Received: 13 July 2020 | Accepted: 14 January 2021

DOI: 10.1002/env.2670

RESEARCH ARTICLE

WILEY

Flexible nonstationary spatiotemporal modeling of high-frequency monitoring data

Christopher J. Geoga[1,2] | **Mihai Anitescu**[2,3] | **Michael L. Stein**[1,3]

[1]Department of Statistics, Rutgers University, New Brunswick, New Jersey, USA

[2]Division of Mathematics and Computer Science, Argonne National Laboratory, Lemont, Illinois, USA

[3]Department of Statistics, University of Chicago, Chicago, Illinois, USA

Correspondence
Christopher J. Geoga, Department of Statistics, Rutgers University, New Brunswick, NJ, USA.
Email: christopher.geoga@rutgers.edu

Funding information
National Science Foundation, Grant/Award Numbers: CNS-1545046, FP061151-01-PR; U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Grant/Award Numbers: DE-AC02-06CH11347, DE-AC02-06CH11357

Abstract

Many physical datasets are generated by collections of instruments that make measurements at regular time intervals. For such regular monitoring data, we extend the framework of half-spectral covariance functions to the case of nonstationarity in space and time and demonstrate that this method provides a natural and tractable way to incorporate complex behaviors into a covariance model. Further, we use this method with fully time-domain computations to obtain bona fide maximum likelihood estimators—as opposed to using Whittle-type likelihood approximations, for example—that can still be computed conveniently. We apply this method to very high-frequency Doppler LIDAR vertical wind velocity measurements, demonstrating that the model can expressively capture the extreme nonstationarity of dynamics above and below the atmospheric boundary layer and, more importantly, the interaction of the process dynamics across it.

KEYWORDS

atmospheric boundary layer, Doppler LIDAR, meteorology, spectral domain

1 | INTRODUCTION

Gaussian process (GP) models are versatile and useful tools for studying spatial and spatiotemporal processes. A GP model is completely specified by postulating a mean and covariance function, so that a collection of data points $\boldsymbol{y} = \{y_j\}_{j=1}^N$ observed at locations $\{\boldsymbol{x}_j\}_{j=1}^N$ in $\mathbb{R}^d$ has the distribution

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where the mean vector $\boldsymbol{\mu}$ is generated by the mean function, so that $\boldsymbol{\mu}_j = \mu(\boldsymbol{x}_j)$, and $\boldsymbol{\Sigma}$ is the covariance matrix induced by the covariance function, which we denote by K, so that $\boldsymbol{\Sigma}_{j,k} = K(\boldsymbol{x}_j, \boldsymbol{x}_k)$. A common estimation problem is to select a

parametric form for K and attempt to recover the parameters from data. For the duration of this paper, we assume $\boldsymbol{\mu} \equiv 0$. While this may seem restrictive, the analysis of anomaly fields instead of raw data is commonplace, and extensions of the methods we introduce here to a nonzero mean would not be overly difficult.

If $K(\boldsymbol{x}, \boldsymbol{x}')$ is a function only of $\boldsymbol{x} - \boldsymbol{x}'$, we call the process model *stationary*, an assumption which is enormously useful for a variety of both theoretical and computational reasons. For the vast majority of real-world processes, however, even among the ones for which a Gaussian process assumption is tolerable, a stationarity assumption is not. With modern computing tools and instrument infrastructure, spatial datasets are commonly large enough and dense enough that one can determine even by visual inspection that a stationary process model would be fundamentally misrepresentative of the dynamics of the process. For this reason, developing natural and flexible nonstationary models to more accurately capture dependence structure in modern datasets is an exigent statistical problem.

Modeling such nonstationarity is difficult, however, in no small part because it is challenging to specify valid covariance models via positive definite functions. In the stationary case, perhaps the easiest way to verify that a continuous function K is a valid real-valued covariance function is to find an integrable and nonnegative function $S(\boldsymbol{f})$ that is symmetric about the origin and for which K—in its single-argument form—is its Fourier transform. One can then express K as

$$K(\boldsymbol{x} - \boldsymbol{x}') = \int e^{i2\pi\langle \boldsymbol{f}, \boldsymbol{x}-\boldsymbol{x}'\rangle} S(\boldsymbol{f}) \mathrm{d}\boldsymbol{f},$$

and by Bochner's theorem then conclude that K is a valid covariance function. The function $S(\boldsymbol{f})$ is called the *spectral density* corresponding to K. In general, it is easier to specify positive and symmetric functions than positive definite ones, and many of the more exotic stationary covariance functions are derived, or at least confirmed to be valid, in the spectral domain. Many theoretical results about GPs that are relevant to the parameter estimation problem, like those about the equivalence and orthogonality of Gaussian measures, are proven and applied in the spectral domain (Ibragimov & Rozanov, 1978; Stein, 1999; Zhang, 2004; Zhang & Zimmerman, 2005), further emphasizing that the spectral domain is arguably often a more productive way to approach studying covariance structure.

Unfortunately, nonstationarity implies that the covariance function $K(\boldsymbol{x}, \boldsymbol{x}')$ does not admit a single-argument representation as $K(\boldsymbol{x} - \boldsymbol{x}')$, and thus the validity check we describe above cannot be used. Accordingly, it is necessary to find other ways to specify valid covariance functions, and there is a wealth of literature in the field that contributes methods and ideas toward solving this problem. Examples include careful stitching together of valid "local" functions (Higdon, 2002; Paciorek & Schervish, 2006; Stein, 2005a), deforming the coordinate system of a stationary process (Anderes & Stein, 2008; Genton & Perrin, 2004; Sampson & Guttorp, 1992), spatial mixtures of different stationary functions (Fuentes & Smith, 2001), latent higher dimensional representations (Apanasovich & Genton, 2010), separable models (Genton, 2007), and many others. For a much more comprehensive list and stimulating discussion of the diverse field of covariance function specification, we refer the reader to Porcu et al. (2020).

Nonstationary models placing less emphasis on explicitly specifying valid covariance functions are also diverse and successful, for example, dynamic models (Katzfuss & Cressie, 2012; Stroud et al., 2001; Wikle & Cressie, 1999); basis expansions (Krock et al., 2020) and closely related multiresolution methods (Katzfuss, 2017; Matsuo et al., 2011; Nychka et al., 2002, 2018); stochastic partial differential equation models (Lindgren et al., 2011); and again many others. Models from categories such as these enjoy the extremely useful property that *process* models automatically yield valid covariance functions, even if evaluating the functions at arbitrary points can be difficult. In many cases these functions are only worked with and specified implicitly, but inference is still direct and feasible (Xu & Wikle, 2007). In some circumstances, these methods are arguably a more natural choice than specifying covariance functions. In other cases, however, as in the data being studied here, we consider specifying reasonable process models to be comparably difficult—if not even more difficult—to the challenge of specifying a reasonable valid covariance function, and so we see value in both approaches.

The model we introduce here represents a broadly applicable approach of specifying marginal quantities and using half-spectral representations to obtain rich classes of fully spatiotemporal covariance functions. This is only one way of many to model nonstationarity. For example, it can be contrasted to the way that Paciorek–Schervish-type model specify nonstationarity using local geometric anisotropies (Paciorek & Schervish, 2006). Unlike some of the more general-purpose methods discussed above, the specific parameterizations for marginal quantities used here are quite specialized, and would likely need to be modified to be applicable to otherly regularly measured data. For this reason, we see this approach as an appropriate choice for problems with regularly measured data where one wishes to incorporate a great deal of process-specific information into a parametric model. We demonstrate that this method can be pushed into very complex

forms while maintaining validity, numerical robustness, and expressiveness. The cost of this flexibility, however, is an admittedly substantial effort required of the practitioner.

2 | HALF-SPECTRAL COVARIANCE FUNCTIONS

A key observation made in Cressie and Huang (1999) and Stein (2005b) is that one can obtain a valid stationary model by specifying a "half-spectral" form of a covariance function for a spatiotemporal random field $Z(t, \boldsymbol{x})$ (whose single-argument covariance function now depends on the pairing of temporal *and* spatial separations), whereupon the stationary space–time covariance function is obtained by performing an inverse Fourier transform of a function $h(f, \boldsymbol{x} - \boldsymbol{x}')$ with respect to time. The function

$$K((t - t', \boldsymbol{x} - \boldsymbol{x}')) := \int_{\Omega} e^{2\pi i f(t-t')} h(f, \boldsymbol{x} - \boldsymbol{x}') \mathrm{d}f \tag{1}$$

gives a valid covariance function for discrete-time data ($\Omega = [-1/2\Delta t, 1/2\Delta t)$ for sampling rate Δt) or continuous-time data ($\Omega = \mathbb{R}$). Stein (2005b) proposes an expression for h of the form

$$h(f, \boldsymbol{x} - \boldsymbol{x}') = S(f) C_f(\boldsymbol{x} - \boldsymbol{x}') e^{ig(f)\boldsymbol{u}^T(\boldsymbol{x}-\boldsymbol{x}')}$$

for a stationary but nonseparable and asymmetric model, where $S(f)$ can be interpreted to be a marginal spectral density in time; C_f—a valid spatial correlation function—is the modulus of the coherence of $Z(\cdot, \boldsymbol{x})$ and $Z(\cdot, \boldsymbol{x}')$; and $g(f)\boldsymbol{u}^T(\boldsymbol{x} - \boldsymbol{x}')$ is the phase relation. Varying C_f in (1) allows flexibility in the nature of nonseparability, with separability being the special case of C_f and $g(f)$ not depending on f.

For measurements made at regular time intervals and a relatively small number of spatial locations, this modeling perspective can be considered as an augmented multiple time series approach for which incorporating new time series (measurements at new spatial locations) is possible in a model-consistent way and without the requirement of any new parameters. It affords the practitioner a convenient and approachable method for modeling space–time dependence primarily by thinking about marginal—or at least lower dimensional—dependence structure. If the data is regular in time, for example, one can visually inspect marginal time spectra and immediately include features of those estimates in a valid, fully spatiotemporal covariance model. See Horrell and Stein (2017) for an extension to more complex stationary models of this form and Guinness and Stein (2013) for a nonstationary model in a similar vein.

We now extend this modeling framework to the case of processes that are nonstationary in space but stationary in time. Because this modeling perspective relies so heavily on specifying marginals, a nonstationary extension is very direct.

