



Spatio-temporal parse network-based trajectory modeling on the dynamics of criminal justice system

Han Yu, Shanhe Jiang & Hong Huang

To cite this article: Han Yu, Shanhe Jiang & Hong Huang (2022) Spatio-temporal parse network-based trajectory modeling on the dynamics of criminal justice system, Journal of Applied Statistics, 49:8, 1979-2000, DOI: [10.1080/02664763.2021.1887101](https://doi.org/10.1080/02664763.2021.1887101)

To link to this article: <https://doi.org/10.1080/02664763.2021.1887101>



Published online: 16 Feb 2021.



Submit your article to this journal [↗](#)



Article views: 98



View related articles [↗](#)



View Crossmark data [↗](#)



Spatio-temporal parse network-based trajectory modeling on the dynamics of criminal justice system

Han Yu ^a, Shanhe Jiang^b and Hong Huang^c

^aDepartment of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, CO, USA;

^bDepartment of Criminal Justice, Wayne State University, Detroit, MI, USA; ^cSchool of Information, University of South Florida, Tampa, FL, USA

ABSTRACT

We extend the existing group-based trajectory modeling by proposing the network-based trajectory modeling based on judicious design and analysis of a spatio-temporal parse network (STPN) as a representation of neighborhood structure that evolves in time. The STPN offers a principled qualitative specification for an explicit paradigm framework to deal with complex real-world problems. The framework is completed by developing a quantitative specification of latent field representation to merge seamlessly on or alongside the established STPN via hierarchical modeling. The models adopt spatial random effects to characterize the heterogeneity and autocorrelation over the locations where nonlinear trajectories were observed. The trajectories are then investigated in the presence of the operational constraints of the dependence structure induced by the spatial and temporal dimensions. With the framework, complex developmental trajectory problems can be discerned, communicated, diagnosed and modeled in a relatively simple way that interpretation is accessible to nontechnical audiences and quickly comprehensible to technically sophisticated audiences. The proposed modeling is applied to address the challenges of the trajectory modeling of nonlinear dynamics arising from a motivating criminal justice empirical process.

ARTICLE HISTORY

Received 4 November 2019

Accepted 1 February 2021

KEYWORDS

Developmental trajectory; group-based trajectory modeling; network-based trajectory modeling; Markov Gaussian random field; spatial-temporal data

1. Introduction

1.1. Developmental trajectory

Today, the problems with non-negligible spatial and temporal components are ubiquitous in behavioral, biological or physical sciences. We are now in the era of an ongoing revolution brought about by information and communication technologies. The deluge of large complex data indexed by space and time, referred to as spatio-temporal (ST) data, are commonly generated or collected in diverse scientific fields as well as their availability is increasing in scientific and public data sources. In the future ubiquitous computing society, people can receive the most appropriate personalized information for action given their particular circumstances at any time and in any space [44]. The presence and coupling of

CONTACT Han Yu  han.yu@unco.edu

spatial and temporal information in the ST data introduce novel problems, challenges, and opportunities for developmental trajectory modeling, which has broad application fields of scientific and commercial significance, such as social sciences, neuroscience, epidemiology, healthcare, agriculture, transportation, and climate science in general and markets, supply chains, social networks and vehicular networks enabling AI in networking in particular.

Developmental trajectory describes the progression of any behavioral, biological or physical phenomenon. Data with a time-based dimension provide the empirical foundation for the analysis of developmental trajectories – the evolution of an outcome of interest over time. Representing, estimating and understanding developmental trajectories are among the most fundamental and empirically important research topics in the social and behavioral sciences and medicine [29,31]. Of the research topics in the longitudinal analysis, many of the most interesting and challenging problems have a qualitative dimension that allows for potential meaningful subgroups within a population based on some similarity measure. For example, psychology, biology and medicine have a long tradition of taxonomic theorizing about distinctive developmental progressions of subcategories. On one hand, the research problems with a taxonomic dimension aim to chart out the distinctive trajectories, to understand what factors account for their distinctiveness and to test whether individuals following the different trajectories also respond differently to a treatment, such as a medical intervention, a major life event, such as the birth of a child, or a political circumstance, such as a distinctive social or economic policy. On the other hand, these subgroups follow distinctive developmental trajectories that are not identifiable *ex ante* based on some measured set of individual or population characteristics (e.g. socio-demographic variable or socio-economic status).

1.2. The group-based trajectory modeling

To analyze the developmental trajectories in criminology, Nagin and Land [30] laid out the statistical method that has come to be called group-based trajectory modeling (GBTM) in the criminology literature to address issues related to the ‘hot topic’ of the time – the criminal career debate. Those issues were: ‘First, is the life course of individual offending patterns marked by distinctive periods of quiescence? Second, at the level of the individual, do offending rates vary systematically with age? In particular, is the age-crime curve single peaked or flat? Third, are chronic offenders different from less active offenders? Do offenders themselves differ in systematic ways?’ The dominant legacy of Nagin and Land [30] was not its answers to the specific questions but the methodology itself. GBTM is one of the few examples of a statistical method with origins in criminology that has come to be widely used by other substantive disciplines in addition to many applications in criminology for the longitudinal study of crime phenomena [3,39]. In clinical psychology, the GBTMs have been applied to understand the etiology and developmental course of a number of different types of disorders, including depression [7,27], inattention/hyperactivity [19], post-traumatic stress disorder [35], substance abuse [17], and conduct disorder [33], to capture heterogeneity in treatment responses to clinical and randomized trials [2,37], and to facilitate causal inference in epidemiological observational studies [14,15,32]. In medicine, the GBTMs have been applied to study the developmental course of psychiatric disorders with target biomarkers such as body mass index [28], cortisol levels [50,51], as well as indicators of disability in elderly populations [12]. Across all application domains,

the group-based trajectory statistical method lends itself to the presentation of findings in the form of easily understood tabular and graphical data summaries. This form of data summaries has the great advantage of being accessible to nontechnical audiences and quickly comprehensible to audiences that are technically sophisticated.

1.3. Extension of the group-based trajectory modeling

The group-based trajectory modeling is a parametric approach based on the data only with a time-based dimension. These observed values are considered as independent realizations of longitudinal trajectories. However, the hypothesis of independence is no longer acceptable in the presence of the incurred spatial interactions when the observed longitudinal values are anchored in space. The space might be geographic space, or socio-economic space, or more commonly network space of a variety of scales. Ignoring the dependencies induced either by the spatial dimension or by the temporal dimension in statistical analyses can suffer losing essential information and bear the risk of spurious conclusions regarding significance statements. In other words, in the study where the data correspond with observing a single process in which dependence occurs due to omitted or unmeasured space-based or time-based processes, one cannot assume that the data are the realization of a collection of independent experiments, the typical artificial assumption most statistical methods rely on. The models without considering spatial component or temporal component may cause omitted-variable bias as a result of attributing the effect of the omitted covariates loaded from the underlying spatio-temporal confounding factors to the estimated effects of the included variables. Spatial and temporal correlations among observations discount the sample size in computing standard errors or posterior standard deviations for targeted structural fixed effects in statistical inference under the departures from the independence assumption. The departures seriously affect standard statistical procedures that assume independent observations [4]. Thus the spatial information, as well as the temporal information, is an indispensable key component in modeling and understanding the observed data either of or from a complex system conceptualized as a network in the presence of operational constraints and environmental disturbances.

In the past empirical literature, the development of most theoretical work on spatial analysis has assumed an ideal space. The ideal spaces are convenient for developing pure theories of spatial stochastic processes. However, they are far from the real world where events are constrained by the networking of objects in space and time. If trajectory modeling and analysis assuming a plane with Euclidean distances is applied to events that occur on and alongside real spatio-temporal networks, then it is likely to lead to false conclusions [34]. Network representations of units in space and time are becoming extremely important in practice since they determine the quality of spatio-temporal modeling for a complex system that involves identifying the spatial and temporal entities or components and their roles, defining basic or novel types of static and nonlinear dynamic relationships among the components, and developing effective approaches for discovering the influences among the components.

To expand the capability of GBTM, a partition-and-group framework is proposed to accommodate space-based groups of trajectories as well as time-based groups of trajectories. One unique quality of such a study that differentiates ST data from other non-ST data in the classical statistical analysis literature is the presence of structural dependencies

among the objects induced by the important and indispensable spatial and temporal dimensions of the real world. Since spatial or temporal dependency is due to the unknown or unobserved latent variables, it is more interesting in modeling the covariance function appropriately to offset the effects of the unobserved variables and get more valid estimate of a key explanatory variable effect on a phenomenon. The spatio-temporal models include the two components, a systematic component with available explanatory variables and the spatial and temporal correlation component, and how the two components interact across processes and scales of variability. Such spatio-temporal studies enable us to predict big gaps of unknown values at unmeasured locations, identify unusual regions, forecast values at future times, and produce maps.

