General spatio-temporal factor models for high-dimensional random fields on a lattice

GENERAL SPATIO-TEMPORAL FACTOR MODELS FOR HIGH-DIMENSIONAL RANDOM FIELDS ON A LATTICE

BY MATTEO BARIGOZZI[1,a], DAVIDE LA VECCHIA[2,b] AND HANG LIU[3,c]

[1] *Department of Economics, University of Bologna,* [a] *matteo.barigozzi@unibo.it*

[2] *Geneva School of Economics and Management, University of Geneva,* [b] *Davide.LaVecchia@unige.ch*

[3] *Faculty of Business in SciTech, School of Management, University of Science and Technology of China,* [c] *hliu01@ustc.edu.cn*

Motivated by the need for analysing large spatio-temporal panel data, we introduce a novel non-parametric methodology for n-dimensional random fields observed across S spatial locations and T time periods. We call it General Spatio-Temporal Factor Model (GSTFM). First, we provide the probabilistic and mathematical underpinning needed for the representation of a random field as the sum of two components: the common component (driven by a small number q of latent factors) and the idiosyncratic component (mildly cross-correlated). We show that the two components are identified as $n \to \infty$. Second, we propose an estimator of the common component and derive its statistical guarantees (consistency and rate of convergence) as $\min(n, S, T) \to \infty$. Third, we propose an information criterion to determine the number of factors. Estimation makes use of Fourier analysis in the frequency domain and thus it fully exploits the information on the spatio-temporal covariance structure of the whole panel. Synthetic data examples illustrate the applicability of GSTFM and its advantages over the extant generalized dynamic factor model that ignores the spatial correlations.

1. Introduction.

1.1. *Big data on spatio-temporal processes.* Many data analysis problems in economics, finance, medicine, environmental sciences, and other scientific areas need to conduct inference on random phenomena observed over time and registered at different locations.

Supervised and unsupervised learning methods for random fields (henceforth, rf) are suitable tools for the statistical analysis of this type of data: they provide an understanding of the key spatial and/or temporal dynamics of the studied phenomena. For instance, rf are routinely applied in medicine for fMRI data analysis (see e.g. Lazar, 2008), in geostatistics for satellite images analysis (see e.g. Cressie and Wikle, 2015), in natural sciences for modeling complex phenomena (see e.g. Vanmarcke, 2010), in economics for the analysis of spatial panel data (see e.g. Baltagi, 2008) just to mention few book-length introductions.

In this paper we consider datasets containing records on spatio-temporal rf over a lattice. We let $(s_1 \ s_2) \in \mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2$ denote the spatial position in and $t \in \mathbb{Z}$ represent the time index—in principle, the dimension of the spatial lattice can be larger than two. For instance, the points $(s_1 \ s_2)$ can be: in geostatistics, geographical regions represented as a network with a given adjacency matrix; in image analysis, the position of pixels in an image. At each $(s_1 \ s_2)$ and time t, the object of interest is the n-dimensional rf: $\boldsymbol{x}_n = \{\boldsymbol{x}_{n\boldsymbol{\varsigma}} = (x_{1\boldsymbol{\varsigma}} \cdots x_{\ell\boldsymbol{\varsigma}} \cdots x_{n\boldsymbol{\varsigma}})^\top, \boldsymbol{\varsigma} = (s_1 \ s_2 \ t)^\top \in \mathbb{Z}^3\}$, for $n \in \mathbb{N}$. Typical inference goals for these types of data include e.g. constructing and analyzing generative models, quantifying spatio-temporal dependency, prediction or image restoration.

MSC2020 subject classifications: 62H25, 62M15, 62M40, 62H20, 60G35.

Keywords and phrases: Factor models, high-dimensional random fields, spectral analysis, functional/phsyical dependence.

One key aspect related to the statistical analysis of this data is that n is of the order of several hundreds and the number of locations and time points may have the same magnitude. A common approach to analyze the resulting large spatio-temporal rf datasets is to resort on standard time series methods, like e.g. spatio-temporal autoregressive models (see e.g. Cressie and Wikle, 2015). Nevertheless, because of the high-dimensionality, standard parametric approaches are not feasible (e.g. in a vector autoregressive model with one time lag for the time series available at each location $\boldsymbol{s}$, the number of parameters is n^2) and alternative non-parametric techniques are needed.

To fight against the curse of dimensionality, one may look at the literature on high-dimensional time series and think of relying on factor models, which can be considered as an unsupervised learning technique. This class of models is based on a *divide and conquer* approach, where a high-dimensional dataset is decomposed into a low-dimensional common component plus an idiosyncratic component, which might be (serially and cross-sectionally) correlated or not (see, e.g., Forni et al., 2000; Bai, 2003; Lam and Yao, 2012). Among the existing approaches to factor analysis, the General Dynamic Factor Model (GDFM) of Forni et al. (2000) defines the most general, non-parametric, factor model which is based on a decomposition of the observations into the sum of two mutually orthogonal (at all leads and all lags) components: the common component (driven by a small number q of factors and their lags) and the idiosyncratic component (both serially and mildly cross-correlated). This decomposition looks attractive since it is able to capture not only contemporaneous correlations but all leading and lagging co-movements in time among the n components of the time series.

In the case of a spatio-temporal rf the set of correlations among its n components is much richer. Indeed, an observation at time t and spatial location $(s_1\ s_2)$ might depend on observations at time t' in the same location, or on observations at the same time but at spatial location $(s_1'\ s_2')$, but also on observations at time t' and spatial location $(s_1'\ s_2')$. Thus, factor models for spatio-temporal rf have to account for this richer correlation structure.

1.2. *Our contributions: the paper in brief.* We introduce the General Spatio-Temporal Factor Model (GSTFM), a new a class of factor models which allows us to analyze high-dimensional spatio-temporal datasets by capturing all relevant correlations, across both time and space. Our results contribute to different streams of literature on rf theory and inference on high-dimensional data.

(i) We derive the decomposition of a spatio-temporal rf into a common component, which depends on q unobservable factors, and an idiosyncratic component, see Theorem 4.1. To obtain this result, we need to tackle an important challenge rooted into probability theory: because of the lack of ordering in $\mathbb{Z}^2$, the GDFM results already available in the literature on high-dimensional time series cannot be applied in our setting. Indeed, the extant results are available for discrete time (regularly spaced) time series indexed by $t \in \mathbb{Z}$ and rely on a generalization of the Wold representation to the case of infinite-dimensional stationary processes as derived by Forni and Lippi (2001) and Hallin and Lippi (2013). Similar concepts are not readily available for a rf indexed in $\mathbb{Z}^3$. As a possible solution, we might specify a notion of spatial past, selecting e.g. the half-plane or the quarter-plane formulation. However, this choice entails the drawback that different versions of the Wold decomposition (see Mandrekar and Redett, 2017) are available, with no obvious indication on which one has to be preferred in our context. To avoid this issue, we resort on the Fourier analysis in the frequency domain. This methodological approach requires a careful extension to rf of the time series notions of canonical isomorphism, dynamic averaging sequences, aggregation space, dynamic eigenvalues and eigenvectors, spatio-temporal linear filters, idiosyncratic variables, and many others. Our theoretical developments would not be justified without these preliminary results. Clearly, our results nest as a special case the GDFM results.

(ii) The mentioned results are at the population level: to apply our methodology we need an estimation procedure of the factor decomposition and, in particular, of the common component. To this end, we derive a complete and operational estimation theory, which contributes to the literature on the statistical analysis of rf. More in detail, we build on Deb, Pourahmadi and Wu (2017) and, making use of a suitable notion of functional dependence measure for spatio-temporal rf, we derive a consistent estimator of a high-dimensional spectral density matrix. We provide its statistical guarantees, proving consistency (with rate) of the proposed estimator. These general results (which are of their own theoretical interest, see Appendix D) substantially extend the applicability of spectral analysis to non-linear, non-Gaussian, or non-strong mixing rf. Thanks to these novel results, we derive the rate of converge of our estimator of the common component of spatio-temporal rf. The asymptotic regime that we consider is very flexible: it simply requires that the number of locations and the time diverge, but does not need a specification of the type of asymptotics (in-fill or a long-span); see Theorems 6.1 and 6.2 for the mathematical detail.

(iii) The above theoretical developments hinge on a central aspect: the selection of the number of latent factors. We take care of that and state a simple and operational criterion, providing its theoretical underpinning in Theorem 7.1.

(iv) We consider the computational aspects needed to implement our methodology by studying synthetic data examples (see the supplementary material for additional numerical exercices), in which the underlying data generating process involves different types of convolutions over the lattice, which in turn imply different levels of spatio-temporal aggregation.

1.3. *Related work.* Modern factor models essentially originate in four pioneering contributions: Geweke (1977), Sargent and Sims (1977), and Chamberlain and Rothschild (1983). The reference factor model for this work is the GDFM introduced by Forni et al. (2000) and Forni and Lippi (2001), where the idiosyncratic components are allowed to be serially and weakly cross-sectionally correlated. A criterion for the number of factors is proposed by Hallin and Liška (2007). The GDFM was then generalized to the presence of a block structure in the data (where blocks can be seen as spatial locations) by Hallin and Liška (2011), and further developed in a predictive context by Forni et al. (2005, 2015, 2017). Successful applications in macroeconomic and financial time series problems are, e.g., in Altissimo et al. (2010); Cristadoro et al. (2005); Hallin and Trucíos (2021); Trucíos et al. (2022).

A second strand of influential papers considers high-dimensional factor models for time series but of the static type in the sense that the factors are assumed to be loaded only contemporaneously by the observed time series, see, e.g., Stock and Watson (2002); Bai (2003); Fan, Liao and Mincheva (2013). Contrary to the common belief that a GDFM as an equivalent static representation with more factors, Gersing, Rust and Deistler (2024) show that, in fact, the GDFM nests the static approach and it is therefore a superior approach. Furthermore, the GDFM decomposition is not a statistical model rather a representation result (Forni and Lippi, 2001; Hallin and Lippi, 2013). Last, Lam and Yao (2012) propose a time series factor model with static loadings and white noise idiosyncratic components, while the GDFM puts no restrictions on their autocorrelations.

Spatial factor models and related techniques for the analysis of large spatial datasets are also available in the statistical literature. For instance, Christensen and Amemiya (2002) introduce a generalized shifted-factor model for purely spatial data; Wang and Wall (2003) study correlations which are caused by a common spatially correlated underlying factors; Park et al. (2009); Bodelet and La Vecchia (2022) propose a semi-parametric (robust) factor model for spatio-temporal data.

Last, a spatio-temporal dataset as the one considered in this paper can in principle be described as a tensor, with one mode corresponding to the time dimension, two to the spatial

dimensions and one to the cross-sectional units. The most natural approach for statistical inference would then be to model data as a 3-fold tensor evolving in time, i.e., a 3-fold time series tensor, so that its moments are estimated by taking time averages. This is in-line with existing tensor factor model approaches (see, e.g., Chen, Yang and Zhang, 2022; Chang et al., 2023), which, however, are only of the static type and do not embed any notion of spatial lag. In the next section we consider the case where we unfold such a tensor into a vector and we show that such approach does not fully capture spatial dependencies.

1.4. *Outline.* The paper has the following structure. In Section 2 we provide a motivating example for the necessity of introducing a new class of dynamic spatio-temporal factor models. In Section 3 we review the main concepts for the spectral analysis of rf and for aggregation. In Sections 4-5 we derive the representation theorem for the GSTFM and we define the spatio-temporal dynamic principal components. In Section 6 we present our estimator and its asymptotic properties. In Section 7 we introduce a criterion to estimate the number of factors. In Section 8, we provide a pseudo code (see Algorithm 1) to implement our method and we show numerical results on simulated data. In Section 9 we conclude. In the supplementary material: we prove all theoretical results (Appendices A, B, C, E, and F), we prove new results for the estimation of the spectral density of a spatio-temporal rf (Appendix D), we give an algorithm to estimate the number of factors (Appendix G), we apply our methodology to fMRI data (Appendix H), and we provide further numerical results (Appendix I).

