the technology space, indicating the way and direction of learning (following the concept of a 'path' known from social network analysis).

$$P = \{\tau_1, \tau_2, \dots, \tau_L\} \in T = \{\tau_1, \tau_2, \dots, \tau_k\} \quad (\text{A11})$$

where  $l = 1, 2, \dots, L$  with  $L$  being the length of the research path. Assuming technology class  $A$  being the present knowledge that is built upon, the agent moves along its research path (e.g.,  $\{A, B, E, H\}$ ) as illustrated in Figure A3).

Between each transition to the next technology class, the research is evaluated to be successful or not. Only in the case of successful research the new technology is acquired; otherwise the agent tries again or eventually chooses an alternate path. How this path is chosen depends on the mode of knowledge creation.



**Figure A3.** Example of research path in a subset of the technology space.

**Mode of collaboration.** There are two different modes of collaboration (the ratio is steered by an external parameter): (i) *research-mode* and (ii) *service-mode*. Technically, following the research-mode, both partners follow the research path as determined by the partner actively looking for a collaboration partner, assuming he is the consortium leader and hence specifies the research direction. However, since the partner has (by definition of partner search) expertise on at least one technology class on that research path, the agent is able to learn from the partner in these classes; in a sense, he receives a certain amount of the partner's expertise—additionally to his own research results (the partner's level of expertise is not reduced by this). By how much the level of expertise is increased (how much is learnt in each technology class) depends on the cognitive distance between the two collaborators, as well as the absorptive capacity (see '*knowledge gain*').

Alternatively, following the service-mode, it is evaluated whether the collaboration partner has a higher expertise in one of the technology classes on the selected research path; if so, the starting TC of the research path to reach the research target is changed to the closest TC to the target TC (but not the target TC itself), where the partner has higher expertise. Hence, the research target is probably reached faster; however, there is potentially less knowledge gain since there are fewer possibilities to gain knowledge in the particular technology classes. In the case that the partner has no higher expertise level in any of the technology classes on the research path, its collaborative strategy is shifted to the research-mode. Although the agent is still not able to learn directly from the partner (less expertise in all relevant TCs), the success of the research is (most likely) positively influenced by the collaboration partner (see '*research success evaluation*').

**Mode of knowledge creation.** In selecting their research strategy, the agents can choose between exploitative and explorative knowledge creation, where exploitative knowledge creation reflects a direct way, and exploration a more indirect way of creating knowledge. Dependent on the research strategy, the agents' set their research path  $P$ . In the case of exploitation, having set the starting point of the research path in the technology space, the agent chooses a research target (a target TC). The selection of the target TC takes place according to a decaying probability function based on the technological distance between the classes in the technology space, such that closer TCs exhibit a higher probability to be chosen as a target. Hence, the distance between the starting TC and target TC indicates the degree of radicality  $\lambda$  of the agent's research endeavour. Next, the agent identifies the set of shortest paths and can either select the shortest weighted path, or one of the shortest paths with respect to the number of TCs on the path. In the case of exploration, the research path is determined by subsequently choosing the next most proximate TC (originating from the agent's current TC), where the length of the path is determined randomly, representing the equivalent of a researching period between one and five years (set by external parameter). The last TC of the research path is specified as the designated target TC.

**Output evaluation.** Whether or not a patent emerges from a knowledge gain is determined by means of econometrically estimated coefficients influencing the patenting propensity of an



individual agent by means of a region-specific probability that is determined by regional characteristics, i.e., human resources, GRP per capita, R&D expenditures per capita, and degree of centrality. Note that only fully accomplished research paths are subject to the evaluation for a patent, where each technology class on the path represents a patent class (analogously to patent documents issued by, for example, the European Patent Office). We estimate a standard Poisson regression model (see [75]) to account for the true integer nature and the distributional assumptions of the number of patents as the dependent variable. The estimated parameters are used to compute the region's predicted empirical probabilities (to receive at least one patent) by means of

$$\hat{\Pr}(y > 0|x) = 1 - \hat{\Pr}(y < 1|x) = 1 - \frac{e^{-x\hat{\beta}}(x\hat{\beta})^m}{m!} \quad \text{with } m = 0 \quad (\text{A12})$$

**Parameters.** The model comprises external system parameters that are not empirically based. A first set of parameters (Table A1) is fixed to values determined by the calibration process and, by this, specifies a baseline scenario.

