Statistical Modeling for Spatio-Temporal Data from Physical Convection-Diffusion Processes

Xiao Liu\textsuperscript{1}, Kyongmin Yeo\textsuperscript{2}, and Siyuan Lu\textsuperscript{2}

\textsuperscript{1}Department of Industrial Engineering, University of Arkansas
\textsuperscript{2}IBM T. J. Watson Research Center

Abstract

This paper proposes a statistical modeling approach for spatio-temporal data arising from a generic class of convection-diffusion processes. Such processes are found in various scientific and engineering domains where fundamental physics imposes critical constraints on how data can be modeled and how statistical models should be interpreted. We employ the idea of spectrum decomposition to approximate the physical processes. However, unlike existing models which often assume constant convection-diffusion and zero-mean source-sink, we consider a more realistic scenario with spatially-varying convection-diffusion and nonzero-mean source-sink. As a result, the temporal evolution of spectrum coefficients is closely coupled with each other, which can be seen as the non-linear transfer or redistribution of energy across multiple scales. Because of the spatially-varying convection-diffusion, the space-time covariance is naturally non-stationary in space. A systematic approach is proposed to integrate the theoretical results into the framework of hierarchical dynamical spatio-temporal models. Statistical inference using computationally efficient algorithms is investigated. Some practical considerations related to computational efficiency are discussed in order to make the proposed approach practical. The advantages of the proposed approach are demonstrated by numerical examples, a case study, and comparison studies. Computer code and data are made available.

Key words: Convection-diffusion processes, spatio-temporal modeling, hierarchical dynamical spatio-temporal models, spectral method, radar-based precipitation nowcasting
1 Introduction

The digitization of physical space has led to an explosive growth of spatio-temporal data from physical convection-diffusion processes. Examples span multi-disciplinary areas including environmental science, meteorology, geology, material science, etc. For such processes, fundamental physics often imposes critical constraints on how the data can be modeled and how the model should be interpreted. This paper proposes a statistical modeling approach for spatio-temporal data arising from a generic class of convection-diffusion processes given by stochastic partial differential equations (SPDE). The proposed approach extends some recent advances in the statistical modeling of spatio-temporal data, and provides a holistic framework that enables the integration of statistics into various scientific and engineering applications.

1.1 Motivation and Background

Convection-diffusion equations have been widely used to describe how physical quantities (e.g., particles, air pollutant, energy, rain cells, heat, etc.) are transferred inside a physical system due to two processes: convection and diffusion. Interdisciplinary examples include, but not limited to, urban air pollution processes, motion of extreme weather systems such as Hurricane Florence in the 2018 Atlantic season, decomposition of polymer in space and time, propagation of smoke from California wildfires, regional epidemics of Influenza A (H3N2) in 2018, as well as the 3D temperature field in data centers. For illustrative purposes, Figure 1 shows some typical spatio-temporal processes governed by convection-diffusion equations.

![Figure 1: Motivating examples of spatio-temporal data arising from physical convection-diffusion processes: (a) urban air pollution; (b) radar images of extreme weather events; (c) polymer decomposition in space and time; (d) satellite images of wildfires.](image)

For spatio-temporal data arising from physical convection-diffusion processes, deterministic physics-based numerical models are not suitable for real-time operations because of the high computational cost as well as their ineffectiveness in utilizing real-time sensor data—a main source of information in the era of Big Data. For example, the state-of-the-art numerical model in industry for constructing the data center temperature field is based on the Navier-Stokes equation with its boundary conditions being established from sensor measurements (Hamann et al., 2010). Other examples include the Numerical Weather Prediction (NWP) and Numerical Air Quality Models which typically run for hours or even days (EPA,
For many complex physical processes, it is not only challenging to build the physics models from the first principles but also difficult to handle the high-degree uncertainty associated with model parameters and model inputs.

Statistical spatio-temporal models, on the other hand, have been proven useful for short-term prediction and space-time interpolation using observations; see e.g., Wikle et al. (2001); Guinness and Stein (2013); Liu et al. (2016); Kuusela and Stein (2017); Reich et al. (2018); Guan et al. (2018). For spatio-temporal data arising from physical processes, however, fundamental physics often imposes critical constraints on how the data can be modeled and interpreted. For example, the space-time covariance structures, to a great extent, are determined by the convection-diffusion equations from which the data are generated. When the critical connection to fundamental physics is missing, the capabilities of pure data-driven approaches may not be fully appreciated by domain experts for monitoring, planning, and “what-if” analyses. In addition, for non-stationary and highly dynamic processes, it is well-known that the specification of space-time covariance structures can be extremely challenging and sometimes unrealistic (Nychka and Wikle, 2002; Cressie and Wikle, 2011; Simpson et al., 2012). As a result, the models are often motivated by considerations of mathematical convenience (e.g., isotropic, space-time separable, etc.), which may not acknowledge the full complexity of the underlying physical processes.

Hence, there is an urgent demand for physics-based statistical models and computationally efficient algorithms tailored for large-scale spatio-temporal data arising from physical processes. When operational insights are required based on fundamental domain knowledge, our engagement with industry over recent years strongly indicates that pure data-driven approaches could be limited even in data-rich scientific and engineering fields. As both the volume and variety of data increase, this issue appears to be more crucial than ever when statistical models are inevitably becoming more complex but seemingly less interpretable. Hence, data intensive physical analytics—an emerging field which resides at the intersection of physical modeling and statistics (Hamann, 2016)—is being actively pursued in both academia and industry.

1.2 Literature Review

The pioneering work of statistical modeling for spatio-temporal data can be found in Banerjee et al. (2004), Schabenberger and Gotway (2005) and Cressie and Wikle (2011). The mainstay approach models the spatio-temporal data by constructing random fields with specified space-time covariance structures. This approach is referred to as the geostatistical paradigm and has gained tremendous popularity and success over the past decades. On the other hand, since the geostatistical paradigm hinges heavily on the specification of space-time covariance structures, such an approach is known to be challenging when dealing with large data arising from non-stationary and highly dynamic processes. The specification of space-time covariance structures can be extremely difficult for complex spatio-temporal processes, if not unrealistic. To elaborate, consider the modeling of hourly urban air quality data collected by monitoring stations (Liu et al., 2016). The space-time correlation of pollutant concentration, which depends on wind, temperature, solar radiation and traffic,
is usually too complex to be directly specified or validated. Similar examples can be found in the modeling of a dynamic weather system where the space-time correlation of radar reflectivity data constantly changes due to highly dynamic factors such as the small-scale localized convection (Liu et al., 2018). Hence, there have been prolonged research interests to provide flexible and effective ways to construct non-stationary covariance functions (Cressie and Huang, 1999; Gneiting, 2002; Fuentes et al., 2005; Ghosh et al., 2010; Reich et al., 2011). But, a known class of covariance functions still appears to be limited. To keep the model mathematically tractable, the space-time covariance functions are typically derived from moments of multivariate probability distributions based on exploratory data analysis with assumptions such as isotropic, stationary and space-time separable (Carroll et al., 1997). Moreover, the geostatistical modeling paradigm can be computationally challenging for large spatio-temporal data sets due to the prohibitive cubic operations $O((T \cdot N)^3)$ with $T$ and $N$ being the number of samples in time and space (Cressie and Wikle, 2011; Banerjee and Fuentes, 2012; Simpson et al., 2012; Yan et al., 2017). For example, a standard dual polarization meteorological Doppler weather radar image contains 230,400 pixels and a new image is generated every 5 minutes. Due to the high computational cost, the geostatistical modeling paradigm, which heavily relies on random fields, becomes less practical for large problems and approximations are commonly used; for example, Gaussian Markov Random Fields representation (Lindgren and Rue, 2011), Nearest-Neighbor Gaussian Process (Datta et al., 2016; Banerjee, 2017), kernel convolution (Higdon, 1998; Kontar et al., 2017), low rank representation (Cressie and Johannesson, 2002; Nychka and Wikle, 2002; Banerjee et al., 2008), approximation of likelihood functions (Stein et al., 2004; Fuentes, 2007; Guinness and Fuentes, 2015), Bayesian inference for latent Gaussian models based on the integrated nested Laplace approximations (Rue et al., 2009; R-INLA, 2019), Matrix-free state-space model (Mondal and Wang, 2019), Vecchia approximations of Gaussian processes (Katzfuss et al., 2019), as well as the multi-resolution approximation (M-RA) of Gaussian processes observed at irregular spatial locations (Katzfuss, 2017).

Progress has been made in integrating physics into the statistical spatio-temporal models, which requires tremendous knowledge on both statistics and domain knowledge. A powerful modeling framework known as the Hierarchical Dynamical Spatio-Temporal Models (DSTM) has been proposed (Wikle and Cressie, 1999; Berliner, 2003; Cressie and Wikle, 2011; Katzfuss et al., 2019). The power of DSTM comes from the dynamical model specification by a series of conditional models that lead to a complex joint space-time covariance structures which can hardly be directly specified. Stroud et al. (2001) investigated a pure data-driven framework for dynamic spatio-temporal modeling. The use of Monte Carlo approaches makes the dynamical model more computable with very large data sets and for non-linear non-Gaussian models (Carlin et al., 1992; Banerjee et al., 2004). In recent years, SPDE based modeling approach of large spatio-temporal data has drawn much attention due to its connection to many physical convection-diffusion processes. Whittle (1963) obtained the spectrum of the stationary solution of a convection-diffusion SPDE, which can be used to establish the space-time covariance through inverse Fourier transform. Brown et al. (2000) proposed a blur-generated non-separable space-time models and showed the
explicit link of the proposed model to a convection-diffusion SPDE in the limiting case. Sigrist et al. (2015) proposed a systematic approach to obtain a Gaussian process by solving a SPDE with constant convection-diffusion. Jones and Zhang (1997) developed statistical models for unequally spaced data arising from continuous stationary space-time processes described by a SPDE with white noise. Spectral density functions were obtained for the stationary space-time process and the inverse Fourier transform was employed to obtain the covariance functions. Lindgren and Rue (2011) investigated the explicit connection between a convection-diffusion SPDE and Gaussian Markov random field. It is noted that, a DSTM can also be motivated by special types of partial differential equations (e.g., diffusion, reaction-diffusion, or simplified convection-diffusion equations) under special conditions (e.g., constant convection and diffusion). A summary of the latest advances in the spatial modeling with SPDE can be found in Cressie and Wikle (2011) and Krainski et al. (2019).

1.3 Overview of the Paper

This paper proposes a statistical modeling approach for spatio-temporal data arising from a generic class of convection-diffusion SPDE with (i) spatially-varying convection-diffusion, and (ii) nonzero-mean spatio-temporal source-sink. We employ the important idea of spectrum decomposition to approximate the spatio-temporal process given by a SPDE. However, unlike the existing results based on constant convection-diffusion and zero-mean source-sink, the proposed model has some major differences: (i) under spatially-varying convection-diffusion, the temporal evolution of spectrum coefficients is coupled with each other, which can be viewed as the non-linear energy transfer between multiple scales; (ii) while constant convection-diffusion and zero-mean source-sink lead to a stationary process in the limiting case, the proposed spatio-temporal model clearly has a non-stationary space-time covariance structure due to the spatially-varying convection-diffusion. We integrate our theoretical results into the DSTM framework, and investigate the statistical inference using computationally efficient algorithms. To our best knowledge, such a statistical modeling approach is not yet available in the literature.

In Section 2, we describe the general modeling framework and present the key theoretical results. Section 3 focuses on real-valued processes under discrete sampling, which is a typical scenario for many scientific and engineering applications. In Section 4, a systematic approach is presented to integrate the established theoretical results into the DSTM framework. Computationally efficient statistical inference methods are also presented in this section. Section 5 provides a numerical example and a case study of radar-based precipitation nowcasting to demonstrate the advantages of the proposed method.

