Geostatistical visualization of ecological interactions in tumors

1st Hunter Bryan Boyce
Program in Biomedical Informatics
Stanford University
Stanford, CA, USA
hboyce@stanford.edu

2nd Parag Mallick
Canary Center at Stanford for Cancer Early Detection
Stanford University
Palo Alto, CA, USA
paragm@stanford.edu

***Abstract*—Recent advances in our understanding of cancer progression have highlighted the roles played by molecular heterogeneity and by the tumor microenvironment in driving drug resistance and metastasis. The coupling of single-cell measurement technologies with algorithms, such as t-sne and SPADE, have enabled deep investigation of tumor heterogeneity. However, such techniques only capture molecular heterogeneity and do not enable the quantification nor visualization of intercellular interactions . They additionally do not allow the visualization of ecological niches that are critical to understanding tumor behavior. Novel computational tools to quantify and visualize spatial patterns in the tumor microenvironment are critically needed. Here, we take a tumor ecology perspective to examine how predation, mutualism, commensalism, and parasitism may impact tumor development and spatial patterning. We additionally quantify local spatial heterogeneity and the emergent global spatial behavior of the models using geostatistics. By visualizing emergent spatial patterns we demonstrate the potential utility of a geostatistical analysis in differentiating amongst cell-cell interactions in the tumor microenvironment. These studies introduce both an ecological framework for characterizing intercellular interactions in cancer and a novel way of quantifying and visualizing spatial patterns in cancer.**

***Index Terms*—Geostatistics; Pathology; Cancer; Tumor Ecology; Agent-Based Modeling, Visualization**

I. Introduction

Recent reports have demonstrated that intratumoral-heterogeneity, including genetic mutation, expression noise, and microenvironmental variation can have a significant impact on tumor progression and invasion [1]–[3]. Single-cell RNA sequencing [4] and technologies such as mass cytometry [5], are promising solutions to profiling heterogeneity within the tumor by quantifying the RNA content within individual cells or by surveying portions of the cellular proteome. Analysis of this data type attempts to visualize cellular phenotypes and their lineages through dimensionality reduction techniques like t-stochastic neighbor embedding (t-SNE) [6], and spanning-tree progression analysis of density-normalized events (SPADE) [7]. Despite being extremely powerful tools to understand heterogeneity, they do not provide any information about tumors spatial patterning nor about what that patterning may imply about the ecological and evolutionary forces driving tumor behavior.

The prevalence of banks of archived tissue samples [8], in concert with the recent emergence of platforms that can measure the expression of dozens of proteins and RNA *in situ* in pathology sections and tumor microarrays [9]–[11] make it possible to examine the interplay between molecular heterogeneity and spatial patterning amongst diverse cell types and molecularly defined subtypes. With the ability to survey the spatial context of cells within tissues comes computational challenges to understand how to quantify spatial interactions and how to appropriately visualize the intercellular and higher-order spatial relationships. Examples of such challenges include finding regions where expression of a protein of interest is concentrated, identifying sudden spatial shifts in protein expression, or quantifying cell-cell interactions.

Visualizing spatial phenomena in tumors has predominantly been limited to a qualitative description of the spatial context of protein expression, cell type, or geostatistical measure [12]–[16]. While these assessments have been shown to have modest clinical utility, they do not readily provide insight into underlying biological processes. Specifically, they ignore the contribution of local spatial phenomena. We hypothesize that concurrently monitoring six geostatistics describing both local spatial patterns and their emergent global behavior will enable visualization and quantification of spatial processes driven by cell-cell interactions. To test this hypothesis, we use agent-based models to grow simple virtual tumors exhibiting four different ecological interspecific interactions: predation, mutualism, commensalism, and parasitism [17], [18]. Using our new visualization approach to simultaneously examine protein expression, and global and local spatial phenomena, we demonstrate how a multi-parametric approach enables a more complete view of the spatial ecology of the tumor.

II. Background

Tumor ecology

The interactions between the diverse types of cells within a tumor (epithelial, stromal, immune) and their environment have multiple analogues in community ecology and can be described by four classes of interactions which have all been reported to exist within tumors [17]–[20]: predation, mutualism, commensalism, and parasitism. Predatory interactions consist of those where one organism targets and kills another organism. In tumors, this behavior is most prominently noted between certain immune cells that target and kill cancer cells

978-1-7281-1867-3/19/$31.00 ©2019 IEEE

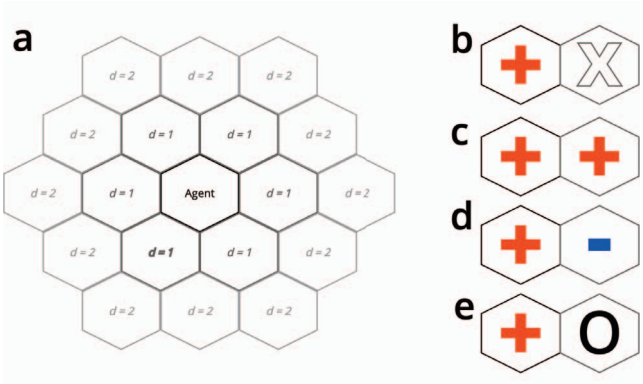

Fig. 1. Agent-based model neighbors and interactions. (a) Any agents within a hexagonal distance $d = 2$ are considered neighbors of the central agent. The change in oncoprotein expression upon ecological interaction in (b) predation, (c) mutualism, (d) parasitism, and (e) commensal models.

[21]. Research suggests that the degree of co-localization between these two cells types can influence the progression of the tumor [14], [16]. Mutualistic relationships are those in which two species both benefit by interacting with one another. This type of cooperation has been observed among tumor subclones that harbored mutations which mutualistically created an environment that favored tumor growth [22]. Commensalism is described as relationships where one species benefits from interaction with a second species that is neither benefited nor harmed. Like mutualism, commensal interactions have been observed between tumor subclones wherein one subclone benefits from consuming a substrate that the other produces [23]. Lastly, parasitic interactions in which one species benefits from harming, but not killing, another species. In cancer, this is observed when tumor cells steer high energy substrates away from nearby healthy tissues in the reverse Warburg effect [24].