1.5. *Notation.* Given a complex matrix $\mathbf{D}$, we denote by $\mathbf{D}^\dagger$ the complex conjugate of the transposed of $\mathbf{D}$, by $\mathbf{D}^\top$ its transposed, by $\bar{\mathbf{D}}$ its complex conjugate, and for a real matrix $\mathbf{D}$ we have $\bar{\mathbf{D}} = \mathbf{D}$ and $\mathbf{D}^\dagger = \mathbf{D}^\top$. A similar notation holds for complex and real vectors. Given a complex scalar z its complex conjugate is denoted as $z^\dagger$. Given two complex row vectors $\boldsymbol{w} = (w_1 \cdots w_m)$ and $\boldsymbol{v} = (v_1 \cdots v_m)$ we let $\langle \boldsymbol{w}, \boldsymbol{v} \rangle = \boldsymbol{w}\boldsymbol{v}^\dagger = \sum_{i=1}^m w_i v_i^\dagger$ and $\|\boldsymbol{w}\| = \sqrt{\langle \boldsymbol{w}, \boldsymbol{w} \rangle}$ is the L_2 or Euclidean norm. Real or integer vectors are always column vectors and given two such vectors $\boldsymbol{w} = (w_1 \cdots w_m)^\top$ and $\boldsymbol{v} = (v_1 \cdots v_m)^\top$ we let $\langle \boldsymbol{w}, \boldsymbol{v} \rangle = \boldsymbol{w}^\top \boldsymbol{v} = \sum_{i=1}^m w_i v_i$ and $\|\boldsymbol{w}\| = \sqrt{\langle \boldsymbol{w}, \boldsymbol{w} \rangle}$. For a complex scalar we have $|z| = \sqrt{z z^\dagger}$. We use the notation $\mathcal{L}$ for the Lebesgue measure either on $\mathbb{R}^d$ or on $\mathbb{C}^d$ or on $\boldsymbol{\Theta} = [-\pi, \pi) \times [-\pi, \pi) \times [-\pi, \pi)$. When no ambiguity can arise, we use the shortcuts $\sum_{\boldsymbol{h}} = \sum_{h_1 \in \mathbb{Z}} \sum_{h_2 \in \mathbb{Z}} \sum_{h_3 \in \mathbb{Z}}$ and $\int_{\boldsymbol{\Theta}} \mathrm{d}\boldsymbol{\theta} = \int_{-\pi}^{\pi} \mathrm{d}\theta_1 \int_{-\pi}^{\pi} \mathrm{d}\theta_2 \int_{-\pi}^{\pi} \mathrm{d}\theta_3$.

2. Motivating example. To motivate our investigation, we illustrate via numerical examples, the inadequacy of the classical GDFM by Forni et al. (2000) in the spatio-temporal setting. Assume we are given realizations of n random variables in $S_1 \times S_2$ spatial locations (therefore, the total number of locations is $S = S_1 S_2$) and T time points. We organize the data into an n-dimensional rf

$$\boldsymbol{x}_n = \{x_{\ell \boldsymbol{\varsigma}},\ \ell = 1, \ldots, n,\ \boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top,\ s_1 = 1, \ldots, S_1,\ s_2 = 1, \ldots, S_2,\ t = 1, \ldots, T\}.$$

Under our GSTFM the ℓ-th component of $\boldsymbol{x}_n$ is such that $x_{\ell \boldsymbol{\varsigma}} = \chi_{\ell \boldsymbol{\varsigma}} + \xi_{\ell \boldsymbol{\varsigma}}$, for $\ell = 1, \ldots, n$. The term $\chi_{\ell \boldsymbol{\varsigma}}$ is called common component and it is a linear combination of q latent rf, with $q \ll n$, located at the same spatial location and time period as well as at neighbouring spatial points and at various lags. The term $\xi_{\ell \boldsymbol{\varsigma}}$ is called idiosyncratic component and is assumed to be weakly cross-sectionally correlated. In Theorem 4.1 we show that under the considered setting the presence of an eigen-gap in the eigenvalues of the spatio-temporal spectral density matrix (see Section 3.2 for a definition) is a key distinctive feature. In particular, the q largest eigeinvalues of the spatio-temporal spectral density matrix diverge as $n \to \infty$ while the remaining $n - q$ stay bounded if and only if the common component $\chi_{\ell \boldsymbol{\varsigma}}$ is driven by q spatio-temporal factors. As $n \to \infty$, we can then disentangle the common and idiosyncratic

components and, consequently, we can identify the GSTFM. This is the main feature of general factor models, sometimes called blessing of dimensionality, as opposed to the curse of dimensionality typically affecting large-dimensional models.

To verify this phenomenon we simulate the common component of the GSTFM with $q = 2, 3$ common factors, loaded according to a quite general and commonly encountered configuration (see Model (a) in (22) for details) of the spatio-temporal dependencies. For ease of simulation, the idiosyncratic component is generated from the standard normal distribution. Then, for different subsets of dimension $m = 1, \dots, n$ we estimate the $m \times m$ spatio-temporal spectral density matrix of $\boldsymbol{x}_m$ and we compute its ten largest eigenvalues, averaged across all frequencies. In Figure 1, we display these eigenvalues as a function of the cross-sectional dimension m: we clearly see that the eigen-gap becomes more and more pronounced as m increases, a manifestation of the blessing of dimensionality.

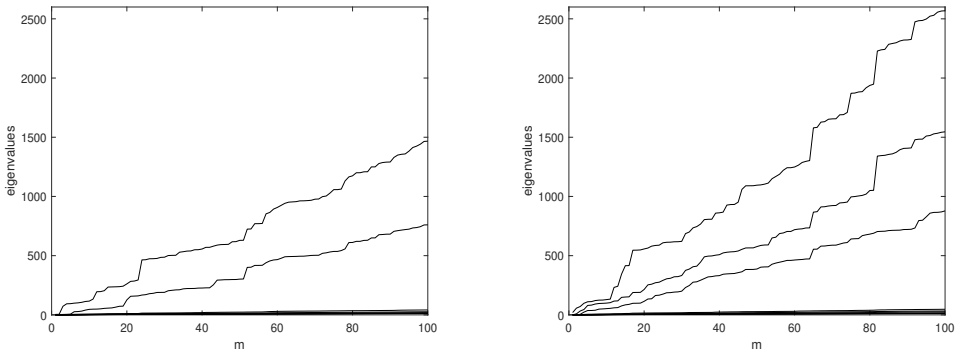

FIG 1. *Spatio-temporal dynamic eigenvalues (averaged over all frequencies) for Model (a) in* (22)*, with* $n = 100$*,* $(S_1, S_2, T) = (10, 10, 100)$*, and* $q = 2$ *(left) or* $q = 3$ *(right). The x-axis represents the dimension of the rf* $m = 1, \dots, n$*.*

If instead we decide to resort on the extant GDFM, then the natural thing to do is to stack at each time point t the data into an nS_1S_2-dimensional time series (the order in which the locations and variables are stacked is irrelevant for this discussion):

$$\mathbf{x}_N = \{x_{it},\ i = 1, \dots, N,\ N = (nS_1S_2),\ t = 1, \dots, T\}.$$

Notice that the rf $\boldsymbol{x}_n$ and the time series $\mathbf{x}_N$ contain the same data points but encoded in different ways. Under the GDFM the i-th component of $\mathbf{x}_n$ is also such that $x_{it} = \chi_{it}^{\text{GDFM}} + \xi_{it}^{\text{GDFM}}$, for $i = 1, \dots, N$, where now χ_{it}^{GDFM} is a linear combination of r latent time series, with $r \ll N$, at the same time period as well as at various lags. Given that the stacking procedure yields a very large dataset, the asymptotic results in Forni et al. (2000) should apply: the eigenvalues of the estimated spectral density of $\mathbf{x}_n$ should display an eigen-gap, between the r-th and $(r+1)$-th eigenvalues, increasing as $N \to \infty$. In fact, if there is no spatial correlation in the data, then we would expect to have $r = q$, as the only correlations left would be cross-sectional and temporal and the GDFM is designed precisely to capture them. But, if there are spatial correlations, then the proposed stacking approach is likely to be flawed: ignoring spatial correlations might lead to detect more factors due to overlooked spatial dependencies. Indeed, if there are ignored spatial correlations, the GDFM might not even be correctly identified. To show this, we consider again the data simulated from the GSTFM and that yield Figure 1. We estimate the spectral density of the stacked vector $\mathbf{x}_m$ for $m = 1, \dots, N$ and, in Figure 2, we display, as functions of m, the ten largest corresponding

eigenvalues averaged over all frequencies. Since now $N \gg n$, we might expect an eigen-gap even more evident than the one clearly visible in Figure 1. In contrast, in Figure 2 no eigen-gap is detectable at all, even for very large cross-sectional dimensions: this means that the true number r of factors cannot be recovered and the GDFM is not identifiable in this setting (for further details on identification of factor models, see Corollary 4.2 and the related discussion).

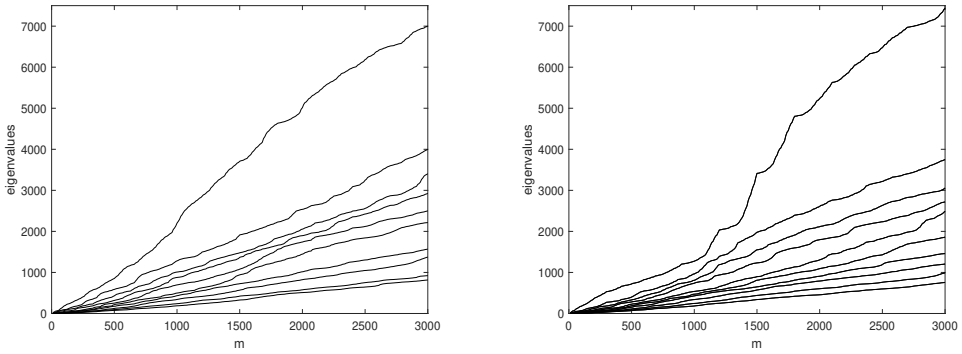

FIG 2. *Temporal dynamic eigenvalues (averaged over all frequencies) for Model (a) in (22), with* $n = 30$, $(S_1, S_2, T) = (10, 10, 100)$, *and* $q = 2$ *(left) or* $q = 3$ *(right). The x-axis represents the dimension of the stacked vector* $m = 1, \ldots, (nS_1S_2)$.

The above arguments illustrate a two-fold statistical problem related to the development of a novel theory of general factor models for spatio-temporal rf. On the one hand, there is the need for a representation theory for large-dimensional rf, which allows us to capture the common spatio-temporal factors that explain the spatio-temporal co-movements of the process. On the other hand, there is the central need for estimation methods which have statistical guarantees and yield an estimate of the common spatio-temporal component converging to true spatio-temporal common component. In the next sections, we illustrate how to solve this statistical problem.

3. Fundamental notions.

3.1. *Random fields.* Our object of interest is the infinite-dimensional rf on a lattice $\boldsymbol{x} = \{\boldsymbol{x}_{(\boldsymbol{s}\ t)} = (x_{1,(\boldsymbol{s}\ t)}\ x_{2,(\boldsymbol{s}\ t)} \cdots x_{\ell,(\boldsymbol{s}\ t)} \cdots)^\top, \boldsymbol{s} \in \mathbb{Z}^2, t \in \mathbb{Z}\}$. We index space-time points as $\boldsymbol{\varsigma} = (\boldsymbol{s}\ t)^\top = (s_1\ s_2\ t)^\top \in \mathbb{Z}^3$, with s_1, s_2, and t allowed to vary independently. We remark that we may directly extend our approach to rf observed in $\mathbb{Z}^d$, with $d \geq 3$. Nevertheless, we prefer to develop our methodology when $d = 3$, as this case covers most real spatio-temporal data and it also allows us to keep the notation simple. So, for any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$, we define the infinite-dimensional random vector $\boldsymbol{x}_{\boldsymbol{\varsigma}} = (x_{1\boldsymbol{\varsigma}}\ x_{2,\boldsymbol{\varsigma}} \cdots x_{\ell\boldsymbol{\varsigma}} \cdots)^\top$ and for any $n \in \mathbb{N}$, we also define the n-dimensional column random vector $\boldsymbol{x}_{n\boldsymbol{\varsigma}} = (x_{1\boldsymbol{\varsigma}} \cdots x_{n\boldsymbol{\varsigma}})^\top$, which is an element of the n-dimensional rf $\boldsymbol{x}_n = \{\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$. Clearly, $\boldsymbol{x}_n$ is a sub-process of $\boldsymbol{x}$.