**Table A1.** Calibrated external system parameters.

Parameter	Description	Type	Calibrated Value
<i>BasePatentProb</i>	Scaling parameter for patent probability (originally estimated econometrically)	$[0, 1] \in \mathbb{R}_0$	1.0
<i>CollabInternalProb</i>	Share of agents performing collaborative research (vs. internal research)	$[0, 1] \in \mathbb{R}_0$	0.5
<i>CollabMemorySize</i>	Length of collaboration memory vector (determines number of former collaboration partners being remembered)	$[1, 10] \in \mathbb{N}$	9
<i>CollabModeProb</i>	Share of agents with service-oriented mode of collaboration (vs. research-mode)	$[0, 1] \in \mathbb{R}_0$	0.8
<i>ResearchStrategyProb</i>	Share of agents with exploitative mode of knowledge creation (vs. explorative)	$[0, 1] \in \mathbb{R}_0$	0.3

Additionally, a second set of external system parameters (Table A2) is purely specified by means of user input (i.e., they are not calibrated and not empirically based).

**Table A2.** Non-calibrated external system parameters.

Parameter	Description	Type	Initialisation Value
<i>ExplorativePathLength</i>	Indicates the maximum length of research project for explorative mode of knowledge creation (1 step = 1 month)	$\text{unif}(0, x), x \in \mathbb{R}_+$	5
<i>Delta</i> [ $\delta$ ]	Determines how much is learnt from the partner in collaborative research (optimal learning distance)	$\mathbb{R}_+$	1
<i>Lambda</i> [ $\lambda$ ]	Determines the degree of radicality in search for a research target technology class (in exploitative research)	$[0, 1] \in \mathbb{R}_0$	1.0

**Partner choice.** A core element of the collaborative research process is the agent's choice of a suitable collaboration partner. Ahead of the general process of partner choice, a collaboration memory

(see ‘*collaboration memory*’ for details) serves to account for re-occurring collaborations with partners of previously successful joint research projects. The general partner choice is organised along three main steps covering the spatial, sectoral, and cognitive dimension to guarantee the best possible real-world collaboration behaviour of agents. For the *spatial dimension*, i.e., the choice of an empirical probable region to look for a suitable partner agent, we draw upon estimated probabilities for a collaboration taking place between two agents located in two regions. Therefore, we estimate a Spatial Interaction Model (SIM; [76] for details) using data from the EUPRO database (see [risis2.eu](http://risis2.eu) for details and access) comprising systematic information on collaborative research projects in EU framework programmes and explicitly take the geographical distance between the regions, and variables indicating a neighbouring region and country, into account in the model. Thus, we receive individual collaboration probabilities for all combinations of regions with respect to their geographical relations. With respect to the *sectoral dimension*, to determine the sector where the partner search is carried out, the proximity between the sectoral classes (based on co-occurrences of IPC patent classes attributed to each NACE class in Dorner and Harhoff [55]) is used to identify a suitable sector; closer sectors exhibit higher probabilities to be chosen. Once, empirically, a suitable region and sector for the partner choice has been identified, the agent looks for a cognitively proximate collaboration partner, i.e., a partner having expertise in one of the technology classes that is on its research path.

**Research path.** A research path  $P$  is a subset of the technology space, indicating the way and direction of learning (following the concept of a ‘path’ known from Social Network Analysis).

$$P = \{\tau_1, \tau_2, \dots, \tau_L\} \in T = \{\tau_1, \tau_2, \dots, \tau_k\} \quad (\text{A13})$$

where  $l = 1, 2, \dots, L$  with  $L$  being the length of the research path and  $T$  the set of all technology classes (see ‘*learning process*’ and ‘*technology space*’ for details).

**Research strategy.** There are two different research strategies: *internal* and *collaborative* research. Depending on the research strategy, different mechanisms are in place regarding the *learning process*, *knowledge gain* and *research success evaluation* (see respective items in Appendix A for details).