While traditional statistical methods typically rely only on the information contained in the data, the nature of the complexity of dependence and the paucity of observations for high granularity of spatio-temporally associated scientific questions requires a framework of integrating data science methods with the wealth of scientific domain expert insight or practical experience, often encoded as domain-based prior models to accelerate scientific discovery from ST data. In this sense, we propose spatio-temporal parse network-based trajectory modeling (NBTM) as an extension to the existing GBTM, raising the novel challenges induced by the coupling of spatial and temporal dependence structure in the today's ubiquitous spatio-temporal data.

We propose a nonparametric framework that begins with judicious design and analysis of a spatio-temporal parse network (STPN) on the positions in the spatio-temporal space that accommodates the nonlinear developmental trajectories observed in a sequence of time points. STPN is represented as a hybrid graph of directed and undirected edges connecting the nodes structured in Figure 1. Each trajectory of the empirical process on STTP is the series of events on the nodes along the sequence of time points grouped inside a blue dashed rectangle in Figure 1 for the dynamics of the empirical process coupled with spatial covariance $C(s_i, s_j)$ which is non-zero if the locations s_i and s_j are neighbors denoted as $s_i \sim s_j$. The role of STPNs is to make transparent what is the investigator's understanding as well as explicit what are scientific queries. STPN enables us to encode an honest assumption for the dynamics of a complex system under study. The more realistic assumption assumes that the data structure is observed from a single experiment on a sample of objects

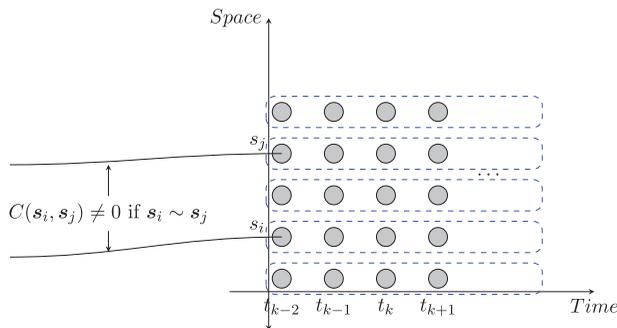


Figure 1. The ensemble of nodes in spacetime that induces STPN to accommodate trajectories with heterogeneity and autocorrelation between paths of nodes enclosed within blue dashed rectangles for trajectories.

structured via a latent spatio-temporal parse network. The stochastic progression of a phenomenon then can be identified with network-constrained trajectories with the dependence structure induced by both the spatial and temporal dimensions. In other words, as stated in the First Law of Geography: ‘Everything is related to everything else, but near things are more related than distant things’, units close to each others in space and time more likely share a set of latent background conditions as well as some graph characteristics which are related to the states of interest, so to have similar incidence. STPN is introduced in the proposed paradigm framework as a base working representation of the complexity of granularity of real life. By base we mean that it is a null hypothetical constraint characterizing current state of knowledge that is explicitly defined as a temporally indexed networking structure transparent for inference and reasoning as well as intuitive contextual interpretation. Alternative hypothetical networks can be proposed against the null for the discovery of important substructures at different granularities through a further analysis of network and assessed against the baseline network to avoid over-partition or under-partition in the network. Network events consist of on-network events and alongside-network events. We call a STPN-constrained event or network event for short to the event that occurs on and alongside a STPN.

The framework is completed by developing a quantitative specification of latent field representation to merge seamlessly on or alongside the established STPN using Bayesian hierarchical models (BHMs) for statistical reasoning under uncertainty. The models adopt spatial random effects to characterize both the autocorrelation and heterogeneity over the locations where developmental trajectories were observed. The trajectories are treated as an empirical process that is investigated in the presence of the operational constraints in the dependence structure induced by the spatial and temporal dimensions of the real world. Stochastic processes play an important role in scientific inference and interpretation in various domain applications since it is on the stochastic processes that scientific theories are postulated [5]. In contrast to the classical statistical modeling that only includes data model and parameter model, stochastic processes are made explicitly in the network-based probabilistic representation with less assumption-laden approach, providing a rich framework to recognize and capture additional sources of uncertainty on the structure of dependence and heterogeneity over and above ordinary statistical unstructured variation. In this sense, the STPN-based model is an expressive probabilistic representation of the spatial dependency structure that evolves through time for the nonlinear trajectories. With the framework, the problems and challenges of complex developmental trajectories can be discerned, communicated, diagnosed and modeled in a relatively simple way that contextual interpretation is accessible to nontechnical audiences and quickly comprehensible to technically sophisticated audiences, which is inherited from the existing GBTM.

Hence the notion of GBTM is appropriately extended to incorporate the spatial dimension of the real world and accommodate the complexity of the spatio-temporal dependence structure. Including additional space-based dimension in the existing GBTM allows for analysis of the interrelationship of heterogeneous trajectories more accurate and tracking the course of an outcome of interest more spatially coherent in a nonlinear dynamic complex system. In contrast to the existing GBTM assuming data structure generated from a repetition of independent experiments on a sample of independent units through time for the sake of simplicity, the proposed NBTM adopts the realistic assumption that the data structure is observed from a single experiment on a sample of STPN networking units.

The role of STPN in the probabilistic modeling is to provide convenient, transparent and explicit means of expressing the substantive assumption, facilitate economical representation of joint probability functions, facilitate efficient inferences from observations, and accelerate scientific discovery with ST data. While the proposed NBTM is developed for the nonlinear dynamics of a complex phenomenon under study as an extension to the existing GBTM by assuming more realistic assumptions, it raises the novel worthwhile challenges induced by the coupling of spatial and temporal dependence structure in the today's ubiquitous spatio-temporal data. The primary purpose of this article is to present a real example from our collaborative applied research in criminal justice to illustrate the challenges of the trajectory modeling arising from the complexity of the nonlinear dynamics of the motivating criminal justice process, along with elaborating the proposed solution to the challenges through a deep Bayesian hierarchical model, given the uncertainties in observations, dynamic process, and associated parameters. The article illustrates the limitations with the existing GBTM and shows the advantages of statistical models that account for spatio-temporal dependence.

The article is structured as follows. Section 2 describes the motivating data. The NBTM is developed in Section 3. Section 4 presents the methodological and empirical results. The paper is concluded with a discussion of the statistical and practical implications in Section 5.

2. The data

The ST data for the research question to the formal social control problem in general motivated the development of a new method to expand and refine the existing GBTM that does not address the spatially heterogeneous and correlated trajectories. We will illustrate the proposed NBTM with the ST data to extend the existing GBTM by incorporating the spatial dimension of the real world. The primary *goal* is to formulate a substantial knowledge-based statistical representation with a focus on the nonlinear dynamics of the procuratorate system through the proposed NBTM to address the novel challenges induced by the coupling of spatial and temporal information in the ST data.

The criminal justice system of the People's Republic of China has been evolving under the prevailing political and economic circumstances from 1949 to 2004 with a major transition in 1979 from planned economy into a market economy through internal reform and opening to integrate into the global market economy. The procuratorate subsystem is of interest in this study due to its significant role in the criminal justice system in China. The procuratorate subsystem is one of the four important components of the criminal justice system with the other three components of the public security, the courts, and the corrections. The major responsibilities of the procuratorate included, but not to be limited to, supervising the law enforcement, making public prosecution on behalf of the State, and investigating criminal cases. In addition to its supervisory role over all aspects of civil government, the government's procurators are responsible for deciding whether someone should be arrested and charged.

Our analysis is based on the data of 2004 First National Economic Census on the inceptions of procuratorate organs for Mainland China, containing information on the 2680 counties of the 31 provinces from 1949 to 2004. Procuratorate organs clustering in a jurisdiction region tend to share similar characteristics. It is beneficial for borrowing

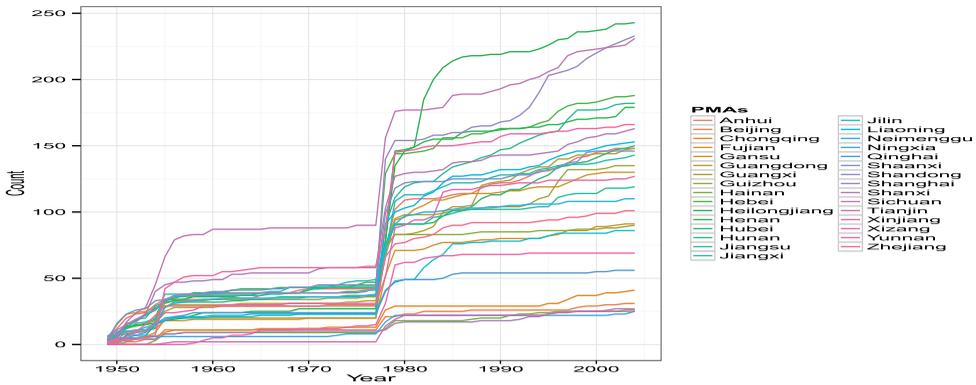


Figure 2. Heterogeneous and auto-correlated developmental trajectories.