Throughout, we let $\mathcal{P} = (\Omega, \mathcal{F}, P)$ be a probability space and let $L_2(\mathcal{P}, \mathbb{C})$ be the linear space of all complex-valued, square-integrable random variables defined on Ω. Then, if $x_{\ell\boldsymbol{\varsigma}} \in L_2(\mathcal{P}, \mathbb{C})$, for any $\ell \in \mathbb{N}$, the process $\{x_{\ell\boldsymbol{\varsigma}}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ is a complex valued scalar random field with finite variance, and for any fixed n the process $\{\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ is a complex valued vector rf with all its elements having finite variance, and the process $\boldsymbol{x} = \{\boldsymbol{x}_{\boldsymbol{\varsigma}}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ is an infinite-dimensional complex valued rf with all its elements having finite variance. Notice that the

space $L_2(\mathcal{P},\mathbb{C})$ is a complex Hilbert space, thus it possesses the usual inner product given by $\text{Cov}(x_{i\boldsymbol{\varsigma}}, x_{j\boldsymbol{\varsigma}'}) = \text{E}[(x_{i\boldsymbol{\varsigma}} - \text{E}(x_{i\boldsymbol{\varsigma}}))(x_{j\boldsymbol{\varsigma}'} - \text{E}(x_{j\boldsymbol{\varsigma}'}))^\dagger]$, where $\text{E}(x_{i\boldsymbol{\varsigma}}) = \int_{\mathbb{C}} u \text{d}P(u)$ represents the expected value taken w.r.t. the probability P.

Last, we define $\boldsymbol{\mathcal{X}}_n = \overline{\text{span}}(\boldsymbol{x}_n)$ as the minimum closed linear subspace of $L_2(\mathcal{P},\mathbb{C})$, containing $\boldsymbol{x}_n$, i.e., the set of all L_2-convergent linear combinations of $x_{\ell\boldsymbol{\varsigma}}$'s. Therefore, a generic element of $\boldsymbol{\mathcal{X}}_n$ is of the form $\zeta_{n\boldsymbol{\varsigma}} = \sum_{\ell=1}^{n} \sum_{\kappa_1\in\mathbb{Z}} \sum_{\kappa_2\in\mathbb{Z}} \sum_{\kappa_3\in\mathbb{Z}} \alpha_{\ell\boldsymbol{\kappa}}\ x_{\ell,\boldsymbol{\varsigma}+\boldsymbol{\kappa}}$ with $\alpha_{\ell\boldsymbol{\kappa}} \in \mathbb{C}$ and $\boldsymbol{\kappa} = (\kappa_1\ \kappa_2\ \kappa_3)^\top \in \mathbb{Z}^3$. Moreover, define $\boldsymbol{\mathcal{X}} = \overline{\text{span}}(\boldsymbol{x})$, which is such that $\boldsymbol{\mathcal{X}} = \overline{\cup_{n=1}^{\infty} \boldsymbol{\mathcal{X}}_n}$ and contains also the limits, as $n \to \infty$, of all L_2-convergent sequences thereof. Hence, both $\boldsymbol{\mathcal{X}}_n$ and $\boldsymbol{\mathcal{X}}$ are Hilbert spaces.

3.2. *Spatio-temporal autocovariance and spectral density matrices.* A spatio-temporal shift between pairs of points $\boldsymbol{\varsigma} = (\boldsymbol{s}\ t)^\top \in \mathbb{Z}^3$ and $\boldsymbol{\varsigma}' = (\boldsymbol{s}'\ t')^\top \in \mathbb{Z}^3$ is defined as a vector $\boldsymbol{h} = (\boldsymbol{\varsigma} - \boldsymbol{\varsigma}') \in \mathbb{Z}^3$ such that $\boldsymbol{h} = ((s_1 - s_1')\ (s_2 - s_2')\ (t - t'))^\top = (h_1\ h_2\ h_3)^\top$. We need to set an origin $\mathbf{0} = (0\ 0\ 0)^\top$, whose location is arbitrary, but once chosen it remains fix. To make our theory insensitive to origin shifts, we impose space-time stationarity, which implies that the first two moments (mean and covariance) of a space-time rf are invariant respect to space-time translation.

To formalize this property, for any $n \in \mathbb{N}$ and any $\boldsymbol{\varsigma}, \boldsymbol{\varsigma}' \in \mathbb{Z}^3$, we define the $n \times n$ autocovariance matrix: $\text{Cov}(\boldsymbol{x}_{n,\boldsymbol{\varsigma}}, \boldsymbol{x}_{n,\boldsymbol{\varsigma}'}) = \text{E}[(\boldsymbol{x}_{n,\boldsymbol{\varsigma}} - \text{E}[\boldsymbol{x}_{n,\boldsymbol{\varsigma}}])(\boldsymbol{x}_{n,\boldsymbol{\varsigma}'} - \text{E}[\boldsymbol{x}_{n,\boldsymbol{\varsigma}'}])^\dagger]$. Notice that the covariance matrix $\text{Cov}(\boldsymbol{x}_{n,\boldsymbol{\varsigma}}, \boldsymbol{x}_{n,\boldsymbol{\varsigma}})$ is non-negative definite; see, e.g., Stein (2012, p.15). Then, we introduce the following assumption of stationarity.

ASSUMPTION 3.1. For any $n \in \mathbb{N}$ the random field $\boldsymbol{x}_n = \{\boldsymbol{x}_{n\boldsymbol{\varsigma}} = (x_{1\boldsymbol{\varsigma}} \cdots x_{n\boldsymbol{\varsigma}})^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ is such that $x_{\ell\boldsymbol{\varsigma}} \in L_2(\mathcal{P},\mathbb{C})$ for any $\ell \leq n$, and, for any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$: (i) $\text{E}(x_{\ell\boldsymbol{\varsigma}}) = 0$ and $\text{Var}(x_{\ell\boldsymbol{\varsigma}}) > 0$; (ii) $\text{Cov}(\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \boldsymbol{x}_{n\boldsymbol{\varsigma}+\boldsymbol{h}}) = \text{E}(\boldsymbol{x}_{n\boldsymbol{\varsigma}} \boldsymbol{x}_{n\boldsymbol{\varsigma}+\boldsymbol{h}}^\dagger) = \boldsymbol{\Gamma}_n^x(\boldsymbol{h})$ for any $\boldsymbol{h} \in \mathbb{Z}^3$.

A few comments on Assumption 3.1. First, the zero-mean assumption can be made without any loss of generality. Second, since, all elements of the rf $\boldsymbol{x}_n$ are in $L_2(\mathcal{P},\mathbb{C})$ then for any fixed n the covariance matrix $\boldsymbol{\Gamma}_n^x(\mathbf{0})$ is finite and so all autocovariances $\boldsymbol{\Gamma}_n^x(\boldsymbol{h})$, $\boldsymbol{h} \neq \mathbf{0}$ are finite too. Third, note that letting $\boldsymbol{\Gamma}_n^x(\boldsymbol{h}) \equiv \boldsymbol{\Gamma}_n^x(h_1, h_2, h_3)$, then the following relations holds

$$\boldsymbol{\Gamma}_n^x(-h_1,-h_2,-h_3) = \boldsymbol{\Gamma}_n^{x\dagger}(h_1,h_2,h_3), \qquad \boldsymbol{\Gamma}_n^x(-h_1,h_2,h_3) = \boldsymbol{\Gamma}_n^{x\dagger}(h_1,-h_2,-h_3),$$
$$\boldsymbol{\Gamma}_n^x(h_1,-h_2,h_3) = \boldsymbol{\Gamma}_n^{x\dagger}(-h_1,h_2,-h_3), \quad \boldsymbol{\Gamma}_n^x(h_1,h_2,-h_3) = \boldsymbol{\Gamma}_n^{x\dagger}(-h_1,-h_2,h_3),$$

which are much weaker requirements than assuming space isotropy—for which we would have $\text{Cov}(\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \boldsymbol{x}_{n\boldsymbol{\varsigma}+\boldsymbol{h}}) = \boldsymbol{\Gamma}_n^x(\|(\pm h_1\ \pm h_2)^\top\|, h_3)$, thus imposing that the second-order moments are invariant under all rigid axes motions (Stein, 2012, p.17). Then, we introduce (see Mandrekar and Redett, 2017, p. 45)

DEFINITION 3.1 (Orthonormal white noise rf). For a given finite integer q, let $\boldsymbol{w} = \{\boldsymbol{w}_{\boldsymbol{\varsigma}} = (w_{1\boldsymbol{\varsigma}} \cdots w_{q\boldsymbol{\varsigma}})^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ be a q-dimensional rf such that $w_{\ell\boldsymbol{\varsigma}} \in L_2(\mathcal{P},\mathbb{C})$ and $\text{E}(w_{\ell\boldsymbol{\varsigma}}) = 0$, for any $\ell = 1,\ldots,q$ and $\boldsymbol{\varsigma} \in \mathbb{Z}^3$. Then we call $\boldsymbol{w}$ an orthonormal white noise rf if, for any $\boldsymbol{\varsigma}, \boldsymbol{\varsigma}' \in \mathbb{Z}^3$, $\text{Cov}\,(\boldsymbol{w}_{\boldsymbol{\varsigma}}, \boldsymbol{w}_{\boldsymbol{\varsigma}'}) = \mathbf{I}_q$ if $\boldsymbol{\varsigma} = \boldsymbol{\varsigma}'$ and it is $\mathbf{0}$ otherwise.

To conduct spectral analysis we add the following

ASSUMPTION 3.2. For any $n \in \mathbb{N}$, the spectral measure of $\boldsymbol{x}_n$ is absolutely continuous (with respect to $\mathcal{L}$ on $\boldsymbol{\Theta}$), so $\boldsymbol{x}_n$ admits an $n \times n$ spectral density matrix given by

$$\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta}) = \sum_{\boldsymbol{h}} \boldsymbol{\Gamma}_n^x(\boldsymbol{h}) e^{-i\langle \boldsymbol{h},\boldsymbol{\theta}\rangle},$$

where $i = \sqrt{-1}$ and $\boldsymbol{\theta} = (\theta_1\ \theta_2\ \theta_3)^\top \in \boldsymbol{\Theta}$.

For a rf the spectral density matrix depends on a vector of frequencies $\boldsymbol{\theta}$ (Brillinger, 1970), and under Assumptions 3.1 and 3.2, $\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta})$ is Hermitian and non-negative definite (Leonenko, 1999, p.13) for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and any $n \in \mathbb{N}$. The lag-$\boldsymbol{h}$ autocovariance matrix is then given by $\boldsymbol{\Gamma}_n^x(\boldsymbol{h}) = \frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} e^{i\langle \boldsymbol{h},\boldsymbol{\theta}\rangle} \boldsymbol{\Sigma}_n^x(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$, for any $\boldsymbol{h} \in \mathbb{Z}^3$.

Let $\boldsymbol{\Sigma}^x(\boldsymbol{\theta})$ denote the infinite matrix having the matrix $\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta})$ as its $n \times n$ top-left submatrix and notice again that as $n \to \infty$ we allow for the possibility of its eigenvalues to diverge (see below), while all its entries are finite by assumption. Moreover, we notice that a q-dimensional orthonormal white noise rf has spectral density $\mathbf{I}_q$. Then we have

DEFINITION 3.2 (Spatio-temporal dynamic eigenvalues). For any $n \in \mathbb{N}$ and $\ell \leq n$, let $\lambda_{n\ell}^x : \boldsymbol{\Theta} \to \mathbb{R}^+$ be defined as the function associating with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the ℓ-th eigenvalue in decreasing order of $\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta})$. We call $\lambda_{n\ell}^x(\boldsymbol{\theta})$ the spatio-temporal dynamic eigenvalues of $\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta})$.