2 The General Modeling Framework

Consider the spatio-temporal data arising from a generic class of convection-diffusion processes:

$$
\frac{\partial}{\partial t} \xi(t, s) = -\tilde{v}_{t,s}^T \nabla \xi(t, s) + \nabla \cdot D_{t,s} \nabla \xi(t, s) - \zeta_{t,s} \xi(t, s) + Q(t, s) + \epsilon(t, s) 
$$ (1)
where $\xi(t, s)$ is a spatio-temporal process in space $s$ and time $t$, $\nabla$ and $\nabla \cdot$ are the gradient and divergence operators, $\vec{v}_{t,s}$, $D_{t,s}$ and $\zeta_{t,s}$ respectively represent the velocity field, diffusivity and decay, $Q(t, s)$ is the source-sink term (i.e., generation-destruction), and $\varepsilon(t, s)$ is a spatio-temporal error process.

The spectral decomposition of $\xi(t, s)$ is given by the linear combination of spatial basis functions and random spectral coefficients (Wikle and Cressie, 1999; Cressie and Wikle, 2011; Sigrist et al., 2015; Mak et al., 2018):

$$\tilde{\xi}(t, s) \approx \sum_{j=1}^{K} \alpha_j(t) f_j(s) \equiv f^T(s) \alpha(t) \quad (2)$$

where $\alpha(t) = (\alpha_1(t), ..., \alpha_K(t))^T$ is a multivariate temporal process of the spectral coefficients, $f_j(s) = \exp(i k_j^T s)$ is the deterministic Fourier basis function, $k_j$ is the spatial wave number, and $f(s) = (f_1(s), ..., f_K(s))^T$. The choice of Fourier spatial basis function is appropriate if $\xi(t, s)$ is periodic. The spectral decomposition (2) provides some key advantages in modeling the spatio-temporal data arising from the physical process (1):

- Because the Fourier basis functions are fixed, the modeling of the spatio-temporal process $\xi(t, s)$ can be converted to the modeling of a multivariate stochastic process $\alpha(t)$. The latter problem is often more tractable.

- If it is possible to find a multivariate stochastic process $\alpha(t)$ such that the approximated process generated by (2) also satisfies the SPDE (1), the fundamental physics in (1) can be naturally built into the statistical model (Wikle and Cressie, 1999; Sigrist et al., 2015). On the other hand, one may choose to model $\alpha(t)$ by pure data-driven approaches, if incorporating physics into the statistical model is not a major modeling concern (Yeo and Melnyk, 2018).

- Once the temporal dynamics of $\alpha(t)$ has been established, the space-time covariance structure of $\tilde{\xi}(t, s)$ can be derived from the spectral decomposition (2) where $\tilde{\xi}(t, s)$ is a linear combination of $\alpha(t)$ and fixed basis functions $f(s)$.

- Dimension reduction is possible by retaining only the low-frequency components in (2), which helps to achieve a tremendous computational advantage without significantly sacrificing modeling accuracy. This is often a critical step which ensures the proposed model to be practical enough for certain applications as shown in the case study.

- The model based on (2) can be integrated into the framework of DSTM (Wikle et al., 1998; Berliner, 2003; Cressie and Wikle, 2011), which enables computationally efficient algorithms for statistical inference, monitoring and prediction.

To realize these modeling advantages, we first show that a multivariate stochastic process $\alpha(t)$ exists such that the approximated process generated by (2) satisfies the SPDE (1). The following assumptions are made: $(A1)$ the velocity field and diffusivity in (1) are spatially-varying but temporally-invariant (i.e., $\vec{v}_{t,s} = \vec{v}_s$, $D_{t,s} = D_s$); $(A2)$ the generation process $Q(t, s)$ in (1) admits a spectral representation $Q(t, s) = f^T(s)\beta(t)$ with
\( \beta(t) = (\beta^{(1)}(t), ..., \beta^{(K)}(t))^T \) dynamically evolving over time; \((A3)\) the error process \( \varepsilon(t, s) = f^T(s)\tilde{\varepsilon}(t) \) is a white-in-time stationary spatial process where \( \tilde{\varepsilon}(t) \) is a \( K \)-dimensional random vector; \((A4)\) the initial condition can be expressed by \( \xi(0, s) = f^T(s)\alpha(0) \) where \( \alpha(0) \sim N(0, \text{diag}\{\hat{h}_0(k_j)\}) \) and \( \hat{h}_0(k_j) \) is the spectral density.

**Proposition 1.** For spatially-varying convection-diffusion \((A1)\), if the initial condition \( \tilde{\xi}(0, s) \), generation process \( Q(t, s) \) and error process \( \varepsilon(t, s) \) are in a space \( S \) spanned by a finite number of Fourier functions \( f^T(s) \) \((A2\sim A4)\), then, there exists a multivariate temporal process \( \alpha(t) \) such that \( \tilde{\xi}(t, s) \) remains in \( S \) and satisfies the SPDE \((1)\):

\[
\dot{\alpha}(t) = G\alpha(t) + \beta(t) + \varepsilon(t). \tag{3}
\]

Here, the \((i, j)\)th entry of the non-diagonal matrix \( G \) is given by

\[
g_{i,j} = C^{-1} \int_S (\mathbf{-k}_j^T D_s k_j - \zeta_{t,s} - i (\tilde{v}_s^T k_j - \nabla \cdot D_s k_j)) f_j(s)f_i^*(s)ds \tag{4}
\]

where \( C = \int f_j(s)f_j^*(s)ds \), \( ^* \) represents the complex conjugation, and the covariance of the error process \( \tilde{\varepsilon}(t) \) is given by \( \text{cov}(\tilde{\varepsilon}(t), \tilde{\varepsilon}(t')) = \delta_{t,t'}\text{diag}\{\hat{h}(k_j)\} \) with \( \delta_{t,t'} \) and \( \hat{h}(\cdot) \) respectively representing the Kronecker delta and spectral density function.

**Remark.** The temporal evolution of \( \alpha(t) \) in \((3)\) has well-defined interpretations: the first term on the right hand side, \( G\alpha(t) \), captures the transition of the spectral coefficients \( \alpha(t) \) due to convection, diffusion and decay in \((1)\) as \( G \) depends on \( \tilde{v}_s, D_s \) and \( \zeta_{t,s} \). The second term, \( \beta(t) \), captures the additional change to \( \alpha(t) \) due to the source-sink \( Q(t, s) \). The third term \( \varepsilon(t) \) introduces uncertainty into the evolution of \( \alpha(t) \) due to the error process \( \varepsilon(t, s) \) in \((1)\).

The process \( \alpha(t) \) defined in \((3)\) differs from the existing result in the literature. Since \( G \) becomes non-diagonal for spatially-varying convection-diffusion, the transition equation \((3)\) clearly implies that the temporal evolution of individual components in \( \alpha(t) \) is coupled: at any time \( t \), the transition of each component in \( \alpha(t) \) depends not only on itself, but also on all other components in \( \alpha(t) \). If the Fourier coefficients are seen as the energy assigned to different frequencies, the coupling of Fourier coefficients can be naturally interpreted as the **energy transfer across multiple scales**, i.e., the spatially-varying convection-diffusion redistributes the energy across different frequencies. Under constant convection-diffusion and zero source-sink (i.e., \( Q(t, s) = 0, \mathbf{\tilde{v}}_{t,s} = \mathbf{\tilde{v}} \) and \( D_{t,s} = D \)), the result in \((3)\) degenerates to the result in Sigrist et al. (2015) where the temporal evolution of each component of \( \alpha(t) \) is completely de-coupled, leading to a stationary solution of \((1)\) in the limiting case; also see Brown et al. (2000). When the velocity field \( \mathbf{\tilde{v}}_s \) and diffusivity \( D_s \) vary in space, the covariance structure in our model is obviously non-stationary in space.

Proposition 1 can be proved using the Galerkin method which converts a continuous differential equation to a discrete problem.
Proof of Proposition 1:

Since \( \vec{v}_j^T \nabla f_j(s) = i \vec{v}_j^T k_j f_j(s) \) and \( \nabla \cdot D_s \nabla f_j(s) = (-k_j^T D_s k_j + i \nabla \cdot D_s k_j) f_j(s) \), substituting the spectral decomposition (2) into the SPDE (1) yields:

\[
\sum_{j=1}^{K} f_j(s) \dot{\alpha}_j(t) = - \sum_{j=1}^{K} i \vec{v}_j^T k_j f_j(s) \alpha_j(t) + \sum_{j=1}^{K} (-k_j^T D_s k_j + i \nabla \cdot D_s k_j) f_j(s) \alpha_j(t) \\
- \zeta_{t,s} \sum_{j=1}^{K} f_j(s) \alpha_j(t) + \sum_{j=1}^{K} f_j(s) \beta_j(t) + \sum_{j=1}^{K} f_j(s) \dot{\varepsilon}_j(t). 
\]

(5)

Using the Galerkin method, we multiply both the left hand side (LHS) and right hand side (RHS) of (5) by the complex conjugate \( f_l^*(s) \) \( l = 1, 2, ..., K \), and integrate the product over the spatial domain. Because the Fourier basis functions are orthogonal, i.e., \( \int f_j(s) f_l^*(s) ds = 0 \) if \( j \neq l \), the LHS of (5) becomes

\[
\int_{\mathbb{S}} \sum_{j=1}^{K} f_j(s) f_l^*(s) \dot{\alpha}_j(t) ds = \dot{\alpha}_l(t) \int_{\mathbb{S}} f_j(s) f_l^*(s) ds \equiv \dot{\alpha}_l(t) C
\]

(6)

for \( l = 1, 2, ..., K \).

Similarly, for any \( l = 1, 2, ..., K \), we multiply the complex conjugate \( f_l^*(s) \) to the RHS of (5) and integrate the product over the spatial domain. Then, the five terms on the RHS are respectively given by

\[
\int_{\mathbb{S}} \sum_{j=1}^{K} i \vec{v}_j^T k_j f_j(s) f_l^*(s) \alpha_j(t) ds = \sum_{j=1}^{K} \left( -i \int \vec{v}_j^T k_j f_j(s) f_l^*(s) ds \right) \alpha_j(t) \equiv \sum_{j=1}^{K} g_{l,j}^{(1)} \alpha_j(t) 
\]

(7)

\[
\int_{\mathbb{S}} \sum_{j=1}^{K} (-k_j^T D_s k_j + i \nabla \cdot D_s k_j) f_j(s) f_l^*(s) \alpha_j(t) ds \\
= \sum_{j=1}^{K} \left( \int (-k_j^T D_s k_j + i \nabla \cdot D_s k_j) f_j(s) f_l^*(s) ds \right) \alpha_j(t) \equiv \sum_{j=1}^{K} g_{l,j}^{(2)} \alpha_j(t) 
\]

(8)

\[
\int_{\mathbb{S}} \sum_{j=1}^{K} \zeta_{t,s} f_j(s) f_l^*(s) \alpha_j(t) ds = \sum_{j=1}^{K} \left( \int \zeta_{t,s} f_j(s) f_l^*(s) ds \right) \alpha_j(t) \equiv \sum_{j=1}^{K} g_{l,j}^{(3)} \alpha_j(t) 
\]

(9)

\[
\int_{\mathbb{S}} \sum_{j=1}^{K} f_j(s) f_l^*(s) \beta_j(t) ds = C \beta_l(t) 
\]

(10)
\[
\int \sum_{j=1}^{K} f_j(s) f_i^*(s) \varepsilon_j(t) ds = C \varepsilon_i(t)
\] (11)

Let \( g_{i,j} = C^{-1}(g_{1,j}^{(1)} + g_{2,j}^{(2)} + g_{3,j}^{(3)}) \), we obtain

\[
\dot{\alpha}_l(t) = \sum_{j=1}^{K} g_{l,j} \alpha_j(t) + \beta_l(t) + \tilde{\varepsilon}_l(t)
\] (12)

which implies \( \dot{\alpha}(t) = G \alpha(t) + \beta(t) + \tilde{\varepsilon}(t) \), as was to be proved \( \blacksquare \)

Once the temporal dynamics of \( \alpha(t) \) has been established, the space-time covariance of \( \tilde{\xi}(t, s) \) is derived from the spectral decomposition (2) where \( \tilde{\xi}(t, s) \) is a linear combination of \( \alpha(t) \) and fixed basis functions \( f(s) \). It follows from (2) that

\[
\text{cov}(\tilde{\xi}(t + \Delta, s), \tilde{\xi}(t, s')) = f^T(s) \text{cov}(\alpha(t + \Delta), \alpha(t)) f(s').
\] (13)

Hence, it is sufficient to obtain the covariance structure of the multivariate temporal process \( \alpha(t) \), which is given in the following proposition.