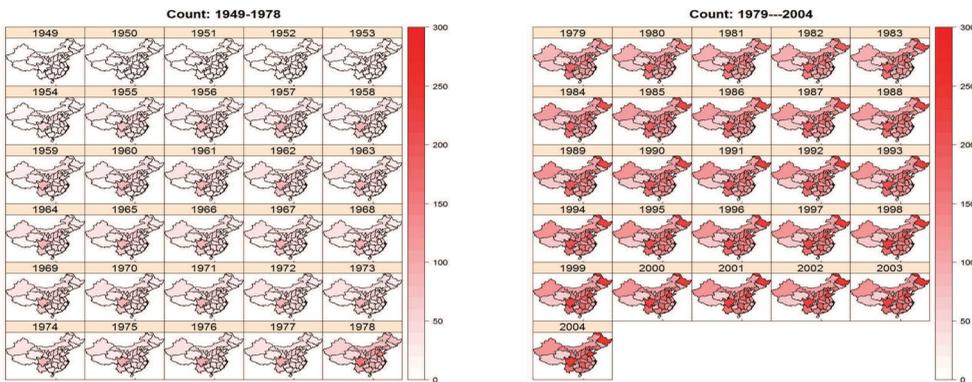


Figure 3. The dynamical geographical mapping of the developmental trajectories.

strength and reducing computationally cost by aggregating the network-constrained rare events over administrative areas. The diffusion of developmental trajectories of the organs is displayed in Figure 2 with increasing within-variability over time. To express data in an appealing and interpretive way, the geographical mapping of the procuratorate organs data are displayed in Figure 3 specific to administrative areas, namely, provinces, before and after the major transition of state policy respectively for diffusion and dynamics. Geographical map can add vital context that is generally lost if data is viewed only through a spread sheet and allow the public to get better data interpretation, gain better insight and stay informed. Heterogeneity coupled with the strong skewness and excess zeros in each of the areas was observed in Figure 4. The source of heterogeneity is often geographically due to socio-demographic and socio-economic origins exclusive to each distinct region in different time periods. The feature of excess zeros is observed in Table 1 that displays the numbers of zeros versus the total counts of the inceptions across areas. The zero inflation implies that there are two hidden competing states in structure, typical and atypical trajectories, conceptualized as activity and quiescence over different periods of time.

Although there are qualitative studies comparing criminal justice systems in the criminology literature, the trajectory modeling of criminal justice systems has been underexplored in the literature. Neither existing large scale complex data nor existing quantitative

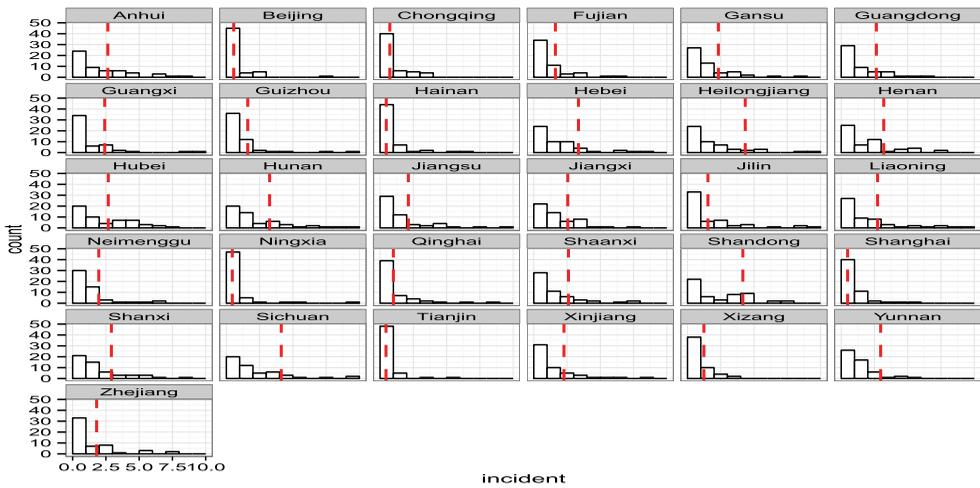


Figure 4. The distribution of increments for each area with its mean count indicated by a red dashed line.

Table 1. The distribution of the zero increments versus the associated total counts in 2004.

s_j	Anhui	Beijing	Chongqing	Fujian	Gansu	Guangdong	Guangxi	Guizhou
Zeros	24	45	40	34	27	29	34	36
Total	148	31	41	91	130	148	135	90
s_j	Hainan	Hebei	Heilongjiang	Henan	Hubei	Hunan	Jiangsu	Jiangxi
Zeros	44	24	24	25	20	20	29	27
Total	26	188	243	179	150	182	119	143
s_j	Jilin	Liaoning	Neimenggu	Ningxia	Qinghai	Shaanxi	Shandong	Shanghai
Zeros	30	47	22	33	28	22	40	21
Total	86	153	110	25	146	233	27	163
s_j	Shanxi	Sichuan	Tianjin	Xinjiang	Xizang	Yunnan	Zhejiang	
Zeros	39	20	48	31	38	26	33	
Total	56	231	25	127	69	166	101	

approaches have been considered to study the progression trajectory of a criminal justice system. To fill the gap, NBTM is proposed to obtain an insight into the most fundamental problems in criminal justice.

3. Spatio-Temporal parse network-Based trajectory modeling

With the observational data, we elaborated the network-based trajectory modeling with a representation that allows us to learn about or predict the spatio-temporal process and focus on new features of the data about which we were unaware.

The NBTM can schematically be represented as a hierarchical model (HM) [5]. For $i = 1, 2, \dots, n$,

$$\text{Data Model : } Y_i(t) | \eta_i(t), \theta_d \sim \mathcal{D}(\eta_i(t), \theta_d), \tag{1}$$

$$\text{Latent Random Field : } \eta_i(t) = \beta_0 + \sum_{m=1}^p \beta_m x_{mi}(t) + \omega_i(t), \tag{2}$$

$$\text{Dynamic Process : } \omega_i(t) = \mathcal{M}_\tau(\omega(t), \Phi_\omega) + \xi_i(t) \tag{3}$$

$$\text{Regularization : } \Phi_\omega \sim \pi_1(\theta_r), \tag{4}$$

$$\text{Residual Process : } \xi_i(t) \sim \pi_2(\theta_u), \tag{5}$$

$$\text{Parameters : } \theta = (\beta_0, \beta, \theta_d, \theta_r, \theta_u) \sim \pi(\beta_0, \beta, \theta_d, \theta_r, \theta_u). \tag{6}$$

$\mathcal{D}(\cdot)$ in data model (1) is some distribution for data $Y_i(t)$ for site s_i at time t from exponential family or mixture distributions given the latent random field $\eta_i(t)$ and parameter θ_d . The latent random field (2) contains the spatio-temporal process $\omega(t) = (\omega_i(t))_{i=1}^n$ where $\mathcal{M}_\tau(\cdot)$ is an evolution operator of lag τ and $\xi_i(t)$ is a noise process specified in model (5). $\omega_i(t)$ is often over-parameterized and regularized by model (4). $\beta = (\beta_1, \dots, \beta_p)$ is the coefficient vector for the vector of p exogenous covariates $x_i(t) = (x_{1i}(t), \dots, x_{pi}(t))$. The HM is completed by the prior distribution $\pi(\cdot)$ specified in model (6) for θ .

3.1. Data models: finite mixture models

We begin with the assumption on the data generating process. For models where latent effects are to some extent identified by the specification of their prior distributions generally involving additional parameters known as hyperparameters, there may be sensitivity in posterior inferences to the assumed priors. While one strategy known as informal sensitivity analysis is to consider a limited range of alternative priors and assess changes in inferences, more formal approaches to robustness are to base on non-parametric priors, or via mixture (‘contamination’) priors. When distinct groups present in the population can not be identified by settled theory, finite mixture models (FMM) are a ‘niche between parametric and nonparametric approach to statistical estimation’[26]. Using FMMs allows a better understanding of the possible relaxations of the baseline assumption with the latent variable(s) as well as the components of the models with substantial scientific knowledge. For a clean and intuitive illustration, we focus only on the alternative models using FMMs for different assumptions concerning mixing distributions, resulting in more flexible sampling distributions. A general form of mixing allows the sampling distribution sufficiently flexible to accommodate any realistic distribution to some desired degree of accuracy.