Agent-based modeling

Agent-based models are often used in cancer to simulate the behavior of individual cells within a larger tissue context [25], [26]. In agent-based models of cancer each individual unit, or agent, often represent single cells within the tumor microenvironment. Parameters for each cell, such as probability of death and division, and rules for cell-cell interactions, allow modeling of the individual cellular behavior that gives rise to heterogeneity in the tumor microenvironment. Agent-based models allow profiling interspecific interactions by measuring local spatial patterns and investigation of emergent behaviors of those interactions.

III. Methods

A. Agent-based Modeling

A spatial, agent-based model was used to simulate a basic tumor growing. The model is a lattice-based cellular automaton model [26], [27] which considers two-dimensional space as a hexagonal grid where each position has six direct neighbors with which it can interact (Figure 1a). The model keeps track of the number of agents, the types of agents, the ecological model, and the width and height of the canvas. The canvas is non-toroidal, such that the top and bottom, and left and right, edges of the canvas do not wrap to each other. The model employs a random activation schedule wherein agents are activated in a random order at each step of the model.

Each agent in the model represents a cell in the tumor microenvironment. The cell can be a cancer cell, a cancer stem cell, stromal cell, or an immune cell. Each cell has a unique id, a hexagonal coordinate position in the model, an instantaneous death probability, and a division probability. Cancer cells have an oncoprotein parameter which represents the measurement of Ki67, a biomarker of proliferative capacity or fitness. Cancer cells also have a counter for the number of divisions remaining before cell death. Cancer stem cells act like cancer cells, except that they have replicative immortality and therefore have an infinite number of divisions remaining. Stromal cells have the same parameters as cancer cells, but with decreased replicative capacity. Immune cells have a probability to target a neighboring cancer cell and a probability to kill the cancer cell it is targeting. Each cancer cell population has rules for how fitness is modified upon interaction with another cellular population that is dictated by the model.

The model is initialized with an initial number of cancer cells, rules for ecological interactions, and a width and height for the canvas. Experimental pathology samples show strong spatial autocorrelation of protein abundance. Specifically, cells that are spatially proximal tend to have correlated expression. This necessitates an oncoprotein initialization strategy that is dependent on the expression of the neighbors of a given cell for determining oncoprotein measurement of that cell. To achieve spatial autocorrelation we utilize a Markov Chain Monte Carlo (MCMC) sampling strategy that incorporates the mean oncoprotein expression of a cellar neighborhood:

$$x' = x_0 + \mathcal{N}\left(0, h^2\right) \tag{1}$$

Where x_0 is the mean oncoprotein of your neighbors and h is the walk rate for the MCMC. We use the Metropolis-Hastings algorithm to step through the chain where we calculate an acceptance probability:

$$\alpha = \min\left(1, \frac{\pi\left(x'\right) q\left(x_0 | x'\right)}{\pi\left(x_0\right) q\left(x' | x_0\right)}\right) \tag{2}$$

Where q is the proposal distribution and π is the target distribution or stationary distribution of the Markov chain we are modeling. We then sample a uniform random variable u and if $u < \alpha$, we accept the proposal x', otherwise we reject and sample the MCMC again. We select a walk rate h for the MCMC such that the proposal acceptance rate is around 90% resulting in an autocorrelated sample. This sampling strategy is used in the initialization of the model and when new agents are added to the model. After oncoprotein initialization for each agent, its fitness is modified using a piecewise generalized logistic function:

$$f(x) = \begin{cases} \frac{1}{1+\exp(-kx-(\mu+d))} + 1 & x \geq \mu \\ \frac{1}{1+\exp(-kx-(\mu-d))} & x < \mu \end{cases} \quad (3)$$

Where x is the oncoprotein of the cell, k is the growth rate of the logistic curve, μ is the mean of the target distribution from the MCMC sampling, and d is how far away from μ that growth occurs. The resulting $f(x)$ is used as a coefficient to the division and death probabilities for the agent. This has the effect of considerably increasing fitness if the oncoprotein expression of a cell increases sufficiently from the mean of the target distribution.

B. Modeling Ecological Interactions

To model predatory interactions (Figure 1b) in the tumor microenvironment, we initialize the tumor model with cancer cells and cancer stem cells in the center of the model canvas and initialize a perimeter of immune cells encircling the tumor. Immune cells are inactive until the tumor reaches a cancer cell count threshold, upon which the immune cells are activated. While the tumor continues to grow, immune cells begin to move with preference toward the weighted center of the cancer cells. When an immune cell neighbors a cancer cell it will target and attach to the cancer with its initialized target probability. If an immune cell is already targeting a cancer cell, it will kill that cell with its initialized kill probability. Immune cells can only target one cancer cell at a time and targeting and killing occur at different steps of the model.

To model mutualistic interactions (Figure 1c) in the tumor microenvironment, we initialize two separate tumors on the canvas, each with cancer cells and cancer stem cells, and the tumors are allowed to grow. If cancer cells from the first tumor neighbor cancer cells from the second tumor, a mutualistic interaction occurs and both cells increase in oncoprotein expression and gain a fitness benefit. This allows greater replicative capacity. This increase in proliferative ability is not compound and is only allowed to happen once.

Modeling commensal interactions (Figure 1d) in the tumor microenvironment is initialized in the same manner as the mutualism model. The difference being, when cancer cells from the first tumor neighbor cancer cells from the second, the first gains a fitness advantage while the other is unchanged. Like the mutualism model, the fitness advantage is not compound.

The parasitic model (Figure 1e) is setup similarly to the mutualism and commensalism models. However, instead of a second tumor we initialize a region representing a stromal region. The replication and death rates of this region are selected such that the size and shape of the region remains roughly constant. When cancer cells from the tumor neighbor cells in the stromal compartment, the cancer cells gain a fitness advantage while the stromal cells have a decrease in proliferative capability. Similar to the previous two models, the change in proliferative capacity is not compound.

In addition to the previously described models, we simulated a baseline model of a single tumor growing with no ecological interactions. The tumor is allowed to grow without external influence for the lifetime of the simulation.

C. Geostatistical Methods

For all equations in this section, x_i is the oncoprotein expression of cell i and $w_{ij} = 1$ if cell i and cell j are neighbors in the model. Otherwise $w_{ij} = 0$. Cells are considered neighbors if they have a Manhattan distance of two or less (Figure 1a).