**Proposition 2.** For a multivariate temporal process \( \alpha(t) \) defined in (3),

\[
\text{cov}(\alpha(t+\Delta), \alpha(t)) = \exp(G \Delta) \left( \exp(Gt) H_0 \exp^*(G^T t) + \int_0^t \exp(G(t - \tau)) H \exp^*(G^T (t - \tau)) d\tau \right)
\] (14)

where \( H = \text{diag}(\hat{h}(k_j)) \) and \( H_0 = \text{diag}(\hat{h}_0(k_j)) \) are respectively the spectral density of \( \tilde{\varepsilon}(t) \) and \( \alpha(0) \). When \( t \to \infty \), the effect of the initial condition \( \alpha(0) \) vanishes and we have

\[
\lim_{t \to \infty} \text{cov}(\alpha(t + \Delta), \alpha(t)) = \exp(G \Delta) \int_0^t \exp(G(t - \tau)) H \exp^*(G^T (t - \tau)) d\tau.
\] (15)

**Proof of Proposition 2:**

Solving the ODE (3) for discrete time points, we have

\[
\alpha(t + \Delta) = \exp(G \Delta) \alpha(t) + q(\Delta)
\] (16)

where

\[
q(\Delta) \sim N \left( \beta(t) \Delta, \int_0^\Delta \exp(G(\Delta - \tau)) H \exp^*(G^T (\Delta - \tau)) d\tau \right).
\] (17)

Hence, the covariance matrix of \( \alpha(t) \) is

\[
\text{var}(\alpha(t)) = \exp(Gt) H_0 \exp^*(G^T t) + \int_0^t \exp(G(t - \tau)) H \exp^*(G^T (t - \tau)) d\tau.
\] (18)
It is also noted that

\[
\text{cov}(\alpha(t + \Delta), \alpha(t)) = \text{cov}(\exp(G\Delta)\alpha(t) + q(\Delta), \alpha(t)) = \exp(G\Delta)\text{var}(\alpha(t)).
\]  

Substituting (18) into (19) immediately implies (14), as was to be shown ■

Remark. In (17), \( q(\Delta) \) is the noise added to the transition of \( \alpha(t) \) due to the noise term in the SPDE (1). To generate some insights on the results presented in Proposition 2, we consider a special case under constant convection, diffusion and decay (i.e., \( \bar{\sigma}_s = \bar{\sigma}, D_s = \text{diag}(d) \) and \( \zeta_s = \zeta \)). Under such a simplified case, \( G \) becomes a diagonal matrix with its \((i,i)\)th entry being \( g_{i,i} = -d k_i^T k_i - \zeta - i\bar{\sigma}_s^T k_i \). Then, it follows from (17) that the \((i,i)\)th entry of the covariance matrix of \( q(\Delta) \) is

\[
h(k_i) \int_0^\Delta \exp(-(d k_i^T k_i + \zeta + i\bar{\sigma}_s^T k_i)(\Delta - \tau)) \exp^*(-(d k_i^T k_i + \zeta + i\bar{\sigma}_s^T k_i)(\Delta - \tau)) \, d\tau
\]

\[
= \frac{h(k_i)}{2d k_i^T k_i + 2\zeta} \left(1 - \exp(-(2d k_i^T k_i + 2\zeta)\Delta)\right).
\]

It is seen from (20) that: (i) the uncertainty associated with \( q(\Delta) \) does not depend on convection; (ii) for small \( \Delta \), \( d \) and \( \zeta \), (20) can be approximated by \( h(k_i)\Delta \) which indicates that the effect of diffusion and decay is small on \( q(\Delta) \) within a small time interval \( \Delta \); (iii) when \( \Delta \to \infty \), (20) becomes \( h(k_i)(2d k_i^T k_i + 2\zeta)^{-1} \) which indicates that the diffusion and decay force the spectrum of the noise to decay.

3 The Special Case: Discrete Real-Valued Processes

In Section 3, we restrict our attention to real-valued processes under discrete sampling, which is the most common scenario in scientific and engineering applications. For example, temperature field in a data center server room is sampled by thermal sensors at discrete times and spatial intervals, radar or satellite images of a weather system are taken at fixed time intervals and over fixed-size image pixels, and so on.

Consider a real-valued process \( \xi(t, s) \) defined on a regular grid of \( N_1 \times N_2 \) spatial locations at time points \( t_1, t_2, ..., t_T \). Without loss of generality, \( N_1 \) and \( N_2 \) are assumed to be even numbers, and the spatial domain is standardized as \( \mathbb{S} = [0, 1]^2 \). The proposed approach is rooted in the 2D discrete Fourier transform (DFT). For a 2D process \( x(t_1, n_2) \) at grid points


\((n_1, n_2), n_1 = 1, 2, ..., N_1\) and \(n_2 = 1, 2, ..., N_2\), the DFT of \(x(n_1, n_2)\) is

\[
X(k_1, k_2) = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1, n_2) e^{-j(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_2)}
\]

\[
= \frac{1}{N_1 N_2} \sum_{n_1} \sum_{n_2} x(n_1, n_2) \cos \left(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_2\right) - j \frac{1}{N_1 N_2} \sum_{n_1} \sum_{n_2} x(n_1, n_2) \sin \left(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_2\right)
\]

\[
\equiv \alpha_{k_1, k_2}^{(R)} - j \alpha_{k_1, k_2}^{(I)}
\]

for \(k_1 = -N_1/2 + 1, -N_1/2 + 2, ..., N_1/2 = -N_2/2 + 1, -N_2/2 + 2, ..., N_2/2\).

The inverse transform is also obtained as:

\[
x(n_1, n_2) = \sum_{k_1=-N_1/2+1}^{N_1/2} \sum_{k_2=-N_2/2+1}^{N_2/2} X(k_1, k_2) e^{j\left(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_1\right)}
\]

\[
= \sum_{k_1} \sum_{k_2} (\alpha_{k_1, k_2}^{(R)} - j \alpha_{k_1, k_2}^{(I)}) \left\{ \cos \left(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_2\right) + j \sin \left(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_2\right) \right\}
\]

\[
= \sum_{k_1} \sum_{k_2} \alpha_{k_1, k_2}^{(R)} \cos \left(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_2\right) + \sum_{k_1} \sum_{k_2} \alpha_{k_1, k_2}^{(I)} \sin \left(\frac{2\pi}{N_1} n_1 k_1 + \frac{2\pi}{N_2} n_2 k_2\right).
\]

Note that, if \(x(n_1, n_2)\) is real, it is well known that (i.e., rotational symmetry)

\[
|X(k_1, k_2)| = |X(-k_1, -k_2)|, \quad \arg X(k_1, k_2) = - \arg X(-k_1, -k_2).
\]

As an illustration, Figure 2 shows a 80 × 80 pixel weather radar image and its 2D DFT. The first row of this figure shows the original radar image and the image recovered by inverse DFT, the second row shows the real and imaginary parts of the FFT \(X(k_1, k_2)\) for \(k_1, k_2 = -39, -38, ..., 40\), and the bottom row presents a zoomed-in view of the real and imaginary parts of \(X(k_1, k_2)\) by only keeping the low-frequency components. For such a real-valued process, the rotational symmetry of the real and imaginary parts of \(X(k_1, k_2)\) is clearly seen, i.e., \(|X(k_1, k_2)| = |X(-k_1, -k_2)|\) and \(\arg X(k_1, k_2) = - \arg X(-k_1, -k_2)\). One key observation from Figure 2 is that the Fourier coefficients corresponding to the high-frequency terms (approximately \(k > 15\)) are practically zero, which strongly suggests the possibility of dimension reduction by keeping only the low-frequency components; also see the case study to be presented in Section 5.2. If it is assumed that the image in Figure 2 is band-limited within \([-15, 15]\), then, the spatial sampling rate in this example (i.e., 1/80) far exceeds the Nyquist rate in this example, which explains why the image can be perfectly reconstructed by inverse FFT.
Figure 2: Illustrations of 2D DFT for real-valued processes (Row 1: original weather radar image and the image recovered by inverse DFT; Row 2: the real and imaginary parts of the FFT, $X(k_1, k_2)$; Row 3: the low-frequency region of the real and imaginary parts of $X(k_1, k_2)$, i.e., a zoomed-in view).
The 2D DFT motivates us to approximate the real-valued spatio-temporal process $\xi(t, s)$ by allowing $\alpha_{k_1, k_2}^{(R)}$ and $\alpha_{k_1, k_2}^{(I)}$ to vary over time. Hence, we have

$$\xi(t, s) \approx \sum_{k_1 = -N_1/2+1}^{N_1/2} \sum_{k_2 = -N_2/2+1}^{N_2/2} \alpha_{k_1, k_2}^{(R)}(t) f_{k_1, k_2}^{(R)}(s) + \alpha_{k_1, k_2}^{(I)}(t) f_{k_1, k_2}^{(I)}(s)$$

(24)

where $f_{k_1, k_2}^{(R)}(s) = \cos(2\pi s_1 k_1 + 2\pi s_2 k_2)$, $f_{k_1, k_2}^{(I)}(s) = \sin(2\pi s_1 k_1 + 2\pi s_2 k_2)$.

In addition, since the DFT of real-valued processes is periodic, we obtain from (23) the following constraints:

$$\alpha_{k_1, k_2}^{(R)}(t) = \alpha_{-k_1+iN_1,-k_2+jN_2}(t), \alpha_{k_1, k_2}^{(I)}(t) = -\alpha_{-k_1+iN_1,-k_2+jN_2}(t), \quad i, j \in \mathbb{N}$$

(25)

and

$$\alpha_{0,0}^{(I)}(t) = \alpha_{0,N_2}^{(I)}(t) = \alpha_{N_1,0}^{(I)}(t) = \alpha_{N_1,N_2}^{(I)}(t) = 0.$$  

(26)

The important constraints shown in (25) and (26) enable us to re-write (24) as

$$\xi(t, s) = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) f_k^{(R)}(s) + 2 \sum_{k \in \Omega_2} \left( \alpha_k^{(R)}(t) f_k^{(R)}(s) + \alpha_k^{(I)}(t) f_k^{(I)}(s) \right)$$

(27)

where $k = (k_1, k_2)$, and the two sets of spatial wave numbers, $\Omega_1$ and $\Omega_2$, are defined as:

$$\Omega_1 = \{(0,0), (0, \frac{N_2}{2}), (\frac{N_1}{2}, 0), (\frac{N_1}{2}, \frac{N_2}{2})\}$$

$$\Omega_2 = \{(k_1, k_2); k_1 = 0, 1, \ldots, \frac{N_1}{2}, k_2 = 0, 1, \ldots, \frac{N_2}{2}\}$$

$$\cup \{(k_1, k_2); k_1 = -1, \ldots, -\frac{N_1}{2}+1, k_2 = -1, \ldots, -\frac{N_2}{2}+1\} \setminus \Omega_1.$$  

(28)

It is easy to see that the dimension of the Fourier coefficients is $N_1 \times N_2$, i.e., the total number of $\alpha^{(R)}$ and $\alpha^{(I)}$ in (27). In spectral analysis, it is a common practice to further drop the terms corresponding to the highest frequencies, $\frac{N_1}{2}$ and $\frac{N_2}{2}$, in order to achieve the rotational symmetry between all pairs in the wave number space. Hence, an alternative approach to (27) is:

$$\xi(t, s) = \left[ \alpha_k^{(R)}(t) f_k^{(R)}(s) \right]_{k=(0,0)} + 2 \sum_{k \in \Omega_3} \left( \alpha_k^{(R)}(t) f_k^{(R)}(s) + \alpha_k^{(I)}(t) f_k^{(I)}(s) \right)$$

(29)

where

$$\Omega_3 = \{(k_1, k_2); k_1 = 0, 1, \ldots, \frac{N_1}{2} - 1, k_2 = 0, 1, \ldots, \frac{N_2}{2} - 1\}$$

$$\cup \{(k_1, k_2); k_1 = -1, \ldots, -\frac{N_1}{2}+1, k_2 = -1, \ldots, -\frac{N_2}{2}+1\} \setminus \{(0,0)\}.$$  

(30)
and the dimension of (29) is reduced to $N_1 \times N_2 - N_1 - N_2 + 1$. For example, if $N_1 = N_2 = 80$ as shown in Figure 2, the dimensions of (27) and (29) are respectively 6,400 and 6,241.