The trajectory can be considered as a realization from a Poisson associated process based on a STPN with the area-specific intensity

$$\lambda_{s_i}(t) = \exp\{\eta(s_i, t)\}, \quad i = 1, 2, \dots, n,$$

where $\eta(s, t)$ is a latent dynamic random field. We will simplify the notation by using the index i to denote area s_i in the following presentation.

We start with the baseline assumption of likelihood function that each spatially indexed incident is conditionally independent in area s_i at time period t given the time-varying area-based intensity rate $\lambda_i(t)$ measuring incident proneness. The conditional distribution of the count of events $Y_i(t)$ occurring in a certain area s_i at a fixed period of time t follows the Poisson distribution

$$Y_i(t)|\lambda_i(t) \sim \text{i.i.d. Poisson}(\lambda_i(t)). \tag{7}$$

Model (7) is specified as a baseline model that encodes a state of minimum information against which other kinds of assumed structural heterogeneous trajectories can be assessed.

In other words, we begin with a model specification under the assumption of complete randomness for trajectories and then investigate various kinds of departures from this hypothesized model by relaxing the assumption. There are often multiple competing alternative model specifications arising either from theoretical propositions or from alternative specifications of the same theory in practically inclined researches.

Although the likelihood function (7) from the exponential family is convenient, it is not flexible for some real-world problems with extra variation. The variance of the Poisson distribution is expected to be equal to the mean count of occurrences. As the mean count increases, the skewness diminishes, and the distribution becomes approximately normal. When the mean count is low, then the data consists of mostly low values and less frequently higher values resulting in a distribution with a long right tail. Counts of zeros become increasingly likely as the mean count approaches zero. The fitting of a single Poisson distribution often forces too much structure on the data with extra-Poisson variation problems such as overdispersion and excess zeros. In practice, count data frequently depart from the Poisson distribution due to a larger frequency of extreme observations resulting in the variance considerably greater than the mean in the observed distribution called overdispersion. Overdispersion reflects some combination of unexplained variation in the observed data for regions and dependence structure underlying the observed data.

A mixture of Poisson distributions to address the overdispersion can be represented as the conditional Poisson-Gamma mixture (also known as negative binomial) model

$$Y_i(t)|\lambda_i(t), \zeta \sim \text{i.i.d. Poisson}(\lambda_i(t)\zeta) \quad \text{and} \quad \zeta \sim \text{Gamma}(a, b). \quad (8)$$

We consider zero-inflation as well when the distribution of counts has a much larger than the expected count of zeros assumed by Poisson distribution. The conditional Poisson-Bernoulli mixture model can be used to capture an excess of zeros that cannot be estimated by Poisson distribution:

$$Y_i(t)|\lambda_i(t), \delta \sim \text{i.i.d. Poisson}((1 - \delta)\lambda_i(t)) \quad \text{and} \quad \delta \sim \text{Bernoulli}(\pi_0), \quad (9)$$

with hyperparameter π_0 as the proportion of extra zeros introduced to capture the two desired states: activity and quiescence. In other words, the model assumes that the zeros have two different origins: ‘structural’ origin and ‘sampling’ origin. Structural zeros are observed due to some specific structure in the data while sampling zeros are due to the usual Poisson distribution whose zeros are assumed to happen by chance. A uniform prior for π_0 was used if no information is available.

3.2. Process models: STPN-based structured trajectory modeling

To accommodate spatially constrained trajectories, we adopt an appropriate spatio-temporal parse network on Figure 1 that is a time-varying spatial network. We adopt a generic structured additive predictor [8] to specify the trajectories over the chain graph whose topology and/or attributes can change with time,

$$\eta_i(t) = \mathbf{X}_i\boldsymbol{\beta} + b_i + \gamma(t) + \delta_i(t), \quad i = 1, \dots, n, \quad (10)$$

where b_i is the overall spatial random effect at site s_i , $\gamma(t)$ the temporally structured trend, and $\delta_i(t)$ the spatio-temporal interaction, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ the vector of linear fixed effects of the vector of covariates $\mathbf{X}_i = (1, X_{i1}, \dots, X_{ip})$.

3.2.1. Spatial component

A variety of specifications have been proposed for the latent level of spatial random effects $\mathbf{b} = (b_1, \dots, b_i, \dots, b_n)^T$. We focus on random effects \mathbf{b} to account for extra-Poisson variation or spatial correlation due to unobserved or unmeasured heterogeneity or structural correlation if the exponential family distribution or associated finite mixture models are used. Due to inherent sampling variability it is not recommended to inspect crude estimates directly, but borrow strength over neighboring regions to get more reliable region-specific estimates. The spatial effect \mathbf{b} in the classical Besag-York-Mollié (BYM) model is decomposed into a sum of an unstructured and a structured spatial component $\mathbf{b} = \mathbf{u} + \mathbf{v}$. While the structured spatial effect \mathbf{u} represents the fact that outcomes among the regions close in space are more correlated than among distant regions, the unstructured spatial effect \mathbf{v} represents the clustering information for the regions in the data. It is considered as a surrogate for unobserved i.i.d. region-specific random effects that can not be captured by the smooth spatial trend. In our application, the random effects account for the possible overdispersion caused by unobserved heterogeneity due to regional differences in local statutes, organization, funding, and policies. Neglecting unobserved heterogeneity may lead to considerably biased estimates for the remaining effects and false standard error estimates.

While unstructured spatial random effect component $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \tau_v^{-1}\mathbf{I})$ with precision τ_v accounts for pure overdispersion due to Poisson distribution assumption, an intrinsic Gaussian Markov random field (IGMRF) prior with precision τ_u [43] was assigned to the structured spatial random effect of a particular region s_i that depends on the effects of all neighboring regions as follows,

$$u_i | \mathbf{u}_{-i}, \tau_u \sim \mathcal{N} \left(\frac{1}{n_{\partial s_i}} \sum_{j \in \partial s_i} u_j, \frac{1}{n_{\partial s_i} \tau_u} \right), \tag{11}$$

where \mathbf{u}_{-i} denotes the vector of all spatial effects excluding site s_i , ∂s_i contains all the direct neighbors of s_i and $n_{\partial s_i} = |\partial s_i|$ is the number of neighbors for site s_i . The IGMRF component tends to produce similar estimates for u_i and u_j if areas s_i and s_j are geographically close in addition to an unstructured spatial random effect component v_i that accounts for independent region-specific noise. We consider the notion of neighbors and neighborhoods to use spatial smoothing techniques with shortest path distances constrained to the established spatio-temporal parse network where the Euclidean distances between two study regions are not appropriate due to the discreteness in the aggregate data. Two regions are treated as neighbors if they share a common boundary. The application is restricted to the prior based on adjacency weights resulting in a sparse structure matrix Q in this paper. The resulting covariance matrix of \mathbf{b} is $Var(\mathbf{b} | \tau_u^{-1}, \tau_v^{-1}) = \tau_u^{-1}Q^{-} + \tau_v^{-1}\mathbf{I}$, where Q^{-} is the generalized inverse of Q .

When the spatial effect b_i is decomposed into the spatially structured component u_i and the unstructured component v_i in the BYM model, it is not clear to see how the spatially structured component is distinguished independently from the unstructured component. This potential confounding makes prior definitions difficult for the hyperparameters of the two random effects. To address the non-identifiability problem, there are various alternative reparameterized models have been proposed using a mixture model structure, in which the precision parameters of the two components are replaced by a common precision

parameter and a mixing parameter. The mixture model structure distributes the variability between the structured and unstructured components. Of the alternative specifications, the Leroux model and the Dean model are commonly used with only one random effect component [6,25,49].

However, two issues should be addressed in the mixture model structure. First, the precision parameter does not represent the marginal precision but is confounded with the mixing parameter if the spatially structured component is not scaled. The effect of any prior assigned to the precision parameter thus depends on the graph structure of the application. Then a given prior is not transferable between different applications if the underlying graph changes. Second, the choice of hyperpriors for the random effects is not straightforward. Simpson et al. [46] proposed a new BYM parameterization that accounts for scaling and provides an intuitive way to define priors by taking the model structure into account. This new model provides a new way to look at the BYM model and a sensible model formulation where all parameters have a clear interpretation. The model structure is similar to the Dean model [6], with the crucial modification that the precision parameter is mapped to the marginal standard deviation. This makes the parameters of the model interpretable and facilitates assignment of interpretable hyperpriors. The framework of penalized complexity (PC) priors is applied to formulate prior distributions for the hyperparameters. The spatial model is thereby seen as a flexible extension of two simpler base models towards which it will shrink, if not indicated otherwise by the data. The upper level base model assumes a constant response surface, while the lower level model assumes a varying response surface over space without spatial autocorrelation.