1) Global Statistics: Global statistics are geostatistics describing the entire model at the current step. We calculate three global statistics at each step that describe how oncoprotein is clustered and the spatial heterogeneity of oncoprotein in the model. To quantify general clustering of oncoprotein at the global level we use the Getis-Ord general G statistic [28]:

$$G = \frac{\sum_i \sum_j w_{ij} x_i x_j}{\sum_i \sum_j x_i x_j} \quad (4)$$

The G statistic is bounded between 0, indicating clustering of low values, and 1 indicating clustering of high values. To quantify spatial heterogeneity we utilize two statistics. First, Moran's I statistic [29]:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \quad (5)$$

Where N is the number of agents in the model, $W = \sum_i \sum_j w_{ij}$, and $z_i = x_i - \bar{X}$. The value of I is between -1 for perfectly dispersed (low autocorrelation) data and $+1$ for perfectly clustered (high autocorrelation) data. The second statistic that we use to quantify spatial heterogeneity is Geary's C [30] :

$$C = \frac{N-1}{2W} \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i (x_i - \bar{X})^2} \quad (6)$$

Where N is the number of agents in the model and $W = \sum_i \sum_j w_{ij}$. Geary's C is inversely related to Moran's I and can take on values between 0 with perfect clustering (high autocorrelation) and an undefined upper bound for increased dispersion (low autocorrelation.

2) Local Statistics: Local geostatistics describe spatial behavior at the agent level. Each of the global geostatistics has a local counterpart. However, instead of having a single statistic for the entire model, each agent in the model takes on a value for these statistics. To quantify the degree of spatial clustering we utilize the local Getis-Ord G_i statistic [28], [31]:

$$G_i = \frac{\sum_j w_{ij} x_j - \bar{X} \sum_j w_{ij}}{\sqrt{\frac{\sum_j x_j^2}{n} - \bar{X}^2} \sqrt{\frac{N \sum_j w_{ij}^2 - \left(\sum_j w_{ij}\right)^2}{N-1}}} \quad (7)$$

Large positive values of G_i correspond to a hotspot of expression at location i and low negative values correspond to a coldspot. To quantify local spatial heterogeneity we again use two different statistics. First, Anselin local Moran's I_i [29], [32]:

$$I_i = \frac{x_i - \bar{X}}{\frac{\sum_j (x_j - \bar{X})^2}{N-1}} \sum_j w_{ij} \left(x_j - \bar{X}\right) \quad (8)$$

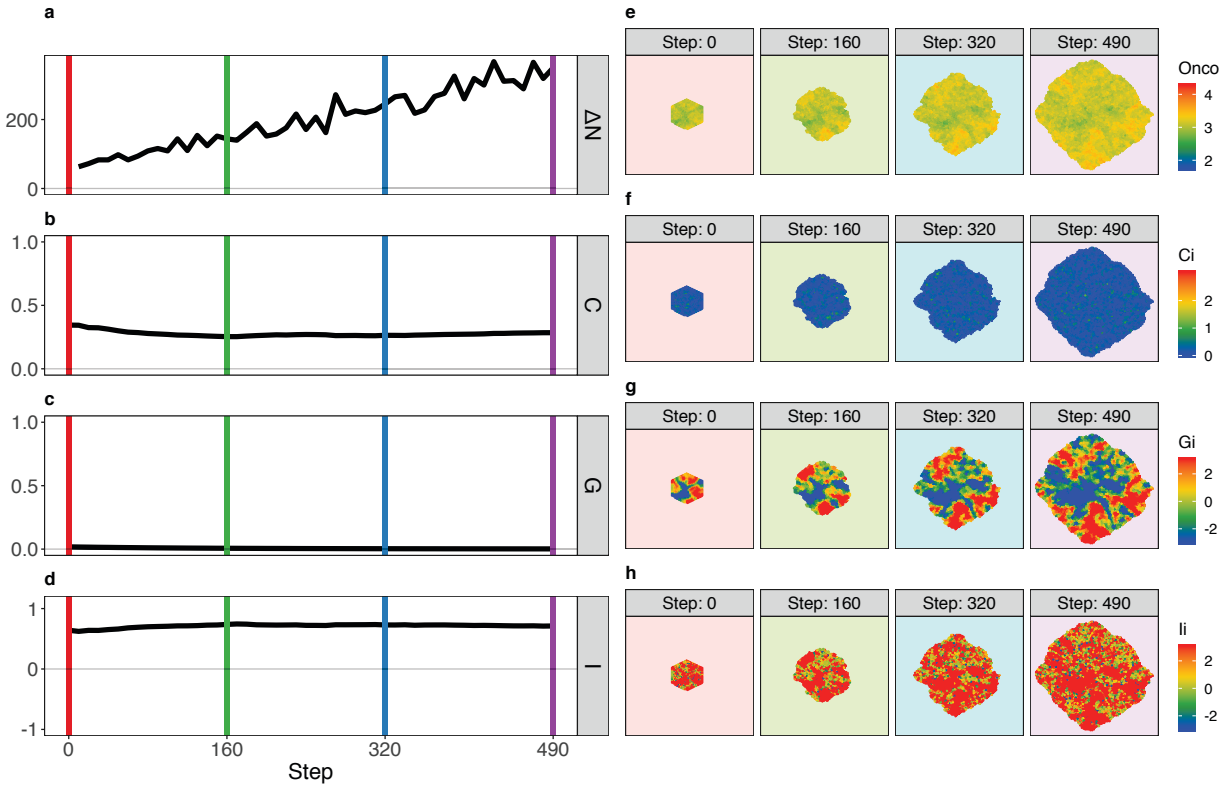

Fig. 2. Baseline model geostatistics. Values for the life of the model for a) the delta number of cancer cells from the step previous; b) Geary's C (Eq. 6); c) Getis-Ord G (Eq. 4); d) Moran's I (Eq. 5). State of the model at time steps denoted by colored vertical lines in a-d for e) oncoprotein expression; f) Local Geary's C_i (Eq. 9); g) Getis-Ord G_i (Eq. 7); h) Anselin Moran's I_i (Eq. 8)

A positive value for I_i indicates that neighbors of agent i are similar to itself (low heterogeneity). A negative value indicates dissimilar neighbors (high heterogeneity). The second statistic for spatial heterogeneity is the local Geary's C_i statistic [30], [32]:

$$C_i = \sum_j w_{ij} \left(x_i - x_j\right)^2 \tag{9}$$

A value close to zero for C_i indicates homogeneous neighbors and large positive values indicate heterogeneous neighbors.