DEFINITION 3.3 (Spatio-temporal dynamic eigenvectors). For any $n \in \mathbb{N}$ and $\ell \leq n$ let $\boldsymbol{p}_{n\ell}^x : \boldsymbol{\Theta} \to \mathbb{C}^n$ be such that, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the row vector $\boldsymbol{p}_{n\ell}^x(\boldsymbol{\theta})$ satisfies (i) $\|\boldsymbol{p}_{n\ell}^x(\boldsymbol{\theta})\| = 1$; (ii) $\boldsymbol{p}_{n\ell}^x(\boldsymbol{\theta}) \boldsymbol{p}_{nj}^{x\dagger}(\boldsymbol{\theta}) = 0$, for $\ell \neq j$; (iii) $\boldsymbol{p}_{n\ell}^x(\boldsymbol{\theta}) \boldsymbol{\Sigma}_n^x(\boldsymbol{\theta}) = \lambda_{n\ell}^x(\boldsymbol{\theta}) \boldsymbol{p}_{n\ell}^x(\boldsymbol{\theta})$. Then, the functions $\{\boldsymbol{p}_{n\ell}^x, \ell = 1, \ldots, n\}$ constitute a set of spatio-temporal dynamic eigenvectors associated with the spatio-temporal eigenvalues $\lambda_{n\ell}^x$ and the rf $\boldsymbol{x}_n$.

Hereafter, we also assume strict positive definiteness of $\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta})$

ASSUMPTION 3.3. For any $n \in \mathbb{N}$ and $\ell \leq n$, and any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\lambda_{n\ell}^x(\boldsymbol{\theta}) > 0$.

REMARK 3.1. Following Forni and Lippi (2001, Lemma 3 and 4), one can easily prove that: (i) the real functions $\lambda_{n\ell}^x$ are Lebesgue-measurable and integrable in $\boldsymbol{\Theta}$, for any given $n \in \mathbb{N}$ and $\ell \leq n$; (ii) $\lambda_{n\ell}^x(\boldsymbol{\theta})$ is a non-decreasing function of n, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. In particular, from (ii) it follows that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\lim_{n\to\infty} \lambda_{n\ell}^x(\boldsymbol{\theta}) = \sup_{n\in\mathbb{N}} \lambda_{n\ell}^x(\boldsymbol{\theta})$, and it is well defined for any $\ell \leq n$.

3.3. *Spatio-temporal linear filters.* First, for any $n \in \mathbb{N}$ and $\ell \leq n$, let us consider the three linear operators, $L_j : \boldsymbol{\mathcal{X}}_n \to \boldsymbol{\mathcal{X}}_n$, $j = 1, 2, 3$, such that, for any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$,

$$L_1\, x_{\ell\boldsymbol{\varsigma}} = x_{\ell(s_1-1\ s_2\ t)}, \quad L_2\, x_{\ell\boldsymbol{\varsigma}} = x_{\ell(s_1\ s_2-1\ t)}, \quad L_3\, x_{\ell\boldsymbol{\varsigma}} = x_{\ell(s_1\ s_2\ t-1)}, \tag{1}$$

so that when L_j is applied to the vector $\boldsymbol{x}_{n\boldsymbol{\varsigma}}$ it shift all its n components along the space or time dimension. L_1 and L_2 act on the (spatial) dimensions of the lattice (see Whittle (1954)), while L_3 is the usual time lag operator. We also set $L \equiv L_1 L_2 L_3$ and $L^{\boldsymbol{\kappa}} \equiv L_1^{\kappa_1} L_2^{\kappa_2} L_3^{\kappa_3}$. The operators are commutative, e.g. $L_1 L_2 L_3\, x_{\ell\boldsymbol{\varsigma}} = L_3 L_1 L_2\, x_{\ell\boldsymbol{\varsigma}}$. In Lemma A.1 in Appendix A, we show that L_j are unitary operators that can be extended to $\boldsymbol{\mathcal{X}}$.

Second, consider a generic n-dimensional row vector of functions $\boldsymbol{f}_n = (f_1\ \cdots f_n)$ with $f_\ell : \boldsymbol{\Theta} \to \mathbb{C}$ being measurable for any $\ell \leq n$ and such that the following conditions hold: (i) $\|\boldsymbol{f}_n\|_{\boldsymbol{\Sigma}_n^x}^2 = \frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} \boldsymbol{f}_n(\boldsymbol{\theta}) \boldsymbol{\Sigma}_n^x(\boldsymbol{\theta}) \boldsymbol{f}_n^\dagger(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} < \infty$, and (ii) $\|\boldsymbol{f}_n\|^2 = \|\boldsymbol{f}_n\|_{\mathbf{I}_n}^2 < \infty$, respectively. The space of such functions is a complex Hilbert space denoted as $L_2^n(\boldsymbol{\Theta}, \mathbb{C}, \boldsymbol{\Sigma}_n^x, \mathbf{I}_n)$, obtained from the intersection of two Hilbert spaces each endowed with inner products derived from one of the two norms defined above.

Third, consider the map $\mathcal{J} : L_2^n(\boldsymbol{\Theta}, \mathbb{C}, \boldsymbol{\Sigma}_n^x, \mathbf{I}_n) \to \boldsymbol{\mathcal{X}}_n$, such that, for any $n \in \mathbb{N}$ and $\ell \leq n$,

$$\mathcal{J}\left[(\delta_{\ell 1} \cdots \delta_{\ell k} \cdots \delta_{\ell n}) e^{i\langle \boldsymbol{\varsigma}, \cdot\rangle}\right] = x_{\ell\boldsymbol{\varsigma}}, \quad \text{for any } \boldsymbol{\varsigma} \in \mathbb{Z}^3, \tag{2}$$

where $\delta_{\ell k} = 1$ if $k = \ell$ and $\delta_{\ell k} = 0$ if $k \neq \ell$ and $e^{i\langle \boldsymbol{\varsigma}, \cdot\rangle}$ stands for the mapping $\boldsymbol{\theta} \mapsto e^{i\langle \boldsymbol{\varsigma}, \boldsymbol{\theta}\rangle}$ from $\boldsymbol{\Theta}$ to $\mathbb{C}$. Thus, we have

$$L\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \mathcal{J}\left[\boldsymbol{\iota}_n e^{i\langle (s_1-1\ s_2-1\ t-1)^\top, \cdot\rangle}\right] = \mathcal{J}\left[\boldsymbol{\iota}_n e^{-i\langle (1\ 1\ 1)^\top, \cdot\rangle} e^{i\langle \boldsymbol{\varsigma}, \cdot\rangle}\right], \tag{3}$$

where $\boldsymbol{\iota}_n$ is an n-dimensional vector of ones. In Lemma A.2 in Appendix A we prove that $\mathcal{J}$ is an isomorphism (that is, $\mathcal{J}$ preserves the inner product and is one-to-one), also called canonical isomorphism, and it can be extended to the Hilbert space of infinite-dimensional functions $\boldsymbol{f}$, with norms $\|\boldsymbol{f}\|^2_{\boldsymbol{\Sigma}^x} = \lim_{n\to\infty} \|\boldsymbol{f}_n\|^2_{\boldsymbol{\Sigma}^x_n} < \infty$ and $\|\boldsymbol{f}\|^2 = \lim_{n\to\infty} \|\boldsymbol{f}_n\|^2 < \infty$. Notice that $\mathcal{J}$ is an isomorphism between the measure spaces $(\boldsymbol{\Theta}, \mathcal{B}(\boldsymbol{\Theta}), \mathcal{L})$ and $(\boldsymbol{\mathcal{X}}_n, \mathcal{B}(\mathbb{C}^n), \mathcal{L})$, where $\mathcal{B}(\boldsymbol{\Theta})$ and $\mathcal{B}(\mathbb{C}^n)$ are the Borel σ-fields on $\boldsymbol{\Theta}$ and $\mathbb{C}^n$, respectively. The definition of $\mathcal{J}$ extends to our setting the classical isomorphism typically applied in time series analysis (see e.g. Brockwell and Davis, 2006, Section 4.8).

Now, for any $\boldsymbol{f}_n \in L_2^n(\boldsymbol{\Theta}, \mathbb{C}, \boldsymbol{\Sigma}_n^x, \mathbf{I}_n)$ consider the Fourier expansion

$$\boldsymbol{f}_n(\boldsymbol{\theta}) = \sum_{\boldsymbol{\kappa}\in\mathbb{Z}^3} \mathbf{f}_{n\boldsymbol{\kappa}} e^{-i\langle \boldsymbol{\kappa}, \boldsymbol{\theta}\rangle}, \tag{4}$$

$$\mathbf{f}_{n\boldsymbol{\kappa}} = \frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} e^{i\langle \boldsymbol{\kappa}, \boldsymbol{\theta}\rangle} \boldsymbol{f}_n(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{5}$$

where the convergence in (4) is in the sense of the L_2-norm, and $\{\mathbf{f}_{n\boldsymbol{\kappa}}, \boldsymbol{\kappa} \in \mathbb{Z}^3\}$ are the Fourier coefficients.

Since also $(\mathbf{f}_{n\boldsymbol{\kappa}} e^{-i\langle \boldsymbol{\kappa}, \cdot\rangle}) \in L_2^n(\boldsymbol{\Theta}, \mathbb{C}, \boldsymbol{\Sigma}_n^x, \mathbf{I}_n)$, then we apply to it the canonical isomorphism $\mathcal{J}$ to map it into elements $\boldsymbol{\mathcal{X}}_n$. This defines the filtered processes associated to $\boldsymbol{f}_n$

$$\underline{\boldsymbol{f}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \mathcal{J}\left[\boldsymbol{f}_n e^{i\langle \boldsymbol{\varsigma}, \cdot\rangle}\right]. \tag{6}$$

Therefore, from (4), (5), and (6), and by linearity of the canonical isomorphism, we have

$$\underline{\boldsymbol{f}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \sum_{\boldsymbol{\kappa}} \mathbf{f}_{n\boldsymbol{\kappa}} L^{\boldsymbol{\kappa}} \boldsymbol{x}_{n\boldsymbol{\varsigma}} = \left\{ \sum_{\boldsymbol{\kappa}} \left[\frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} e^{i\langle \boldsymbol{\kappa}, \boldsymbol{\theta}\rangle} \boldsymbol{f}_n(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \right] L^{\boldsymbol{\kappa}} \right\} \boldsymbol{x}_{n\boldsymbol{\varsigma}}, \tag{7}$$

which defines an n-dimensional linear spatio-temporal filter. Note that $\underline{\boldsymbol{f}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} \in \boldsymbol{\mathcal{X}}_n$ is the isomorphic map of $(\boldsymbol{f}_n e^{i\langle \boldsymbol{\varsigma}, \cdot\rangle}) \in L_2^n(\boldsymbol{\Theta}, \mathbb{C}, \boldsymbol{\Sigma}^x, \mathbf{I}_n)$, that is multiplications become convolutions via the isomorphism $\mathcal{J}$ and viceversa via $\mathcal{J}^{-1}$. Hereafter, the composition of two linear filters, is denoted as

$$\underline{\boldsymbol{g}}_n(L) \star \underline{\boldsymbol{f}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \left\{ \sum_{\boldsymbol{\kappa}} \left[\frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} e^{i\langle \boldsymbol{\kappa}, \boldsymbol{\theta}\rangle} \boldsymbol{g}_n(\boldsymbol{\theta}) \boldsymbol{f}_n(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \right] L^{\boldsymbol{\kappa}} \right\} \boldsymbol{x}_{n\boldsymbol{\varsigma}}. \tag{8}$$

3.4. *Aggregation, common, and idiosyncratic components.* Our approach is based on spatio-temporal aggregation of rf. To introduce the key tools needed for our development, we start by

DEFINITION 3.4 (Spatio-temporal aggregation of rf). For any $n \in \mathbb{N}$, consider an n-dimensional row vector of functions $\boldsymbol{a}_n \in L_2^n(\boldsymbol{\Theta}, \mathbb{C}, \boldsymbol{\Sigma}_n, \mathbf{I}_n)$. The sequence $\{\boldsymbol{a}_n, n \in \mathbb{N}\}$ is a spatio-temporal dynamic averaging sequence (STDAS) if

$$\lim_{n\to\infty} \|\boldsymbol{a}_n\| = \lim_{n\to\infty} \left(\frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} \boldsymbol{a}_n(\boldsymbol{\theta}) \boldsymbol{a}_n^\dagger(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \right)^{1/2} = 0.$$

Moreover, we say that y is an aggregate if for any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$ there exists a STDAS $\{\boldsymbol{a}_n, n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} \underline{\boldsymbol{a}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = y_{\boldsymbol{\varsigma}}$ in mean-square and $y_{\boldsymbol{\varsigma}} \in \boldsymbol{\mathcal{X}}$. We denote the set of all aggregates by $\mathcal{G}(\boldsymbol{x})$ and we refer to it as the aggregation space of $\boldsymbol{\mathcal{X}}$.