Riebler et al. [41] introduced a modified BYM model that splits the variability independently over the spatial random effects to address the identifiability and scaling

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}} \left(\sqrt{1 - \phi} \mathbf{v} + \sqrt{\phi} \mathbf{u}_* \right)$$

with covariance matrix $\text{Var}(\mathbf{b} | \tau_b, \phi) = \tau_b^{-1} [(1 - \phi) \mathbf{I} + \phi \mathbf{Q}_*^-]$, where \mathbf{Q}_*^- is the precision of the scaled \mathbf{u}_* . In the design of the spatial random effects, the hyperparameters can be seen independently from each other to improve parameter control. Furthermore, the scaled spatial component facilitates assignment of meaningful hyperpriors and make these transferable between spatial applications with different graph structures. The hyperparameters themselves are used to define flexible extensions of simple base models. Consequently, penalized complexity priors for these parameters can be derived based on the information-theoretic distance from the flexible model to the base model, giving priors with clear interpretation.

3.2.2. Nonlinear dynamics

The temporally structured effect $\gamma(t)$ is specified dynamically using a set of temporal random effects to capture the nonlinear trend with an intrinsic conditional autoregressive (ICAR) prior on P-splines. P-splines assume that the unknown functions can be approximated by basis functions representation of B-splines, $\sum_{k=1}^K c_k B_k(t)$, where $K = m + q$ and q is the degree of a polynomial spline defined based on a set of $m + 1$ knots $t_{\min} = k_0 < k_1 < \dots < k_m = t_{\max}$ and $\mathbf{c} = (c_1, c_2, \dots, c_K)^T$ is the vector of unknown coefficients. The second-order random walks of the coefficients was used with smoothing priors

with precision τ_γ ,

$$c_k | c_{k-1}, \dots, c_1, \tau_\gamma \sim N(2c_{k-1} - c_{k-2}, \tau_\gamma^{-1}), \quad k = 3, 4, \dots, K,$$

and diffuse priors $\pi(c_1) \sim \text{const.}$ and $\pi(c_2) \sim \text{const.}$ for initial values [1,24].

3.2.3. Space–Time interactions

In addition to the main spatial and temporal effects, an interaction between area and time is modeled through addition of an interaction term $\delta_i(t)$ that combines spatial and temporal structured effects defined on Figure 1. The interaction would explain differences in the developmental trajectory for different areas. The parameter vector δ follows a Gaussian distribution with a precision matrix $\tau_\delta \mathbf{R}_\delta$, $\delta \sim \mathcal{N}(\mathbf{0}, (\tau_\delta \mathbf{R}_\delta)^{-})$, where τ_δ is unknown scalar while \mathbf{R}_δ is the structure matrix, identifying the type of temporal and/or spatial dependence between the elements of \mathbf{R}_δ . There exist various specifications for the structure matrix \mathbf{R}_δ . Knorr-Held (2000) proposed four ways to define the structure matrix [23].

Type I interaction assumes that the two unstructured effects v_i (spatial effect) and γ_t (temporal effect) interact. The structure matrix for this type is expressed as $\mathbf{R}_\delta = \mathbf{R}_v \otimes \mathbf{R}_\gamma = \mathbf{I} \otimes \mathbf{I} = \mathbf{I}$. Since both v and γ do not have a spatial or temporal structure, an identically independent non-informative normal model for $\delta_{it} \sim N(0, \tau_\delta^{-1})$ is used.

Type II interaction is specified as the interaction between the structured temporal main effect $\gamma(t)$ and the unstructured spatial effect v_i with interaction structure matrix $\mathbf{R}_\delta = \mathbf{R}_v \otimes \mathbf{R}_\gamma$, where $\mathbf{R}_v = \mathbf{I}$ and \mathbf{R}_γ is a neighborhood structure that can be defined through a random walk. Thus, a random walk across time for each area independently from all other areas is assumed for δ_{it} . This implies the parameter vector $(\delta_{i1}, \dots, \delta_{iT})$ has an autoregressive structure on the time component for i^{th} area, which is independent from the ones of other elements.

Type III interaction is specified between the unstructured temporal effect γ_t and the structured spatial main effect u_i . The structure matrix is written as $\mathbf{R}_\delta = \mathbf{R}_\gamma \otimes \mathbf{R}_u$, where $\mathbf{R}_\gamma = \mathbf{I}$ and \mathbf{R}_u is specified as a CAR neighborhood structure. Thus, the parameters of the i^{th} time point in $\{\delta_{i1}, \dots, \delta_{iT}\}$ have a spatial structure independent from the other time points.

Type IV interaction combines spatial and temporal structured effects, namely, u_i and $\gamma(t)$. The resulting interaction matrix can be written as $\mathbf{R}_\delta = \mathbf{R}_u \otimes \mathbf{R}_\gamma$, which is the most complex interaction structure. The specification of the spatio-temporal interaction combines RW2 for temporal dependence and the BYM model for spatial dependence. Type IV interaction structure is more meaning than any of the types I, II and III interaction structures as a representation of neighborhood structure that evolves in time inherent in a complex phenomenon.

The proposed NBTM with type IV interaction is obtained by assigning priors to all the hyperparameters as an extra layer of modeling complexity and computational difficulty. To ensure the computationally tractability and avoid being technically involved though it might be argued that this is not a good strategy, the model specifications are completed by starting with heuristic weakly informative inverse Gamma $IG(a, b)$ priors for all the precisions to obtain a data-driven amount of smoothness using small values of a and b as a convenient default in the open source software for practical analysis [21,22]. Small values for a and b correspond to an approximate uniform distribution for $\log(\tau^2)$ with τ^2

equivalent to the smoothing parameters in the frequentist approach that controls the trade-off between smoothness and flexibility. In addition to the common inverse gamma priors that lead to Gibbs sampling updates, several other alternative prior specifications for the precision have been suggested as default priors in the statistics literature: scale-dependent priors [20], penalized complexity priors [46], and half-normal, half-Cauchy or approximate uniform priors for precision [9,10,16], among others. The selection of appropriate hyperpriors for the precision parameters is an important topic in all kinds of Bayesian regression models. The further discussion, such as axiomatic reasoning, about the suitability of the specifications is beyond the scope of this paper. As a sensitivity check, the model should be re-estimated with different choices for the parameters a and b .

4. Results

We considered the standard Poisson regression models, the Poisson-Gamma mixture regression models and the Poisson-Bernoulli mixture regression models each with group-based trajectory modeling (GBTM), spatial modeling (BYM), network-based trajectory modeling (NBTM) and network-based trajectory modeling with covariates (NBTMC) for the data. We end up with 12 models in total to compare. Bayesian inference was performed by Markov chain Monte Carlo (MCMC) algorithm implemented in R. We used 520,000 MCMC iterations with a burn-in phase of 20,000 and a thinning parameter of 500. The simulated values do not show anomalies, upward or downward trends but look like a random scatter around a stable mean value, the chain appears to have reached convergence. We observed that for any lag the autocorrelation is close to zero, suggesting that the simulated values can be considered almost independent.

Model evaluation is conducted by comparing the relative performance of the competing candidate models via the information-theoretic approach. The most commonly used measure of model fit based on the deviance for Bayesian hierarchical models is the deviance information criterion (DIC) used for discriminating between competing data distributions and predictor structure of the distribution parameters [21], proposed by Spiegelhalter et al. [46]. DIC is a generalization of the Akaike information criterion (AIC). DIC is decomposed into two terms: $DIC = \bar{D}(\theta) + p_D$, where $\bar{D}(\theta)$ is the posterior mean deviance measuring the fit to the data and p_D measures the complexity of the model. We ran three independent MCMC simulations as suggested by Gilks et al. [11] to avoid the possibility of the lack of convergence when only one chain is run. We saw that the differences between DIC values never exceed one unit.

The final results are summarized in Table 2 that are quite encouraging for NBTM. While the comparison displays a clear advantage of GBTM over spatial modeling, NBTM significantly outperforms GBTM in the three sampling distributions. With the same structure of spatial and/or temporal components, a naive data generating distribution directly from the exponential family is inadequate to accommodate the features of the conditional sampling distribution of complex dynamics given the spatio-temporal process and the Poisson-Bernoulli mixture model shows superior fit compared to the other models from Table 2. Since the Poisson-Bernoulli mixture model accommodates zero inflation in the data better than any other model being considered, this indicates that the distribution of the observed trajectories was zero-inflated due to the presence of both structural and

Table 2. Summary of DICs.

Data generating distribution	Modeling	DIC	p_D
Poisson	GBTM	34157.78	19.85
	BYM	52575.92	30.77
	NBTM	3200.56	50.22
	NBTMC	3002.66	50.95
Poisson-Gamma	GBTM	34157.97	20.10
	BYM	52575.92	30.78
	NBTM	3200.02	49.97
	NBTMC	3001.64	50.39
Poisson-Bernoulli	GBTM	33826.37	25.76
	BYM	50888.18	35.86
	NBTM	3117.02	45.68
	NBTMC	2900.64	28.88

Table 3. Variances of spatial effects.