IV. Results and Discussion

An effective approach for measuring and visualizing spatial heterogeneity in the context of interspecific ecological interactions in tumors should: 1) exhibit sensitivity to true variation in a growing tumor over time; 2) be insensitive to minor variations in a broadly unchanging tumor; 3) not present multiple dimensions of redundant information; and 4) differentiate among distinct ecological interactions. Below we examine the performance of our approach to measure and visualize tumor ecology and heterogeneity across a series of virtual tumors exhibiting a range of ecological patterns.

A. Baseline model shows stability in the measurement of global behavior

The global spatial behavior of the baseline model shows stability through the lifetime of the simulation (Figure 2b-d). This is expected behavior from a tumor in which there are no interspecific interactions affecting subpopulations. The only global variable that is increasing over time is ΔN (Figure 2a) which is the result of a larger cancer cell population that is actively dividing.

The local spatial behavior shows a wide distribution of values for spatial clustering and spatial autocorrelation (Figure 2g,h). However, the patterning of the local spatial phenomenon is relatively unchanged as the tumor grows with specific patterns growing proportional to each other with respect to overall tumor size. The value of C_i (Figure 2f) is close to zero for all cells in the model throughout the life of the simulation. As this statistic has no dependence on the mean, it much more sensitive to the presence of an interface between two different cell types.

B. Predatory behavior results in continually varying geostatistics

Visualizing the global behavior of the predation model (Figure 3a-d) shows a global response that is very sensitive to the state of the model. In the initial stages of the model, the tumor is allowed to grow unimpeded with no interaction with

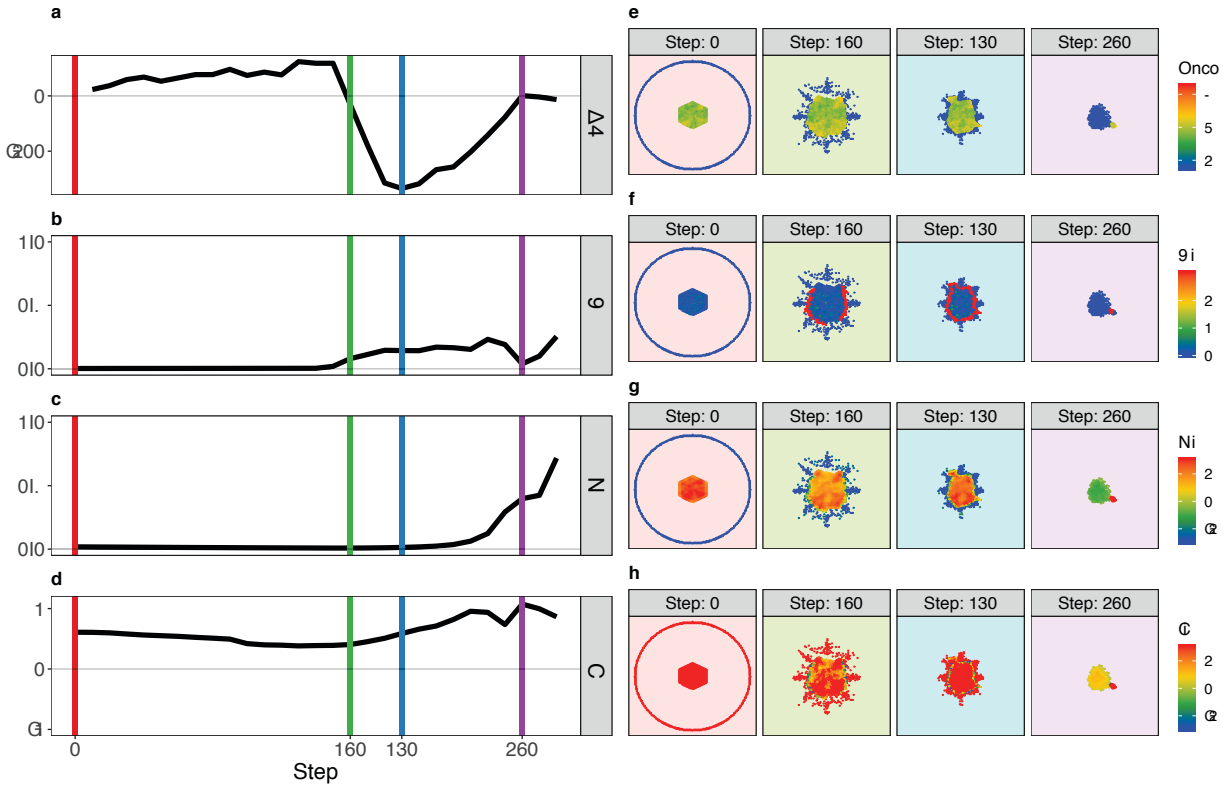

Fig. 3. Predation model geostatistics. Values for the life of the model for a) the delta number of cancer cells from the step previous; b) Geary's C (Eq. 6); c) Getis-Ord G (Eq. 4); d) Moran's I (Eq. 5). State of the model at time steps denoted by colored vertical lines in a-d for e) oncoprotein expression; f) Local Geary's C_i (Eq. 9); g) Getis-Ord G_i (Eq. 7); h) Anselin Moran's I_i (Eq. 8)

immune cells. This is behavior is reflected in global geostatistics that remain stable during this initial period. As the tumor grows, ΔN steadily grows relative to the size of the tumor until predatory interspecific interactions occur and growth drops precipitously at step 160 of the model (Figure 3a). This is met with an increase spatial heterogeneity as immune cells are interfacing with cancer cells (Figure 3b,f). As the immune cells successfully kill and target cancer cells through step 260, we observe a decrease in spatial heterogeneity and an increase in spatial clustering as marked by a decrease in C and an increase in G and I (Figure 3b-d). This can be explained by a spatially constrained cancer cell population with high autocorrelation. When the immune cells target the remaining cancer cells after step 260, we observe another increase in spatial heterogeneity.

The local geostatistics in the predation model exhibit patterns dependent on area. During the initial predatory interaction in steps 160-190, we observe a marked increase in local spatial heterogeneity at the interaction interface (Figure 3f). As the space the cancer cells reside in is constrained, it becomes more of a spatial hotspot relative to the immune cell population surrounding it (Figure 3g). Overall, we can utilize local Geary's C_i as a proxy for the presence of an ecological interaction and we observe a hotspot response that is dependent on area constraints.