Intuitively, the aggregation via a STDAS corresponds to averaging an infinite-dimensional rf both in the cross-section and in the space-time dimensions, simultaneously. Notice that,

because of the definition of $\boldsymbol{\mathcal{X}}$, any aggregate, i.e., any element of $\mathcal{G}(\boldsymbol{x})$, has variance either finite strictly positive or zero. By generalizing to rf the definitions given by Forni and Lippi (2001) and Hallin and Lippi (2013), we have

DEFINITION 3.5 (Idiosyncratic and common components). We say that an infinite-dimensional rf $\boldsymbol{w}$ with elements $w_{\ell\boldsymbol{\varsigma}} \in \boldsymbol{\mathcal{X}}$ for any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$ and $\ell \in \mathbb{N}$ (i) is idiosyncratic if for any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$ and any STDAS $\{\boldsymbol{a}_n, n \in \mathbb{N}\}$, $\lim_{n\to\infty} \underline{\boldsymbol{a}}_n(L)\boldsymbol{w}_{n\boldsymbol{\varsigma}} = 0$ in mean-square; (ii) is common if it is not idiosyncratic, i.e., if there exist $\boldsymbol{\varsigma} \in \mathbb{Z}^3$ and STDAS $\{\boldsymbol{a}_n, n \in \mathbb{N}\}$, $\lim_{n\to\infty} \underline{\boldsymbol{a}}_n(L)\boldsymbol{w}_{n\boldsymbol{\varsigma}} = y^a_{\boldsymbol{\varsigma}}$ in mean-square such that $y^a_{\boldsymbol{\varsigma}} \in \boldsymbol{\mathcal{X}}$ and $0 < \text{Var}(y^a_{\boldsymbol{\varsigma}}) < \infty$.

Hereafter, we also refer to the components w_ℓ of $\boldsymbol{w}$ as idiosyncratic or common if $\boldsymbol{w}$ is idiosyncratic or common, respectively. Note that if $\boldsymbol{w}$ is idiosyncratic then $\mathcal{G}(\boldsymbol{w}) = \{0\}$, that is it contains only the zero element. In contrast, if $\boldsymbol{w}$ is common, by aggregating it we get a rf with finite and strictly positive variance, in other words, $\mathcal{G}(\boldsymbol{w})$ contains only non-degenerate rf. This yields

DEFINITION 3.6 (Common factors). Given an n-dimensional rf $\boldsymbol{x}_n$ with elements $x_{\ell\boldsymbol{\varsigma}} \in \boldsymbol{\mathcal{X}}$ for any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$ and $\ell \leq n$, we say that a scalar rf w is a common factor if there exists a STDAS $\{\boldsymbol{a}_n, n \in \mathbb{N}\}$ such that $w_{\boldsymbol{\varsigma}} = \lim_{n\to\infty} \underline{\boldsymbol{a}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}$ in mean-square, $w_{\boldsymbol{\varsigma}} \in \boldsymbol{\mathcal{X}}$, and $0 < \text{Var}(w_{\boldsymbol{\varsigma}}) < \infty$.

Clearly, by comparing Definition 3.6 with Definition 3.5(ii) we see that the common factors are elements of the aggregation space of the common components. Last, since $\mathcal{G}(\boldsymbol{x})$ in Definition 3.4 is a closed subspace of $\boldsymbol{\mathcal{X}}$, we provide the following

DEFINITION 3.7 (Canonical decomposition). For any $\ell \in \mathbb{N}$ and any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$, the orthogonal decomposition: $x_{\ell\boldsymbol{\varsigma}} = \text{proj}(x_{\ell\boldsymbol{\varsigma}}|\mathcal{G}(\boldsymbol{x})) + \delta_{\ell\boldsymbol{\varsigma}}$ is called the canonical decomposition of the rf $x_{\ell\boldsymbol{\varsigma}}$.

4. The General Spatio-Temporal Factor Model as a representation result. We show that any n dimensional rf $\boldsymbol{x}_n$, satisfying Assumptions 3.1-3.3, can be summarized by its projection, $\boldsymbol{\chi}_n$, on a q-dimensional sub-space generated by q cross-sectional and spatio-temporal aggregations of the components of $\boldsymbol{x}_n$, where q is a given finite positive integer independent of n. The rf $\boldsymbol{\chi}_n$ is such that as $n \to \infty$ it survives under cross-sectional and spatio-temporal aggregation, i.e., it converges in mean-square to a nonzero finite-variance rf. The rf $\boldsymbol{\xi}_n = \boldsymbol{x}_n - \boldsymbol{\chi}_n$ instead vanishes under cross-sectional and space-time aggregation as $n \to \infty$. Hence, according to Definition 3.5, the elements $\boldsymbol{\chi}_n$ are common rf and the elements of $\boldsymbol{\xi}_n$ are idiosyncratic. Intuitively, the distinct asymptotic behavior of the two components under aggregation means that if any pervasive signal is present in $\boldsymbol{x}_n$, an aggregation operation should help recovering it as $n \to \infty$ and the signal will appear in the elements of $\boldsymbol{\chi}_n$.

Denote the linear sub-space generated by all components of an idiosyncratic rf (which are scalars) as $\boldsymbol{\mathcal{X}}^{\text{idio}} \subseteq \boldsymbol{\mathcal{X}}$ and the linear sub-space generated by all components of a common rf as $\boldsymbol{\mathcal{X}}^{\text{com}} \subseteq \boldsymbol{\mathcal{X}}$. Then, by Definition 3.5, we have ($\oplus$ denotes the direct sum of two sub-spaces)

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}^{\text{com}} \oplus \boldsymbol{\mathcal{X}}^{\text{idio}}. \tag{9}$$

Therefore, any element of a stationary rf $\boldsymbol{x}_n$ satisfying Assumption 3.1 can be decomposed as the sum of its linear projection onto the the common sub-space, $\boldsymbol{\mathcal{X}}^{\text{com}}$, plus its linear projection onto the orthogonal complement, $\boldsymbol{\mathcal{X}}^{\text{idio}}$. This is the essence of our dynamic spatio-temporal factor decomposition, see Hallin and Lippi (2013) in the GDFM case.

If we now assume the existence of a spectral density matrix as required by Assumption 3.2, and of a finite number of factors q independent of n, we can further characterize the decomposition implied by (9) by means of the following

DEFINITION 4.1 (q-General Spatio-Temporal Factor Model). Let q be a non-negative integer. We say that the rf $\boldsymbol{x} = \{x_{\ell\boldsymbol{\varsigma}}, \ell \in \mathbb{N}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ with $x_{\ell\boldsymbol{\varsigma}} \in L_2(\mathcal{P}, \mathbb{C})$ is represented as a q-General Spatio-Temporal Factor Model (q-GSTFM) if $L_2(\mathcal{P}, \mathbb{C})$ contains: (a) an orthonormal q-dimensional white noise rf $\boldsymbol{u} = \{\boldsymbol{u}_{\boldsymbol{\varsigma}} = (u_{1\boldsymbol{\varsigma}} \cdots u_{q\boldsymbol{\varsigma}})^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$; (b) an infinite-dimensional rf $\boldsymbol{\xi} = \{\xi_{\ell\boldsymbol{\varsigma}}, \ell \in \mathbb{N}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$; both fulfilling Assumptions 3.1 and 3.2 and such that:
(i) for any $\ell \in \mathbb{N}$ and any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$

$$x_{\ell\boldsymbol{\varsigma}} = \chi_{\ell\boldsymbol{\varsigma}} + \xi_{\ell\boldsymbol{\varsigma}}, \tag{10}$$

$$\chi_{\ell\boldsymbol{\varsigma}} = \underline{\boldsymbol{b}}_\ell(L)\boldsymbol{u}_{\boldsymbol{\varsigma}} = \sum_{\boldsymbol{\kappa}} \sum_{j=1}^{q} \mathrm{b}_{\ell j,\boldsymbol{\kappa}} u_{j,\boldsymbol{\varsigma}-\boldsymbol{\kappa}}, \tag{11}$$

thus defining an infinite-dimensional rf $\boldsymbol{\chi} = \{\chi_{\ell\boldsymbol{\varsigma}}, \ell \in \mathbb{N}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$;
(ii) letting $b_{\ell j}(\boldsymbol{\theta}) = \sum_{\boldsymbol{\kappa}} \mathrm{b}_{\ell j,\boldsymbol{\kappa}} e^{-i\langle \boldsymbol{\kappa}, \boldsymbol{\theta}\rangle}$, $j = 1, \ldots, q$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and $\boldsymbol{b}_\ell(\boldsymbol{\theta}) = (b_{\ell 1}(\boldsymbol{\theta}) \cdots b_{\ell q}(\boldsymbol{\theta}))$, it holds that $\|\boldsymbol{b}_\ell\|^2 = \frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} \boldsymbol{b}_\ell(\boldsymbol{\theta}) \boldsymbol{b}_\ell^\dagger(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} < \infty$;
(iii) for any $\ell \in \mathbb{N}$, $j = 1, \ldots, q$, and $\boldsymbol{\varsigma}, \boldsymbol{\varsigma}' \in \mathbb{Z}^3$ such that $\boldsymbol{\varsigma} \neq \boldsymbol{\varsigma}'$, it holds that $\mathrm{E}[\xi_{\ell\boldsymbol{\varsigma}} u_{j\boldsymbol{\varsigma}'}] = 0$.
Furthermore, for any $n \in \mathbb{N}$ consider the n-dimensional sub-processes $\boldsymbol{\chi}_n = \{\boldsymbol{\chi}_{n\boldsymbol{\varsigma}} = (\chi_{1\boldsymbol{\varsigma}} \cdots \chi_{n\boldsymbol{\varsigma}})^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ and $\boldsymbol{\xi}_n = \{\boldsymbol{\xi}_{n\boldsymbol{\varsigma}} = (\xi_{1\boldsymbol{\varsigma}} \cdots \xi_{n\boldsymbol{\varsigma}})^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$, with j-th largest dynamic spatio-temporal eigenvalues $\lambda^\chi_{nj}(\boldsymbol{\theta})$ and $\lambda^\xi_{nj}(\boldsymbol{\theta})$, respectively, then
(iv) $\inf\{M : \mathcal{L}[\boldsymbol{\theta} : \lim_{n\to\infty} \lambda^\xi_{n1}(\boldsymbol{\theta}) > M] = 0\} < \infty$;
(v) $\lim_{n\to\infty} \lambda^\chi_{nq}(\boldsymbol{\theta}) = \infty$, $\mathcal{L}$-a.e. in $\boldsymbol{\Theta}$.

Hereafter, we refer to the infinite-dimensional rf $\boldsymbol{\chi}$ and $\boldsymbol{\xi}$ as the common and idiosyncratic components of the representation (10). Indeed, part (iv) implies that $\boldsymbol{\xi}$ is idiosyncratic, since, as proved in Proposition B.7 in Appendix B.2, a rf is idiosyncratic if and only if its dynamic spatio-temporal eigenvalues are essentially bounded functions. Moreover, since by part (iii) $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$ have orthogonal elements, then $\boldsymbol{\chi}$ cannot be idiosyncratic, hence it is common. Since the elements of $\boldsymbol{\chi}$ and $\boldsymbol{\xi}$ belong to $\boldsymbol{\mathcal{X}}^{\text{com}}$ and $\boldsymbol{\mathcal{X}}^{\text{idio}}$, respectively, the q-GSTFM representation of Definition 4.1 and (9) yield equivalent decompositions.