Spatial effect	Mean	Sd	2.5%	50%	97.5%	Min	Max
Structured	0.1703	0.0827	0.0696	0.1498	0.3843	0.0414	0.6149
Unstructured	0.4595	0.1525	0.2387	0.4312	0.8321	0.1705	1.2645

sampling zeroes. That zero-inflated models fit better than their corresponding non-zero-inflated counterparts match the results from the exploratory data analysis. The structural zeros and sampling count distribution in a Poisson–Bernoulli mixture model can provide a more meaningful and precise interpretation of the clusters, activity and quiescence, underlying the observed feature of diverging trajectories in Figure 2. Under the same sampling distribution, the NBTMC fits the best based on the DIC in Table 2. The Poisson regression model and the Poisson-gamma mixture models with only pure spatial component are inferior to the other models in terms of DIC. This suggests the significance of spatio-temporal effect on the response.

The DIC values for the 12 models under consideration indicate a clear preference for the NBTMC-based Poisson–Bernoulli mixture model with the lowest DIC. The NBTMC-based Poisson–Bernoulli mixture model was selected to account for the excess structural zeros as well as the structural effect on the trajectories. We primarily focus on the results related to the research questions addressed by GBTM as discussed in the Introduction.

Figure 5 shows the relative estimated variation of the spatial effects: the range of the estimated unstructured effects clearly exceeds that of the structured effects. This implies that the spatial heterogeneity appears to be caused more by local circumstances of the locations than the influences from their neighbors responsible for the clear bifurcation in the increasingly divergent trajectories observed in Figure 2, with one cluster of trajectories illustrating a positive impact of temporal components (in year) while the other cluster of trajectories showing poor relation to time. The increasing dispersion is as a result of the unbalanced inherent latent socio-economic factors in addition to environmental conditions across the country: the regions differ in their local statutes, organization, funding, and policies regarding their justice systems. The mean of the variance for unstructured random effects is about 2.7 times that of the variance for structured random effects. When comparing the standard deviations of the effects, the standard deviation for unstructured random effects is 1.8 times that of the variance for structured random effects. When comparing the effects on 2.5% quantile to 97.5% quantile, the unstructured random effects have 95%

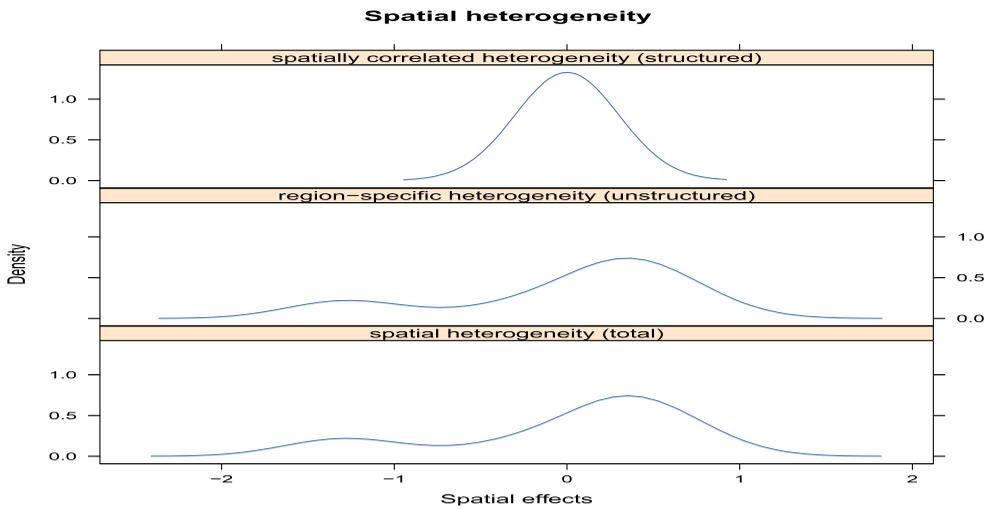


Figure 5. Estimated spatial effects: structured spatially correlated heterogeneity (top), unstructured spatial heterogeneity (middle) and total spatial heterogeneity (bottom).

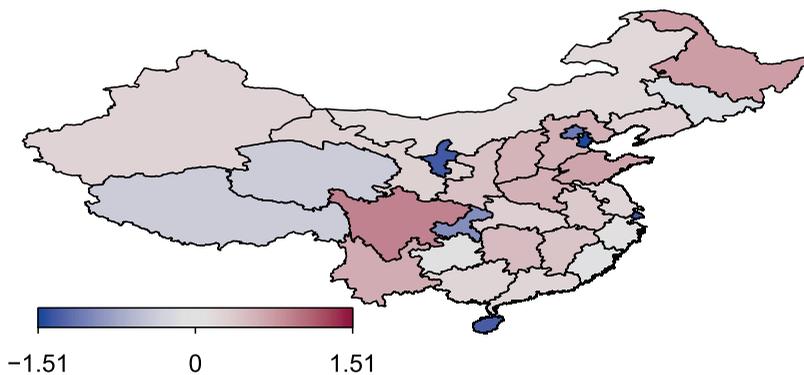


Figure 6. Geographical mapping of the posterior mean estimates of total spatial effects.

credible interval [0.2387, 0.8321] with median 0.4312 compared to the 95% credible interval [0.0696, 0.3843] with median 0.1498 for structured random effects. The geographical mapping of the estimated spatial effect in Figure 6 further indicates the spatial distribution of growing disparity among the regions.

In addition, the NBTM unveils a new interesting logistic growth phenomenon for the system dynamics [50]. The logistic growth curve is restricted by the realistic resources saturated in environmental constraints that are associated with the different regimes regarding political, social and economic changes in a period. The logistic growth theoretically originates from the study of human population dynamics. It has since been used to explain many biological and chemical phenomena. Following a scientific revolution of a system, progress generally follows a sigmoid curve: it starts with a period of fast progress, which gradually stabilizes as the system reaches the limitations of resource constraints, called carrying capacity that the environmental resources can support, and then further improvements become incremental. Figure 7 charts out the pronounced progression curve which

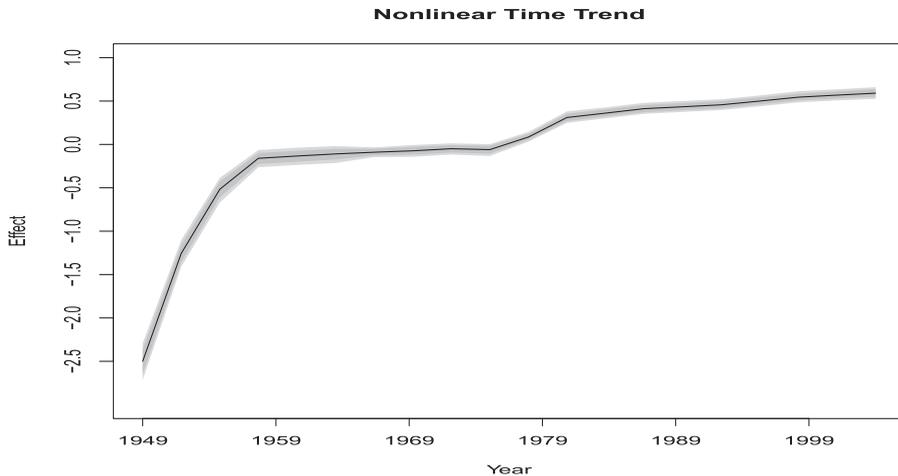


Figure 7. The posterior mean estimate of the dynamic effect revealing the phases of growth.

is significant in the sense that at least the 80% and 95% pointwise credible bands do not cover the zero line fully for distinct time-based clusters by the characteristic points of inflection. The result indicates that the growth rates of the developmental trajectories are time-varying: the rate of growth is faster for the inception of a new or innovative system in the beginning period and slows down later due to underlying competition for active and quiescent sources. When the system reaches an equilibrium state, namely, a saturation level characteristic of the environment in a period, the growth rate is close to zero. In contrast to parametric GBTM marking the distinctive periods of quiescence in the life course to address the 'hot topic' of the time about the criminal career debate, nonparametric NBTM identifies the abrupt transitions by the characteristic points of inflection for the distinctive time-based clusters in the spatio-temporal process between different regimes, revealing the phases of inception, expending, disruption, developing, and formulating: The estimated developmental trajectory climbs up as a J curve for the first 9 years ended at the year 1958 (1949–1958), reflecting a new system expansion. Then the dynamics has almost no significant influence for the next 20 years (1959–1979) until the end of the disruptive innovation in year 1979 for another logistic growth. After the disruptive innovation of open system there is a rapid increase for a few years. Finally, the chart shows the stable dynamics again in the range of observations.