As shown in Figure 2, the Fourier coefficients corresponding to high frequency components are often close to zero (i.e., the energy is concentrated in the low-frequency region). Hence, the approximation (29) is well justified. In fact, in the case study shown in Section 5.2, only the low-frequency terms are retained in order to gain a tremendous computational advantage which barely sacrifices any modeling accuracy.

For the same radar image data shown in Figure 2, Figure 3 illustrates the Fourier basis functions at selected frequencies. The left panel shows $f^{(R)}_{40,40}(s)$ with $k_1 = k_2 = 40$ (the last entry in $\Omega_1$), the middle panel shows $f^{(R)}_{1,0}(s)$ with $k_1 = 1$ and $k_2 = 0$ (the 40th entry in $\Omega_2$), and the right panel shows $f^{(I)}_{1,10}(s)$ with $k_1 = 1$ and $k_2 = 10$ (the 50th entry in $\Omega_2$).

Proposition 3 establishes the temporal dynamics of $\alpha^{(R)}(t)$ and $\alpha^{(I)}(t)$ such that the approximate process by (27) satisfies the SPDE (1). In the case that the process is approximated by (29), the extension of Proposition 3 is trivial and thus omitted.

**Proposition 3.** Let $\alpha(t) = (\{\alpha^{(R)}_k(t); k \in \Omega_1\}, \{\alpha^{(R)}_k(t); k \in \Omega_2\}, \{\alpha^{(I)}_k(t); k \in \Omega_2\})^T$ be a collection of spectral coefficients; let $\xi(0, s)$ be the initial condition where $\alpha(0) \sim N(0, H_0)$ with $H_0 = \text{diag}(\tilde{h}_0(k))$ being a spectral density; let $e^{(R)}(s)$ be the initial condition where $\alpha(0) \sim N(0, H_0)$ with $H_0 = \text{diag}(\tilde{h}_0(k))$ being a spectral density; let $e^{(I)}(s)$ be the initial condition where $\alpha(0) \sim N(0, H_0)$ with $H_0 = \text{diag}(\tilde{h}_0(k))$ being a spectral density; let

$$Q(t, s) = \sum_{k \in \Omega_1} \beta^{(R)}_k(t) f^{(R)}_k(s) + 2 \sum_{k \in \Omega_2} \left( \beta^{(R)}_k(t) f^{(R)}_k(s) + \beta^{(I)}_k(t) f^{(I)}_k(s) \right)$$

be the spectral representation of the deterministic source-sink process, and let

$$e(t, s) = \sum_{k \in \Omega_1} \varepsilon^{(R)}_k(t) f^{(R)}_k(s) + 2 \sum_{k \in \Omega_2} \left( \varepsilon^{(R)}_k(t) f^{(R)}_k(s) + \varepsilon^{(I)}_k(t) f^{(I)}_k(s) \right)$$

be the spectral representation of the stochastic source-sink process.
be the spectral representation of the white-in-time stationary error process and \( \tilde{e}(t) = (\{ \tilde{e}^{(R)}(t); \mathbf{k} \in \Omega_1 \}, \{ \tilde{e}^{(I)}(t); \mathbf{k} \in \Omega_2 \})^T \) is a random noise vector with spectral density \( \mathbf{H} = \text{diag}(h(\mathbf{k})) \). Then, the temporal dynamics of \( \alpha_k^{(R)}(t) \) and \( \alpha_k^{(I)}(t) \) is given as follows such that the approximate process (27) satisfies the real-valued SPDE (1):

For \( \mathbf{k}' \in \Omega_1 \),

\[
\dot{\alpha}_{k'}^{(R)}(t) = C^{-1} \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) \Psi_{1,5,9}(\mathbf{k}, \mathbf{k}') + 2C^{-1} \sum_{k \in \Omega_2} \alpha_k^{(R)}(t) \Psi_{1,5,9}(\mathbf{k}, \mathbf{k}') + 2C^{-1} \sum_{k \in \Omega_2} \alpha_k^{(I)}(t) \Psi_{2,6,10}(\mathbf{k}, \mathbf{k}') + \beta_k^{(R)}(t) + \tilde{e}_{k'}^{(R)}(t) \tag{33}
\]

For \( \mathbf{k}' \in \Omega_2 \),

\[
\dot{\alpha}_{k'}^{(R)}(t) = (2C)^{-1} \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) \Psi_{1,5,9}(\mathbf{k}, \mathbf{k}') + C^{-1} \sum_{k \in \Omega_2} \alpha_k^{(R)}(t) \Psi_{1,5,9}(\mathbf{k}, \mathbf{k}') + C^{-1} \sum_{k \in \Omega_2} \alpha_k^{(I)}(t) \Psi_{2,6,10}(\mathbf{k}, \mathbf{k}') + \beta_k^{(R)}(t) + \tilde{e}_{k'}^{(R)}(t) \tag{34}
\]

\[
\dot{\alpha}_{k'}^{(I)}(t) = (2C)^{-1} \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) \Psi_{3,7,11}(\mathbf{k}, \mathbf{k}') + C^{-1} \sum_{k \in \Omega_2} \alpha_k^{(R)}(t) \Psi_{3,7,11}(\mathbf{k}, \mathbf{k}') + C^{-1} \sum_{k \in \Omega_2} \alpha_k^{(I)}(t) \Psi_{4,8,12}(\mathbf{k}, \mathbf{k}') + \beta_k^{(I)}(t) + \tilde{e}_{k'}^{(I)}(t) \tag{35}
\]

where \( \Psi_{1,5,9} = \Psi_1 + \Psi_5 + \Phi_9 \), \( \Psi_{2,6,10} = \Psi_2 + \Psi_6 + \Phi_{10} \), \( \Psi_{3,7,11} = \Psi_3 + \Psi_7 + \Phi_{11} \), \( \Psi_{4,8,12} = \Psi_4 + \Psi_8 + \Phi_{12} \), and

\[
\Psi_1(\mathbf{k}, \mathbf{k}') = \int \tilde{\sigma}_s^T \tilde{k}_f^{(I)}(s) \tilde{f}_k^{(R)}(s) ds, \quad \Psi_2(\mathbf{k}, \mathbf{k}') = -\int \tilde{\sigma}_s^T \tilde{k}_f^{(R)}(s) \tilde{f}_k^{(I)}(s) ds,
\]

\[
\Psi_3(\mathbf{k}, \mathbf{k}') = \int \tilde{\sigma}_s^T \tilde{k}_f^{(I)}(s) \tilde{f}_k^{(I)}(s) ds, \quad \Psi_4(\mathbf{k}, \mathbf{k}') = -\int \tilde{\sigma}_s^T \tilde{k}_f^{(R)}(s) \tilde{f}_k^{(I)}(s) ds,
\]

\[
\Psi_5(\mathbf{k}, \mathbf{k}') = \int (\tilde{k}_s^T \tilde{D}_s \tilde{k}_f^{(I)}(s) - \nabla \cdot \tilde{D}_s \tilde{k}_f^{(I)}(s)) \tilde{f}_k^{(R)}(s) ds,
\]

\[
\Psi_6(\mathbf{k}, \mathbf{k}') = \int (\tilde{k}_s^T \tilde{D}_s \tilde{k}_f^{(I)}(s) - \nabla \cdot \tilde{D}_s \tilde{k}_f^{(I)}(s)) \tilde{f}_k^{(R)}(s) ds,
\]

\[
\Psi_7(\mathbf{k}, \mathbf{k}') = \int (\tilde{k}_s^T \tilde{D}_s \tilde{k}_f^{(I)}(s) - \nabla \cdot \tilde{D}_s \tilde{k}_f^{(I)}(s)) \tilde{f}_k^{(I)}(s) ds,
\]

\[
\Psi_8(\mathbf{k}, \mathbf{k}') = \int (\tilde{k}_s^T \tilde{D}_s \tilde{k}_f^{(I)}(s) - \nabla \cdot \tilde{D}_s \tilde{k}_f^{(I)}(s)) \tilde{f}_k^{(I)}(s) ds,
\]

\[
\Psi_9(\mathbf{k}, \mathbf{k}') = -\int \zeta_s \tilde{f}_k^{(R)}(s) \tilde{f}_k^{(I)}(s) ds, \quad \Psi_{10}(\mathbf{k}, \mathbf{k}') = -\int \zeta_s \tilde{f}_k^{(I)}(s) \tilde{f}_k^{(R)}(s) ds, \quad \Psi_{11}(\mathbf{k}, \mathbf{k}') = -\int \zeta_s \tilde{f}_k^{(R)}(s) \tilde{f}_k^{(I)}(s) ds,
\]

\[
\Psi_{12}(\mathbf{k}, \mathbf{k}') = -\int \zeta_s \tilde{f}_k^{(I)}(s) \tilde{f}_k^{(I)}(s) ds,
\]

\[
\tilde{k} = 2\pi k.
\]

Proof of proposition 3:

Note that, \( \tilde{\sigma}_s^T \tilde{k}_f^{(I)}(s) = -\tilde{\sigma}_s^T \tilde{k}_f^{(I)}(s), \) \( \tilde{\sigma}_s^T \nabla \tilde{k}_f^{(I)}(s) = \tilde{\sigma}_s^T \tilde{k}_f^{(R)}(s) = \tilde{\sigma}_s^T \tilde{k}_f^{(R)}(s) = \nabla \cdot \tilde{D}_s \tilde{k}_f^{(R)}(s) = -\tilde{k}_s^T \tilde{D}_s \tilde{k}_f^{(R)}(s) = \tilde{k}_s^T \tilde{D}_s \tilde{k}_f^{(R)}(s) + \nabla \cdot \tilde{D}_s \tilde{k}_f^{(R)}(s). \)
Then, based on the Galerkin projection, we respectively multiply the LHS of the SPDE (1) by \( \tilde{f}_{k'}^{(R)}(s) \) and \( \tilde{f}_{k'}^{(I)}(s) \) for \( k' \in \Omega_1 \cup \Omega_2 \), and integrate the product over the spatial domain. Then, the LHS of (1) becomes

\[
\int_\mathcal{S} \frac{\partial}{\partial t} \xi(t,s) \tilde{f}_{k'}^{(R)}(s)ds
\]

\[
= \int_\mathcal{S} \left( \sum_{k \in \Omega_1} \hat{\alpha}_k^{(R)}(t) \tilde{f}_{k}^{(R)}(s) + 2 \sum_{k \in \Omega_2} (\hat{\alpha}_k^{(R)}(t) \tilde{f}_{k}^{(R)}(s) + \hat{\alpha}_k^{(I)}(t) \tilde{f}_{k}^{(I)}(s)) \right) \tilde{f}_{k'}^{(R)}(s)ds \quad (36)
\]

\[
= \begin{cases} 
C \hat{\alpha}_k^{(R)}(t) & \text{for } k' \in \Omega_1 \\
2C \hat{\alpha}_k^{(R)}(t) & \text{for } k' \in \Omega_2
\end{cases}
\]

and

\[
\int_\mathcal{S} \frac{\partial}{\partial t} \xi(t,s) \tilde{f}_{k'}^{(I)}(s)ds
\]

\[
= \int_\mathcal{S} \left( \sum_{k \in \Omega_1} \hat{\alpha}_k^{(R)}(t) \tilde{f}_{k}^{(R)}(s) + 2 \sum_{k \in \Omega_2} (\hat{\alpha}_k^{(R)}(t) \tilde{f}_{k}^{(R)}(s) + \hat{\alpha}_k^{(I)}(t) \tilde{f}_{k}^{(I)}(s)) \right) \tilde{f}_{k'}^{(I)}(s)ds \quad (37)
\]

\[
= \begin{cases} 
0 & \text{for } k' \in \Omega_1 \\
2C \hat{\alpha}_k^{(I)}(t) & \text{for } k' \in \Omega_2
\end{cases}
\]