Theorem 1. *Let $\{S_{\boldsymbol{x}}(f\ ;\ \boldsymbol{\theta})\}$ be integrable temporal spectral densities indexed by spatial locations $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}$, and $\left\{C_f(\boldsymbol{x}, \boldsymbol{x}'\ ;\ \boldsymbol{\theta})\right\}$ be real-valued correlation functions indexed by frequency f and parameters $\boldsymbol{\theta}$. If $g(f)$ is an odd function, then*

$$K((t, \boldsymbol{x}), (t', \boldsymbol{x}')) = \int_{\Omega} e^{i2\pi f(t-t') + ig(f)\boldsymbol{u}^T(\boldsymbol{x}-\boldsymbol{x}')} \sqrt{S_{\boldsymbol{x}}(f\ ;\ \boldsymbol{\theta}) S_{\boldsymbol{x}'}(f\ ;\ \boldsymbol{\theta})} C_f(\boldsymbol{x}, \boldsymbol{x}'\ ;\ \boldsymbol{\theta}) \mathrm{d}f \tag{2}$$

defines a valid covariance function for a spatially nonstationary real-valued space–time process $Z(t, \boldsymbol{x})$ with finite second moments.

Proof. As in the proof provided in Stein (2005b), we consider the matrix-valued function (suppressing the dependence on parameters $\boldsymbol{\theta}$) given by

$$\boldsymbol{\Phi}(f) := \left\{ \sqrt{S_{\boldsymbol{x}_j}(f) S_{\boldsymbol{x}_k}(f)} C_f(\boldsymbol{x}_j, \boldsymbol{x}_k) e^{ig(f)\boldsymbol{u}^T(\boldsymbol{x}_j - \boldsymbol{x}_k')} \right\}_{j,k=1}^{N} \tag{3}$$

for arbitary spatial locations $\left\{\boldsymbol{x}_j\right\}_{j=1}^{N}$ and argue via several applications of the Schur product theorem. We observe that one can write $\boldsymbol{\Phi}(f) = \boldsymbol{S}(f) \circ \boldsymbol{C}(f) \circ \boldsymbol{\mathcal{P}}(f)$, a Hadamard product of the time-spectra term, correlation function, and phase term, and we check that all three are positive semidefinite and Hermitian. $\boldsymbol{C}(f)$ is by definition positive definite, and

we note that $\boldsymbol{S}(f)$ is positive semidefinite by recognizing the rank one structure of $\boldsymbol{S}(f) = \boldsymbol{s}(f)\boldsymbol{s}(f)^T$ with $\boldsymbol{u}_j(f) = \sqrt{S_{\boldsymbol{x}_j}(f)}$. Similarly, $\mathcal{P}(f)$ has a Hermitian rank-one symmetric factorization since $e^{ig(f)\boldsymbol{u}^T(\boldsymbol{x}-\boldsymbol{x}')} = e^{ig(f)\boldsymbol{u}^T\boldsymbol{x}}e^{-ig(f)\boldsymbol{u}^T\boldsymbol{x}'}$. By the Schur product theorem, then, we conclude that $\boldsymbol{\Phi}(f)$ is positive semidefinite for all f. Since $\sqrt{S_{\boldsymbol{x}_j}(f)S_{\boldsymbol{x}_k}(f)} \leq \max_j S_{\boldsymbol{x}_j}(f)$ is integrable and both C_f and the complex exponential are bounded, $\boldsymbol{\Phi}(f)$ is integrable. We thus conclude that $\boldsymbol{\Phi}(f)$ is integrable, Hermitian, and positive semidefinite for each f. Since $\boldsymbol{\Phi}(f) = \overline{\boldsymbol{\Phi}(-f)}$ as well, its Fourier transform will be real-valued.

To complete the proof, we pick an arbitrary collection of space–time coordinates $\left\{(t_j, \boldsymbol{x}_j)\right\}_{j=1}^N$ and define the matrix-valued function $\boldsymbol{O}(f)$ of componentwise Fourier transforms in time at the appropriate value, so that $\boldsymbol{O}(f)_{j,k} = e^{i2\pi f(t_j - t_k)}$. Letting $\boldsymbol{\Sigma}$ denote the matrix generated by K, we note that it can be expressed as

$$\boldsymbol{\Sigma} = \int_\Omega \{\boldsymbol{O}(f) \circ \boldsymbol{\Phi}(f)\}\, \mathrm{d}f, \tag{4}$$

which by componentwise inspection we immediately observe to be real-valued and integrable by the previous paragraph. Further, just like $\mathcal{P}(f)$ above, $\boldsymbol{O}(f)$ admits a similar rank-one representation, from which we see that it is Hermitian and positive semidefinite. By another application of the Schur product theorem, then, we observe that Equation (4) is a real-valued mixture of Hermitian positive semidefinite matrices for arbitrary coordinates $\left\{(t_j, \boldsymbol{x}_j)\right\}_{j=1}^N$ and is thus a valid real-valued covariance matrix. ∎

Several comments are in order. First, the proof of Theorem 1 demonstrates that this argument extends to a much broader class of functions than the one that will be studied here. For example, an algebraic representation of $\boldsymbol{S}(f) = \boldsymbol{Q}(f)\boldsymbol{A}(f)\boldsymbol{Q}(f)^T$ where $\boldsymbol{A}$ is positive semidefinite could provide an even richer nonparametric representation for marginal and cross spectra of a spatiotemporal process. Building rank-deficient positive semidefinite matrices is easy, and per the argument of the proof above, such an extension would also yield an equally valid real-valued covariance function. Even beyond that, this model as stated here is itself a minor extension of the much more general and abstract covariance functions presented in Porcu, Mateu, and Christakos (2009) and Porcu, Gregori, and Mateu (2009).

Second, temporal stationarity—or at most a restrictive form of nonstationarity—is required by this model as it is stated above. Namely, the model requires a spectral density (as we have defined it in terms of a single-argument representation for the equivalent time domain representation) along that dimension. This unfortunately precludes several common nonstationary spectral-domain models, such as processes that are locally stationary but whose spectra evolve slowly and continuously with time (as in Poppick et al., 2016), which would substantially complicate the numerics required to build covariance matrices for maximum likelihood. Further, as we will discuss below, performing space–time domain computations for this model with minimal numerical concerns is only made feasible by the use of the fast Fourier transform algorithm (Cooley & Tukey, 1965). This means that the covariance function can only conveniently be evaluated at the inverse Fourier frequencies, and so other relaxations of temporal stationarity like deformation-type methods would also be difficult to apply.

Finally, we note that the above argument still applies if C_f is a bounded covariance function and not a correlation function, and that requiring it to be a correlation function primarily serves to mitigate identifiability problems. Moreover, C_f need not be stationary. As will be demonstrated later in this work, between the space-dependent marginal time spectra $S_{\boldsymbol{x}}$ and a nonstationary coherence function $C_f(\boldsymbol{x}, \boldsymbol{x}')$, it is easily possible to compose this model with other nonstationary models to obtain an even more flexible valid covariance function.

Despite the difficulty discussed above, there are several restrictive ways to relax the temporal stationary of this model. While slightly more complex and involved ideas are possible, in this work we only employ the simplest possible extension in the form of a nonstationary scale.

Corollary 1. *If K is the function given in (2), $\eta \geq 0$, and $\lambda(\boldsymbol{x}, t) \geq 0$ everywhere, then the function*

$$\tilde{K}((\boldsymbol{x}, t), (\boldsymbol{x}', t')) := \lambda(\boldsymbol{x}, t)\lambda(\boldsymbol{x}', t')K((\boldsymbol{x}, t), (\boldsymbol{x}', t')) + \eta_{st}^2 \mathbf{1}_{\{(\boldsymbol{x},t)=(\boldsymbol{x}',t')\}} \tag{5}$$

is also a valid covariance function.

With the further addition of a "temporal nugget" $\eta_t^2 \mathbf{1}_{\{t=t'\}}$, in the same spirit as the "spatial nugget" of Gneiting (2002), this model is precisely the formulation we use here to explore our data.

For a relatively small number of spatial locations, using the FFT to numerically evaluate these integrals for all pairs of spatial locations and then to assemble covariance matrices is straightforward. With little numerical work, then, one can write simple models for marginal—or at least lower dimensional—behavior and obtain all the function values necessary to build a valid space–time covariance matrix in $\mathcal{O}(n_s^2\, n_t \log_2 n_t)$ time complexity, where n_s is the number of unique spatial locations and $n_t = (t_{\text{end}} - t_1)/\Delta t$, for measurement times $\{t_1, t_1 + \Delta t, t_1 + 2\Delta_t, \ \dots \ , t_{\text{end}}\}$. In particular, setting $\Delta t = 1$, the integral approximation for a given function h by

$$\int_{-1/2}^{1/2} e^{2\pi i f t} h(f) df \approx \frac{1}{M} \sum_{j=0}^{M-1} e^{2\pi i f_j t} h(f_j)$$

for Fourier frequencies $\{f_j\}$ can be evaluated simultaneously and efficiently at all necessary time lags t via the inverse FFT algorithm of length M. This is effectively an application of trapezoidal integration, and by also taking sufficiently long FFTs to avoid edge effects, these approximations can easily be made accurate to double precision for kernels that can be checked with a closed form expression. In our experience, taking an FFT length of $M = 7 \cdot n_t$ is more than sufficient. To the degree that any linear algebra for large ill-conditioned matrices is exact in finite precision arithmetic, then, the evaluation of the likelihood with covariance matrices assembled from this approximation is exact. As further evidence of this, the log-likelihood for the data that will be introduced below FFT sizes ranging from $5 \cdot n_t$ to $21 \cdot n_t$ all agree to at least single precision in our experimentation, suggesting that this numerical scheme provides effectively exact function evaluation and thus provides a way to minimize the exact likelihood.

Despite the ostensible complexity of evaluating this function at all necessary locations to assemble a covariance matrix, in our use case with on the order of 30 spatial locations and one thousand time locations, evaluating the function at all necessary space–time locations and aggregating the values into a large dictionary-like data structure takes less than 1 s on a Dell Latitude E5470 with an Intel Core i5-6200U@2.3GHz CPU, a mobile processor not designed for heavy computation, and poses no numerical or computational challenge. Once these values (and the values for the derivatives of (5) with respect to kernel parameters) have been computed, covariance matrices can be assembled and the likelihood and its derivatives can be evaluated as usual. In fact, since updating values of a matrix with this kernel is then done by accessing the dictionary structure as opposed to performing any actual numerical computation, the additional cost of assembling the kernel is often much smaller than the time saved by the faster elementwise assignments to the memory buffer representing the covariance matrix.