Furthermore, the statistical patterns can be easily understood when the substantive map of spacetime is exploited to add vital context to express data for the diffusion and growth phenomenon. The distinctive time-based clusters are criminologically in line with the substantial contextual knowledge about the disruptive social innovations in the years of the survey. Since 1949, the PRC has experienced distinct political and economic circumstances. In 1949, the revolutionary innovation took place that the newly founded communist government declared that all the laws from the old regime were repealed and all the legal organs from the old regime were dissolved. The central government set up the Supreme People's Procuratorate and other legal organs. Following the trend, local governments at the province, prefecture and county levels also set up people's procuratorates and other legal agencies. Starting in the second half year of 1957, the disruptive Anti-Rightist

Campaign started. Law and criminal justice organs were seriously criticized not to obey the leadership of the Chinese Communist Party and almost stagnated during 1958–1965. From 1966 to 1976, Chinese People’s Liberation Army controlled courts, procuratorates and public security organs. When another disruptive innovation took place that the third plenary session of the 11th Central Committee of Communist Party of China was held in 1978, market reform started and rule by law was stressed to move the focus of the criminal justice system from class struggle to economic development. The development of criminal justice system was resumed and many more new procuratorates were created since 1978. Given a new distinct circumstance during 1979–2004 in contrast to 1949–1978, factors such as urbanization, population, economic level and crime rate may be expected to affect the developmental trajectory. In the course of moving toward a free market economy during 1979–2004, China was experiencing new types of crimes as well as crimes of a magnitude that did not exist before 1978. The new findings not only point to avenues of further inquiry in the domain but also suggest the substantial methodological value of the NBTM approach in understanding the dynamics of a system since it sheds light on a variety of subjects of interest to academics. With observational data, we never select the ‘true’ model and do not claim that the representation we proposed is optimal or unique. Instead, we make the claim that the proposed NBTM is effective, understandable, and transparent in the sense of reasoning under uncertainty compared to the existing GBTM. The modeling allows us to learn about or predict the spatio-temporal process and focus on new features of the data about which we were unaware.

5. Discussion

We propose NBTM augmented with additional STPN as structural constraints to accommodate complex growth characteristics in spacetime for revealing and understanding system dynamics. Methodologically, the proposed modeling is an extension and refinement of the existing parametric group-based trajectory modeling that was laid out by Nagin and Land to address the issues related to the ‘hot topic’ of criminal career debate [29,30]. It moved the formation of complexity and comprehensibility of trajectory modeling further in spatial and temporal dimensions within the scope of computational tractability provided by current algorithm and technology with mixture modeling for every space-based, time-based or non-ST feature-based cluster of trajectories. First, the nonparametric representation of NBTM can unveil the presence of more interpretable nonlinear growth dynamics with the context of STPN in contrast to the polynomial representation of GBTM specified for the presence of single peaked or flat of developmental trajectories. Second, the proposed NBTM conducts sensitivity analysis by a more general parameterization of mixing that enables the sampling distribution sufficiently flexible to approximate any realistic distribution to some desired degree of accuracy. The modeling can be considered as the hierarchical mixtures-of-experts (HMEs), combining aspects of finite mixture models and generalized nonlinear models. The mixture of generalized mixed effects models provides a comparatively fast learning and good generalization for nonlinear complex problems [26].

The modeling has practical implications in statistical learning. STPNs add vital context to express data dynamically with contextual understanding and better interpretation of growth and diffusion phenomena. With a marriage between graph theory and probability theory of hierarchical modeling, what is the investigator’s understanding

(known or assumed) of dynamics is transparent and what are the scientific queries (association or causation) is explicit for a spatio-temporal process under study. The dynamic graph is represented as a hybrid chain graph that respects the asymmetry in temporally directed relationships augmented primarily with spatially symmetrical relationships in undirected Markov network that denotes the existence of unobserved common causes by the Common Cause Principle of Reichenbach [40] where certain patterns of dependency, void of temporal information, are conceptually characteristic of certain causal directionalities [36]. When the structure of spatio-temporal parse network is integrated with the dependence structure of associated events in statistical hierarchical modeling, further causal inference methodology can be established, incorporating the benefits of machine learning with statistical inference that we have been working on. Although the method is elaborated on the application in the trajectory modeling of criminal system, we believe that the proposed NBTM with data-driven shrinkage towards the existing GBTM is transferable between the criminology and other substantive disciplines where the existing GBTM have successfully applied in the form of graphical and tabular data summaries accessible to nontechnical audiences and quickly comprehensible to technically sophisticated audiences.

A few limitations of this study were noted for future research. First, the spatio-temporal analysis on networking event should include the analysis of a network itself, such as geographical network analysis [13], communication network analysis [19], and circuit network analysis [48]. From the methodological perspective, the hierarchical architecture of the proposed STPN-based probabilistic representation allows the approaches developed from high-dimensional statistics and machine learning to combine with the techniques developed from the emerging network science. To search among different granularities for true discoveries, strong predictive power and interpretability of substantive spatio-temporal parse networks, statistical hypothesis significance testing can be introduced as an assessment of an alternative network against the proposed null network to address the over-partition or under-partition in discovering important substructures. However, the analysis of a network itself is non-trivial and often requires domain expertise since a network space does not imply a space consisting of networks [42] like a function space in mathematics [38]. In the network-constrained trajectory statistical analysis, the computation based on the measure of the shortest path distance is much more difficult than that of Euclidean distance because it requires the management of network topology. We used a meaningful level of granularity of network in our demonstration, where we exploited a technique of lowering the resolution of the representation of the spatio-temporal random effects to make model fitting faster. Our research can be improved by taking further analysis of the network itself that is beyond the scope of this paper. We intend this paper to contribute a first step toward micro-scale spatial analysis by spatio-temporal parse network-based trajectory modeling and develop further micro-scale spatial analysis in the future work to address the challenge of real-time spatial analysis.

Second, the practical problems in applied statistics are, by and large, computational in nature [45]. The scalability of model fitting is challenging in the spatio-temporal parse network-based trajectory modeling from simple to adequate graph support for inference and learning with high-dimensional data. Given that many situations on multivariate (conditional) density approximation require a reasonably large number of components and each component will have a very large number of parameters, efficient algorithms that can handle very high-dimensional spaces will be required for inference methods and is

still under ongoing research and development to the potential to address the simultaneous challenges encountered by all the fields. In practice, it requires much more CPU time and memory cost to implement the computational methods in a statistical model constrained alongside a more refined dynamic network graph in software.

Third, specification of hyperpriors in hierarchical models is still an area of methodological research. Further work needs to be done on the selection of good, sensible default hyperpriors for precision parameters specific to the proposed models with the spatial-temporal interaction for different goals, in which is technically involved.

Acknowledgments

We thank two referees, the Associate Editor and the Editor-in-Chief for their careful review which was very helpful in improving upon the initial submission.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Han Yu  <http://orcid.org/0000-0003-0639-4445>

References

- [1] A. Brezger and S. Lang, *Generalized additive regression based on Bayesian P-splines*, *Comput. Stat. Data Anal.* 50 (2006), pp. 967–991.
- [2] C.H. Brown, W. Wang, S.G. Kellam, B.O. Muthen, H. Petras, P. Toyinbo, J. Poduska, N. Ialongo, P.A. Wyman, P. Chamberlain, Z. Sloboda, D.P. MacKinnon, A. Windham, and The Prevention Science and Methodology Group, *Methods for testing theory and evaluating impact in randomized field trials: intent-to-treat analyses for integrating the perspectives of person, place, and time*, *Drug Alcohol Depend.* 95 (2008), pp. S74–S104.
- [3] S.D. Bushway and D. Weisburd, *Acknowledging the centrality of quantitative criminology in criminology and criminal justice*, *Criminologist* 31 (2006), pp. 1–3.
- [4] P. Congdon, *Applied Bayesian Modelling*, 2nd ed. John Wiley & Sons, Ltd., West Sussex, 2014.
- [5] N. Cressie and C.K. Wikle, *Statistics for Spatio-Temporal Data*, John Wiley & Sons, Inc., Hoboken, NJ, 2011.
- [6] C.B. Dean, M.D. Ugarte, and A.F. Militino, *Detecting interaction between random region and fixed age effects in disease mapping*, *Biometrics* 57 (2001), pp. 197–202.
- [7] M.C. Dekker, R.F. Ferdinand, N.D.J. van Lang, I.L. Bongers, I.L. J.D. van der Ende, and F.C. Verhulst, *Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome*, *J. Child. Psychol. Psychiatry* 48 (2007), pp. 657–666.
- [8] L. Fahrmeir, T. Kneib, and S. Lang, *Penalized structured additive regression for spacetime data: A Bayesian perspective*, *Statist. Sinica* 14 (2004), pp. 731–761.
- [9] A. Gelman, *Analysis of variance: why it is more important than ever (with discussion)*, *The Annals of Statistics* 33 (2005), pp. 1–53.
- [10] A. Gelman, *Prior distributions for variance parameters in hierarchical models*, *Bayesian Analysis* 1 (2006), pp. 515–534.
- [11] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, London, 1996. 131–143.
- [12] T.M. Gill, E.A. Gahbauer, L. Han, and H.G. Allore, *Trajectories of disability in the last year of life*, *N. Engl. J. Med.* 362 (2010), pp. 1173–1180.
- [13] P. Haggett and R.J. Chorley, *Network Analysis in Geography*, St. Martin's Press, London, 1969.