Similarly, we respectively multiply the RHS of the SPDE (1) by \( \tilde{f}_{k'}^{(R)}(s) \) and \( \tilde{f}_{k'}^{(I)}(s) \) for \( k' \in \Omega_1 \cup \Omega_2 \), and integrate the product over the spatial domain. Then, the first term \( -\tilde{v}_s^T \nabla \xi(t,s) \) on the RHS of (1) becomes:

\[-\tilde{v}_s^T \nabla \xi(t,s) = \sum_{k \in \Omega_1} \tilde{v}_s^T \tilde{k}\tilde{f}_{k}^{(I)}(s)\alpha_k^{(R)}(t) + 2 \sum_{k \in \Omega_2} \left\{ \tilde{v}_s^T \tilde{k}\tilde{f}_{k}^{(I)}(s)\alpha_k^{(R)}(t) - \tilde{v}_s^T \tilde{k}\tilde{f}_{k}^{(R)}(s)\alpha_k^{(I)}(t) \right\}. \quad (38)\]

Hence, for any \( k' \in \Omega_1 \cup \Omega_2 \), we have

\[-\int_\mathcal{S} \tilde{v}_s^T \nabla \xi(t,s) \tilde{f}_{k'}^{(R)}(s)ds = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t)\Psi_1(k,k') + 2 \sum_{k \in \Omega_2} \left\{ \alpha_k^{(R)}(t)\Psi_1(k,k') + \alpha_k^{(I)}(t)\Psi_2(k,k') \right\} \quad (39)\]

\[-\int_\mathcal{S} \tilde{v}_s^T \nabla \xi(t,s) \tilde{f}_{k'}^{(I)}(s)ds = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t)\Psi_3(k,k') + 2 \sum_{k \in \Omega_2} \left\{ \alpha_k^{(R)}(t)\Psi_3(k,k') + \alpha_k^{(I)}(t)\Psi_4(k,k') \right\} \quad (40)\]
The second term $\nabla \cdot D_s \nabla \zeta(t, s)$ on the RHS of (1) becomes:

$$\nabla \cdot D_s \nabla \zeta(t, s) = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) (-k^T D_s k f_k^{(R)} - \nabla \cdot D_s k f_k^{(I)})$$

$$+ 2 \sum_{k \in \Omega_2} \alpha_k^{(R)}(t) (-k^T D_s k f_k^{(R)} - \nabla \cdot D_s k f_k^{(I)}) + 2 \sum_{k \in \Omega_2} \alpha_k^{(I)}(t) (-k^T D_s k f_k^{(I)} - \nabla \cdot D_s k f_k^{(R)}).$$

(41)

Hence, for any $k' \in \Omega_1 \cup \Omega_2$, we have

$$\int_S \vec{v}_s^T \nabla D_s \nabla \zeta(t, s) f_k^{(R)}(s) ds = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) \Psi_5(k, k') + 2 \sum_{k \in \Omega_2} \left\{ \alpha_k^{(R)}(t) \Psi_5(k, k') + \alpha_k^{(I)}(t) \Psi_6(k, k') \right\}$$

(42)

$$\int_S \vec{v}_s^T \nabla D_s \nabla \zeta(t, s) f_k^{(I)}(s) ds = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) \Psi_7(k, k') + 2 \sum_{k \in \Omega_2} \left\{ \alpha_k^{(R)}(t) \Psi_7(k, k') + \alpha_k^{(I)}(t) \Psi_8(k, k') \right\}.$$  

(43)

The third term $-\zeta(t, s)$ on the RHS of (1) becomes:

$$-\zeta(t, s) = -\zeta(t, s) \left( \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) f_k^{(R)}(s) + 2 \sum_{k \in \Omega_2} (\alpha_k^{(R)}(t) f_k^{(R)}(s) + \alpha_k^{(I)}(t) f_k^{(I)}(s)) \right).$$

(44)

Hence, for any $k' \in \Omega_1 \cup \Omega_2$, we have

$$\int_S -\zeta(t, s) f_k^{(R)}(s) ds = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) \Psi_9(k, k') + 2 \sum_{k \in \Omega_2} \left\{ \alpha_k^{(R)}(t) \Psi_9(k, k') + \alpha_k^{(I)}(t) \Psi_{10}(k, k') \right\}.$$  

(45)

$$\int_S -\zeta(t, s) f_k^{(I)}(s) ds = \sum_{k \in \Omega_1} \alpha_k^{(R)}(t) \Psi_{11}(k, k') + 2 \sum_{k \in \Omega_2} \left\{ \alpha_k^{(R)}(t) \Psi_{11}(k, k') + \alpha_k^{(I)}(t) \Psi_{12}(k, k') \right\}.$$  

(46)

as was to be proved $\blacksquare$

**Remark (nonlinear energy transfer between multiple scales):** In Proposition 3, $\Psi_1, ..., \Psi_4$ determine the transition of Fourier mode $\alpha_k^{(R)}(t) - j \alpha_k^{(I)}(t)$ due to convection under the velocity field $\vec{v}_s$. In a special case when $\vec{v}_s$ is constant and the source-sink process is ignored, the magnitude (“energy”) of each Fourier mode is conserved under $\Psi_1, ..., \Psi_4$ only with shifted argument. This can be clearly shown by letting $\vec{v}_s = \vec{v}$, $\beta_k^{(R)} = \beta_k^{(I)} = \varepsilon_k^{(R)}(t) = \varepsilon_k^{(I)}(t) = 0$. Then, for $k' \in \Omega_2$, $\Psi_1 = \Psi_4 = 0$, $\Psi_2 = -\Psi_3 = -2C' \pi \vec{v} k'$, and it follows from (34) and (35) that

$$\dot{\alpha}_k^{(R)}(t) = -2C' \pi \alpha_k^{(I)}(t) \vec{v} k', \quad \dot{\alpha}_k^{(I)}(t) = 2C' \pi \alpha_k^{(R)}(t) \vec{v} k'.$$  

(47)
Hence,
\[
\frac{d}{dt} \left\{ (\alpha_{k'}^{(R)}(t))^2 + (\alpha_{k'}^{(I)}(t))^2 \right\} = 2\alpha_{k'}^{(R)}(t)\dot{\alpha}_{k'}^{(R)}(t) + 2\alpha_{k'}^{(I)}(t)\dot{\alpha}_{k'}^{(I)}(t)
\]
\[
= 4C\pi \bar{v} k' (-\alpha_{k'}^{(I)}(t)\alpha_{k'}^{(R)}(t) + \alpha_{k'}^{(I)}(t)\alpha_{k'}^{(R)}(t)) = 0
\]

which implies the conservation of energy at each Fourier mode \( \alpha_k^{(R)}(t) - j\alpha_k^{(I)}(t) \) under the special case with constant convection-diffusion and zero source-sink.

However, \( \bar{v}_s \) varies in space in our paper. As a result, the evolution of multiple Fourier modes must interact with each other as clearly shown in (33)\textendash(35) (also see the discussions under Proposition 1). Hence, although the convection under a spatially-varying velocity field \( \bar{v}_s \) does not increase nor decrease the total energy within the system, it re-distributes the energy to different Fourier modes which can be viewed as the non-linear energy transfer between multiple scales.

4 Integration with the DSTM Framework

The results obtained in Sections 2 and 3 enable us to embed the physics, governed by the convection-diffusion process (1), into the framework of Hierarchical Spatio-Temporal Models (DSTM) described in Wikle et al. (1998); Cressie and Wikle (2011); Katzfuss et al. (2019). The DSTM has been shown to be effective in handling non-stationary, irregular complex processes with either continuous, discrete or multivariate data, and provides computational advantages for inference, monitoring, prediction and dynamic monitoring network design. The DSTM framework consists of a data model, a process model, and a parameter model. The data model can be directly obtained from the spectral decomposition (2) and is given by:

Data Model: \( Y(t) = F\alpha(t) + V(t) \) (49)

where \( Y(t) = (Y_{s_1}(t), ..., Y_{s_N}(t))^T \) is a \( N \times 1 \) vector that contains the data observed at time \( t \), \( \alpha(t) \) is a \( K \times 1 \) vector that contains the spectral coefficients, \( F = (f(s_1), f(s_2), ..., f(s_N))^T \) is a \( N \times K \) observation matrix of Fourier functions, and the noise term \( V(t) \) is added to capture the unexplained variation such as observation error.

The process model is the temporal process \( \alpha(t) \) which is motivated by the physical process (1) and can be obtained from (16) as follows:

Process Model: \( \alpha(t) = \exp(G\Delta)\alpha(t - \Delta) + \beta(t - \Delta)\Delta + W_\alpha(t) \) (50)

where \( W_\alpha(t) \sim N(0, \int_0^\Delta \exp(G(\Delta - \tau))H \exp^*(G^T(\Delta - \tau))d\tau) \) obtained from (17).

If necessary, a parameter model can also be included to incorporate any prior information on model parameters from other sources, especially from physical model outputs which are widely available for many spatio-temporal modeling problems. For example, when modeling the temperature field in a data center, information on the velocity field (i.e., air flow) and diffusion matrix can usually be derived from the existing CFD model and flow sensors.
Another example is the modeling of urban air pollution process or the motion of extreme weather systems, information on wind field is usually available from Numerical Weather Prediction (NWP) models (Liu et al., 2016, 2018; Zhuk et al., 2017).

In the convection-diffusion process (1), \( Q(t, s) \) is the deterministic source (i.e., generation) and sink (i.e., destruction) of the quantify of interest. As discussed in Section 2, \( Q(t, s) \) admits a spectral representation \( Q(t, s) = f^T(s)\beta(t) \) with \( \beta(t) = (\beta^{(1)}(t), ..., \beta^{(K)}(t))^T \) dynamically evolving over time. If an AR(1) process is considered for \( \beta(t) \), we obtain

**Proposition 4.** Under the assumptions A1∼A4 made in Proposition 1 and consider an AR(1) process for \( \beta(t) \) such that \( \beta(t) = M\Delta\beta(t - \Delta) + W(t) \) where \( W(t) \sim N(0, \tau_\beta^2 I_K) \) and \( M\Delta = \rho\Delta I_K \), a dynamical model with local linear growth can be written as:

\[
\begin{align*}
Y(t) &= (F, 0_{N,K})\theta(t) + V \\
\theta(t) &= \tilde{G}\theta(t - \Delta) + W(t)
\end{align*}
\]

(51)

where

\[
\tilde{G} = \left( \begin{array}{c}
\frac{\exp(G\Delta)}{0_K} \\
\frac{\Delta I_K}{M\Delta}
\end{array} \right)
\]

(52)

and \( \theta(t) = (\alpha(t), \beta(t))^T \), \( V \sim N(0, \tau_\beta^2 I_N) \), \( W(t) = (W_\alpha(t), W_\beta(t))^T \).