3 | DOPPLER LIDAR WIND MEASUREMENT DATA

The US Department of Energy's Atmospheric Radiation Measurement (ARM) program was created with the goal to provide long-term in situ and remotely sensed observations in various climate regimes to improve the understanding of processes impacting atmospheric radiation (Stokes & Schwartz, 1994). The Southern Great Plains (SGP) observatory is the first ARM site and the largest climate research facility in the world, equipped with more than 50 instruments and providing continuous measurements ranging from basic meteorology and radiation to cloud and aerosol properties at several locations in the north-central Oklahoma and south Kansas region (Mather & Voyles, 2013; Sisterson et al., 2016). In this work, we study measurements of the vertical component of the wind field at very high spatiotemporal resolution as observed by the 1.5 μm pulsed Doppler LIDAR (DL) deployed at the SGP central facility (CF).[1]

The principle of operation of the DL is similar to that of the Doppler radars in that they both transmit pulses of energy into the atmosphere and measure the returned signal scattered back by clear-air irregularities (Gage & Balsley, 1978; Muradyan & Coulter, 2020). In other words, scattering is assumed to originate from atmospheric particulates moving at the same speed as the wind. The ARM DL has a full upper hemispheric scanning capability, measuring the one-dimensional velocity projections across a range of angles in order to fully resolve the three-dimensional wind field (Newsom, 2012). The general DL scanning strategy for this purpose at the SGP CF is to perform 8-beam plan-position-indicator scans once every approximately 15 min, and all sampling between those times is done in a vertically staring mode. When the DL is pointing vertically, it provides height- and time-resolved measurements of the vertical velocity.

[1]The ARM processed/quality-controlled DL fixed-beam stare measurements are freely available (Newsom & Krishnamurthy, 2010). The raw data were used in this work, which can be obtained upon request at `https://www.arm.gov` using the "Ask us" page.

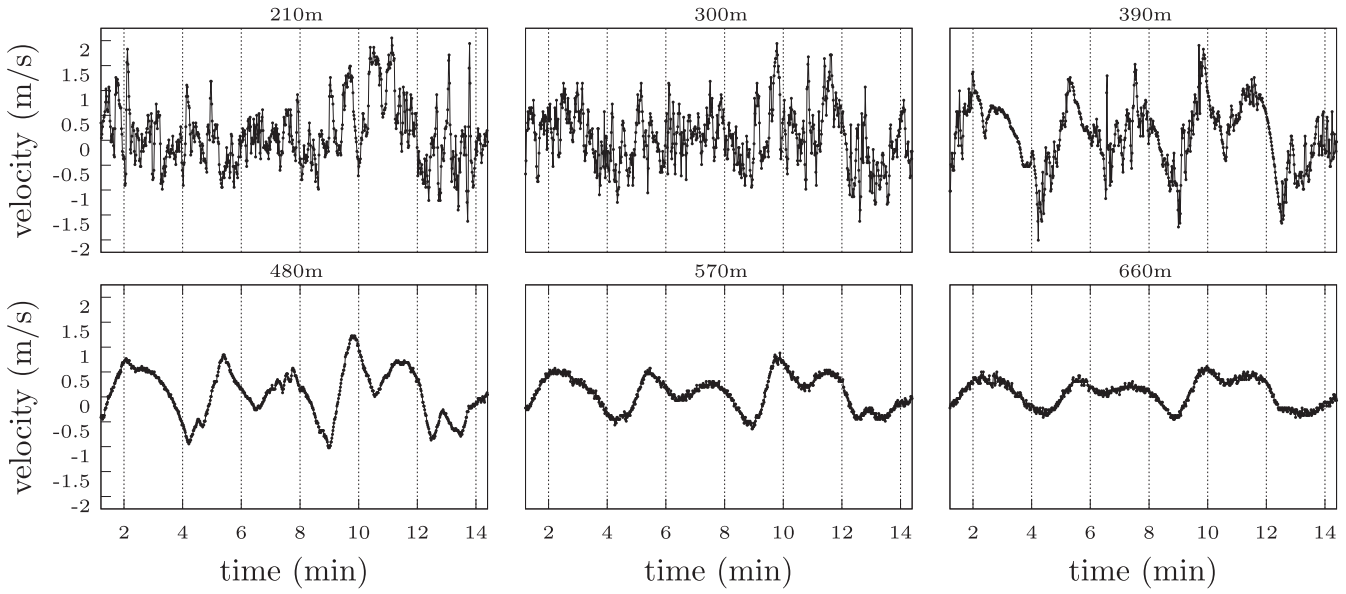

FIGURE 1 Fifteen minutes of Doppler LIDAR vertical wind velocity measurements at several different range gates on June 2, 2015 at 1400UTC, shortly after sunrise

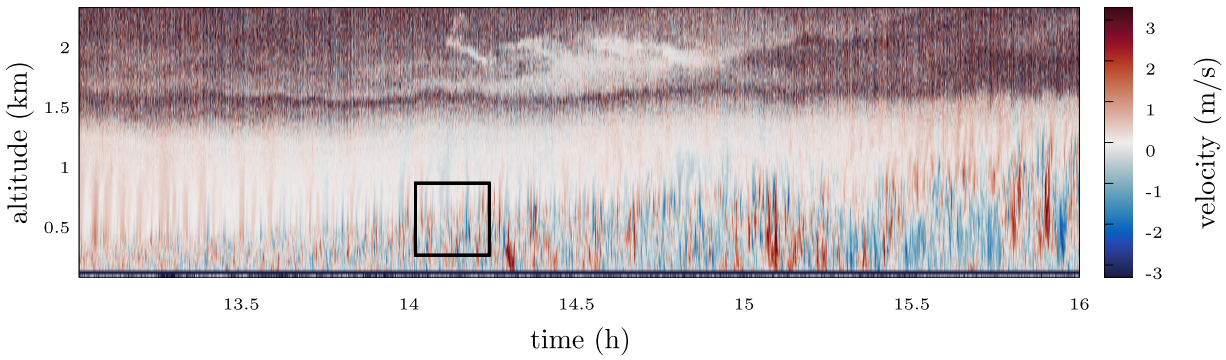

FIGURE 2 An alternative view of the process from 1300 to 1600 UTC on June 24, 2015, a day with particularly high activity, demonstrating the rising mixing height after sunrise and the dynamic domain of data reliability. The color scale for this image is highly truncated, as the unreliable measurements at altitudes with low SNR range sporadically between 20 and −20 m/s. The boxed subdomain shows the size of the spatiotemporal regions that will be individually fitted in later sections

As we focus exclusively on vertical measurements here, the resulting data from this instrument is multiple gappy time series of the vertical wind velocity components at 30-m vertical intervals (commonly referred to as "range gates"), which extend to a maximum of 9.6 km range and are sampled at approximately 1.2 s intervals. For more details about the instrument's theory of operation, functionality, and configuration, see Newsom (2012). Figure 1 shows an example collection of time series plots of the data at discrete range gates over the duration of 15 min on June 2, 2015, while Figure 2 shows the time-height cross-section of similar measurements from about 200 m to 2.25 km above ground level (AGL).

As one would expect, this data exemplify extreme nonstationarity that no sensible model could ignore. In space, the primary source of nonstationarity is the height of the atmospheric boundary layer (ABL), which is the layer directly in contact with the earth's surface through turbulent mixing processes (Tucker et al., 2009). As can be seen in Figure 2, the ABL height increases during the day via solar heating of the surface, resulting in a well-mixed daytime convective boundary layer (BL). In the absence of solar radiation at night, the convection is stopped by heat loss to space, resulting in a thin nocturnal BL with typically 100–300 m depth (Cushman-Roisin, 2019). Above the BL height, the concentrations of particulates such as aerosols and cloud particles, which are the source of scattering, dramatically

decrease, resulting in a substantial reduction in the backscatter signal and thus measurement quality. Typically, only the lowest 2–3 km of the atmosphere yields high-quality measurements (Newsom, 2012). In addition, the minimum range for the DL is approximately 100 m, and even if data is available below that height, the lowest range gates are suspect and are recommended not to be used (personal communication with instrument mentor). As the maximum operational range of the ARM DL is up to 9.6 km, this is a substantial truncation, but inspection of the data on its full domain would involve a great deal of preprocessing measurement artifacts, which is outside the scope of this work.[2]

In order to study the dependence structure across the BL and capture the dynamics above and below it simultaneously without having to contend too deeply with measurement quality concerns or computational issues, we focus this work on the time 1400 UTC—shortly after the local sunrise in northern Oklahoma—when the ABL height has grown to approximately 400–500 m on average. This part of the day is convenient as the ABL height is well above the minimum altitude of DL operation, so that plenty of high-quality measurements are available below it as well as above it. For the purpose of manageable data sizes for exact linear algebra, we focus on altitudes between about 200 and 850 m in order to have large spatial samples of range gates both above and below the BL while also being able to work with full measurement segments between 3D scans. For reference, considering range gates 7–28 and a full time segment of vertical measurements (about 14 min) amounts to approximately 17 thousand data points.

This data source has several features and properties that make it interesting to study. For one, the spatial component is vertical and not horizontal, and as such, the process exhibits much sharper changes in dynamics along its spatial index. Just as time should not be treated as just another spatial dimension, we believe that the vertical dimension of many physical processes should be treated with the same care when the relevant physical dynamics are so different, as they are here. The spatial nonstationarity that this data exhibit is very sharp and likely not even once differentiable, which is unusual to observe on such a fine spatial scale and with such high correlation across the spatial transition. The measurements at 300 and 480 m in Figure 1, for example, clearly have very strong coherence at low frequencies despite how different finer-scale behavior is at each altitude. Any locally stationary model would necessarily be throwing out a great deal of information of that kind, which is arguably of the highest scientific interest. In some cases, the creation of intricate nonstationary models has been criticized for not actually capturing more behavior, as it is frequently the case that most information that can be learned about a spatial or spatiotemporal process corresponds to high frequency behavior of the spectral density (or is otherwise "microergodic"), which thus largely corresponds to local information (Stein, 1999). This data are an example of a situation when that type of pessimism is unnecessary.

Finally, the measurements used in this work are made at a much higher frequency than most other sources of environmental data, allowing the resolution of many fine-scale properties that even hourly data would necessarily miss. At the time scales of these measurements, a great deal can be inferred about smoothness properties of the field, as well as the study of wind gusts, turbulence, and other small-scale meteorological phenomena, which will especially be critical in the lower portion of the ABL. In general, the features that require high-frequency resolution to observe also make the data more challenging to study. Things like measurement quantization when the wind velocity is very low at night, for example, are real concerns, as well as distinguishing real short time-scale velocity fluctuations from instrument noise. As will be clear, this work is only scratching the surface of what can be studied with this type of high-frequency and high-resolution monitoring data, and it is precisely its associated challenges that make it interesting.

4 | A HALF-SPECTRAL MODEL FOR DOPPLER LIDAR DATA

One of the strongest features of the data, as can be seen in Figure 2, is the strong low-frequency coherence across the entire domain, but especially above the ABL. Another is that the scale of the measurements appears to decay as the altitude increases above the ABL height. A major goal of this modeling effort is to capture that behavior, and in order to do that with acceptable parameter economy and interpretability, we introduce first a common functional form that we use extensively in several parts of this model. Taking inspiration from the frequency response function of the Butterworth filter from signal processing (Butterworth, 1930), whose signature feature is being flat in some neighborhood of zero and

[2]https://plot.dmf.arm.gov/PLOTS/sgp/sgpdlfpt/20150602/sgpdlfptC1.b1.fpts_12-15hour.20150602.png, for example, gives an example of ARM's automatic visualization of DL measurements from 1200 to 1500 UTC on June 2, 2015, demonstrating the vertical extent of measurements with good SNR.

then decaying algebraically, we define the utility function $B(z)$ as

$$B(z; \xi_0, \xi_1, \xi_2) := e^{\xi_0}\left[1 + \left(\frac{z}{\xi_1}\right)^{2\xi_2}\right]^{-1}.$$

As we demonstrate below, this function is convenient in three different parts of the specification of the model given by Equation (5).