- [14] A.M. Haviland, D.S. Nagin, and P.R. Rosenbaum, *Combining propensity score matching and group-based trajectory analysis in an observational study*, *Psychol. Methods* 12 (2007), pp. 247–267.
- [15] A.M. Haviland, P.R. Rosenbaum, D.S. Nagin, and R.E. Tremblay, *Combining group-based trajectory modeling and propensity score matching for causal inferences in nonexperimental longitudinal data*, *Dev. Psychol.* 44 (2008), pp. 422–436.
- [16] J.S. Hodges, *Richly Parameterized Linear Models: Additive, and Spatial Models Using Random Effects*, Time Series, Chapman & Hall/CRC, 2013.
- [17] M.C. Hu, B. Muthen, C. Schaffran, P.C. Griesler, and L.D.B. Kande, *Developmental trajectories of criteria of nicotine dependence in adolescence*, *Drug Alcohol Depend.* 98 (2008), pp. 94–104.
- [18] J. M. Jester, J. T. Nigg, A. Buu, L. I. Puttler, J. M. Glass, M. M. Heitzeg, H. E. Fitzgerald, and R. A. Zucker, *Trajectories of childhood aggression and inattention/hyperactivity: differential effects on substance abuse in adolescence*, *J. Am. Acad. Child. Adolesc. Psychiatry* 47 (2008), pp. 1158–1165.
- [19] G. Kesidis, *An Introduction to Communication Network Analysis*, John Wiley & Sons, Inc., Hoboken, NJ, 2007.
- [20] N. Klein and T. Kneib, *Scale-dependent priors for variance parameters in structured additive distributional regression*, *Bayesian Anal.* 11 (2016), pp. 1071–1106.
- [21] N. Klein, T. Kneib, S. Klasen, and S. Lang, *Bayesian structured additive distributional regression for multivariate responses*, *Journal of the Royal Statistical Society Series C, Royal Statistical Society*, 64 (2015), pp. 569–591.
- [22] N. Klein, T. Kneib, S. Lang, and A. Sohn, *Bayesian structured additive distributional regression with an application to regional income inequality in Germany*, *Ann. Appl. Stat.* 9 (2015), pp. 1024–1052.
- [23] L. Knorr-Held, *Bayesian modelling of inseparable space-time variation in disease risk*, *Statist. Med.* 19 (2000), pp. 2555–2567.
- [24] S. Lang and A. Brezger, *Bayesian P-splines*, *J. Comput. Graph. Stat.* 13 (2004), pp. 183–212.
- [25] B.G. Leroux, X. Lei, and N. Breslow, *Estimation of disease rates in small areas: a new mixed model for spatial dependence*, in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, M.E. Halloran and D. Berry, eds, Springer, New York, 2000, pp. 179–191.
- [26] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, 2000.
- [27] P.A. Mora, I.M. Bennett, I.T. Elo, L. Mathew, J.C. Coyne, and J.F. Cullhane, *Distinct trajectories of perinatal depressive symptomatology: evidence from growth mixture modeling*, *Am. J. Epidemiol.* 169 (2009), pp. 24–32.
- [28] S. Mustillo, C. Worthman, A. Erkanli, G. Keeler, A. Angold, and E.J. Costello, *Obesity and psychiatric disorder: developmental trajectories*, *Pediatrics* 111 (2003), pp. 851–859.
- [29] D.S. Nagin, *Group-based trajectory modeling: an overview*, *Ann. Nutr. Metab.* 65 (2014), pp. 205–210.
- [30] D.S. Nagin and K.C. Land, *Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model*, *Criminology* 31 (1993), pp. 327–362.
- [31] D.S. Nagin and C.L. Odgers, *Group-based trajectory modeling (nearly) two decades later*, *J. Quant. Criminol.* 26 (2010), pp. 445–453.
- [32] C. L. Odgers, A. Caspi, D. S. Nagin, A. R. Piquero, W. S. Slutske, B. J. Milne, N. Dickson, R. Poulton, and T. E. Moffitt, *Is it important to prevent early exposure to drugs and alcohol among adolescents?*, *Psychol. Sci.* 19 (2008a), pp. 1037–1044.
- [33] C. L. Odgers, T. E. Moffitt, J. M. Broadbent, N. Dickson, R. J. Hancox, H. Harrington, R. Poulton, M. R. Sears, W. M. Thomson, and A. Caspi, *Female and male antisocial trajectories: from childhood origins to adult outcomes*, *Dev. Psychopathol.* 20 (2008b), pp. 673–716.
- [34] A. Okabe and K. Sugihara, *Spatial Analysis Along Networks: Statistical and Computational Methods*, John Wiley & Sons Ltd., West Sussex, 2012.
- [35] H.K. Orcutt, D.J. Erickson, and J. Wolfe, *The course of PTSD symptoms among Gulf war veterans: A growth mixture modeling approach*, *J. Trauma Stress* 17 (2004), pp. 195–202.
- [36] J. Pearl, *Causality: Models Reasoning and Inference*, 2nd ed. Cambridge University Press, New York, 2009.

- [37] J.E. Peer and W.D. Spaulding, *Heterogeneity in recovery of psychosocial functioning during psychiatric rehabilitation: an exploratory study using latent growth mixture modeling*, *Schizophr. Res.* 93 (2007), pp. 186–193.
- [38] W.J. Pervin, *Foundation of General Topology*, Academic Press, New York, 1964.
- [39] A.R. Piquero, *The long view of crime: a synthesis of longitudinal research*, in *Taking stock of developmental trajectories of criminal activity over the life course*, A. Liberman, ed., Springer, New York, 2008, pp. 23–78.
- [40] H. Reichenbach, *The Direction of Time*, University of California Press, Berkeley, CA, 1956.
- [41] A. Riebler, S.H. Sørbye, D. Simpson, and H. Rue, *An intuitive Bayesian spatial model for disease mapping that accounts for scaling*, *Stat. Methods Med. Res.* 25 (2016), pp. 1145–1165.
- [42] S. Riviere and D. Schmitt, *Two-dimensional line space Voronoi diagram*, *International Symposium on Voronoi Diagrams in Science and Engineering*, 2007, pp. 168–175.
- [43] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, Monographs on Statistics & Applied Probability. Chapman and Hall/CRC, Boca Raton, FL, 2005.
- [44] K. Sakamura and N. Koshizuka, *Ubiquitous computing technologies for ubiquitous learning*, *IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'05)*, (2005), pp. 11–20.
- [45] D. Simpson, F. Lindgren, and H. Rue, *In order to make spatial statistics computationally feasible, we need to forget about the covariance function*, *Environmetrics* 23 (2012), pp. 65–74.
- [46] D. Simpson, H. Rue, A. Riebler, G.T. Martins, and H.S. Sørbye, *Penalising model component complexity: A principled, practical approach to constructing priors*, *Stat. Sci.* 32 (2017), pp. 1–28.
- [47] D. Spiegelhalter, N. Best, B. Carlin, and A. Van Der Linde, *Bayesian measures of model complexity and fit (with discussion)*, *J. R. Stat. Soc. Ser. B* 64 (2002), pp. 583–639.
- [48] W.D. Stanley, *Network Analysis with Applications*, 4th ed. Prentice-Hall, Upper Saddle River, NJ, 2003.
- [49] H.S. Stern and N. Cressie, *Posterior predictive model checks for disease mapping models*, *Stat. Med.* 19 (2000), pp. 2377–2397.
- [50] A. Tsoularis, *Analysis of logistic growth models*, *Res. Lett. Inf. Math. Sci.* 2 (2001), pp. 23–46.
- [51] I. Van Bokhoven, S.H.M. Van Goozen, H. Van Engeland, B. Schaal, L. Arseneault, J. Séguin, D.S. Nagin, F. Vitaro, and R.E. Tremblay, *Salivary cortisol and aggression in a population-based longitudinal study of adolescent males*, *J. Neural. Transm.* 112 (2005), pp. 1083–1096.
- [52] M.J. Van Ryzin, M. Chatham, E. Kryzer, D.A. Kertes, and R. Gunnar, *Identifying atypical cortisol patterns in young children: the benefits of group-based trajectory modeling*, *Psychoneuroendocrinology* 34 (2009), pp. 50–61.