It is easy to see that \( W(t) \sim N(0, \Sigma_W) \), and

\[
\Sigma_W = \left( \int_0^\Delta \exp(G(\Delta - \tau))H \exp^*(G^T(\Delta - \tau))d\tau \right) \tau_\beta^2 I_K.
\]

(53)

For large spatio-temporal data, the traditional geostatistical modeling paradigm is limited due to the prohibitive \( O((TN)^3) \) operations associated with factorizing the dense covariance matrices. The conditioning structure of the proposed DSTM reduces the cubic time complexity to \( O(TN^3) \) if the Kalman filter and FFBS (Forward Filtering and Backward Sampling) are used (Carter and Kohn, 1994). More importantly, since the data process (49) is based on the spectral representation of a spatial process, Sigrist et al. (2015) noted that the Kalman filter can be efficiently performed in the spectral space for gridded data by converting \( Y(t) \) from the physical space to \( \tilde{Y}(t) \) in the spectral space using FFT which requires \( O(TN\log(N)) \). The dynamical model (51) in the spectral space is given by

\[
\begin{align*}
\tilde{Y}(t) &= \tilde{F}\theta(t) + \tilde{V} \\
\theta(t) &= \tilde{G}\theta(t - \Delta) + W(t)
\end{align*}
\]

(54)

where \( \tilde{Y}(t) \) is the Fourier transform of \( Y(t) \) obtained from FFT, \( \tilde{F} = (I_K, 0_K) \), and \( V \sim N(0, \tau_\beta^2 I_K) \).
Note that, the dimensions of \( \tilde{Y}(t), \theta(t) \) and \( \hat{F} \) are respectively \( K \times 1, 2K \times 1 \) and \( K \times 2K \). In practice, it is possible to only keep the low frequency components and make \( K \) much smaller than \( N \) (see Figure 2). Hence, the choice of \( K \) in (2) controls the dimension as well as the computational cost of the problem.

**Proposition 5 (Kalman filter in the spectral space).** Consider a dynamical model specified by (54). Let \( m_{t|t-\Delta} \) and \( m_{t|t} \) respectively be the one-step-ahead predictive and filtering means of \( \theta(t) \), let \( Q_{t|t-\Delta} \) and \( Q_{t|t} \) respectively be the one-step-ahead predictive and filtering covariance matrices of \( \theta(t) \), and let \( m_{0|0} \sim N(\bar{y}(0), Q_{0|0}) \), then, the following statements hold:

The one-step-ahead predictive mean and covariance matrix of \( \theta(t) \) are

\[
\begin{align*}
    m_{t|t-\Delta} &= \tilde{G} m_{t-\Delta|t-\Delta} \\
    Q_{t|t-\Delta} &= Q_{t-\Delta|t-\Delta} \tilde{G} + \Sigma \nonumber
\end{align*}
\]

The filtering mean and covariance matrix of \( \theta(t) \) are

\[
\begin{align*}
    m_{t|t} &= \tilde{G} m_{t|t-\Delta} + 
    \begin{pmatrix}
        Q_{t|t-\Delta}^{[1:K,1:K]} (Q_{t|t-\Delta}^{[1:K,1:K]} + \tau_{\tilde{V}} I_K)^{-1} (\tilde{Y}(t) - m_{t|t-\Delta}^{[1:K]}) \\
        0_K
    \end{pmatrix} \\
    Q_{t|t} &= Q_{t|t-\Delta} + 
    \begin{pmatrix}
        Q_{t|t-\Delta}^{[1:K,1:K]} (Q_{t|t-\Delta}^{[1:K,1:K]} + \tau_{\tilde{V}} I_K)^{-1} Q_{t|t-\Delta}^{[1:K,1:K]} \\
        0_K
    \end{pmatrix} \nonumber
\end{align*}
\]

where \([1:K,1:K]\) returns the first \( K \) rows and \( K \) columns of a matrix.

The iterations above can be easily obtained from the classical Kalman Filter (Petris et al., 2009) by noting that \( Q_{t|t} \) and \( Q_{t|t-\Delta} \) are diagonal, \( \tilde{G} Q_{t-\Delta|t-\Delta} \tilde{G}^T = Q_{t-\Delta|t-\Delta} \tilde{G}^T \), \( \tilde{F} Q_{t|t-\Delta} \tilde{F}^T = Q_{t|t-\Delta}^{[1:K,1:K]} \), and \( \tilde{F} m_{t|t-\Delta} = m_{t|t-\Delta}^{[1:K]} \). Hence, \( Q_{t|t-\Delta}^{[1:K,1:K]} + \tau_{\tilde{V}} I_K \) is diagonal, and the computational cost of the spectral Kalman Filter is linear in both \( T \) and \( K \), i.e., \( O(TK) \).

The dynamical model (54) contains a large number unknown parameters to be estimated. These parameters include: (i) the velocity field \( \tilde{v}_s \), diffusivity \( D_s \), and decay \( \zeta_s \) which vary in space; (ii) \( \rho_\Delta \) and \( \tau_\beta \); (iii) the spectral density \( H = \text{diag}(h(k_j)) \) of \( \tilde{e}(t) \); and (iv) \( \tau_{\tilde{V}} \). Here, \( \tilde{v}_s, D_s, \zeta_s \) and \( \rho_\Delta \) are needed for computing \( \tilde{G} \) in (52), \( \rho_\Delta \) and \( H \) are needed for evaluating the covariance matrix \( \Sigma \), and \( \tau_{\tilde{V}} \) determines the covariance matrix of \( \tilde{V} \).

The log-likelihood function can be constructed based on the one-step-ahead predictive distribution of \( \tilde{Y}(t) \) obtained from the Kalman Filter iterations (West and Harrison, 1997; Petris et al., 2009):

\[
\begin{align*}
    l &= -\frac{1}{2} \sum_{t=1}^{T} \log \det(Q_{t|t-\Delta}^{[1:K,1:K]} + \tau_{\tilde{V}} I_K) \\
    &\quad -\frac{1}{2} (\tilde{Y}(t) - m_{t|t-\Delta}^{[1:K]})^T (Q_{t|t-\Delta}^{[1:K,1:K]} + \tau_{\tilde{V}} I_K)^{-1} (\tilde{Y}(t) - m_{t|t-\Delta}^{[1:K]}). \tag{57}
\end{align*}
\]
Note that, the problem apparently becomes intractable if we allow the velocity field \( \vec{v}_s \), diffusivity \( D_s \), and decay \( \zeta_s \) to arbitrarily vary in the spatial domain. Hence, it is necessary to impose some assumptions (such as spatial smoothness) on \( \vec{v}_s \), \( D_s \) and \( \zeta_s \). In this paper, we consider a flexible locally weighted mixture of linear regression model (Stroud et al., 2001) for the velocity field \( \vec{v}_s \), although other modeling approaches are possible given the problem of interest. Let \( v_s^{(x)} \) and \( v_s^{(y)} \) respectively be the horizontal and vertical components of the velocity vector \( \vec{v}_s \) at \( s \) and let \( v^{(\max)} > 0 \) be some specified maximum speed magnitude, we assume

\[
v_s^{(x)} = v^{(\max)} \tanh \left\{ \sum_{j=1}^{J} \pi_j(s) b_j^T(s) \gamma_j^{(x)} \right\}, \quad v_s^{(y)} = v^{(\max)} \tanh \left\{ \sum_{j=1}^{J} \pi_j(s) b_j^T(s) \gamma_j^{(y)} \right\}
\]

(58)

where \( b_j(s) \) is a column vector of known basis functions, \( \gamma_j^{(x)} \) and \( \gamma_j^{(y)} \) are column vectors of unknown parameters, and \( \pi_j \) is a non-negative kernel centered at chosen locations. Note that, the hyperbolic tangent, \( \tanh(\cdot) \), is used to bound both \( v_s^{(x)} \) and \( v_s^{(y)} \) within \([-v^{(\max)}, v^{(\max)}] \).

Let \( \mathbf{v}^{(x)} = (v_s^{(x)}, ..., v_{s_N}^{(x)})^T \, \mathbf{v}^{(y)} = (v_s^{(y)}, ..., v_{s_N}^{(y)})^T \), \( B_j = (b_j(s_1), ..., b_j(s_N)) \) and \( \pi_j = (\pi_j(s_1), ..., \pi_j(s_N)) \), we have

\[
\mathbf{v}^{(x)} = \tanh \left\{ \left( \text{diag}(\pi_1) B_1, ..., \text{diag}(\pi_J) B_J \right) \begin{pmatrix} \gamma_1^{(x)} \\ \vdots \\ \gamma_J^{(x)} \end{pmatrix} \right\}
\]

\[
\mathbf{v}^{(y)} = \tanh \left\{ \left( \text{diag}(\pi_1) B_1, ..., \text{diag}(\pi_J) B_J \right) \begin{pmatrix} \gamma_1^{(y)} \\ \vdots \\ \gamma_J^{(y)} \end{pmatrix} \right\}
\]

(59)

Similarly, it is possible to model the decay \( \zeta_s \) based on the same idea. The diffusivity \( D_s \) usually depends on the velocity field \( \vec{v}_s \), through some functions motivated by fundamental physics. For some applications, prior information about \( \vec{v}_s, D_s \), and \( \zeta_s \) might be available. For example, in the modeling weather systems such as tropical thunderstorms, the velocity, diffusivity and decay can be generated from Numerical Weather Prediction models. For another example, in the modeling of the temperature field in a DC computer room, the air flow can be obtained from the Navier-Stokes equations. Hence, it is sometimes possible to treat \( \vec{v}_s, D_s \) and \( \zeta_s \) as known, or, incorporate the prior knowledge on these parameters in a Bayesian framework.

5 Numerical Example and Case Study

A numerical example is firstly presented to investigate the basic properties and generate some critical insights of the proposed approach. A case study is followed to illustrate the application of the proposed method on the modeling of weather radar image data for extreme tropical thunderstorms.
5.1 Numerical Example

The numerical example is based on a simulated spatio-temporal data set shown in Figure 4. The data are generated from the SPDE (1) on a $20 \times 20$ gridded spatial area for $t = 0, 1, \ldots, 10$. Both the initial condition ($t = 0$) and velocity field $\vec{v}_s$ are shown on the left panel of Figure 4. The velocity vectors vary over the spatial domain, and a small vortex is created in the southeast region to test the capabilities of the proposed model in capturing such complexities. The diffusivity $D_s$ is set to an identify matrix. The decay $\zeta$ is set to a constant 0.9. The error process $\varepsilon(t, s)$ is generated by $f^T(s)\tilde{\varepsilon}(t)$ where $\varepsilon(t)$ is a white-in-time but spatially correlated Gaussian noise with the spectral density $0.05I_K$. A spatially-varying source-sink term $Q(s, t)$ is also included in the process.

![Figure 4: The data set (simulated from (1)) used for the numerical example. The panel on the left shows the initial condition and the spatially-varying velocity field. The figures on the right show the simulated images for $t = 1, 2, \ldots, 10$.](image)

The model parameters are estimated by MLE. Here, the velocity field is modeled by (59) with $J = 4$ and $v^{(\text{max})} = 0.19$, which is approximately one fifth of the width of the spatial domain in the horizontal or vertical direction. Note that, in the absence of any prior knowledge on the velocity field, the four kernels are uniformly allocated over the spatial domain:
When prior knowledge on the velocity field is available as in some applications, such knowledge can guide us to appropriately place the kernels. In radar-based precipitation nowcasting, for example, relevant information on wind field can be obtained from Numerical Weather Predictions and more kernels can be placed to areas where the variability of wind is expected to be high.