Let us next specify a form for the nonstationary scale parameters, which we choose to represent in a flexible form as

$$\lambda(x, t) = B((x - \beta)_+; 0, \phi_1, \phi_2)\sum_{j=1}^{4} w_j(t)\theta_{t_j}, \tag{6}$$

where $\theta_{t_j} \geq 0$, $\{t_j\}$ are knots in time chosen evenly across the domain, and $w_j(t)$ are time-dependent weights that are normalized to sum to one, with unnormalized forms given by $\tilde{w}_j(t) := e^{-|t-t_j|/50\Delta t}$ (for reference, the time window for the data that will ultimately be fitted here is about $775\Delta t$). Here and for the rest of the work, we use the notation x for the altitude in place of $\boldsymbol{x}$, our earlier signifier for an arbitrary location in $\mathbb{R}^d$. The spatial dependence of the scale comes in the form of the Butterworth-type function B, which starts to decay once the altitude x, in units of meters, is above the ABL height, which we denote with β and is itself a model parameter. Due to the structure we see across several days of data, we prefer this representation for the spatial dependence on scale to other similarly economical approximations.

To introduce spatial nonstationarity of components in the marginal spectrum and coherence while maintaining interpretability, we implement a very simple space-dependent parameter mapping given by, for example, with a parameter denoted ψ,

$$\psi(x) := (1 - w(x))\psi_0 + w(x)\psi_1, \tag{7}$$

where

$$w(x) := (1 + e^{-\tau(x-\beta)})^{-1}, \tag{8}$$

and the subscripts 0 and 1 denote in this case the regimes of being below and above the ABL respectively, so that effectively the spatially indexed parameters are logistic interpolants between two values. The inverse scale parameter τ of this logistic function is informative about the sharpness of the transition from below-ABL and above-ABL dynamics, and is very important to this model. For one reason, fixing a constant ABL height β for 15 min is obviously not correct from a brief inspection of Figure 2. Since that is a form of temporal nonstationarity that this model cannot easily account for, we introduce this τ parameter to at least allow the model the ability to communicate about the sharpness of this transition and capture better the dynamics at range gates near the mean height of the ABL.

Equipped with this parametric form of spatial dependence, we define the marginal spectral densities with

$$S_x(f) = (1 + B(\sin \pi f; \xi_0(x), \xi_1, \xi_2))(e^{\rho(x)}\sin^2 \pi f + 1)^{-\nu(x)-1/2}, \tag{9}$$

where the spatial dependence of the low frequency multiplier $1 + B$ comes in the form of the parameter $\xi_0(x)$, and ξ_1 and ξ_2 are fixed across space. Here and in the coherence function given below, we use $\sin \pi f$ instead of f so that the periodic extension of the model is smooth with respect to f at the endpoints, which we consider to be a desirable form of model consistency for discretely observed continuous-time processes. Finally, as a small technical note, for time series measured at nonunit time steps, a change of variables in the Fourier transform suggests that a Δt rescaling should be applied to S_x. We do not do that here and instead elect to simply treat the time sampling as unit sampling and scale nonparametric spectral density estimators consistently with that in the following section, implicitly absorbing this $\Delta t \approx 1.2$ s into the θ_j parameters.

This spectral density corresponds closely to a filtered Matérn spectral density for continuously indexed processes, which is best characterized by its algebraic decay, so that $S(\boldsymbol{f}) \sim \|\boldsymbol{f}\|_2^{-2\nu-d}$ in d dimensions. In discrete time, ν is not so directly interpretable since even negative values yield spectral densities corresponding to nonpathological processes. But the algebraic tail decay is nonetheless convenient and at least heuristically informative about smoothness that would

TABLE 1 A summary of all the parameters used in the covariance model as parameterized in (5)

β	Boundary layer height
τ	Shape parameter for boundary layer regime transition rate
θ_{t_j}	Local scale parameters
ϕ_j	Scale decay with altitude over boundary layer
$\xi_j(x)$	Space-dependent shape parameters for temporal low-frequency multiplier $1 + B(f)$
$\rho(x)$	Space-dependent temporal range-type parameter
$\nu(x)$	Space-dependent temporal smoothness-type parameter
$\zeta_j(x)$	Space-dependent shape parameters for coherence function $\gamma(x)$
α	Phase function rate-type parameter
η_{st}	Spatiotemporal nugget size
η_t	Purely temporal nugget size

be implied by a continuous version of the model, and using $\sin \pi f$ instead of f extends this functional form nicely to a function that is smooth on the unit torus. The extra multiplier of $1 + B$ provides the model the ability to add extra low frequency energy without affecting the higher frequencies, which is a pronounced feature of the data and can be seen in the diagnostic figures of the next section.

The coherence function C_f is crucial to this model, and is extremely different above and below the ABL. To be as flexible as possible, we choose C_f to be a nonstationary Paciorek–Schervish-type correlation function with an extension to variable smoothness (Stein, 2005a) given by

$$C_f(x, x') := \frac{\sqrt[4]{\gamma(x)\gamma(x')}}{\gamma(x, x')} \mathcal{M}_{(\nu_s(x)+\nu_s(x'))/2}\left(\frac{|x - x'|}{\gamma(x, x')}\right), \tag{10}$$

where $\gamma(x, x') := \sqrt{(\gamma(x) + \gamma(x))/2}$, and we parameterize

$$\gamma(x)(f) := B(\sin \pi f; \zeta_0(x), \zeta_1(x), \zeta_2(x)),$$

using ζ instead of ξ to distinguish between the parameters of $B(f)$ for S_x and C_f. Unfortunately the spatially indexed smoothnesses for C_f are not smoothly interpolated like other spatially indexed parameters in order to avoid taking derivatives of second-kind Bessel functions (refer to the Discussion section for more information) in the derivatives of C_f, and are instead discontinuously parameterized as

$$\nu_s(x) = \mathbf{1}\{x \le \beta\}\frac{1}{2} + (1 - \mathbf{1}\{x \le \beta\})\frac{3}{4}.$$

We acknowledge that it is somewhat noteworthy to have fixed smoothnesses for this correlation function. Considering that there are only approximately 20 spatial locations being considered in this model, however, our experimentation indicates that the effect of varying these parameters on the likelihood is modest.

Since the true phase dynamics of this process are almost certainly driven by nonlinearity that no Gaussian process model could capture, we include only the simplest form of $g(f)$ in this model, setting $g(f) := \alpha \sin \pi f$, with no spatial variation. With the inclusion of our space–time and temporal nuggets, we now have a complete covariance function of the form (5). Table 1 provides a final summary and reference for the interpretation of each parameter in the model.

With this model fully introduced, it merits comment that since it will be fitted with maximum likelihood estimators, parameters informing the high frequency behavior will be highly prioritized over parameters that inform low frequency behavior. Because of the human tendency to focus on lower frequency behavior when performing visual inspection of simulations or other diagnostics, however, we have included several parameters and degrees of freedom here that do not heavily affect the likelihood. The $\{\xi(x)\}$ parameters, for example, are a good example of extra degrees of freedom that do not strongly affect the likelihood and are completely confined to the very lowest of frequencies. Their effect on visual diagnostics for conditional simulations, however, are very noticeable. For certain purposes, a smaller model could be used

with similarly satisfying results, and the supplemental information compares visual diagnostics of the full model shown here and a reduced one. Since this work is an investigation into how well a Gaussian process model can capture the very complex dynamics of vertical wind velocity, however, we proceed with a maximalist approach here.

5 | ESTIMATION RESULTS

In this section, we discuss the results of applying maximum likelihood estimation for the above model to six different days of data in June 2015 with comparable meteorological conditions. Log-likelihoods for Gaussian process models are almost always nonconvex, even in the simplest case of standard parameterizations of stationary Matérn covariances. As such, although it is not often discussed in the statistical literature, the optimization of these functions with respect to parameters is a serious challenge even in the most simple cases. Considering the size and complexity of the model described here, it comes as no surprise that the optimization of this model is particularly difficult, and so we provide a detailed discussion of the optimization in the Appendix for interested readers. Briefly, the parameters here were estimated using a trust region method that employs high-quality randomized estimators for the gradient of the likelihood and the expected Fisher information matrix (see Geoga et al., 2019), which one might consider to be a trust region analog to Fisher scoring (Osborne, 1992). Another advantage of this model that is useful at this juncture is that Whittle-type estimators for marginal quantities provide convenient and high-quality initializers. Using an intermediate estimation process like that means that one can start at a naive initialization and reliably obtain 25-dimensional estimators like those reported here.

As described in Section 3, we restrict the estimation here to approximately 14-min time segments (about 780 measurements) and range gates between 210 m (the seventh gate) and either 690 m (the 23rd gate) or 840 m (the 28th gate) based on the height of the ABL so that both above- and below-mixing height dynamics have reasonable representation. These time segments correspond to measurements between the temporal measurement gaps noted previously, and the resulting data sizes are near the limit of the computational abilities of the hardware used for the estimation. We do stress, however, that the methods described here could equally well be applied to gappy regular monitoring data by virtue of the fact that performing the computations in the time domain means that an FFT is being applied to the analytical spectrum, not the data itself.

Table 2 summarizes the point estimates for each day and provides the implied standard deviation from the expected Fisher matrix. Using this matrix as a proxy for the precision matrix of the MLE is a nontrivial approximation, considering that the necessary conditions for that to even hold asymptotically—for example, the log-likelihood looking quadratic near the MLE—are likely not met, or at least are not met for this data size. Nonetheless, as the expected Fisher matrix is the inverse of the Godambe information matrix for the score equations, we believe it to be nonetheless at least reasonably informative about the uncertainty of each point estimate (Heyde, 2008; Stein et al., 2013).

One main takeaway from these tables is that the parameters seem to be estimated reasonably precisely and have believable values, although of interest is the fact that for most days, there is at least one parameter that is not well resolved, and it is not always the same parameter. In some sense we find this encouraging, as it is likely an unreasonable expectation that for a priori domain choices every parameter is resolvable across any nontrivial collection of days due to the domain variability discussed earlier. Robustness to this issue seems to be a more important and realistic priority than full resolution at every day and in all conditions. Further, the inconsistencies between which parameters are not well resolved serve as at least an indirect indication that a model much simpler than the one used here would likely sacrifice some degrees of freedom that maximum likelihood does consider to be valuable in at least some circumstances.