**Research success evaluation.** The evaluation of the research success is a necessary intermediate step for every transition from one technology class to another, i.e., to determine whether a new TC on the path is reached. In the case of *internal research*, the research success is determined by means of a scaled composite indicator (interpretable as success probability) depending on (i) agent-specific characteristics  $c_{im}$  (overall expertise and R&D quota) and (ii) technological proximity between the involved technology classes  $p_{\tau_l, \tau_{l+1}}$ . This can be formalised as

$$P(\text{success}) = \gamma A + (1 - \gamma) p_{\tau_l, \tau_{l+1}} \quad (\text{A14})$$

with  $A = \sum \frac{1}{m} 1(c_{im} > \text{median}(c_{.m}))$  and  $p_{\tau_l, \tau_{l+1}} \in [0, 1]$ , where  $m$  indicates the number of agent-specific characteristics included. In the case of *collaborative research*, the evaluation of the research success is similar to the case of internal research, however, differs regarding the agent-specific characteristics; such that the maximum value of either agent in the collaboration partnership is used for the evaluation of the research success, and hence, increasing the probability of success. Recalling the formula of the research success  $P(\text{success}) = \gamma A + (1 - \gamma) p_{\tau_l, \tau_{l+1}}$ ,  $A$  is now defined as

$$A = \sum \frac{1}{m} 1(\max(c_{im}, c_{jm}) > \text{median}(c_{.m})) \quad (\text{A15})$$

**Technology space and technology class (TC).** The technology space is defined as a network comprising a set of technology classes, with the links indicating the technological proximity between these classes (see Figure A3 for an illustrative example). The network is constructed by extracting patent data from the European Patent Office for the EU-27 countries including the United Kingdom

and Norway from 2012 to 2016 and determining the co-occurrences of IPC patent classes (3-digit) on patent documents. Formally, the technology space  $S$  can be defined as symmetric matrix

$$S = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}_{k \times k} \quad (\text{A16})$$

where  $p_{..}$  denote the Jaccard coefficient as a measure of proximity (derived from the co-occurrences of IPC patent classes), and  $k$  is the number of technology classes considered. In Figure A4, an example of an exemplary subset of the technology space is illustrated. The nodes indicate technology classes and the links represent connectivity between these nodes.

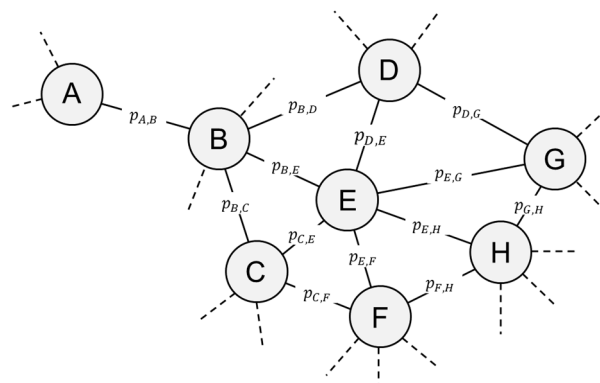
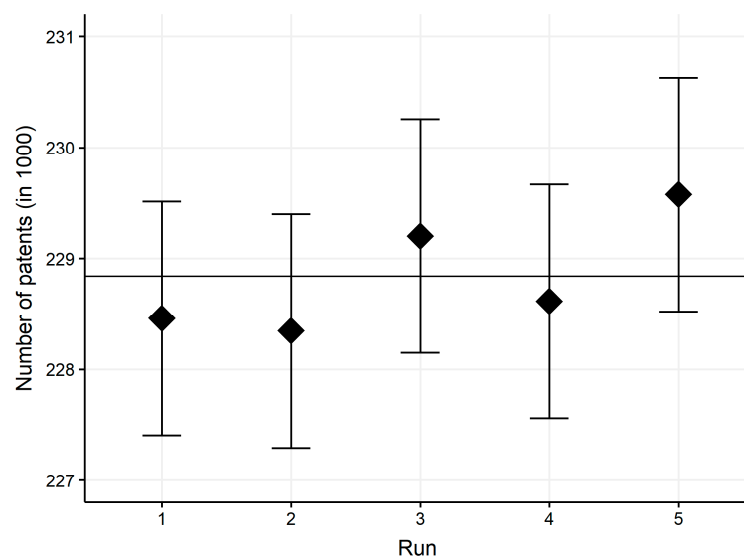


Figure A4. Exemplary subset of the technology space.