Figure 5 shows the actual and estimated velocity fields. The centers of the four kernels in (59) are also shown in the figure. We see that the proposed approach accurately estimates the velocity field which varies in the spatial domain. In particular, the small vortex in the southeast region is successfully captured, demonstrating the capabilities of the proposed approach in modeling spatio-temporal data arising from a physical process (1) under spatially-varying velocity field.

![Figure 5: Actual (left) and estimated (right) velocity fields](image)

We further compare the proposed approach with two major existing approaches: OF (Optical Flow) and COTREC (Tracking Radar Echoes by Correlation). OF is a typical gradient-based method which assumes that the image brightness level is time-invariant and constructs the velocity field based on the first-order variation (Horn and Schunck, 1981). In meteorology, gradient-based methods are referred to as the method of Lagrangian Persistence, and are now the building block of many storm-tracking algorithms (RMI, 2008). On the other hand, COTREC is a typical pattern-based method based on the concept of area tracking. Tracking areas (image pixel arrays) are defined around all pixel grid points, and corresponding areas are searched in the next image by maximizing the cross-correlation between areas. Then, the velocity field can be constructed given the spatial lags between areas and the time lag between two images. Unlike gradient-based methods, pattern-based methods assume that the shape of image brightness patterns within defined areas do not
change over short time intervals (Li and Lai, 2004). Figure 6 shows the box plot of the error associated with the estimated velocity vectors in both horizontal and vertical directions. It is clearly seen that the proposed approach provides more accurate predictions. This is mainly due to the fact that the underlying dynamics of the physical process (1) is built into the proposed statistical model. However, the intensity-invariant assumption, required by OF, is violated because a source-sink and a decay term are included when the data are simulated (which is often the case in reality). As for the COTREC method, the shape-invariant assumption can hardly be justified for spatially-varying velocity field. For example, the vortex included in the velocity field constantly changes the pattern of the images.

The Kalman Filter in spectral domain (Proposition 5) is used to obtained the filtered $\theta$, which is a $361 \times 1$ vector in the dynamical model (54). Then, inverse Fourier transform is applied to construct the filtered images, and the reconstructed images at times $t = 2, 4, 6, 8, 10$ are shown in Figure 7.

### 5.2 Case Study: Radar-Based Weather System Modeling

Following the numerical example, a case study is presented in which the proposed method is used to model a highly chaotic weather system—a tropical thunderstorm—based on weather radar images. Weather radar echoes, correlated in both space and time, provide a rich source of information for short-term precipitation nowcasting. In the meteorological community, the methods which are used to perform spatio-temporal extrapolation/advection of radar reflectivity (echo) field are collectively known as the radar-based Quantitative Precipitation Forecasts (QPF) (RMI, 2008).

Figure 8 (the top row) shows a sequence of radar images generated by a dual polarization
Meteorological Doppler Weather Radar (MDWR) system. These images, generated at 5-min intervals and on a Cartesian 2D grid of 200×200 pixels, contain the standard Constant Altitude Plan Position Indicator (CAPPI) reflectivity data at 1 km above the mean sea level. The legend shows the precipitation rate (mm/hr) converted from the CAPPI data using Marshall-Palmer relationship. The centers of two neighboring pixels are approximately 0.5km apart, and the spatial area covered by an image is approximately 10,000km². It is possible to see that the thunderstorm system is moving from west to east driven by a velocity (wind) field, which is not directly observable. As tropical storms are highly chaotic with small-scale localized convective cells, radar images are noisy and non-smooth as clearly shown in Figure 8 (the top row). Hence, we perform FFT on the original images and reconstruct the low-pass filtered images by only retaining the low-frequency components (with angular wavenumber less than or equal 8 radians per unit distance in both directions). The low-pass filtered images are shown in the bottom row of Figure 8.

The low-pass filtered images, which are much smoother, are used for estimating the unknown model parameters including the wind field. As discussed in Section 2, one of the main advantages of the spectral-decomposition-based method is that it allows us to achieve a significant dimension reduction by discarding high-frequency components. For example, in this case study, the original dimension of \( \theta \) in (54) is 10,000 × 1 if all frequency terms are retained, while the dimension of \( \theta \) becomes 289×1 by only keeping low-frequency components with angular wavenumber (≤ 8 radians per unit distance).

In addition, to further accelerate the computation for (streaming) spatio-temporal radar reflectivity data, the dynamical model (54) naturally motivates us to adopt the Sequential Monte Carlo (SMC) or Particle Filter which has been proven useful for on-line filtering with unknown parameters. In this case study, we implement the SMC algorithm in the spectral domain and leverage the computing power of Graphics Processing Units (GPU). For GPU computing, the SMC algorithm must satisfy the following three criteria: 1) every thread must execute the same instructions, 2) minimum, ideally zero, data transfer to and from...
GPU during the simulation process, 3) high throughput computing, i.e., a large number of simultaneous threads with minimal interactions between threads (Terenin et al., 2019). Hence, for the classical SMC algorithm, while the propagation and weighting steps satisfy all three criteria, the resampling step clearly violates the last criterion as collective operations, such as summation, is required across all particles. Murray et al. (2016) proposed a Metropolis type of resampler which does not require collective operations across particles. By integrating the Metropolis resampling algorithm into the classical SMC algorithm in the spectral domain, the following SMC algorithm is adopted in this case study:

\[
\text{Data: } Y(t) \xrightarrow{\text{FFT}} \tilde{Y}(t) \text{ for } t = 1, \ldots, T \text{ and load } \{\tilde{Y}(t)\}_{t=1:T} \text{ on all threads} \\
\text{for } i=1,\ldots,n \text{ do} \\
\quad // \text{ this for loop is completely parallelized on GPU with no interaction between threads} \\
\quad \theta_i(0) \sim [\pi_0(\theta(0))] // \text{ initialize particle } i \\
\quad \omega_i(0) \leftarrow n^{-1} // \text{ initialize weight } i \\
\text{for } t=1,\ldots,T \text{ do} \\
\quad a_t \leftarrow \text{Metropolis}(\omega(t-1)) // \text{ The Metropolis algorithm is completely parallelized on} \\
\quad \text{GPU with no interaction between threads (Murray et al., 2016)} \\
\quad \text{for } i=1,\ldots,n \text{ do} \\
\quad \quad // \text{ this for loop is completely parallelized on GPU with no interaction across} \\
\quad \quad \text{threads} \\
\quad \quad \theta_i(t) \sim [\theta(t)\theta_i(t-1)] // \text{ particle propagation} \\
\quad \quad \omega_i(t) \leftarrow \text{likelihood}(\tilde{Y}(t) - (I_K,0_K)\theta_i(t)) // \text{ Bootstrap particle filter}
\]

Figure 9 shows both the observed and filtered images at selected times, \( t = 1, 3, 5 \).
The estimated spatially-varying wind field is also included in the filtered images. It can be seen that the estimated wind field clearly shows a west-to-east motion of the weather system, which is consistent with our initial observation. It is important to note that, for tropical areas where trade winds from both hemispheres meet (known as the Inter-tropical Convergence Zone), the wind fields near the equator are typically highly variable in space and is captured by the proposed approach.

Figure 9: Observed and filtered images at selected times, $t = 1, 3, 5$, with the estimated spatially-varying wind field.

For tropical thunderstorms, the growth and decay of radar reflectivity become more prominent due to the presence of many small-scale localized convective storm cells embedded in the storm systems. More importantly, the strong solar heating of land areas in tropical areas causes a phenomenon known as the convective heating. Convective heating causes the land areas to become heated more than its surroundings, and leads to significant evaporation that creates the small-scale localized convective weather cells. For such a weather system, heavy thunderstorms can develop, grow, and dissipate very suddenly in a random manner. The lifetime of convective thunderstorm cells can be as short as tens of minutes, posing a tremendous challenge to the spatio-temporal modeling of radar reflectivity data. Note that, although it has been pointed out that the errors in the linear extrapolation of the radar echo field assuming a persistent reflectivity level are mainly due to the growth and decay of reflectivity (Browning et al., 1982), many operational radar-based QPF systems do not take into account the rapid small-scale growth and decay, such as the GANDOLF system in UK (Bowler et al., 2004) and the SWIRLS system developed in Hong Kong (Li and Lai, 2004).
In addition, the rapid growth-decay of weather systems also makes the Numerical Weather Prediction (NWP) models ineffective in predicting the exact location and intensity of individual thunderstorms (RMI, 2008).

The proposed spatio-temporal model, on the other hand, has the capability to capture the growth-decay of radar reflectivity in space and time. Note that, the proposed approach is motivated from the SPDE (1) with a source-sink component $Q(s,t)$, and the dynamical model (54) incorporates the source-sink component as an AR(1) process. Figure 10 shows the estimated spatial growth-dissipation of the weather system at selected times, $t = 2, 4, 6$. These images are reconstructed through inverse Fourier transform based on the filtered $\tilde{\theta}(t)$.

![Figure 10: Observed images (top) and estimated growth-dissipation (bottom) of the weather system in space and at selected times, $t = 2, 4, 6$.](image)

The dynamical model (54) also enables us to perform short-term spatio-temporal extrapolation of radar reflectivity field, i.e., short-term precipitation nowcasting. Note that, in tropical areas, radar images are not used for mid-range or long-term predictions as tropical thunderstorm systems develop quickly and are of very short lifespan. For example, the official lead time for the warnings of heavy storms is usually between 15 and 45 minutes in tropical countries such as Singapore (NEA, 2017). Figure 11 shows both the actual and 5-, 15- and 25-minute-ahead nowcasting based on our model. The spatio-temporal extrapolation of the weather system reasonably matches with the observed reflectivity fields. The nowcasting accuracy gradually deteriorates as the prediction horizon increases, which is expected given the chaotic nature of tropical weather systems.
Finally, the proposed method is applied to a number of 199 tropical storm events recorded in 2010 and 2011. Comparison studies are performed between the proposed method and two other methods in the literature. The first method is a well-known radar-based short-term precipitation prediction approach known as COTREC (Tracking Radar Echoes by Correlation), and has been widely implemented by the meteorological community (Li and Lai, 2004). The second method is a recently proposed spatio-temporal conditional autoregressive model under the Lagrangian integration scheme (Liu et al., 2018). We respectively denote these two methods by COTREC2004 and LGK2018.

Figure 12 shows the box plot of the Mean Squared Error (MSE) associated with the extrapolated radar reflectivity fields based on the 199 storm events. The horizontal axis shows the nowcasting horizon, which ranges from 5 minutes to 30 minutes. It is seen that the proposed method outperforms both COTREC2004 and LGK2018. This advantage is primarily due to the improvement of wind field estimation by the proposed method. Both COTREC2004 and LGK2018 estimate the wind field using pattern-based methods which require the shape of radar images within defined tracking areas do not change over short time intervals, which is often violated for highly dynamic tropical thunderstorms.

6 Conclusions
This paper proposed a statistical modeling approach for spatio-temporal data arising from a class of convection-diffusion processes. Motivated by existing results in the liter-
Figure 12: Box plot of the MSE of the (low-pass) radar images based on 199 storm events. The horizontal axis indicates the nowcasting horizon ranging from 5 to 30 minutes.

ature, the proposed approach is rooted in the spectrum decomposition of spatio-temporal processes. However, unlike existing results, this paper obtained the statistical model by considering spatially-varying convection-diffusion and nonzero-mean source-sink. Due to the spatially-varying convection-diffusion, the temporal evolution of individual spectrum coefficients is coupled which extends the existing results to a more complicated but realistic scenario. This phenomenon is known as the non-linear energy transfer across multiple scales due to spatially-varying convection. As a result, the proposed spatio-temporal model has a non-stationary covariance structure which has been established by our theoretical investigation. A dynamical model, in the spectral domain, has been constructed by integrating our theoretical findings into the framework of DSTM. The conditional structure of DSTM leads to tremendous advantages in statistical inference and computation. The proposed approach has been illustrated by a numerical example as well as a case study based on real datasets. The advantages of the proposed method have been clearly demonstrated through a comparison study based on 199 storm events. Note that, since the convection-diffusion processes can be widely found in interdisciplinary domains including environmental sciences, image analysis, geology, material science, reliability engineering, etc., the proposed method has the potential to impact a spectrum of scientific and engineering applications. The R code is available at GitHub: https://github.com/dnncode/Spatio-Temporal-Model-for-SPDE.