Figure 3 shows the correlation (not covariance) matrix for the MLE on June 2. As one might expect, the parameters determining the effectively nonparametric scale function are highly correlated and are inversely correlated with the smoothness, as is the range-type parameter ρ. These types of correlation structure are quite common in stationary Matérn-type models. Moreover, there is strong negative correlation between the latter two parameters used in the parameterization of B, which is the same phenomenon as above. Of note, though, is that fixing one of those two parameters (or profiling one in terms of the other) resulted in substantially lower likelihoods and much more challenging optimization. Similarly, correlation of the $\{\zeta_j(x)\}$ parameters above and below the ABL is reasonably strong, likely due to the way that they are combined for range gates near the mean ABL height, although that is obviously another case where reducing the model would be wrong. Otherwise, correlations are not concerningly high in our assessment, and this further suggests that the model is formulated well enough to avoid identifiability problems despite the high parameter count.

Figures 4 and 5 show the model-implied marginal spectra and coherences for several specific distance respectively, where coherences are shown in terms of their real and imaginary parts. The marginal spectra in general seem well

TABLE 2 Results from maximum likelihood estimation of the 6 days that were studied

	02	03	06	20	24	28
θ_{t_1}	1.06 (0.03)	0.72 (0.09)	1.03 (0.03)	1.13 (0.02)	1.65 (0.03)	1.54 (0.04)
θ_{t_2}	1.18 (0.03)	0.62 (0.09)	1.23 (0.03)	0.92 (0.03)	1.78 (0.03)	1.23 (0.04)
θ_{t_3}	1.27 (0.03)	0.64 (0.09)	1.63 (0.03)	1.16 (0.02)	1.86 (0.03)	1.25 (0.04)
θ_{t_4}	1.19 (0.03)	0.82 (0.09)	1.98 (0.03)	1.26 (0.02)	2.26 (0.03)	0.87 (0.04)
ρ_0	2.72 (0.14)	0.35 (0.50)	3.48 (0.14)	2.37 (0.11)	2.76 (0.13)	2.86 (0.16)
ν_0	1.14 (0.06)	7.07 (2.53)	1.02 (0.04)	1.43 (0.07)	1.21 (0.05)	1.35 (0.07)
ρ_1	5.53 (0.18)	2.31 (0.55)	6.76 (0.12)	8.93 (0.15)	4.89 (0.18)	6.25 (0.19)
ν_1	2.14 (0.18)	20.05 (9.77)	3.14 (0.19)	0.37 (0.03)	2.11 (0.17)	2.35 (0.25)
ζ_{00}	12.46 (0.49)	13.64 (0.18)	0.50 (2.7e4)	17.47 (0.89)	11.89 (0.29)	11.48 (0.44)
ζ_{01}	0.02 (5.6e−3)	0.01 (2.1e−3)	0.03 (5.2e−3)	2.1e−4 (1.2e−4)	0.07 (0.01)	0.02 (7.7e−3)
ζ_{02}	0.84 (0.05)	0.77 (0.04)	1.35 (0.06)	0.68 (0.03)	1.09 (0.06)	0.73 (0.05)
ζ_{10}	16.79 (0.50)	13.57 (0.21)	18.13 (0.33)	24.81 (0.84)	14.95 (0.31)	17.25 (0.49)
ζ_{11}	0.01 (2.5e−3)	10.66 (3.52)	5.3e−4 (1.9e−4)	1.5e−5 (1.0e−5)	0.04 (4.5e−3)	4.4e−3 (1.1e−3)
ζ_{12}	1.38 (0.09)	69.03 (23.14)	0.67 (0.06)	1.02 (0.05)	1.86 (0.12)	0.97 (0.10)
β	4.6e2 (2.15)	6.6e2 (4.10)	4.7e2 (3.18)	6.8e2 (1.23)	5.8e2 (3.49)	5.5e2 (3.94)
τ	0.06 (2.1e−3)	0.04 (2.4e−3)	0.02 (7.8e−4)	0.03 (4.5e−4)	0.03 (1.3e−3)	0.04 (1.3e−3)
ξ_{00}	3.58 (0.86)	18.98 (8.80)	1.02 (0.39)	14.94 (5.07)	0.47 (0.22)	37.07 (22.65)
ξ_{01}	7.08 (1.92)	9.21 (4.72)	0.02 (0.23)	38.48 (14.24)	1.57 (0.61)	38.31 (24.50)
ξ_2	6.55 (2.93)	0.85 (0.13)	12.98 (36.07)	1.09 (0.18)	5.99 (4.90)	0.83 (0.17)
ξ_1	0.03 (1.9e−3)	0.04 (8.4e−3)	0.02 (2.3e−3)	6.8e−3 (2.0e−3)	0.06 (5.6e−3)	2.2e−3 (1.4e−3)
ϕ_1	1.1e2 (17.67)	1.6e2 (17.63)	2.3e2 (48.73)	4.10 (0.95)	1.1e2 (28.44)	3.6e2 (45.75)
ϕ_2	0.64 (0.07)	1.20 (0.27)	1.03 (0.33)	0.45 (0.04)	0.51 (0.06)	1.14 (0.39)
α	2.1e−4 (3.1e−4)	1.2e−4 (1.3e−4)	1.8e−3 (3.4e−4)	6.9e−4 (2.4e−4)	4.2e−4 (1.9e−4)	1.1e−5 (2.7e−4)
η_{st}	0.04 (3.7e−4)	0.11 (1.0e−3)	0.02 (2.2e−4)	0.07 (7.3e−4)	0.06 (6.3e−4)	0.07 (5.5e−4)
η_t	0.01 (6.9e−4)	0.05 (2.0e−3)	7.5e−3 (4.3e−4)	0.03 (1.4e−3)	0.02 (1.1e−3)	0.03 (1.1e−3)

Note: Standard deviations obtained via the stochastic expected Fisher matrix are provided in parentheses.

captured by the model (shown here with the nugget terms added and using θ_{t_1} for the scale parameter), and the two-regime nonstationarity seems justifiable after inspecting the spectra of individual range gates that are shown faintly in grey. Maximum likelihood is well known to prioritize capturing high-frequency behavior as accurately as possible over low-frequency behavior, and this observation in an earlier stage of the model formulation was a primary motivator for the extra low-frequency term $(1 + B(f))$ in (9), which costs a few extra parameters but substantially improves the model's ability to capture the behavior at the lowest frequencies.

For measurements at this frequency and a reasonably small time segment, coherences are unfortunately difficult to estimate without either very strong and potentially confounding bias or intractable variability. We attempt to strike a balance between the two in Figure 5, although the interpretation is still difficult. Even bearing this difficulty in mind, however, the results suggest that the model is not perfectly specified, in particular above the ABL and in the form of the spatial dependence. Noting that these plots again show frequency on the log scale, we see that the coherence estimates are not discernibly different from noise at even relatively low frequencies, and so assessing the quality of the estimated coherence is challenging. More heuristically, however, the results do seem plausible, and important features such as near-unit coherence at low frequencies above the ABL are captured. While it is tempting to consider a more complex parametric form in the hope of representing these coherences better, inspection of similar figures for other days in the supplementary material suggest that much of what one might interpret as structure in these estimates does not consistently appear

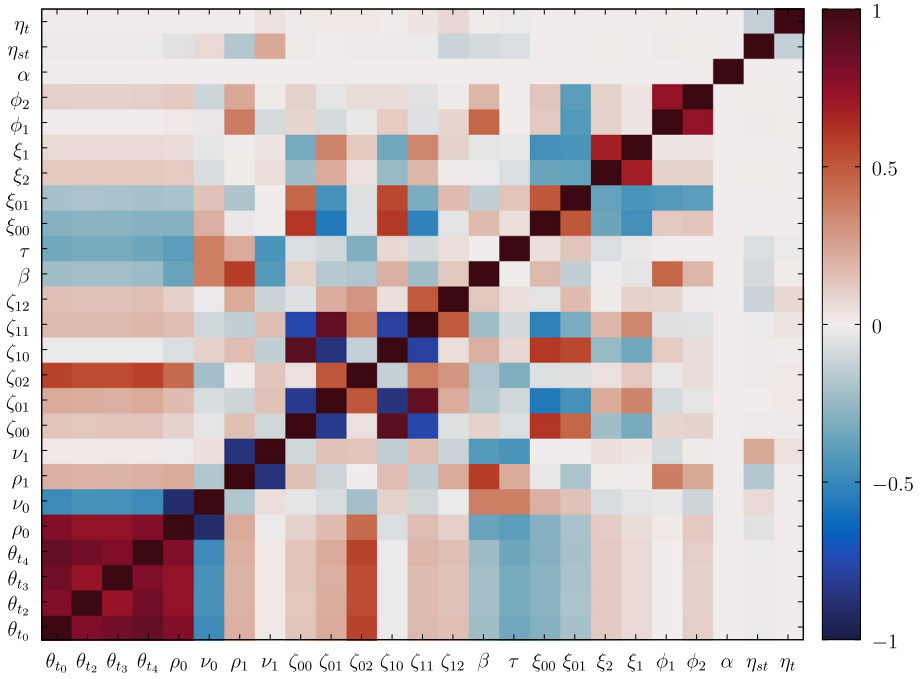

FIGURE 3 The correlation matrix of the MLE of the model for the data on June 2

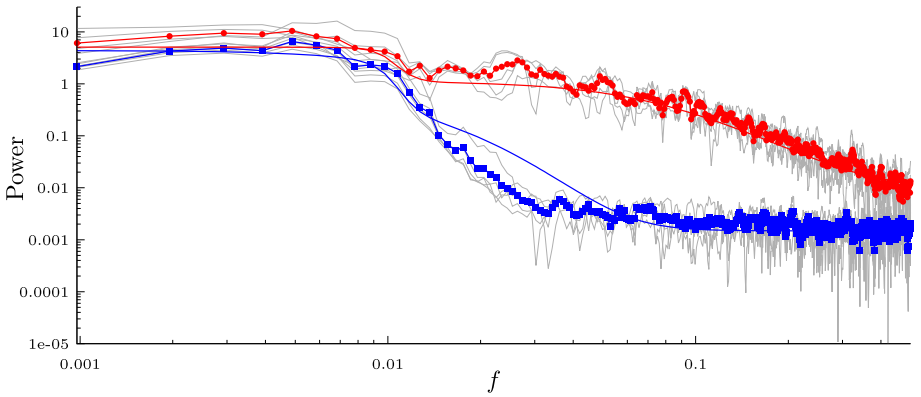

FIGURE 4 Marginal spectra for gates above the ABL (blue/squares) and below the ABL (red/circles) on June 2. Analytical estimates for these spectra (including nugget parameters) are overlaid as lines of corresponding colors. Estimators from individual range gates are shown faintly in gray

across days, and a more complex form for the coherence might result in overfitting that would make multiday extensions more challenging, rather than less so.