C. Geostatistics stabilize to baseline after initial mutual interaction

The end stage global spatial behavior of the mutualism model looks very similar to the initialization state (Figure 4 b-d). Initially, there is relatively low spatial heterogeneity in the two tumors as evidenced by a low value for Geary's C and a high value for Moran's I (Figure 4b,d). As the model progress, ΔN steadily increases in line with the size of the tumors until step 130 when the two tumors begin a mutualistic relationship. Increased spatial heterogeneity is observed through increased local Geary's C_i at the interface between the two tumors (Figure 4f) and is visualized globally by increases in Geary's C and Moran's I. The ΔN stalls during this initial interaction as the space available for each tumor to grow into has been decreased (Figure 4e-h). However, after an increase in proliferative capacity due to the mutualistic interaction, ΔN continues an upward trajectory. While we observe a similar increase in measured global spatial heterogeneity as the predation model, we observe global spatial clustering and local spatial patterns that differ greatly between the two models.

For the rest of the model's life we observe a return to the baseline low heterogeneity global geostatistics even though there is a clear increase in local spatial heterogeneity. The increase in local heterogeneity is matched by the same increase in overall deviation of oncoprotein expression from the mean,

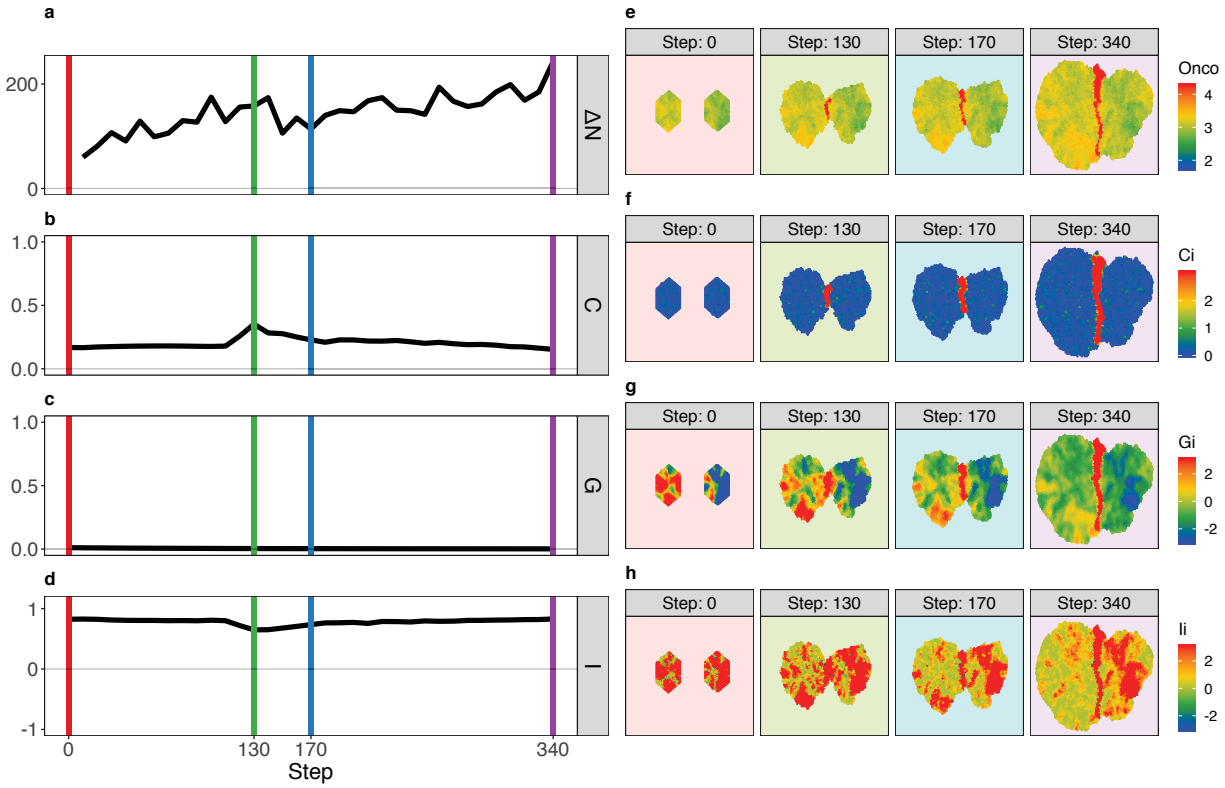

Fig. 4. Mutualism model geostatistics. Values for the life of the model for a) the delta number of cancer cells from the step previous; b) Geary's C (Eq. 6); c) Getis-Ord G (Eq. 4); d) Moran's I (Eq. 5). State of the model at time steps denoted by colored vertical lines in a-d for e) oncoprotein expression; f) Local Geary's C_i (Eq. 9); g) Getis-Ord G_i (Eq. 7); h) Anselin Moran's I_i (Eq. 8)

resulting in global geostatistics that behave as if there is low heterogeneity (Eqs. 5 and 6). This model highlights the utility of visualizing the global and local geostatistics simultaneously to accurately capture the spatial phenomena occurring in the tumor.

D. Commensal interaction results in a new geostatistical baseline

The commensalism model displays similar global spatial behavior to the mutualism up through the initial interspecific interaction with low spatial heterogeneity that increases when the two tumors meet as evidenced by an increase in Geary's C and decrease in Moran's I (Figure 5b,d). However, as the model progresses, we do not observe the same return to baseline heterogeneity that the mutualism model displayed, but instead a new global pattern that more accurately represents the local spatial heterogeneity (Figure 5e-h) and is able to discriminate from the mutualistic interactions. This can be explained by the smaller region affected by an increase in oncoprotein expression in the commensal model (Figure 5e) relative to the mutual model (Figure 4e), resulting in a smaller increase to the mean oncoprotein in the model leading to an increase in Geary's C (Eq. 6) and decrease in Moran's I (Eq. 5). Geostatistics that have no dependence on the mean, like local Geary's C (Eq. 9) exhibit spatial phenomena (Figure 5f) similar to the mutual model. Counterintuitively, in this scenario, a smaller local spatial effect results in a larger shift in global spatial behavior and, once again, highlights the utility of visualizing local and global phenomena simulatenously.