References
Banerjee, S. (2017), “High-Dimensional Bayesian Geostatistics,” Bayesian Analysis, 12, 583–614.
Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data, 2nd ed.*, Boca Raton, Florida: CRC Press.

Banerjee, S. and Fuentes, M. (2012), “Bayesian Modeling for Large Spatial Datasets.” *WIREs Computational Statistics*, 4, 59–66.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), “Gaussian Predictive Proess Models for Large Spatial Data Sets,” *Journal of the Royal Statistical Society: Series B*, 70, 825–848.

Benjamin, S. G., Brown, J. M., Brunet, G., Lynch, P., Saito, K., and Sclatter, T. W. (2019), “100 Years of Progress in Forecasting and NWP Applications,” *American Meteorological Society*.

Berliner, L. M. (2003), “Physical-Statistical Modeling in Geophysics,” *Journal of Geophysical Research-Atmospheres*, 108, 3–10.

Bowler, N. E. H., Pierce, C. E., and Seed, A. (2004), “Development of a precipitation nowcasting algorithm based upon optical flow techniques,” *Journal of Hydrology*.

Brown, P. E., Karesen, K. F., Roberts, G. O., and Tonellato, S. (2000), “Blur-Generated Non-Separable Space-Time Models,” *Journal of the Royal Statistical Society: Series B*, 62, 847–860.

Browning, K. A., Collier, C. G., Lark, P. R., Menmuir, P., Monk, G. A., and Owens, R. G. (1982), “On the Forecasting of Frontal Rain Using a Weather Radar Network,” *Monthly Weather Review*.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992), “A Monte Carlo Approach to Non-normal and Nonlinear State-Space Modeling,” *Journal of the American Statistical Association*, 87, 493–500.

Carroll, R., Chen, E., Li, T., Newton, H., Schmiediche, H., and Wang, N. (1997), “Ozone exposure and population density in Harris county,” *Journal of the American Statistical Association*, 92, 392–404.

Carter, C. K. and Kohn, R. (1994), “On Gibbs Sampling for State Space Models,” *Biometrika*, 81, 541–553.

Cressie, N. and Huang, H. C. (1999), “Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions,” *Journal of the American Statistical Association*, 94, 1330–1340.

Cressie, N. and Johannesson, G. (2002), “Fixed Rank Kriging for Very Large Spatial Data Sets,” *Journal of the Royal Statistical Society: Series B*, 70, 209–226.

Cressie, N. and Wikle, C. (2011), *Statistics for Spatio-Temporal Data*, Hoboken, New Jersey: John Wiley & Sons.
Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016), “Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets,” *Journal of the American Statistical Association*, 111, 800–812.

EPA (2017), “Environmental Protection Agency: Revisions to the Guideline on Air Quality Models: Enhancements to the AERMOD Dispersion Modeling System and Incorporation of Approaches To Address Ozone and Fine Particulate Matter,” *Federal Register*, 82, 5182–5235.

Fuentes, M. (2007), “Approximate Likelihood for Large Irregularly Spaced Spatial Data,” *Journal of the American Statistical Association*, 102, 321–331.

Fuentes, M., Chen, L., Davis, J. M., and Lackmann, G. M. (2005), “Modeling and Predicting Complex Space-Time Structures and Patterns of Coastal Wind Fields,” *Environmetrics*, 16, 449–464.

Ghosh, S. K., Bhave, P. E., Davis, J. M., and Lee, H. (2010), “Spatio-Temporal Analysis of Total Nitrate Concentrations using Dynamic Statistical Models,” *Journal of the American Statistical Association*, 105, 538–551.

Gneiting, T. (2002), “Nonseparable, Stationary Covariance Functions for Space-Time Data,” *Journal of the American Statistical Association*, 97, 590–600.

Guan, Y., Johnson, M., Katzfuss, M., Mannshardt-Hawk, E., Messier, K. P., Reich, B. J., and Song, J. J. (2018), “Fine-scale spatiotemporal air pollution analysis using mobile monitors on Google Street View vehicles,” *arXiv:1810.03576v2*, 00, 0000–0000.

Guinness, J. and Fuentes, M. (2015), “Likelihood Approximations for Big Nonstationary Spatial-Temporal Lattice Data,” *Statistica Sinica*.

Guinness, J. and Stein, M. (2013), “Interpolation of Nonstationary High Frequency Spatial-Temporal Temperature Data,” *Annals of Applied Statistics*.

Hamann, B. K. P. and Schunck, B. G. (1981), “Determining optical flow,” *Artificial Intelligence*, 17, 185–203.
Jones, R. H. and Zhang, Y. (1997), “Models for Continuous Stationary Space-Time Processes,” in *Modelling Longitudinal and Spatially Correlated Data*, eds. Gregoire, T. G., Brillinger, D. R., Diggle, P. J., Russek-Cohen, E., Warren, W. G., and Wolfinger, R. D., New York: Springer, pp. 289–298.

Katzfuss, M. (2017), “A multi-resolution approximation for massive spatial datasets,” *Journal of the American Statistical Association*, 112, 201–214.

Katzfuss, M., Stroud, J. R., and Wikle, C. K. (2019), “Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models,” *arXiv:1704.06988v2*, 00, 0000–0000.

Kontar, R., Zhou, S. Y., Sankavaram, C., Du, X., and Zhang, Y. (2017), “Nonparametric Modeling and Prognosis of Condition Monitoring Signals Using Multivariate Gaussian Convolution Processes,” *Technometrics (to appear)*.

Krainski, E. T., Gomez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2019), *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*, Boca Raton: Chapman and Hall/CRC.

Kuusela, M. and Stein, M. (2017), “Locally stationary spatio-temporal interpolation of Argo profiling float data,” in *Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences*, vol. 474.

Li, P. and Lai, S. T. (2004), “Short-range quantitative precipitation forecasting in Hong Kong,” *Journal of Hydrology*.

Lindgren, F. and Rue, H. (2011), “An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: the Stochastic Partial Differential Equation Approach,” *Journal of the Royal Statistical Society: Series B*, 73, 423–498.

Liu, X., Gopal, V., and Kalagnanam, J. (2018), “A Spatio-Temporal Modeling Framework for Weather Radar Image Data in Tropical Southeast Asia,” *The Annals of Applied Statistics (To Appear)*, 00, 00–00.

Liu, X., Yeo, K. M., Hwang, Y. D., Singh, J., and Kalagnanam, J. (2016), “A Statistical Modeling Approach for Air Quality Data Based on Physical Dispersion Processes and Its Application to Ozone Modeling,” *The Annals of Applied Statistics*, 10, 756–785.

Mak, S., Sung, C. L., Wang, X. J., Yeh, S. T., Chang, Y. H., Joseph, V. R., Yang, V., and Wu, C. F. J. (2018), “An Efficient Surrogate Model for Emulation and Physics Extraction of Large Eddy Simulations,” *Journal of the American Statistical Association*, 113, 1443–1456.

Mondal, D. and Wang, C. (2019), “A matrix-free method for spatial-temporal Gaussian state-space models,” *Statistica Sinica (to appear)*, 00, 0000–0000.
Murray, L. M., Lee, A., and Jacob, P. E. (2016), “Parallel Resampling in the Particle Filter,” *Journal of Computational and Graphical Statistics*, 25, 789–805.

NEA (2017), *Challenges in Weather Forecasting*, [http://www.nea.gov.sg/training-knowledge/weather-climate/weather-forecast/challenges-in-weather-forecasting/](http://www.nea.gov.sg/training-knowledge/weather-climate/weather-forecast/challenges-in-weather-forecasting/), National Environmental Agency Singapore.

Nychka, D. and Wikle, C. Royle, J. A. (2002), “Multiresolution Models for Nonstationary Spatial Covariance Functions,” *Statistical Modeling*, 2, 315–331.

Petris, G., Petrone, S., and Campagnoli, P. (2009), *Dynamic Linear Models with R*, Springer.

R-INLA (2019), *The R-INLA Project*, [http://www.r-inla.org/](http://www.r-inla.org/).

Reich, B., Guinness, J., Vandekar, S., Shinohara, T., and Staicu, A. M. (2018), “Fully Bayesian Spectral Methods for Imaging Data,” *Biometrics*.

Reich, B. J., Eidsvik, J. Guindani, M., Nail, A. J., and Schmidt, A. M. (2011), “A Class of Covariate-Dependent Spatiotemporal Covariance Functions for the Analysis of Daily Ozone Concentration,” *The Annals of Applied Statistics*, 5, 2425–2447.

RMI (2008), *Quantitative Precipitation Forecasts based on radar observations: principles, algorithms and operational systems*, Royal Meteorological Institute of Belgium.

Rue, H., Martino, S., and Chopin, N. (2009), “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations,” *Journal of the Royal Statistical Society, B*.

Schabenberger, O. and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC.

Sigrist, F., Kunsch, H. R., and Stahel, W. A. (2015), “Stochastic Partial Differential Equation based Modelling of Large Space-Time Data Sets,” *Journal of the Royal Statistical Society: Series B*, 77, 3–33.

Simpson, D., Lindgren, F., and Rue, H. (2012), “In Order to Make Spatial Statistics Computationally Feasible, We need to Forget about the Covariance Function,” *Environmetrics*, 23, 65–74.

Stein, M. L., Chi, Z., and Welty, L. J. (2004), “Approximating Likelihoods for Large Spatial Data Sets,” *Journal of the Royal Statistical Society: Series B*, 66, 275–296.

Stroud, J. R., Muller, P., and Sanso, B. (2001), “Dynamic Models for Spatiotemporal Data,” *Journal of the Royal Statistical Society: Series B*, 63, 673–689.

Terenin, A., Dong, S. F., and Draper, D. (2019), “GPU-Accelerated Gibbs Sampling: A Case Study of the Horseshoe Probit Model,” *Statistics and Computing*, 29, 301–310.
West, M. and Harrison, P. (1997), *Bayesian Forecasting and Dynamical Models*, Springer.

Whittle, P. (1963), “Stochastic Processes in Several Dimensions,” *Bulletin of the International Statistical Institute*, 40, 974–994.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998), “Hierarchical Bayesian Space-Time Models,” *Environmental and Ecological Statistics*, 5, 117–154.

Wikle, C. K. and Cressie, N. (1999), “A Dimension-Reduced Approach to Space-Time Kalman Filtering,” *Biometrika*, 86, 815–829.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001), “Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds,” *Journal of the American Statistical Association*, 96, 382–397.

Yan, H., Paynabar, K., and Shi, J. (2017), “Real-time Monitoring of High-Dimensional Functional Data Streams via Spatio-Temporal Smooth Sparse Decomposition,” *Technometrics* (to appear).

Yeo, K. and Melnyk, I. (2018), “Deep learning algorithm for data-driven simulation of noisy dynamical system,” *arXiv:1802.08323*.

Zhuk, S., Tchrakian, T., Akhriev, A., Lu, S. Y., and Hamann, H. F. (2017), “Where computer vision can aid physics: dynamic cloud motion forecasting from satellite images,” *CoRR abs/1710.00194*, arXiv:1710.00194.