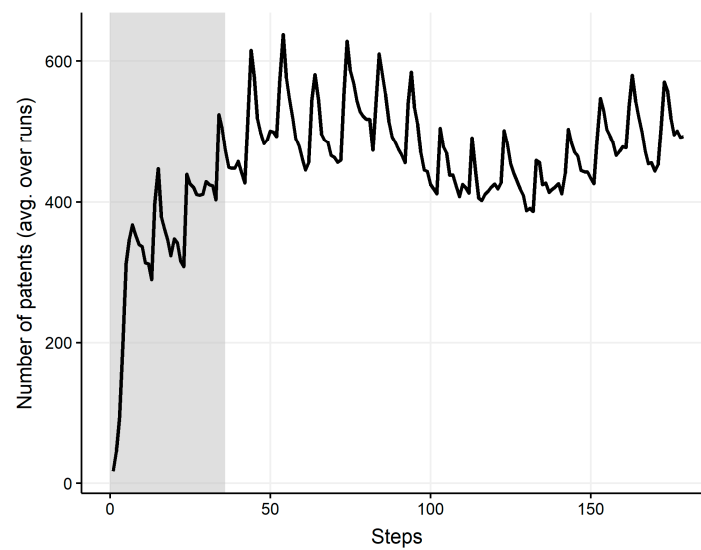
## Appendix B. Robustness Checks

The robustness checks present basic model behaviour over time and model runs to demonstrate the dynamics and robustness of the simulation model with respect to the variability of model outcomes. The results of the robustness checks presented cover 180 steps (equals 15 years) and five simulation runs.



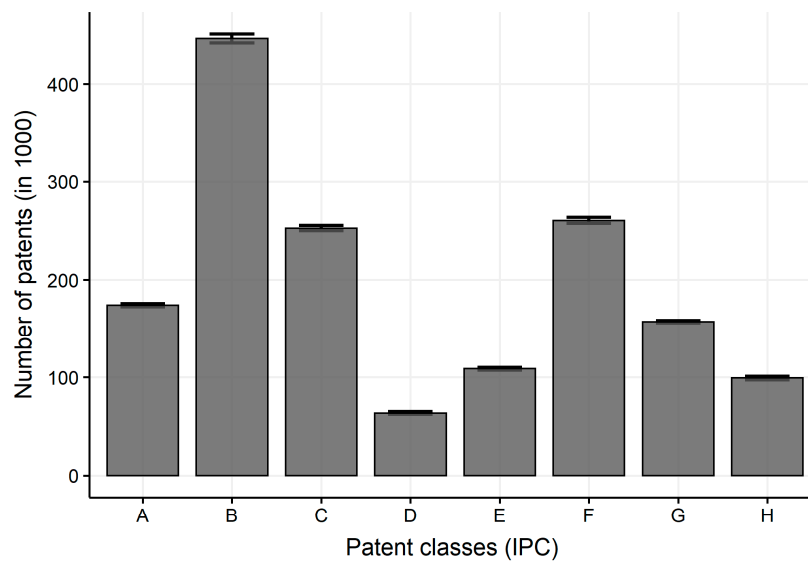
Notes: Error bars indicate an interval of twice the standard deviation; horizontal line denotes the mean value of the runs

Figure A5. Distribution of patents over runs.



Note: The grey area indicates an adjustment phase of 36 steps (equals 3 years)

**Figure A6.** Number of patents per step (average over runs).



Notes: Error bars indicate an interval of twice the standard deviation; patent classes following the International Patent Classification (IPC): A...Human Necessities, B...Performing Operations and Transporting, C...Chemistry and Metallurgy, D...Textiles & Paper, E...Fixed Constructions, F...Mechanical Engineering, Lighting, Heating, Weapons and Blasting, G...Physics, H...Electricity.

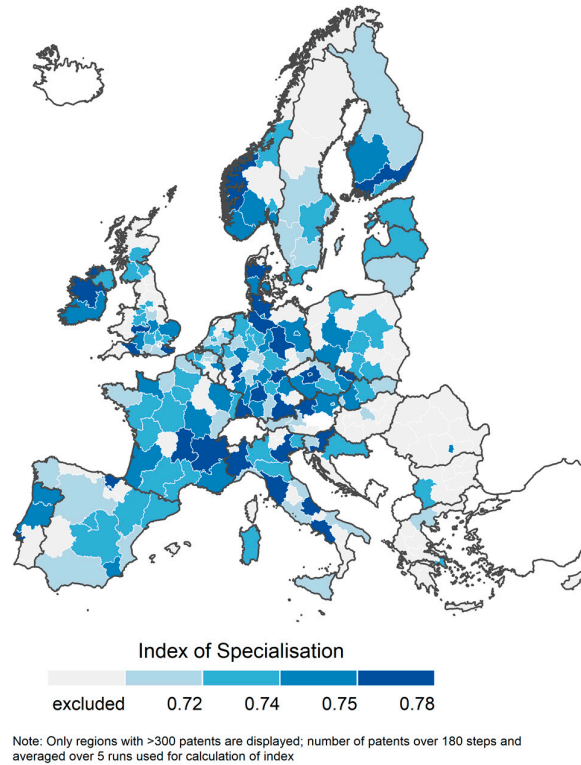
**Figure A7.** Sectoral distribution of patents (average over runs).