There are essentially two ways to obtain a q-GSTFM. On the one hand, one may assume that the rf spatio-temporal dynamics can be modeled as in (10)-(11), mimicking the approach in Forni et al. (2000). On the other hand, one may find a set of very mild assumptions such that a spatio-temporal rf can be represented as in (10)-(11), extending to the rf setting the results of Forni and Lippi (2001). In what follows, we consider the latter approach, which is more general and powerful than the former one: indeed, (10)-(11) is a representation which holds under Assumptions 3.1-3.3, only, and it is not a model imposed by the statistician. The main result of this section is the following

THEOREM 4.1. *Under Assumptions 3.1-3.3, the rf $\boldsymbol{x}$ is represented as a q-GSTFM as in Definition (4.1) if and only if (i) $\inf\{M : \mathcal{L}[\boldsymbol{\theta} : \lim_{n\to\infty} \lambda^x_{nq+1}(\boldsymbol{\theta}) > M] = 0\} < \infty$; (ii) $\lim_{n\to\infty} \lambda^x_{nq}(\boldsymbol{\theta}) = \infty$, $\mathcal{L}$-a.e. in $\boldsymbol{\Theta}$.*

It follows that the presence of an eigen-gap in the dynamic spatio-temporal eigenvalues of the infinite-dimensional rf $\boldsymbol{x}$ is a necessary and sufficient condition for the q-GSTFM to hold. To this end no assumption is needed other than Assumptions 3.1-3.3, which are very mild. Notice also that the case $q = 0$ is possible, in which case $\boldsymbol{x}$ is purely idiosyncratic.

An immediate consequence of Theorem 4.1 is the following

COROLLARY 4.2. *If $\boldsymbol{x}$ is represented as a q-GSTFM as in Definition (4.1), then $\overline{\text{span}}(\boldsymbol{\chi}) = \overline{\text{span}}(\boldsymbol{u}) = \mathcal{G}(\boldsymbol{x})$ and $\chi_{\ell\boldsymbol{\varsigma}} = \text{proj}(x_{\ell\boldsymbol{\varsigma}}|\mathcal{G}(\boldsymbol{x}))$. Moreover, the number of factors q, the common component $\boldsymbol{\chi}$, and the idiosyncratic component $\boldsymbol{\xi}$, are uniquely identified.*

From Corollary 4.2 and Definitions 3.6 and 3.7, it follows that $\boldsymbol{u}$ is a vector of common factors, and the factor decomposition implied by (9) is equivalent to the canonical decomposition. Uniqueness of the latter implies uniqueness of the q-GSTFM. In other words the q-GSTFM is identified. It has to be stressed though that, since the definition of common and idiosyncratic components hold in the limit as $n \to \infty$ (see Definition 3.5), identification is achieved only asymptotically. Indeed, as shown later, if n is fixed no consistency result can be derived when we estimate the model.

The q-GSTFM in Definition 4.1 has two distinctive properties. First, differently from the GDFM, the common component in (11) accounts for the spatio-temporal dependence. The q-dimensional rf of factors $\boldsymbol{u}$ is loaded by each element of $\boldsymbol{x}$ dynamically in time (possibly in a causal way, see Remark 6.3) and in space, since the filters depend on both dimensions. This means that, being a rf, a common shock to $\boldsymbol{x}$ can impact different points in space heterogeneously at the same time and at different points in time and it can impact also the variables observed in a given point in space at different times. Second, we do not impose any specific structure on the second moment of the of vector of idiosyncratic components $\boldsymbol{\xi}$ whose elements can be both cross-sectionally and spatio-temporally cross-auto-correlated, as long as part (iv) is satisfied.

We conclude with a series of remarks about Definition 4.1 and Theorem 4.1.

REMARK 4.1. Because of Remark 3.1, the function $\lim_{n\to\infty} \lambda^{\chi}_{nq}$ is the q-largest dynamic spatio-temporal eigenvalue of the infinite-dimensional rf $\boldsymbol{\chi}$, and, similarly the function $\lim_{n\to\infty} \lambda^{\xi}_{n1}$ is the largest dynamic spatio-temporal eigenvalue of the infinite-dimensional rf $\boldsymbol{\xi}$. Now, by part (v) of Definition 4.1 the former is to be intended as an extended function in the sense that its value is infinite but measurable (Royden and Fitzpatrick, 1988, p. 55), while, by part (iv) the latter is instead an essentially bounded function (Rudin, 1987, p. 66). It follows also that there exists a finite $C > 0$ independent of $\boldsymbol{\theta}$ such that: $\lim_{n\to\infty} \lambda^{\xi}_{n1}(\boldsymbol{\theta}) \leq C$, $\mathcal{L}$-a.e. in $\boldsymbol{\Theta}$.

Moreover, by the monotone convergence theorem, Remark 3.1, and part (iv) in Definition 4.1, we have

$$\lim_{n\to\infty} \int_{\boldsymbol{\Theta}} \lambda^{\xi}_{n1}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} \lim_{n\to\infty} \lambda^{\xi}_{n1}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \leq C.$$

This, in turn implies that the idiosyncratic covariance matrix $\boldsymbol{\Gamma}^{\xi}_{n}(\mathbf{0}) = \mathrm{E}(\boldsymbol{\xi}_{n\boldsymbol{\varsigma}} \boldsymbol{\xi}^{\dagger}_{n\boldsymbol{\varsigma}})$ has largest eigenvalue μ^{ξ}_{n1} such that

$$\lim_{n\to\infty} \mu^{\xi}_{n1} = \lim_{n\to\infty} \max_{\boldsymbol{b}:\boldsymbol{b}^{\top}\boldsymbol{b}=1} \boldsymbol{b}^{\top} \left(\int_{\boldsymbol{\Theta}} \boldsymbol{\Sigma}^{\xi}_{n}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \right) \boldsymbol{b} \leq \lim_{n\to\infty} \int_{\boldsymbol{\Theta}} \lambda^{\xi}_{n1}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \leq C.$$

The latter condition is the usual assumption made in the vector static factor model literature to characterize an idiosyncratic component (see, e.g., Forni et al., 2009). Notice, however, that (v) in Definition 4.1 in general does not imply that the common covariance matrix $\boldsymbol{\Gamma}^{\chi}_{n}(\mathbf{0}) = \mathrm{E}(\boldsymbol{\chi}_{n\boldsymbol{\varsigma}} \boldsymbol{\chi}^{\dagger}_{n\boldsymbol{\varsigma}})$ has eigenvalues diverging as $n \to \infty$, for the effect of common factors might be just lagged and not contemporaneous, in which case only the products $\boldsymbol{\Gamma}^{\chi}_{n}(\boldsymbol{\kappa}) \boldsymbol{\Gamma}^{\chi}_{n}(\boldsymbol{\kappa})$ for $\boldsymbol{\kappa} \neq \mathbf{0}$ will display diverging eigenvalues. This aspect has been studied by Lam and Yao (2012) in the vector case.

REMARK 4.2. A sufficient condition for part (ii) in Definition 4.1 to hold is to ask for square summability of the coefficients of the linear filter $\underline{\boldsymbol{b}}_{\ell}(L)$, i.e., to assume $\sum_{\boldsymbol{\kappa}} |\mathrm{b}_{\ell j,\boldsymbol{\kappa}}|^2 \leq C$ for some finite $C > 0$ independent of j. Indeed, by definition

$$\|\boldsymbol{b}_{\ell}\|^2 = \frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} \sum_{j=1}^{q} \sum_{\boldsymbol{\kappa}} |\mathrm{b}_{\ell j,\boldsymbol{\kappa}}|^2 \mathrm{d}\boldsymbol{\theta} \leq \max_{j=1,\ldots,q} \sum_{\boldsymbol{\kappa}} |\mathrm{b}_{\ell j,\boldsymbol{\kappa}}|^2 \leq C.$$

REMARK 4.3. Theorem 4.1 characterizes the class of rf which admit the q-GSTFM in Definition 4.1. In general, neither the q-dimensional orthonormal white noise rf $\boldsymbol{u}$ nor the filters $\underline{\boldsymbol{b}}_\ell(L)$ in (11) are identified. Indeed, if (10) and (11) hold, then infinitely many other equivalent representations of $\chi_{\ell\boldsymbol{\varsigma}}$ are obtained by setting $\chi_{\ell\boldsymbol{\varsigma}} = \underline{\boldsymbol{m}}_\ell(L)\boldsymbol{z}_{\ell\boldsymbol{\varsigma}}$ for a q-dimensional rf $\boldsymbol{z}$ such that $\boldsymbol{z}_{\ell\boldsymbol{\varsigma}} = \underline{\boldsymbol{D}}(L)\boldsymbol{u}_{\ell\boldsymbol{\varsigma}}$, $\boldsymbol{m}_\ell(\boldsymbol{\theta}) = \boldsymbol{b}_\ell(\boldsymbol{\theta})\boldsymbol{D}^\dagger(\boldsymbol{\theta})$, with $\boldsymbol{D}(\boldsymbol{\theta})$ which is $q \times q$ and such that $\|\boldsymbol{D}\|^2 = \frac{1}{8\pi^3}\int_{\boldsymbol{\Theta}} \boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{D}^\dagger(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} < \infty$ and $\boldsymbol{D}^\dagger(\boldsymbol{\theta})\boldsymbol{D}(\boldsymbol{\theta}) = \mathbf{I}_q$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. It follows that $\boldsymbol{z}$ is also a q-dimensional orthonormal white noise rf. In fact, if $\boldsymbol{D}^\dagger(\boldsymbol{\theta})$ were not orthogonal, as assumed, but just invertible, we could still find equivalent representations of $\chi_{\ell\boldsymbol{\varsigma}}$ where, however, $\boldsymbol{z}$ would no more be a white noise rf, but it is a rf autocorrelated in both the spatial and time dimension.

REMARK 4.4. The proof of Theorem 4.1 is rather technical and lengthy and is given in Appendix B. The necessary condition part ("only if") is is easy to prove (see Appendix B.3). Indeed, by Weyl's inequality (see Appendix B.1), it is straightforward to see that if (iv) and (v) in Definition 4.1 hold then (i) and (ii) in Theorem 4.1 hold. The sufficient condition part ("if") is more difficult to prove and it based on a series of intermediate results. In a nutshell, in the proof we proceed by first constructing a q-dimensional orthonormal white noise vector rf, $\boldsymbol{z}$, say (see Proposition B.5). Then, we show $\overline{\text{span}}(\boldsymbol{z}) = \mathcal{G}(\boldsymbol{x})$ (see Proposition B.6). It follows that the canonical projection $x_{\ell\boldsymbol{\varsigma}} = \text{proj}(x_{\ell\boldsymbol{\varsigma}}|\mathcal{G}(\boldsymbol{x})) + \delta_{\ell\boldsymbol{\varsigma}}$, is such that $\delta_{\ell\boldsymbol{\varsigma}}$ is idiosyncratic (Propositions B.7 and B.8), hence $\text{proj}(x_{\ell\boldsymbol{\varsigma}}|\mathcal{G}(\boldsymbol{x}))$, being orthogonal to $\delta_{\ell\boldsymbol{\varsigma}}$, is common. The proof is completed by means of the arguments in the following remark on the identifiability of the white noise.

5. Recovering the common component - Population results. From Corollary 4.2, it follows that, given an observed n-dimensional rf $\boldsymbol{x}_n$, common factors are obtained as limits as $n \to \infty$ of aggregates of $\boldsymbol{x}_n$ and the common component is obtained by projecting $\boldsymbol{x}_n$ onto such factors. Indeed, projecting onto the aggregation space of $\boldsymbol{x}$ or onto the aggregation space of the common component is equivalent, since the aggregation space of the idiosyncratic component contains only the zero element.