For a more quantitative investigation of what we gain by using such a complex model, we now compare with the Paciorek–Schervish model using as closely analogous parameterizations as possible, given by

$$K_{\mathrm{PS}}((x,t),(x',t')) = \lambda(x,t)\lambda(x',t')c((x,t),(x',t'))\mathcal{M}_{1/2}\left(\sqrt{Q((x,t),(x',t'))}\right) + \eta_{st}^2\mathbf{1}_{\{(x,t)=(x',t')\}} + \eta_t^2\mathbf{1}_{\{t=t'\}}.$$

The scale λ is exactly the same as (6), and the local anisotropy is modeled here with $\mathbf{\Gamma}_\theta(x,t) = \mathrm{diag}(e^{\gamma(x)}, e^{\rho(x)})$, where the spatially indexed parameters $\gamma(x)$ and $\rho(x)$ are also specified as in (7) with parameters β and τ, and

$$\begin{aligned}
\mathbf{\Gamma}((x,t),(x',t')) &= \frac{1}{2}\left(\mathbf{\Gamma}_\theta(x,t) + \mathbf{\Gamma}_\theta(x',t')\right), \\
c((x,t),(x',t')) &= \frac{|\mathbf{\Gamma}_\theta(x,t)|^{1/4}|\mathbf{\Gamma}_\theta(x',t')|^{1/4}}{|\mathbf{\Gamma}_\theta((x,t),(x',t'))|^{1/2}}, \quad \text{and} \\
Q((x,t),(x',t')) &= (x-x',t-t')^T\mathbf{\Gamma}_\theta((x,t),(x',t'))^{-1}(x-x',t-t')
\end{aligned}$$

are the mixed local anisotropy, normalizing correlation constant, and the quadratic form argument respectively (Paciorek & Schervish, 2006). With a total of 14 parameters, the maximized likelihood of this model was approximately 860 units

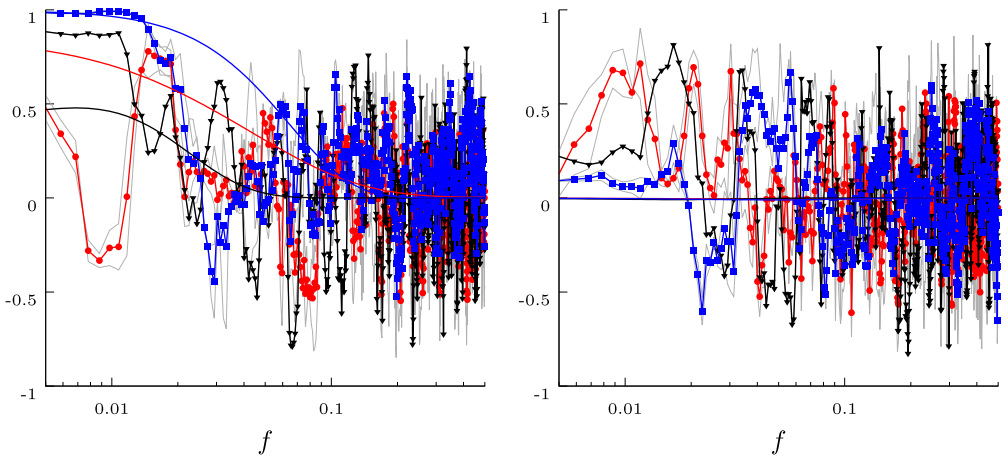

FIGURE 5 Average real (left) and imaginary (right) parts of the complex coherence between measurements three range gates apart above the ABL (blue/squares), below it (red/circles), and twelve gates apart across it (black/triangles), with individual estimators in faint grey. Analytical estimates of each part from the MLE are overlaid in lines of the corresponding color

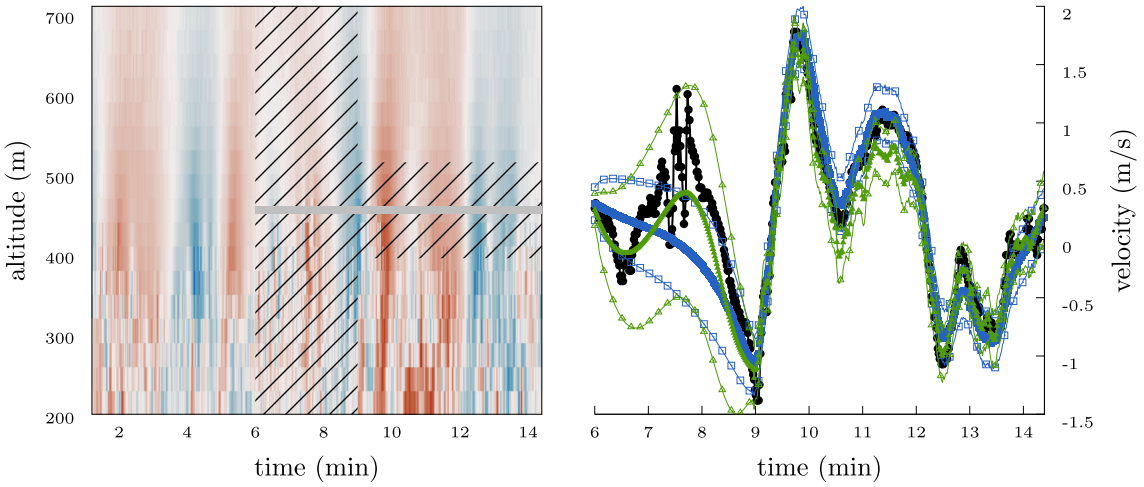

FIGURE 6 (Left) The subsetted data of June 2 for fitting, removed data (black stripes), and interpolation range gate (gray), and (right) Spatiotemporal interpolants with pointwise 95% confidence intervals for the two models discussed above—spatiotemporal Paciorek–Schervish (blue/squares) and half-spectral (green/triangles)—shown together with the actual measurements (black/circles)

lower than that of the half-spectral model. This likelihood difference is substantial, although we do remark that the performance of this model is nonetheless impressive considering how general it is, and that while it is not as easily hyperspecialized as the half-spectral model, it is direct to incorporate at least a few process-specific components.

To further justify the complexity of the model described here, Figure 6 shows a comparison of spatiotemporal interpolants from all three of the above models for the data on June 2. The setting for this study was to remove a section of the domain (shown with black stripes in Figure 6), fit the remaining data, and interpolate at the gate shown by the gray solid line. We further provide 95% confidence intervals for all point estimates using the models' covariance matrices for their respective interpolants. As one might predict from a visual inspection, the prediction RMSE of the half-spectral model is appreciably lower (5.865 vs. 4.245), and so again the half-spectral model is a clear winner. This is particularly noticeable in the portion of the domain where all spatial measurements are removed, highlighting the fact that a representation of nonstationarity via local geometric anisotropies is arguably a less natural fit to this data than the more heavily marginal approach used here.

Finally, Figure 7 provides side-by-side cross-section plots (reminiscent of Figure 2) of the measurement data and a simulation from the fitted model for each of the 6 days. While Table 2 is valuable in assessing the quality of the estimates, for a model of this complexity we also believe that assessing the plausibility of simulations is important. In general, we find the simulations to be satisfying in that regard, although several model misspecifications are uniquely visible in this format. Interestingly, having a fixed estimator for the ABL height β and a smearing-type parameter τ does seem to effectively create simulations where β does not look fixed so long as τ is relatively small. It does, however, highlight that

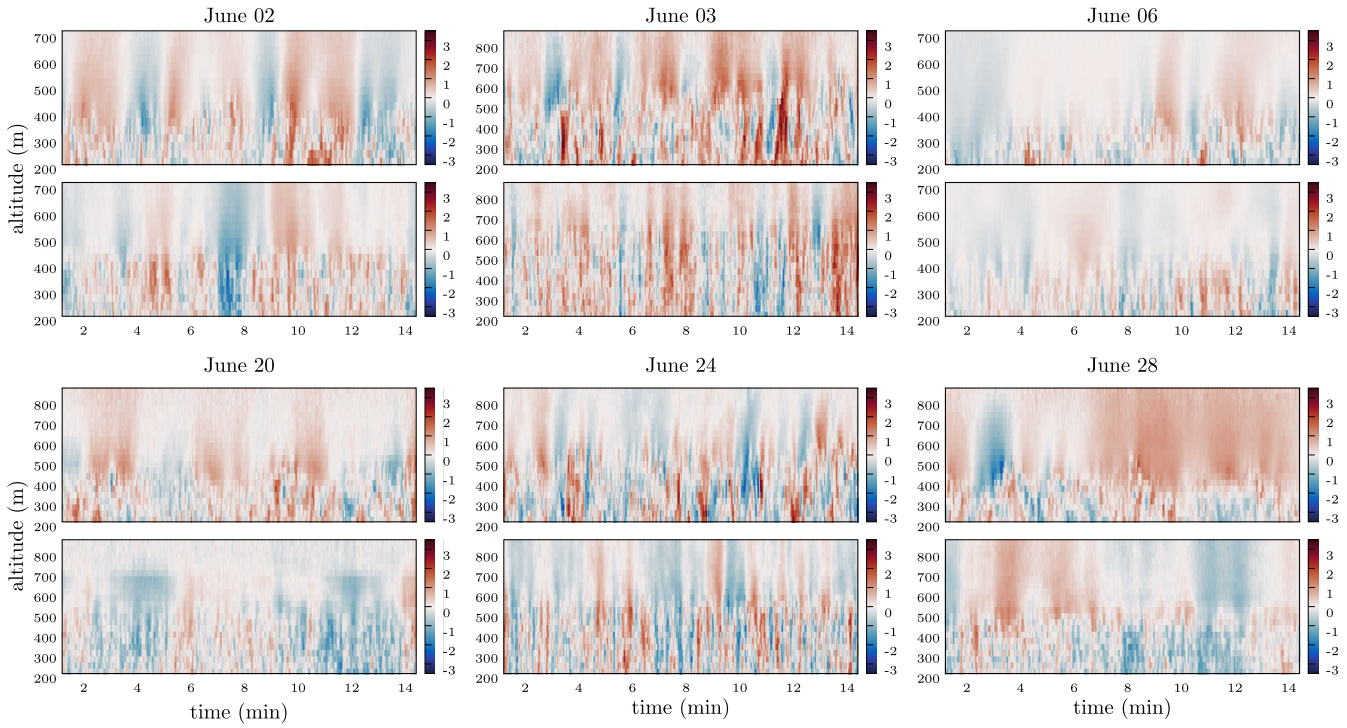

FIGURE 7 Side-by-side comparisons of the data used for estimation (above) and a simulation from the GP model with estimated parameters (below), all in units of m/s, for each fitted day

a fixed β means that the estimate will always be near the ABL's highest point in the data. This makes sense considering that maximum likelihood will certainly penalize underestimation of uncertainty more harshly than overestimation, and the same idea applied to covariance suggests that maximum likelihood will tend to underestimate it sooner than overestimate it. The extra variability below the ABL means that the estimate for β on June 28, for example, is off by as much as 100 m from what one might obtain via more sophisticated and targeted methods, for example, those of Tucker et al. (2009).