### Appendix C. Supplementary Material

The Index of Specialisation assesses the degree of specialisation of each region (relatively to the other regions); however, the index does not indicate in which sectors the regions are specialised. The Index of Specialisation is defined as

$$S_i = \frac{1}{2} \sum_{k=1}^m |y_{ik} - \bar{y}_k| \quad (\text{A17})$$

where  $y_{ik} = x_{ik} / \sum_{k=1}^m x_{ik}$  and  $\bar{y}_k = \sum_{i=1}^n x_{ik} / \sum_{i=1}^n \sum_{k=1}^m x_{ik}$  and  $x$  indicates patents, and  $i$  and  $k$  refer to the region  $i = 1, \dots, n$  and sector  $k = 1, \dots, m$ , respectively. The index ranges from 0 to 1, where 1 indicates full specialisation and 0 implies diversification.



**Figure A8.** Sectoral specialisation of regions (natural breaks).

**Table A3.** Region codes (selected).

NUTS-2 Code	Region Name	NUTS-2 Code	Region Name
AT32	Salzburg	FR43	Franche-Comté
AT33	Tirol	FR52	Bretagne
BE10	Région de Bruxelles-Capitale	FR71	Rhône-Alpes
BE22	Prov. Limburg	FR72	Auvergne
CZ08	Moravskoslezsko	FR82	Provence-Alpes-Côte d'Azur
DE13	Freiburg	IE02	Southern and Eastern Ireland
DE21	Oberbayern	ITC4	Lombardia
DE23	Oberpfalz	ITF1	Abruzzo
DE24	Oberfranken	ITH3	Veneto
DE30	Berlin	ITI4	Friuli-Venezia Giulia
DE93	Lüneburg	NL33	Zuid-Holland
DEA2	Köln	SE11	Stockholm
DEB1	Koblenz	SE21	Småland med öarna
DK04	Midtjylland	SE23	Västsverige
EL30	Αττική	SE31	Norra Mellansverige
EL52	Κεντρική Μακεδονία	SI04	Zahodna Slovenija
ES30	Comunidad de Madrid	SK04	Východné Slovensko
ES41	Castilla y León	UKI3	Inner London—West
ES51	Cataluña	UKI4	Inner London—East
			Outer London—West and North
ES61	Andalucía	UKI7	West
ES62	Región de Murcia	UKK2	Dorset and Somerset
FR10	Île de France		