In this section we prove that among all possible q dimensional aggregates which we can project $\boldsymbol{x}_n$ on, the first q-dynamic spatio-temporal principal components, obtained from the spectral density matrix $\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta})$ (see Definition 5.1 below), are the optimal ones in the sense that they are those with largest variance, hence satisfying a discrete version of the optimality property of the Karhunen-Loève expansion (see, e.g, Hallin, Hörmann and Lippi, 2018, for the GDFM). In Theorem 5.1 we prove that by projecting $\boldsymbol{x}_n$ onto such aggregates we can recover the common component $\boldsymbol{\chi}_n$ in the limit $n \to \infty$.

The canonical decomposition in Definition 3.7 is optimal in the sense that, by definition of linear projection, it minimizes the variance of the residual idiosyncratic term. However, to achieve such decomposition in practice we need to define a basis for the space of aggregates $\mathcal{G}(\boldsymbol{x})$ to project an the elements of $\boldsymbol{x}$ onto. Therefore, given a q-GSTFM in Definition 4.1 all we need to do is to find a q-dimensional rf of common factors, which, because of Definition 3.6 belong to $\mathcal{G}(\boldsymbol{x})$, have finite and strictly positive variance. Moreover, we shall require this q-dimensional rf of factors to be an orthonormal white noise rf.

The definition of common factors holds asymptotically, but in practice we deal with a given fixed n, then, for such given n and any $j = 1, \dots, q$, we should look for those weights $\{\boldsymbol{\alpha}_{nj\boldsymbol{\kappa}}, \boldsymbol{\kappa} \in \mathbb{Z}^3\}$, such that $\underline{\boldsymbol{\alpha}}_{nj}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \sum_{\boldsymbol{\kappa}} \boldsymbol{\alpha}_{nj\boldsymbol{\kappa}}\boldsymbol{x}_{n,\boldsymbol{\varsigma}-\boldsymbol{\kappa}}$, has maximum variance. In view of the canonical isomorphism in (6), we shall then consider the equivalent maximization problem in the frequency domain, i.e., for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we shall solve:

$$\max_{\boldsymbol{\alpha}_{nj}} \boldsymbol{\alpha}_{nj}(\boldsymbol{\theta})\boldsymbol{\Sigma}_n^x(\boldsymbol{\theta})\boldsymbol{\alpha}_{nj}^\dagger(\boldsymbol{\theta}) \text{ s.t. } \|\boldsymbol{\alpha}_{nj}\| = 1, \quad \boldsymbol{\alpha}_{nj}\boldsymbol{\alpha}_{nk}^\dagger = 0, \quad j,k = 1,\dots,q, \ j \neq k. \tag{12}$$

Notice that the objective function is the variance of the discrete Fourier transform of $\underline{\boldsymbol{\alpha}}_{nj}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}$. For any given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the solution of (12) is clearly given by the eigenvector $\boldsymbol{p}^x_{nj}(\boldsymbol{\theta})$ of the spectral density matrix $\boldsymbol{\Sigma}^x_n(\boldsymbol{\theta})$ corresponding to the j-th largest eigenvalue $\lambda^x_{nj}(\boldsymbol{\theta})$, see Definition 3.3. For any $j = 1, \dots, q$, to this solution corresponds a scalar filtered rf $\{\underline{\boldsymbol{p}}^x_{nj}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ which has spectral density $\lambda^x_{nj}(\boldsymbol{\theta})$ and variance $\int_{\boldsymbol{\Theta}} \lambda^x_{nj}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$. So the first, $j = 1$, dynamic spatio-temporal principal component has largest variance as expected. Moreover, these rf are orthogonal contemporaneously and at any spatio-temporal shift, indeed, for $j \neq j'$, we have $\mathrm{E}[(\underline{\boldsymbol{p}}^x_{nj}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}})(\underline{\boldsymbol{p}}^x_{nj'}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}'})^\dagger] = 0$ for all $\boldsymbol{\varsigma}, \boldsymbol{\varsigma}' \in \mathbb{Z}^3$. Notice that if $\mathrm{E}(\boldsymbol{x}_{n\boldsymbol{\varsigma}}) = \boldsymbol{\mu}_n \neq \mathbf{0}$ then the filtered process should be defined as $\underline{\boldsymbol{p}}^x_{nj}(L)(\boldsymbol{x}_{n\boldsymbol{\varsigma}} - \boldsymbol{\mu}_n)$, hence, they always have zero-mean.

However, the q filtered processes defined by solving (12) cannot be directly used as a basis for $\mathcal{G}(\boldsymbol{x})$ for two reasons. First, they are not white noise rf. Second, and most importantly, as $n \to \infty$, their variance is not finite, indeed, under a q-GSTFM, we know that $\lim_{n\to\infty} \lambda^x_{nj}(\boldsymbol{\theta}) = \infty$ for all $j = 1, \dots, q$. Therefore, we need to rescale and whiten those rf. This is accomplished by means of the following (recall the notation in (8)).

DEFINITION 5.1 (Normalized dynamic spatio-temporal principal components). For any $n \in \mathbb{N}$ and $\ell \leq n$, the filtered rf processes

$$\psi^n_{\ell\boldsymbol{\varsigma}} = [\underline{\lambda}^x_{n\ell}(L)]^{-1/2} \star \underline{\boldsymbol{p}}^x_{n\ell}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \left\{ \sum_{\boldsymbol{\kappa}} \left[\frac{1}{8\pi^3} \int_{\boldsymbol{\Theta}} e^{i\langle \boldsymbol{\kappa}, \boldsymbol{\theta} \rangle} [\lambda^x_{n\ell}(\boldsymbol{\theta})]^{-1/2} \boldsymbol{p}^x_{n\ell}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \right] L^{\boldsymbol{\kappa}} \right\} \boldsymbol{x}_{n\boldsymbol{\varsigma}}$$

form a set of normalized dynamic spatio-temporal principal components associated with $\boldsymbol{x}_{n\boldsymbol{\varsigma}}$.

Notice that this definition makes sense since Assumption 3.3 implies that $[\lambda^x_{n\ell}(\boldsymbol{\theta})]^{-1}$ is finite for any $n \in \mathbb{N}$ and $\ell \leq n$. Now, for any $n \in \mathbb{N}$, define the rf $\boldsymbol{\psi}^n = \{\boldsymbol{\psi}^n_{\boldsymbol{\varsigma}} = \left(\psi^n_{1\boldsymbol{\varsigma}} \cdots \psi^n_{q\boldsymbol{\varsigma}}\right)^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$. Then, from Definition 5.1 we have:

$$\boldsymbol{\psi}^n_{\boldsymbol{\varsigma}} = [\underline{\boldsymbol{\Lambda}}_n(L)]^{-1/2} \star \underline{\boldsymbol{P}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \tag{13}$$

where the linear spatio-temporal filters $\underline{\boldsymbol{\Lambda}}_n(L)$ and $\underline{\boldsymbol{P}}_n(L)$ are, respectively, obtained from the $q \times q$ diagonal matrix $\boldsymbol{\Lambda}_n(\boldsymbol{\theta})$ having as entries the dynamic spatio-temporal eigenvalues $\lambda^x_{nj}(\boldsymbol{\theta})$, for $j = 1, \dots, q$, and the $q \times n$ matrix $\boldsymbol{P}_n(\boldsymbol{\theta}) = (\boldsymbol{p}^{x\top}_{n1}(\boldsymbol{\theta}) \cdots \boldsymbol{p}^{x\top}_{nq}(\boldsymbol{\theta}))^\top$ having as rows the q corresponding dynamic spatio-temporal eigenvectors. Now, let $\boldsymbol{\Phi}_n(\boldsymbol{\theta})$ be the $n - q \times n - q$ diagonal matrix having as entries the dynamic spatio-temporal eigenvalues $\lambda^x_{nj}(\boldsymbol{\theta})$, for $j = q + 1, \dots, n$, and let $\boldsymbol{Q}_n(\boldsymbol{\theta})$ be the $n - q \times n$ matrix having as rows the $n - q$ corresponding dynamic spatio-temporal eigenvectors. Then, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\Sigma}^x_n(\boldsymbol{\theta}) = \boldsymbol{P}^\dagger_n(\boldsymbol{\theta})\boldsymbol{\Lambda}_n(\boldsymbol{\theta})\boldsymbol{P}_n(\boldsymbol{\theta}) + \boldsymbol{Q}^\dagger_n(\boldsymbol{\theta})\boldsymbol{\Phi}_n(\boldsymbol{\theta})\boldsymbol{Q}_n(\boldsymbol{\theta})$. Therefore, since $\mathbf{I}_n = \boldsymbol{P}^\dagger_n(\boldsymbol{\theta})\boldsymbol{P}_n(\boldsymbol{\theta}) + \boldsymbol{Q}^\dagger_n(\boldsymbol{\theta})\boldsymbol{Q}_n(\boldsymbol{\theta})$, from (13) we immediately see that $\boldsymbol{\psi}^n$ has spectral density $\mathbf{I}_q$, hence it is a q-dimensional orthonormal white noise rf as required. By letting $n \to \infty$, we obtain from $\boldsymbol{\psi}^n$ the basis for $\mathcal{G}(\boldsymbol{x})$ we are looking for. This is formalized by means of the following

THEOREM 5.1. *For any $n \in \mathbb{N}$ and $\ell \leq n$, denote by $\underline{\boldsymbol{\pi}}_{n\ell}(L)$ the ℓ-th q-dimensional row of $\underline{\boldsymbol{P}}^\dagger_n(L)$. Suppose that (i) and (ii) of Theorem 4.1 and Assumptions 3.1-3.3 hold. Then, for all $\boldsymbol{\varsigma} \in \mathbb{Z}^3$, $\lim_{n\to\infty} \underline{\boldsymbol{\pi}}_{n\ell}(L) \star \underline{\boldsymbol{\Lambda}}^{1/2}_n(L)\boldsymbol{\psi}^n_{\boldsymbol{\varsigma}} = \lim_{n\to\infty} \underline{\boldsymbol{\pi}}_{n\ell}(L) \star \underline{\boldsymbol{P}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \chi_{\ell\boldsymbol{\varsigma}}$ in mean-square.*

This result provides the basis for our estimation approach. It implies that if, for a given $n \in \mathbb{N}$, we knew the spectral density matrix of $\boldsymbol{x}_n$, then, for any $\ell \leq n$ and $\boldsymbol{\varsigma} \in \mathbb{Z}^3$, an estimator of the common component would be:

$$\chi^{(n)}_{\ell\boldsymbol{\varsigma}} = \underline{\boldsymbol{\pi}}_{n\ell}(L) \star \underline{\boldsymbol{P}}_n(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \underline{\boldsymbol{K}}^x_{n\ell}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \text{ say.} \tag{14}$$

This is a consistent estimator since as $n \to \infty$ it converges in mean-square to the unobservable common component χ_ℓ. Notice that since we are dealing with projections the rescaling by means of the eigenvalues introduced in Definition 5.1 is actually not needed in practice, as we just need the eigenvectors.

REMARK 5.1. For any $n \in \mathbb{N}$, let $\boldsymbol{P}_n^{\chi}(\boldsymbol{\theta})$ be the $q \times n$ matrix having as rows the spatio-temporal dynamic eigenvectors of the spectral density matrix of the common component $\boldsymbol{\Sigma}_n^{\chi}(\boldsymbol{\theta})$. Let also $\underline{\boldsymbol{P}}_n^{\chi}(L)$ the associated linear spatio-temporal filter and for any $\ell \leq n$, denote by $\underline{\boldsymbol{\pi}}_{n\ell}^{\chi}(L)$ the ℓ-th q-dimensional row of $\underline{\boldsymbol{P}}_n^{\chi\dagger}(L)$. Then, since $\mathrm{rk}(\boldsymbol{\Sigma}_n^{\chi}(\boldsymbol{\theta})) = q$ for all $n \in \mathbb{N}$ and $\mathcal{L}$-a.e. in $\boldsymbol{\Theta}$, we immediately see that, for any $\ell \leq n$ and $\boldsymbol{\varsigma} \in \mathbb{Z}^3$, we can always write:

$$\chi_{\ell\boldsymbol{\varsigma}} = \underline{\boldsymbol{\pi}}_{n\ell}^{\chi}(L) \star \underline{\boldsymbol{P}}_n^{\chi}(L)\boldsymbol{\chi}_{n\boldsymbol{\varsigma}} = \underline{\boldsymbol{\pi}}_{n\ell}^{\chi}(L) \star \underline{\boldsymbol{P}}_n^{\chi}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}} = \underline{\boldsymbol{K}}_{n\ell}^{\chi}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}, \text{ say,} \tag{15}$$

because $\mathrm{Cov}(\boldsymbol{\chi}_{n\boldsymbol{\varsigma}}, \boldsymbol{\xi}_{n\boldsymbol{\varsigma}'}) = \mathbf{0}$ for all $\boldsymbol{\varsigma}, \boldsymbol{\varsigma}' \in \mathbb{Z}^3$. This, together with (14), implies that, as $n \to \infty$, the coefficients of $\underline{\boldsymbol{K}}_{n\ell}^{x}(L)$ converge in mean-square to the coefficients of $\underline{\boldsymbol{K}}_{n\ell}^{\chi}(L)$.