E. Geostatistics continually change after initial parasitic interaction

The parasitism model was initiated similarly to the mutualism and commensalism models and, as a result, we observe similar behavior through the initial interspecific parasitic interaction (Figure 6). However, we observe a much more gradual increase in global spatial heterogeneity (Figure 6b,d) in contrast to the mutual and commensal models. This gradual behavior persists throughout the life of the model and never stabilizes. This continuously increasing heterogeneity is due to the change in fitness as a result of a change in oncoprotein expression is not symmetric about the mean(Eq. 3). More specifically, a positive change in oncoprotein expression that results in a doubling of replicative capacity has the effect of almost removing the ability to replicate when the change is negative. This results in global visual patterns that are very sensitive to the number of cells on either side of the parasitic relationship at a particular step (Figure 6e). While this may be a peculiarity of how this model was written, it elucidates an interesting relationship between local cellular types and global patterns.

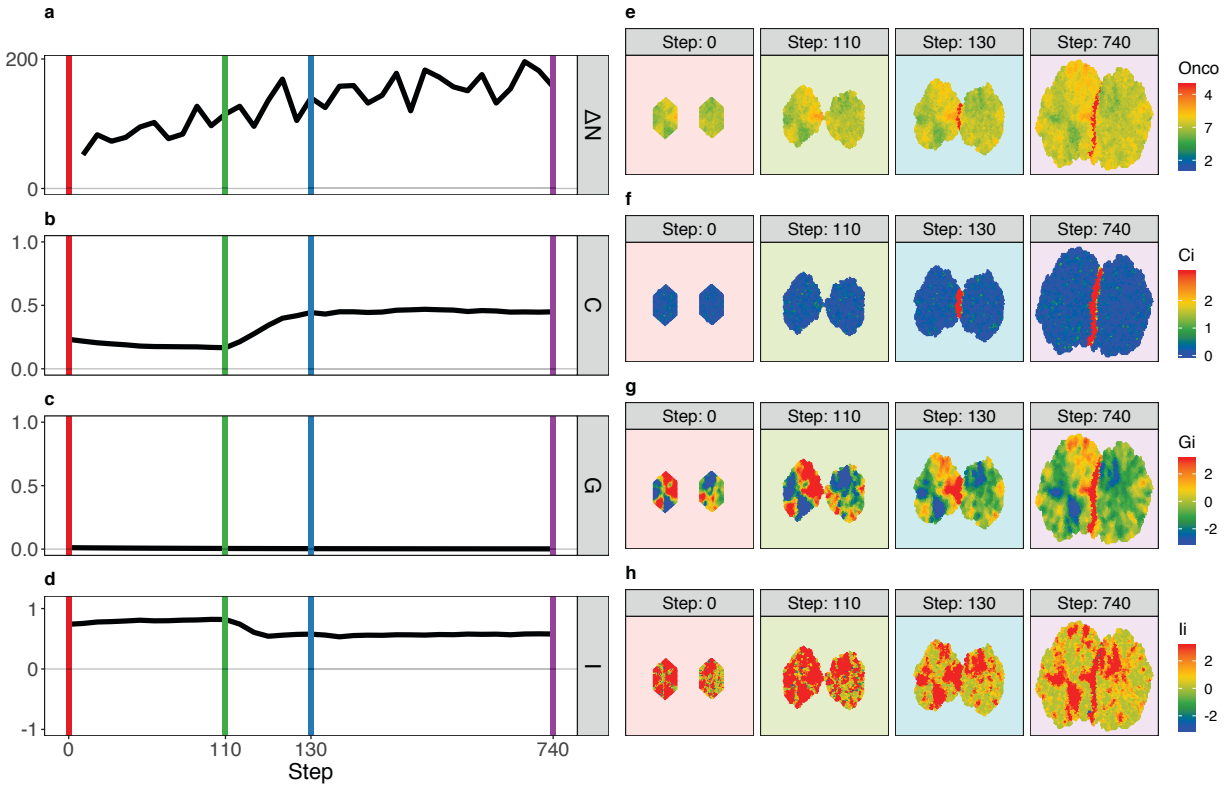

Fig. 5. Commensalism model geostatistics. Values for the life of the model for a) the delta number of cancer cells from the step previous; b) Geary's C (Eq. 6); c) Getis-Ord G (Eq. 4); d) Moran's I (Eq. 5). State of the model at time steps denoted by colored vertical lines in a-d for e) oncoprotein expression; f) Local Geary's C_i (Eq. 9); g) Getis-Ord G_i (Eq. 7); h) Anselin Moran's I_i (Eq. 8)

V. Limitations

Overall, we find that our proposed approach is able to effectively capture the heterogeneity and changes therein across a range of tumors exhibiting diverse ecological behaviors. However, we note that the methods have further room for improvement. One area that remains an ongoing challenge is with regards to edge effects. Relatively poor estimation of geostatistics at the outside borders of regions may arise from the same underlying stationary spatial process. However, as we are focusing on interactions between cells which, for the most part, occur in interior areas of the simulated tumors this is likely not a critical barrier in the short term. Along a similar vein, as the simulation progresses and the tumors become larger, the internal area grows exponentially faster than the border area; lessening the influence of border effects on determination of geostatistics. One additional challenge in visualizing tumor heterogeneity may arise when using a two-dimensional measurement to describe a subset of a three-dimensional process. Notably, the calculation of each geostatistic is agnostic to the dimensionality of space and the weighting matrix only captures which cells are neighbors and not their respective spatial distances. For profiling the ecological interactions of concern, a two-dimensional model is likely a sufficient representation. We additionally note that our tumor growth model is extremely simple. We do not fully describe vascularization or any of the other factors that might significantly alter tumor development and ecological interactions. Critically, our study is focused on the visualization and quantification aspects of tumor ecologies rather than on the details of tumor development. Consequently, our model is of sufficient complexity to test our approach to visualize tumor heterogeneity.

VI. Conclusion

We have introduced a new method that uses a comprehensive geostatistical survey approach to leverage and measure spatial heterogeneity by examining ecological interactions, heterogeneity of tumors, and their local and global spatial patterns. The method was tested by growing virtual tumors using an agent-based model and modeling ecological interspecific interactions between cell types. Our approach is both sensitive to changing spatial patterns and stable with respect to stationary spatial processes. We also observe that quantifying and visualizing global and local spatial phenomena simultaneously is able to discriminate between different ecological relationships that qualitative descriptions of single geostatistical measures are not.