## References

- Malmberg, A.; Sölvell, Ö.; Zander, I. Spatial clustering, local accumulation of knowledge and firm competitiveness. *Geogr. Ann. Ser. B Hum. Geogr.* **1996**, *78*, 85–97. [[CrossRef](#)]
- Audretsch, D.B.; Feldman, M.P. Knowledge Spillovers and the Geography of Innovation. In *Handbook of Regional and Urban Economics*; Elsevier: Cambridge, UK, 2004; pp. 2713–2739.
- Acs, Z.J.; Anselin, L.; Varga, A. Patents and innovation counts as measures of regional production of new knowledge. *Res. Policy* **2002**, *31*, 1069–1085. [[CrossRef](#)]
- Tödting, F.; Trippel, M. One size fits all? Towards a differentiated regional innovation policy approach. *Res. Policy* **2005**, *34*, 1203–1219.
- Wanzenböck, I.; Scherngell, T.; Brenner, T. Embeddedness of regions in European knowledge networks: A comparative analysis of inter-regional R&D collaborations, co-patents and co-publications. *Ann. Reg. Sci.* **2014**, *53*, 337–368.
- Scherngell, T. *The Geography of Networks and R & D Collaborations*; Springer: Heidelberg/Berlin, Germany; New York, NY, USA, 2013.
- Fischer, M.M.; Varga, A. Spatial knowledge spillovers and university research: Evidence from Austria. *Ann. Reg. Sci.* **2003**, *37*, 303–322. [[CrossRef](#)]
- Rodríguez-Pose, A.; Crescenzi, R. Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe. *Reg. Stud.* **2008**, *42*, 51–67. [[CrossRef](#)]
- Neves, P.C.; Sequeira, T.N. Spillovers in the Production of Knowledge: A meta-regression Analysis. *Res. Policy* **2018**, *47*, 750–767. [[CrossRef](#)]
- Jaffe, A.B. Real effects of academic research. *Am. Econ. Rev.* **1989**, *79*, 957–970.
- Paci, R.; Marrocu, E.; Usai, S. The complementary effects of proximity dimensions on knowledge spillovers. *Spat. Econ. Anal.* **2014**, *9*, 9–30. [[CrossRef](#)]
- Marrocu, E.; Paci, R.; Usai, S. Productivity growth in the old and new Europe: The role of agglomeration externalities. *J. Reg. Sci.* **2013**, *53*, 418–442. [[CrossRef](#)]
- Ó hUallacháin, B.; Leslie, T.F. Rethinking the regional knowledge production function. *J. Econ. Geogr.* **2007**, *7*, 737–752. [[CrossRef](#)]
- Ponds, R.; Oort, F.V.; Frenken, K. Innovation, spillovers and university–industry collaboration: An extended knowledge production function approach. *J. Econ. Geogr.* **2009**, *10*, 231–255. [[CrossRef](#)]
- Greunz, L. Geographically and technologically mediated knowledge spillovers between European regions. *Ann. Reg. Sci.* **2003**, *37*, 657–680. [[CrossRef](#)]
- Moreno, R.; Paci, R.; Usai, S. Geographical and sectoral clusters of innovation in Europe. *Ann. Reg. Sci.* **2005**, *39*, 715–739. [[CrossRef](#)]
- Breschi, S.; Lissoni, F. Knowledge spillovers and local innovation systems: A critical survey. *Ind. Corp. Chang.* **2001**, *10*, 975–1005. [[CrossRef](#)]
- Miguélez, E.; Moreno, R. Research networks and inventors' mobility as drivers of innovation: Evidence from Europe. *Reg. Stud.* **2013**, *47*, 1668–1685. [[CrossRef](#)]
- Breschi, S.; Camilla, L. Co-invention networks and inventive productivity in US cities. *J. Urban Econ.* **2016**, *92*, 66–75. [[CrossRef](#)]
- Maggioni, M.A.; Nosvelli, M.; Uberti, T.E. Space versus networks in the geography of innovation: A European analysis. *Pap. Reg. Sci.* **2007**, *86*, 471–493. [[CrossRef](#)]
- Batty, M. A Generic Framework for Computational Spatial Modelling. In *Agent-Based Models of Geographical Systems*; Springer Science & Business Media: Dordrecht, The Netherlands, 2012; pp. 19–50.
- Crooks, A.; Castle, C.; Batty, M. Key challenges in agent-based modelling for geo-spatial simulation. *Comput. Environ. Urban Syst.* **2008**, *32*, 417–430. [[CrossRef](#)]
- Heppenstall, A.J.; Crooks, A.T.; See, L.M.; Batty, M. *Agent-Based Models of Geographical Systems*; Springer Science & Business Media: Dordrecht, The Netherlands, 2011.
- Ausloos, M.; Dawid, H.; Merlone, U. Spatial interactions in agent-based modeling. In *Complexity and Geographical Economics*; Springer: Cham, Germany, 2015; pp. 353–377.
- Dawid, H. Agent-based models of innovation and technological change. *Handb. Comput. Econ.* **2006**, *2*, 1235–1272.