In summary, this model does reasonably well in capturing both quantitative and qualitative features of the process's covariance structure in the marginal spectra. Moreover, visual assessment of the simulations suggests that at least qualitative features of the coherence are reasonably well captured. Considering that the quality of the fits and simulations seems reasonably high across all days, we find this to be generally encouraging. See the supplemental material for analogs of all diagnostic figures shown here for all six of the days studied.

6 | DISCUSSION

We have presented here a very flexible modeling framework for regularly measured but nonstationary data. This method provides a convenient and powerful way to think almost exclusively about single-dimension marginal properties of random fields (like spectra and coherences) and obtain fully nonseparable, asymmetric, and nonstationary covariance functions. While the parameter economy of this method leaves something to be desired, few other covariance functions are sufficiently flexible that one can specify a 25-parameter model that composes two different and convenient representations of nonstationarity and for which true maximum likelihood estimation can be performed. Moreover, despite the complexity of this model, very good initializations can be obtained using Whittle-type estimators for marginal quantities, providing a substantial practical gain.

On a slightly less significant note, we point out that this method is also uniquely capable of providing estimates for the smoothness ν and its uncertainty. Derivatives of the modified second-kind Bessel function $\mathcal{K}_\nu$ with respect to the smoothness ν do exist, and there are series expansions that theoretically facilitate computing them (Olver et al., 2010), but we are not aware of any code to do that, and in our own experimentation it is a daunting numerical challenge. In the Fourier domain, however, the derivatives are straightforward. Because of this, the derivative of covariance matrices with

respect to ν are obtained completely analytically, and so there is no need to be concerned with derivative approximations like finite differencing for the kernel function or covariance matrix with respect to ν.

With that said, however, this method of course has drawbacks and is somewhat limited in the way that it can incorporate temporal nonstationarity (or, more generally, nonstationarity in whatever marginal dimension is being transformed over). Further, while some forms of irregularity in time can be mitigated with proper padding and FFT-based interpolation, complete data irregularity—like randomly sampled points—would make this covariance function much more onerous to evaluate. In that sense, we see this method as being complementary to other methods that are more robust in that respect, such as deformation methods (Anderes & Stein, 2008) or convolution-based methods (Higdon, 2002; Paciorek & Schervish, 2006).

The data introduced here—and high-frequency vertical measurements of meteorological processes in general, of which Doppler LIDAR measurements are but one of many—provide a unique opportunity to study space–time random fields. Unlike the well-studied Irish wind data (de Luna & Genton, 2005; Gneiting, 2002; Haslett & Raftery, 1989; Horrell & Stein, 2017; Stein, 2005a), this data source provides signed measurements at such a high rate in both space and time that it is arguably a different regime in both the relevant atmospheric physics and the relevant statistical concerns. The nonstationarity in altitude—a rarely considered spatial dimension for several environmental processes that in many large-scale physical processes exhibits very unique behavior—is sufficiently sharp that there is strong dependence at two altitudes with obviously different marginal behaviors, and thus provides an exceptional testing ground for nonstationary models where they are the most necessary.

With regard to the estimation procedure, it is worth considering that the analytical estimates of the marginal quantities via the MLE are quite different than those obtained via approximate likelihoods in the spectral domain. In many spectral domain likelihood approximations, one is effectively performing estimation via nonlinear least squares, which is a much more direct attempt at curve fitting. Maximum likelihood, on the other hand, will absolutely sacrifice the very low frequency representation of marginal spectra, for example, if changing such a parameter would increase the likelihood at higher spatiotemporal frequencies. As such, while one should of course expect the estimated quantities S_x and C_f to be reasonably reflective of marginal spectra and coherences, the estimation procedure itself is not attempting to compartmentalize parameters for their respective marginal components or provide the best visual fit of a parametric curve to a nonparametric estimator. In this sense, we find the visual agreement of Figures 4 and 5 to be satisfying, even if a simpler curve-fitting estimation method could likely provide a more impressive visualization.

A major objective of this work was to attempt to write a parametric model that can do a reasonably good job of representing the covariance structure on different days so that eventually those models may be stitched together to form one that is cyclostationary in time on the timescale of weeks. A myriad of other concerns become relevant at that timescale; rain, clouds, and even migratory birds and insect swarms (Muradyan & Coulter, 2020) can introduce confounding structure that this model is not adequately equipped to capture. Nonetheless, however, we are encouraged by what has been achieved here, and are hopeful that, going forward, modifications for cyclostationarity will be made possible. One notable concern from the estimation even across just June 2015, however, is that the day-to-day variation seems high enough that a purely periodic model for many of the parameters (e.g., ν_1) would likely miss a lot of subtle features of the data that are arguably the most interesting, and perhaps most informative about the micrometeorology of the area. A Bayesian perspective may be more appropriate.

More generally, a primary objective for scientists when studying this data is to provide an estimator for the ABL height, which we have denoted as β. Many works in the meteorological literature provide time series-based methods for estimating this quantity—for example, Tucker et al. (2009)—but these methods often do not fully exploit spatiotemporal structure, work on a coarser time scale to provide estimates, and assimilate estimators from multiple data sources. While those estimators are obviously better than the estimate for β that is obtained via fitting the model described here, we believe that this work demonstrates that it should be possible to bridge the gap between what the meteorological community does to estimate this quantity and what the spatiotemporal statistics community might do. A Bayesian approach that puts a Matérn-plus-drift prior on the time-varying process $\beta(t)$, for example, might reasonably provide meaningful posterior estimates for the ABL height on the time scale of *seconds*, as opposed to the time scale of 15 min in Tucker et al. (2009). Moreover, by virtue of being a fully spatiotemporal model, such a procedure could exploit much more structure in the data while also mitigating some potentially nonphysical artifacts that occur from estimators that primarily exploit only time dependence, such as large discontinuities. We believe that these and other similar questions are exigent and will likely be fruitful topics of future work.

ACKNOWLEDGMENTS

The authors would like to thank Paytsar Muradyan for her invaluable support in understanding and working with the Doppler LIDAR data and for her guidance in the writing of Section 3. They are also grateful to Lydia Zoells for her careful copy editing. This material was based on the work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contracts DE-AC02-06CH11347 and DE-AC02-06CH11357. The authors acknowledge partial NSF funding through awards FP061151-01-PR and CNS-1545046 to MA.

REFERENCES

Anderes, E. B., & Stein, M. L. (2008). Estimating deformations of isotropic Gaussian random fields on the plane. *Annals of Statistics*, *36*(2), 719–741.

Apanasovich, T. V., & Genton, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, *97*(1), 15–30.

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98.

Butterworth, S. (1930). In the theory of filter amplifires. *Experimental Wireless & the Wireless Engineer*, *7*, 536–541.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, *19*(90), 297–301.

Cressie, N., & Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, *94*(448), 1330–1339.

Cushman-Roisin, B. (2019). Atmospheric boundary layer. *Environmental Fluid Mechanics*, 165–186. New York, John Wiley & sons.

de Luna, X., & Genton, M. G. (2005). Predictive spatio-temporal models for spatially sparse enviromental data. *Statistica Sinica*, *15*(2), 547–568.

Fuentes, M., & Smith, R. (2001). *A new class of nonstationary models, Technical report*. Tech. report at North Carolina State University, Institute of Statistics.

Gage, K. S., & Balsley, B. B. (1978). Doppler radar probing of the clear atmosphere. *Bulletin of the American Meteorological Society*, *59*(9), 1074–1094.

Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics*, *18*(7), 681–695.

Genton, M. G., & Perrin, O. (2004). On a time deformation reducing nonstationary stochastic processes to local stationarity. *Journal of Applied Probability*, *41*(1), 236–249.

Geoga, C. J., Anitescu, M., & Stein, M. L. (2019). Scalable gaussian process computations using hierarchical matrices. *Journal of Computational and Graphical Statistics*, *29*(2), 227–237.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, *97*(458), 590–600.

Guinness, J., & Stein, M. L. (2013). Interpolation of nonstationary high frequency spatial-temporal temperature data. *Annals of Applied Statistics*, *7*(3), 1684–1708.

Haslett, J., & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *38*(1), 1–21.

Heyde, C. C. (2008). *Quasi-likelihood and its application: A general approach to optimal parameter estimation*. Springer Science & Business Media.

Higdon, D. (2002). *Space and space-time modeling using process convolutions*. In C. W. Anderson, V. Barnett, P. C. Chatwin, & A. H. El-Shaarawi (Eds.), *Quantitative methods for current environmental issues* (pp. 37–56). Springer.

Horrell, M. T., & Stein, M. L. (2017). Half-spectral space-time covariance models. *Spatial Statistics*, *19*, 90–100.

Hutchinson, M. (1990). A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics – Simulation and Computation*, *19*(2), 433–450.

Ibragimov, I. A., & Rozanov, Y. A. (1978). *Gaussian random processes, Stochastic Modelling and Applied Probability*. Springer-Verlag.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, *112*(517), 201–214.

Katzfuss, M., & Cressie, N. (2012). Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics*, *23*(1), 94–107.

Krock, M., Kleiber, W., & Becker, S. (2020). Nonstationary modeling with sparsity for spatial data via the basis graphical lasso. *Journal of Computational and Graphical Statistics*. https://doi.org/10.1080/10618600.2020.1811103.

Lindgren, F., Rue, H., & Lindström, J. (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 (Statistical Methodology)*, *73*(4), 423–498.

Mather, J. H., & Voyles, J. W. (2013). The arm climate research facility: A review of structure and capabilities. *Bulletin of the American Meteorological Society*, *94*(3), 377–392.

Matsuo, T., Nychka, D. W., & Paul, D. (2011). Nonstationary covariance modeling for incomplete data: Monte Carlo EM approach. *Computational Statistics & Data Analysis*, *55*(6), 2059–2073.

Muradyan, P., & Coulter, R. (2020). *Radar Wind Profiler (RWP) and Radio Acoustic Sounding System (RASS) Instrument Handbook (DOE/SC-ARM-TR-044)*, DOE Office of Science Atmospheric Radiation Measurement (ARM) Program, United States. Retrieved from https://www.arm.gov/publications/tech_reports/handbooks/rwp_handbook.pdf

Newsom, R., & Krishnamurthy, R. (2010), Doppler Lidar (DLFPT), *Technical report*, Atmospheric Radiation Measurement (ARM) user facility. https://doi.org/10.5439/1025185. Retrieved from https://www.archive.arm.gov/metadata/html/sgpdlfptS01.b1.html

Newsom, R. K. (2012). Doppler lidar (Dl) handbook, Technical Report DOE/SC-ARM/TR-101, DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States). Retrieved from https://www.osti.gov/biblio/1034640

Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

Nychka, D., Hammerling, D., Krock, M., & Wiens, A. (2018). Modeling and emulation of nonstationary Gaussian fields. *Spatial Statistics*, *28*, 21–38.