Presented here is a unimodal analysis for a single protein and could be used to profile the spatial behavior of a cancer biomarker in the tissue. For example the spatial heterogeneity of EGFR in lung cancer, in addition to its expression profile,

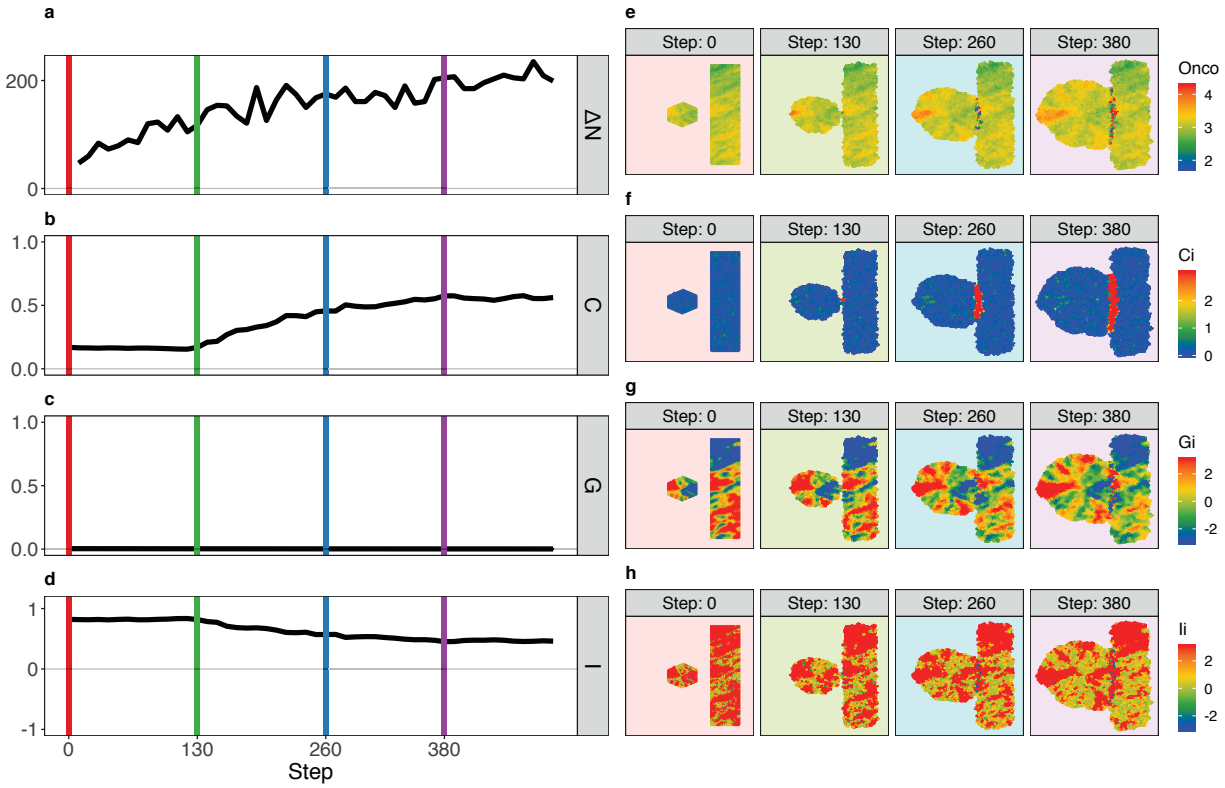

Fig. 6. Parasitism model geostatistics. Values for the life of the model for a) the delta number of cancer cells from the step previous; b) Geary's C (Eq. 6); c) Getis-Ord G (Eq. 4); d) Moran's I (Eq. 5). State of the model at time steps denoted by colored vertical lines in a-d for e) oncoprotein expression; f) Local Geary's C_i (Eq. 9); g) Getis-Ord G_i (Eq. 7); h) Anselin Moran's I_i (Eq. 8)

may result in a novel spatial biomarker. This model and visualization framework can be further extended into a multiparametric approach, allowed by novel hyperplex immunostaining platforms [9]–[11], that spans many proteins across multiple functions. We can measure the spatial patterning of cancer biomarkers, immune cell markers, and stromal markers to better understand the effectiveness of the immune cells, how the tumor responds to the immune presence, and the effect of this interaction on the surrounding tissue simultaneously. This depth of analysis can be used to test complex hypotheses of tumor evolution such as characterizing the spatial context differences between a successful and an ineffective immune response. Overall, this framework allows a deep analysis and visualization of local spatial patterns and their emergent global spatial behavior.

Acknowledgment

We would like to thank B Anana for careful reading of the manuscript. This work was partially funded by the National Library of Medicine T15 LM 007033, DARPA Deep Purple, DARPA Physics of AI, and GE Blue Skies.

References

[1] R. Mark, R. Gillies, R. Gatenby, and A. Anderson, "Impact of metabolic heterogeneity on tumor growth, invasion, and treatment outcomes," *Cancer Res*, vol. 75, no. 8, pp. 1567–1579, 2015.

[2] F. Fu, M. Nowak, and S. Bonhoeffer, "Spatial heterogeneity in drug concentrations can facilitate the emergence of resistance to cancer therapy," *PLoS computational biology*, 2015.

[3] M. Nicholas and C. Swanton, "Biological and therapeutic impact of intratumor heterogeneity in cancer evolution." *Cancer Cell*, vol. 27, no. 1, pp. 15–26, 2015.

[4] X. Pan, R. E. Durrett, H. Zhu, Y. Tanaka, Y. Li, X. Zi, S. L. Marjani, G. Euskirchen, C. Ma, R. H. Lamotte, I. H. Park, M. P. Snyder, C. E. Mason, and S. M. Weissman, "Two methods for full-length RNA sequencing for low quantities of cells and single cells." *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 2, pp. 594–9, 2013.

[5] S. C. Bendall, E. F. Simonds, P. Qiu, E.-a. D. l. Amir, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, R. S. Balderas, S. K. Plevritis, K. Sachs, D. Pe'er, S. D. Tanner, and G. P. Nolan, "Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum." *Science*, vol. 332, no. 6030, pp. 687–96, 2011.

[6] L. Maaten and H. G. of machine learning research, "Visualizing data using t-SNE," *Journal of machine learning research*, 2008.

[7] B. Anchang, T. D. Hart, S. C. Bendall, P. Qiu, Z. Bjornson, M. Linderman, G. P. Nolan, and S. K. Plevritis, "Visualization and cellular hierarchy inference of single-cell data using SPADE." *Nat Protoc*, vol. 11, no. 7, pp. 1264–79, 2016.