26. Gilbert, N.; Pyka, A.; Ahrweiler, P. Innovation networks—a simulation approach. *J. Artif. Soc. Soc. Simul.* **2001**, *4*, 1–13.
27. Vermeulen, B.; Pyka, A. The role of network topology and the spatial distribution and structure of knowledge in regional innovation policy: A calibrated agent-based model study. *Comput. Econ.* **2018**, *52*, 773–808. [[CrossRef](#)]
28. Wang, Z.; Yao, Z.; Gu, G.; Hu, F.; Dai, X. Multi-agent-based simulation on technology innovation-diffusion in China. *Pap. Reg. Sci.* **2014**, *93*, 385–408.
29. Beckenbach, F.; Briegel, R.; Daskalakis, M. *Behavioral Foundation and Agent Based Simulation of Regional Innovation Dynamics*; University of Kassel: Kassel, Germany, 2007.
30. Paier, M.; Dünser, M.; Scherngell, T.; Martin, S. Knowledge creation and research policy in science-based industries: An empirical agent-based model. In *Innovation Networks for Regional Development*; Springer: Cham, Switzerland, 2017; pp. 153–183.
31. März, S.; Friedrich-Nishio, M.; Grupp, H. Knowledge transfer in an innovation simulation model. *Technol. Forecast. Soc. Chang.* **2006**, *73*, 138–152. [[CrossRef](#)]
32. Pyka, A.; Kudic, M.; Müller, M. Systemic interventions in regional innovation systems: Entrepreneurship, knowledge accumulation and regional innovation. *Reg. Stud.* **2019**, *53*, 1321–1332. [[CrossRef](#)]
33. Sebestyén, T.; Varga, A. Knowledge networks in regional development: An agent-based model and its application. *Reg. Stud.* **2019**, *53*, 1333–1343. [[CrossRef](#)]
34. Savin, I.; Egbetokun, A. Emergence of innovation networks from R&D cooperation with endogenous absorptive capacity. *J. Econ. Dyn. Control* **2016**, *64*, 82–103.
35. Thiriot, S.; Kant, J.-D. Using Associative Networks to Represent Adopters Beliefs in a Multiagent Model of Innovation Diffusion. *Adv. Complex Syst.* **2008**, *11*, 261–272. [[CrossRef](#)]
36. Mueller, M.; Bogner, K.; Buchmann, T.; Kudic, M. The effect of structural disparities on knowledge diffusion in networks: An agent-based simulation model. *J. Econ. Interact. Coord.* **2017**, *12*, 613–634. [[CrossRef](#)]
37. Bathelt, H.; Malmberg, A.; Maskell, P. Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Prog. Hum. Geogr.* **2004**, *28*, 31–56. [[CrossRef](#)]
38. Nikolic, I.; Van Dam, K.H.; Kasmire, J. Practice. In *Agent-Based Modelling of Socio-Technical Systems*; Van Dam, K.H., Nikolic, I., Lukszo, Z., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 73–140.
39. Feldman, M.P. *The Geography of Innovation*; Springer Science & Business Media: Dordrecht, The Netherlands, 1994; Volume 2.
40. Boschma, R.; Frenken, K. The emerging empirics of evolutionary economic geography. *J. Econ. Geogr.* **2011**, *11*, 295–307. [[CrossRef](#)]
41. Cooke, P. Regional innovation systems, clusters, and the knowledge economy. *Ind. Corp. Chang.* **2001**, *10*, 945–974. [[CrossRef](#)]
42. Ponds, R.; Van Oort, F.; Frenken, K. The geographical and institutional proximity of research collaboration. *Pap. Reg. Sci.* **2007**, *86*, 423–443. [[CrossRef](#)]
43. Macal, C.M.; North, M.J. Tutorial on agent-based modelling and simulation. *J. Simul.* **2010**, *4*, 151–162. [[CrossRef](#)]
44. EC. Regions in the European Union. In *Nomenclature of territorial Units for Statistics*; NUTS 2013/EU-28; Publication Office of the European Union Luxembourg: Luxembourg, 2015.
45. March, J.G. Exploration and exploitation in organizational learning. *Organ. Sci.* **1991**, *2*, 71–87. [[CrossRef](#)]
46. Scherngell, T.; Barber, M.J. Spatial interaction modelling of cross-region R&D collaborations: Empirical evidence from the 5th EU framework programme. *Pap. Reg. Sci.* **2009**, *88*, 531–546.
47. Hoekman, J.; Frenken, K.; Tijssen, R.J. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Res. Policy* **2010**, *39*, 662–673. [[CrossRef](#)]
48. Rallet, A.; Torre, A. *On Geography and Technology: The Case of Proximity Relations in Localized Innovation Networks*; Clusters and Regional Specialisation: On Geography, Technology and Networks; Pion: London, UK, 1998.
49. Storper, M.; Venables, A.J. Buzz: Face-to-face contact and the urban economy. *J. Econ. Geogr.* **2004**, *4*, 351–370. [[CrossRef](#)]
50. Nooteboom, B. Innovation, learning and industrial organisation. *Camb. J. Econ.* **1999**, *23*, 127–150. [[CrossRef](#)]
51. Cohen, W.M.; Levinthal, D.A. Absorptive capacity: A new perspective on learning and innovation. *Adm. Sci. Q.* **1990**, *35*, 128–152. [[CrossRef](#)]