Nychka, D., Wikle, C., & Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, *2*(4), 315–331.

Olver, F. W., Lozier, D. W., Boisvert, R. F., & Clark, C. W. (2010). *NIST handbook of mathematical functions hardback and CD-ROM*. Cambridge University Press.

Osborne, M. R. (1992). 'Fisher's method of scoring. *International Statistical Review/Revue Internationale de Statistique*, *60*(1), 99–117.

Paciorek, C. J., & Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, *17*(5), 483–506.

Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Temperatures in transient climates: Improved methods for simulations with evolving temporal covariances. *Annals of Applied Statistics*, *10*(1), 477–505.

Porcu, E., Furrer, R., & Nychka, D. (2020). 30 Years of space-time covariance functions. *WIREs Computational Statistics*, e1512.

Porcu, E., Gregori, P., & Mateu, J. (2009). Archimedean spectral densities for nonstationary space-time geostatistics. *Statistica Sinica*, *19*(1), 273–286.

Porcu, E., Mateu, J., & Christakos, G. (2009). Quasi-arithmetic means of covariance functions with potential applications to space-time data. *Journal of Multivariate Analysis*, *100*(8), 1830–1844.

Sampson, P. D., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, *87*(417), 108–119.

Sisterson, D. L., Peppler, R. A., Cress, T. S., Lamb, P. J., & Turner, D. D. (2016). The arm southern great plains (Sgp) site. *Meteorological Monographs*, *57*, 6.1–6.14.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. Springer Science & Business Media.

Stein, M. L. (2005a), *Nonstationary spatial covariance functions, Technical report*.

Stein, M. L. (2005b). Statistical methods for regular monitoring data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(5), 667–687.

Stein, M. L., Chen, J., & Anitescu, M. (2013). Stochastic approximation of score functions for Gaussian processes. *Annals of Applied Statistics*, *7*(2), 1162–1191.

Stokes, G. M., & Schwartz, S. E. (1994). The atmospheric radiation measurement (Arm) program: Programmatic background and design of the cloud and radiation test bed. *Bulletin of the American Meteorological Society*, *75*(7), 1201–1222.

Stroud, J. R., Müller, P., & Sansó, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(4), 673–689.

Tucker, S. C., Senff, C. J., Weickmann, A. M., Brewer, W. A., Banta, R. M., Sandberg, S. P., Law, D. C., & Hardesty, R. M. (2009). Doppler lidar estimation of mixing height using turbulence, shear, and aerosol profiles. *Journal of Atmospheric and Oceanic Technology*, *26*(4), 673–688.

Wächter, A., & Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, *106*(1), 25–57.

Wikle, C. K., & Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, *86*(4), 815–829.

Xu, K., & Wikle, C. K. (2007). Estimation of parameterized spatio-temporal dynamic models. *Journal of Statistical Planning and Inference*, *137*(2), 567–588.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, *99*(465), 250–261.

Zhang, H., & Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, *92*(4), 921–936.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Geoga CJ, Anitescu M, Stein ML. Flexible nonstationary spatiotemporal modeling of high-frequency monitoring data. *Environmetrics*. 2021;e2670. https://doi.org/10.1002/env.2670

APPENDIX

Details and discussion on optimization

As discussed earlier, the numerical optimization of the log-likelihood is a nonconvex problem, and as such, is very difficult even in the simplest cases. As a first step toward improving the quality of the optimization, we opt to use second-order methods, albeit using the expected Fisher matrix as an approximation to the observed information matrix (in the spirit of Fisher scoring). For data sizes on the order of 17,000 observations, the linear algebra required to evaluate the exact likelihood is substantial, and both its gradient and Hessian (and expected Fisher matrix) require matrix–matrix operations, which at these data and parameter sizes would make the estimation infeasible. In order to avoid all matrix–matrix operations, we used "symmetrized" and unbiased randomized sample average approximation (SAA) estimators (Hutchinson, 1990) for both the gradient and expected Fisher information matrix. The gradient of the log-likelihood is given elementwise by

$$-2(\nabla \ell(\boldsymbol{\theta}))_j = \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) - \boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{y}.$$

This can be approximated by substituting the symmetrized unbiased SAA trace estimator, so that

$$-2(\nabla \ell(\boldsymbol{\theta}))_j \approx S^{-1}\sum_{s=1}^{S} \boldsymbol{u}_s^T\boldsymbol{W}^{-1}\boldsymbol{\Sigma}_j\boldsymbol{W}^{-T}\boldsymbol{u}_s - \boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{y}, \tag{A1}$$

where $\boldsymbol{\Sigma}_j := \frac{\partial}{\partial\theta_j}\boldsymbol{\Sigma}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma} = \boldsymbol{W}\boldsymbol{W}^T$, and each $\boldsymbol{u}_s$ is a vector of i.i.d. Rademacher variables, fixed for the duration of the optimization. The expected Fisher matrix, with entries given by

$$\mathcal{I}_{j,k} = \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_k),$$

has a similar symmetrized SAA estimator that is more complicated to express; see Geoga et al. (2019) for full details. These symmetrized estimators have variability that is decoupled from the condition number of the matrix whose trace is being estimated, and they enjoy much more stability than traditional Hutchinson-type trace estimators (Geoga et al., 2019; Hutchinson, 1990). Finally, they can be computed in a single pass with only one memory buffer for the derivative matrices.

Using approximations of these derivatives precludes using powerful and vetted professional optimization software, such as the freely available Ipopt (Wächter & Biegler, 2006), which does not allow for stopping criteria any coarser than the size of the gradient. In particular, optimization methods often make progress by solving quadratic approximations of the objective function, that, in the framework of Nocedal and Wright (2006), may look like

$$\min_{\boldsymbol{p}\in\mathbb{R}^n} m_k(\boldsymbol{p}) = f_k + \nabla f_k^T\boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T\boldsymbol{B}_k\boldsymbol{p} \quad \text{s.t.} \quad \|\boldsymbol{p}\| \le \Delta_k. \tag{A2}$$

While we show here the formalism of trust-region methods, which we subsequently employ, a very similar setup exists for line-search methods, where the norm constraint on p is removed at the expense of requiring $\boldsymbol{B}_k$ to be positive definite (the choice of identity resulting in steepest descent, for example). Substituting into (A2) the randomized approximation of ∇f_k (A1) and the analogous approximation of the expected Fisher matrix for $\boldsymbol{B}_k$ (which is also not the Jacobian of the randomized gradient approximation) will thus become problematic for off-the-shelf algorithms such as Ipopt since their control structure assumes that the linear term includes the true gradient. Moreover, gradient-based stopping criteria may not be possible in our case, since only an approximation for ∇f is provided. We are thus forced to work instead with relative change in the objective function as a stopping criterion. For these reasons, we used a hand-written trust region method adapted directly from Nocedal and Wright (2006), with a subproblem of the type (A2). We made this choice since, in our experience, trust regions methods are more robust to nonconvexity and Hessian ill-conditioning than line search methods, and anecdotally they seem to work better than line search methods for maximum likelihood problems in general. In our experimentation in toy problems with exact derivatives, our implementation reliably does as well as Ipopt for Gaussian process maximum likelihood problems in terms of the terminal likelihood.

While sacrificing higher quality convergence criteria is unfortunate, the computational gain of using the approximated derivatives is quite substantial. In this particular case, it changed the expected run-time of the estimation for a single day of measurements (which we need to do six times) from the order of a week to 7 or 8 h, and is arguably the only setting in which the computations are feasible at all with desktop-level compute power. Evaluating the expected Fisher matrix once with exact linear algebra—even with enough memory (above 50 GB) to store all 25 of the pre-solved matrices $\boldsymbol{\Sigma}^{-1}\frac{\partial}{\partial\theta_j}\boldsymbol{\Sigma}$ and completely remove redundant computations—would take approximately eight hours on a high-end CPU.

For optimization problems such as the this one, assessing convergence quality is a fundamental concern. A full discussion of that assessment is out of the scope of this manuscript, and we refer readers to a dedicated text on numerical optimization, for example, Nocedal and Wright (2006). What does require discussion here is the statistical implications of these issues. It is well known that even the simplest covariance functions can lead to likelihood surfaces with very challenging features. Figure 2 of Stein et al. (2013), for example, shows that the likelihood surface is incredibly flat along a particular nonconvex part of the domain, which in general is likely to be in regions of parameter space that would result in similar interpolation (Stein, 1999). While it is difficult to make a mathematically precise claim on this topic, we believe a similar phenomenon is likely occurring for this much more complex model, and we consider the issue of being somewhere in such a nonconvex region, but potentially far away from the true MLE, to be the primary concern when assessing the quality of the estimation result.

With this in mind, then, and considering the intrinsic difficulty of nonconvex optimization, we observe that it is not possible to confidently assert that the parameters we present in this work are "the" MLE. Fortunately, the model we introduce here is somewhat uniquely suited to some additional heuristic ways to check the quality of parameter estimates: comparison with the marginal nonparametric estimators. As the figures in the main body and the exhaustive collection of figures in the supplementary information demonstrate, the parametric forms for marginal spectra and coherences agree at least reasonably well with the nonparametric estimators in all cases. This is not proof that the estimates we show here are the true MLE, but it is encouraging, at least if one assumes that the parameterization is sufficiently sensible that there are not distant parameters which yield the same marginal parametric functions. Further, nonexhaustive but diverse exploratory heuristic investigations into estimator quality—exploring different initializations, restarting at perturbed estimators, global optimization in a region around current estimators, and checking for convergence at stricter tolerances—suggest that the minimizers presented here are reliable. This is of course still not proof, but it is reason to be hopeful.

As a final detail, to enable unconstrained optimization, the model was fitted with transformed variables, given that many need to be positive. This was necessary due to the randomized derivative estimators, since several parameters being reasonably close to zero means that even a very accurate randomized gradient and expected Fisher matrix can cause domain issues. All point estimates and standard deviations provided in this section, however, are provided for the model as parameterized above, obtained by untransforming the computed MLE and evaluating the likelihood and its derivatives once in that domain.

The code used to perform the estimation is spread out over the software packages `HalfSpectral.jl`[3] for the half-spectral kernel function and `GPMaxlik.jl`[4] for efficient optimization of the likelihood, both written in the Julia programming language (Bezanson et al., 2017). Example files in `HalfSpectral.jl` contain specific scripts used to obtain the estimates below, although due to the size of the dataset, we do not include the data itself in these repositories. For access to the data, see the footnotes to Section 3.

[3]https://git.sr.ht/~cgeoga/HalfSpectral.jl.
[4]https://git.sr.ht/~cgeoga/GPMaxlik.jl.