[8] N. Blow, "Tissue issues," *Nature*, vol. 448, no. 7156, pp. 959–960, 8 2007. [Online]. Available: https://doi.org/10.1038/448959a

[9] M. J. Gerdes, C. J. Sevinsky, A. Sood, S. Adak, M. O. Bello, A. Bordwell, A. Can, A. Corwin, S. Dinn, R. J. Filkins, D. Hollman, V. Kamath, S. Kaanumalle, K. Kenny, M. Larsen, M. Lazare, Q. Li, C. Lowes, M. C. C, M. Elizabeth, M. C. Montalto, Z. Pang, J. Rittscher, S. Alberto, B. D. Sarachan, M. L. Seel, A. Seppo, K. Shaikh, Y. Sui, J. Zhang, and F. Ginty, "Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue," *Proc Natl Acad Sci*, vol. 110, no. 29, pp. 11 982–11 987, 2013.

[10] M. Angelo, S. C. Bendall, R. Finck, M. B. Hale, C. Hitzman, A. D.

Borowsky, R. M. Levenson, J. B. Lowe, S. D. Liu, S. Zhao, Y. Natkunam, and G. P. Nolan, "Multiplexed ion beam imaging of human breast tumors." *Nat. Med.*, vol. 20, no. 4, pp. 436–42, 2014.
[11] Y. Goltsev, N. Samusik, K. Julia, S. Bhate, M. Hale, G. Vazquez, S. Black, and G. P. Nolan, "Deep profiling of mouse splenic architecture with CODEX multiplexed imaging." *Cell*, vol. 174, no. 4, pp. 968–981.e15, 2018.
[12] L. Keren, M. Bosse, D. Marquez, R. Angoshtari, S. Jain, S. Varma, S. R. Yang, A. Kurian, D. Van Valen, R. West, S. C. Bendall, and M. Angelo, "A structured Tumor-Immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging." *Cell*, vol. 174, no. 6, pp. 1373–1387.e19, 2018.
[13] Y. Yuan, H. Failmezger, O. M. Rueda, H. Ali, S. Gräf, S. F. Chin, R. F. Schwarz, C. Curtis, M. J. Dunning, H. Bardwell, N. Johnson, S. Doyle, G. Turashvili, E. Provenzano, S. Aparicio, C. Caldas, and F. Markowetz, "Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling." *Sci Transl Med*, vol. 4, no. 157, p. 157ra143, 2012.
[14] S. Nawaz, A. Heindl, K. Koelble, and Y. Yuan, "Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer." *Mod. Pathol.*, vol. 28, no. 6, pp. 766–77, 2015.
[15] J. L. Carstens, P. de Sampaio, D. Yang, S. Barua, H. Wang, A. Rao, J. P. Allison, L. V. S, and R. Kalluri, "Spatial computation of intratumoral t cells correlates with survival of patients with pancreatic cancer," *Nature* , vol. 8, p. 15095, 2017.
[16] C. C. Maley, K. Koelble, R. Natrajan, A. Aktipis, and Y. Yuan, "An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer." *Breast Cancer Res.*, vol. 17, no. 1, p. 131, 2015.
[17] L. M. Merlo, J. W. Pepper, B. J. Reid, and C. C. Maley, "Cancer as an evolutionary and ecological process." *Nat. Rev. Cancer*, vol. 6, no. 12, pp. 924–35, 2006.
[18] S. Nawaz and Y. Yuan, "Computational pathology: Exploring the spatial dimension of tumor ecology." *Cancer Lett.*, vol. 380, no. 1, pp. 296–303, 2016.
[19] K. S. Korolev, J. B. Xavier, and J. Gore, "Turning ecology and evolution against cancer," *Nat Rev Cancer*, vol. 14, no. 5, p. nrc3712, 2014.
[20] D. P. Tabassum and K. Polyak, "Tumorigenesis: it takes a village." *Nat. Rev. Cancer*, vol. 15, no. 8, pp. 473–83, 2015.
[21] K. Murphy and C. Weaver, *Janeway's Immunobiology*, 9th ed. Taylor & Francis Group, 2017.
[22] M. Greaves and C. C. Maley, "Clonal evolution in cancer," *Nature*, vol. 481, pp. 306–313, 2012.
[23] M. Archetti, D. A. Ferraro, and G. Christofori, "Heterogeneity for IGF-II production maintained by public goods dynamics in neuroendocrine pancreatic cancer," *PNAS*, 2015.
[24] S. Pavlides, W. Diana, R. cycle, N. Flomenberg, A. K. Witkiewicz, P. G. Frank, M. C. Casimiro, C. Wang, P. Fortina, S. Addya, R. G. Pestell, M. U. E, F. Sotgia, and M. P. Lisanti, "The reverse warburg effect: aerobic glycolysis in cancer associated fibroblasts and the tumor stroma," *Cell cycle*, 2009.
[25] P. M. Altrock, L. L. Liu, and F. Michor, "The mathematics of cancer: integrating quantitative models." *Nat. Rev. Cancer*, vol. 15, no. 12, pp. 730–45, 2015.
[26] J. Metzcar, Y. Wang, R. Heiland, and P. Macklin, "A review of Cell-Based computational modeling in cancer biology." *JCO Clin Cancer Inform*, vol. 3, no. 3, pp. 1–13, 2019.
[27] J. Lowengrub, H. Frieboes, F. Jin, Y. L. Chuang, X. Li, P. Macklin, S. Wise, and V. Cristini, "Nonlinear modelling of cancer: bridging the gap between cells and tumours." *Nonlinearity*, vol. 23, no. 1, pp. R1–R9, 2010.
[28] A. Getis and J. Ord, "The analysis of spatial association by use of distance statistics," *Geographical Analysis*, vol. 24, no. 3, pp. 189–206, 1992.
[29] P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, 1950.
[30] R. Geary, "The contiguity ratio and statistical mapping," *The incorporated statistician*, 1954.
[31] J. Ord and A. Getis, "Local spatial autocorrelation statistics: Distributional issues and an application," *Geogr Anal*, vol. 27, no. 4, pp. 286–306, 1995.
[32] L. Anselin, "Local indicators of spatial associationLISA," *Geographical analysis*, vol. 27, no. 2, pp. 93–115, 1995.