52. Cohen, W.M.; Levinthal, D.A. Agent-Based Simulation of innovation diffusion: A review. *Cent. Eur. J. Oper. Res.* **2012**, *20*, 183–230.
53. Golub, G.H.; Van Loan, C.F. *Matrix Computations*, 4th ed.; Johns Hopkins University Press: Baltimore, MA, USA, 2013.
54. Lovelace, R.; Ballas, D. Modelling commuter patterns: A spatial microsimulation approach for combining regional and micro level data. In *ERSA Conference Papers*; European Regional Science Association: Bratislava, Slovakia, 2012.
55. Dorner, M.; Harhoff, D. A novel technology-industry concordance table based on linked inventor-establishment data. *Res. Policy* **2018**, *47*, 768–781. [[CrossRef](#)]
56. Thiele, J.C.; Kurth, W.; Grimm, V. Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *J. Artif. Soc. Soc. Simul.* **2014**, *17*, 11. [[CrossRef](#)]
57. McKay, M.D.; Beckman, R.J.; Conover, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **2000**, *42*, 55–61. [[CrossRef](#)]
58. Schmoch, U. Concept of a technology classification for country comparisons. In *Final Report to the World Intellectual Property Organisation (WIPO)*; 2008; Available online: [http://www.world-intellectual-property-organization.com/edocs/mdocs/classifications/en/ipc\\_ce\\_41/ipc\\_ce\\_41\\_5-annex1.pdf](http://www.world-intellectual-property-organization.com/edocs/mdocs/classifications/en/ipc_ce_41/ipc_ce_41_5-annex1.pdf) (accessed on 30 July 2020).
59. Scherngell, T.; Lata, R. Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. *Pap. Reg. Sci.* **2013**, *92*, 555–577. [[CrossRef](#)]
60. Maggioni, M.A.; Uberti, T.E. Knowledge networks across Europe: Which distance matters? *Ann. Reg. Sci.* **2009**, *43*, 691–720. [[CrossRef](#)]
61. Griliches, Z. *Patent Statistics as Economic Indicators: A Survey*; National Bureau of Economic Research: Cambridge, MA, USA, 1990.
62. Jaffe, A.B.; Trajtenberg, M. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*; MIT Press: Cambridge, MA, USA, 2002.
63. Gertler, M.S. Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *J. Econ. Geogr.* **2003**, *3*, 75–99. [[CrossRef](#)]
64. Moulaert, F.; Sekia, F. Territorial innovation models: A critical survey. *Reg. Stud.* **2003**, *37*, 289–302. [[CrossRef](#)]
65. Paci, R.; Usai, S. Technological enclaves and industrial districts: An analysis of the regional distribution of innovative activity in Europe. *Reg. Stud.* **2000**, *34*, 97–114. [[CrossRef](#)]
66. Glaeser, E.L.; Kohlhase, J.E. Cities, regions and the decline of transport costs. In *Fifty Years of Regional Science*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 197–228.
67. Beaudry, C.; Schiffauerova, A. Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. *Eur. Manag. J.* **2011**, *29*, 362–376. [[CrossRef](#)]
68. Van der Panne, G. Agglomeration externalities: Marshall versus Jacobs. *J. Evol. Econ.* **2004**, *14*, 593–604. [[CrossRef](#)]
69. Marshall, A. *Principles of Economics*; Macmillan: London, UK, 1890.
70. Jacobs, J. *The Economy of Cities*; Random House: New York, NY, USA, 1969.
71. Marshall, A. *Industry and Trade*; Macmillan: London, UK, 1920.
72. Wanzenböck, I.; Piribauer, P. R&D networks and regional knowledge production in Europe: Evidence from a space-time model. *Pap. Reg. Sci.* **2018**, *97*, S1–S24.
73. Sebestyén, T.; Varga, A. Research productivity and the quality of interregional knowledge networks. *Ann. Reg. Sci.* **2013**, *51*, 155–189. [[CrossRef](#)]
74. Scherngell, T. The Geography of R&D Collaboration Networks. In *Handbook of Regional Science*; Fischer, M.M., Nijkamp, P., Eds.; Springer: Berlin, Germany, 2019.
75. Long, S.J.; Freese, J. *Regression Models for Categorical Dependent Variables Using Stata*; Stata Press: College Station, TX, USA, 2006.
76. Fischer, M.M.; Wang, J. *Spatial Data Analysis: Models, Methods and Techniques*; Springer Science & Business Media: Berlin, Germany, 2011.