REMARK 5.2. In general, the dynamic spatio-temporal eigenvectors are complex vectors. However, for any $n \in \mathbb{N}$, we know that $\mathbf{I}_n = \boldsymbol{P}_n^{\top}(\boldsymbol{\theta})\bar{\boldsymbol{P}}_n(\boldsymbol{\theta}) + \boldsymbol{Q}_n^{\top}(\boldsymbol{\theta})\bar{\boldsymbol{Q}}_n(\boldsymbol{\theta})$, and that the spectral density matrix is Hermitian, i.e., $\bar{\boldsymbol{\Sigma}}_n^{x}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_n^{x\top}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_n^{x}(-\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}_n(\boldsymbol{\theta})$ is a real matrix. Therefore, we can always impose $\boldsymbol{p}_{n\ell}^{x}(-\boldsymbol{\theta}) = \bar{\boldsymbol{p}}_{n\ell}^{x}(\boldsymbol{\theta})$ for all $\ell \leq n$. This implies that $\int_{\boldsymbol{\Theta}} e^{i\langle\boldsymbol{\kappa},\boldsymbol{\theta}\rangle}[\lambda_{n\ell}^{x}(\boldsymbol{\theta})]^{-1/2}\boldsymbol{p}_{n\ell}^{x}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ is always a real number, and, thus, the normalized dynamic spatio-temporal principal components are real rf, see also Hallin, Hörmann and Lippi (2018).

6. Recovering the common component - Estimation.

6.1. *Estimation in practice.* The population results derived in Section 4 show that the spatio-temporal common component χ_ℓ can be recovered as $n \to \infty$ from a sequence of projections, see Theorem 5.1. The filters needed to define this projection are given in Definition 5.1 and depend on the dynamic spatio-temporal eigenvalues and eigenvectors of the unknown spectral density matrix.

Let us assume now to be given a finite n-dimensional realization $\boldsymbol{x}_n$ of the infinite-dimensional rf $\boldsymbol{x}$ over $S_1 \times S_2$ points on a 2-dimensional lattice and over T time periods. In order to proceed we need to fix the origin of the lattice, because of stationarity this can be chosen arbitrarily in any location of $\mathbb{Z}^2$. Here we adopt the convention that the point $(s_1\ s_2) = (1\ 1)$ corresponds to the South-West corner of the given lattice. Then, index s_1 grows by moving East while the index s_2 grows by moving North. With this definition of the spatial coordinates, our observations are collected into the $n \times S_1S_2T$-dimensional matrix: $\{x_{\ell\boldsymbol{\varsigma}} = x_{\ell(s_1\ s_2\ t)},\ \ell = 1,\dots,n,\ s_1 = 1,\dots,S_1,\ s_2 = 1,\dots,S_2,\ t = 1,\dots,T\}$.

If the spatio-temporal dynamic eigenvalues of $\boldsymbol{x}_n$ satisfy Theorem 4.1, then, according to the q-GSTFM, for all $\ell = 1,\dots,n$, $s_1 = 1,\dots,S_1$, $s_2 = 1,\dots,S_2$, and $t = 1,\dots,T$ we can write $x_{\ell(s_1\ s_2\ t)} = \chi_{\ell(s_1\ s_2\ t)} + \xi_{\ell(s_1\ s_2\ t)}$, where χ_ℓ is the common component and ξ_ℓ is idiosyncratic. For any given n, we denote as $\boldsymbol{\chi}_n$ and $\boldsymbol{\xi}_n$ the n-dimensional rf of the common and idiosyncratic components. Throughout this section we assume that the number of factors, q, driving the common component is known (see Section 7) and we now describe our estimation strategy.

Let $\boldsymbol{\varsigma}_1 = (s_{11}\ s_{12}\ t_1)^{\top}$ and $\boldsymbol{\varsigma}_2 = (s_{21}\ s_{22}\ t_2)^{\top}$, then an estimator of $\boldsymbol{\Sigma}_n^{x}(\boldsymbol{\theta})$ is

$$\widehat{\boldsymbol{\Sigma}}_n^{x}(\boldsymbol{\theta}) = \frac{1}{S_1S_2T}\sum_{\boldsymbol{\varsigma}_1,\boldsymbol{\varsigma}_2=(1\ 1\ 1)^{\top}}^{(S_1\ S_2\ T)^{\top}} \boldsymbol{x}_{n\boldsymbol{\varsigma}_1}\boldsymbol{x}_{n\boldsymbol{\varsigma}_2}^{\top} K_1\left(\frac{s_{11}-s_{21}}{B_{S_1}}\right) K_2\left(\frac{s_{12}-s_{22}}{B_{S_2}}\right) K_3\left(\frac{t_1-t_2}{B_T}\right) e^{-i\langle\boldsymbol{\varsigma}_1-\boldsymbol{\varsigma}_2,\boldsymbol{\theta}\rangle}, \tag{16}$$

with $K_1(\cdot), K_2(\cdot)$, and $K_3(\cdot)$ being kernel functions and B_{S_1}, B_{S_2}, and B_T being bandwidths, whose properties are discussed later.

In agreement with the population results of Theorem 5.1, the common component is estimated by projecting $\boldsymbol{x}_n$ onto the space spanned by linear filters generated by the q leading spatio-temporal dynamic eigenvectors. For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ let us denote by $\widehat{\boldsymbol{P}}_n(\boldsymbol{\theta})$ the $q \times n$ matrix having as rows the spatio-temporal dynamic eigenvectors of $\widehat{\boldsymbol{\Sigma}}_n^x(\boldsymbol{\theta})$ and, for any $\ell = 1, \dots, n$, let $\widehat{\boldsymbol{\pi}}_{n\ell}(\boldsymbol{\theta})$ the ℓ-th q-dimensional row of $\widehat{\boldsymbol{P}}_n^\dagger(\boldsymbol{\theta})$, and, in agreement with (14) define $\widehat{\boldsymbol{K}}_{n\ell}^x(\boldsymbol{\theta}) = \widehat{\boldsymbol{\pi}}_{n\ell}(\boldsymbol{\theta})\widehat{\boldsymbol{P}}_n(\boldsymbol{\theta})$, generating the linear filter $\underline{\widehat{\boldsymbol{K}}}_{n\ell}^x(L)$. Now, since $\underline{\widehat{\boldsymbol{K}}}_{n\ell}^x(L)$ is in general infinite and two-sided, but $\boldsymbol{x}_{n\boldsymbol{\varsigma}}$ is not available for $\boldsymbol{\varsigma} < (1\ 1\ 1)$ and $\boldsymbol{\varsigma} > (S_1\ S_2\ T)$, we consider instead a truncated linear filter, whose definition depends on the space-time location $\boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top$ in correspondence of which the filter is applied to $\boldsymbol{x}_n$. Namely, we consider

(17)
$$\widehat{\underline{\boldsymbol{K}}}_{n\ell}^{x,\boldsymbol{\varsigma}}(L) = \frac{1}{8\pi^3} \sum_{\kappa_1=\underline{\kappa}_1(s_1)}^{\overline{\kappa}_1(s_1)} \sum_{\kappa_2=\underline{\kappa}_2(s_2)}^{\overline{\kappa}_2(s_2)} \sum_{\kappa_3=\underline{\kappa}_3(t)}^{\overline{\kappa}_1(t)} \left(\int_{\boldsymbol{\Theta}} \widehat{\boldsymbol{K}}_{n\ell}^x(\boldsymbol{\theta}) e^{i\langle (\kappa_1\ \kappa_2\ \kappa_3)^\top, \boldsymbol{\theta} \rangle} \mathrm{d}\boldsymbol{\theta} \right) L_1^{\kappa_1} L_2^{\kappa_2} L_3^{\kappa_3},$$

where, for some integers $M_{S_1} < S_1, M_{S_2} < S_2$, and $M_T < T$, we defined

(18)
$$\begin{aligned} \underline{\kappa}_1(s_1) &= \max\{s_1 - S_1, -M_{S_1}\}, & \overline{\kappa}_1(s_1) &= \min\{s_1 - 1, M_{S_1}\}, \\ \underline{\kappa}_2(s_2) &= \max\{s_2 - S_2, -M_{S_2}\}, & \overline{\kappa}_2(s_2) &= \min\{s_2 - 1, M_{S_2}\}, \\ \underline{\kappa}_3(t) &= \max\{t - T, -M_T\}, & \overline{\kappa}_3(t) &= \min\{t - 1, M_T\}. \end{aligned}$$

For any given $\ell = 1, \dots, n$ and any $\boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top$ such that $s_1 =, 1 \dots, S_1$, $s_2 =, 1 \dots, S_2$, and $t = 1, \dots, T$, the common component is then estimated as

(19)
$$\widehat{\chi}_{\ell\boldsymbol{\varsigma}}^{(n)} = \widehat{\underline{\boldsymbol{K}}}_{n\ell}^{x,\boldsymbol{\varsigma}}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}.$$

REMARK 6.1. In practice all estimated quantities in the frequency domain, as $\widehat{\boldsymbol{\Sigma}}_n^x(\boldsymbol{\theta})$, $\widehat{\boldsymbol{P}}_n(\boldsymbol{\theta})$, and $\widehat{\boldsymbol{K}}_{n\ell}^x(\boldsymbol{\theta})$, should be computed only for a finite number of frequencies, defined as $\boldsymbol{\theta}_{\boldsymbol{h}} = (\theta_{1,h_1}\ \theta_{2,h_2}\ \theta_{3,h_3})^\top$, with $\theta_{1,h_1} = \pi h_1/B_{S_1}$, $\theta_{2,h_2} = \pi h_2/B_{S_2}$, and $\theta_{3,h_3} = \pi h_3/B_T$, for integers $h_1 = -B_{S_1}, \dots, B_{S_1}$, $h_2 = -B_{S_2}, \dots, B_{S_2}$, and $h_3 = -B_T, \dots, B_T$. For simplicity, in this and the following sections we implicitly assume the identities $\int_{\boldsymbol{\Theta}} \mathrm{d}\boldsymbol{\theta} \equiv \sum_{|h_1| \le B_{S_1}} \sum_{|h_2| \le B_{S_2}} \sum_{|h_3| \le B_T}$, $8\pi^3 \equiv (2B_{S_1} + 1)(2B_{S_2} + 1)(2B_T + 1)$, and $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \equiv \max_{|h_1| \le B_{S_1}} \max_{|h_2| \le B_{S_2}} \max_{|h_3| \le B_T}$.

REMARK 6.2. Our approach allows us to directly estimate the common components. Differently, if one were interested in estimating the dynamic spatio-temporal principal components as given by Definition 5.1, one would need an estimator of the spatio-temporal dynamic eigenvectors matrix $\boldsymbol{P}_n(\boldsymbol{\theta})$ that is smooth with respect to $\boldsymbol{\theta}$. To this end one may consider employing the $\sqrt{2}$-*estimation* procedure, see Motta, Wu and Pourahmadi (2023).

6.2. *Assumptions.* For estimation we need to add few more assumptions. First, the GSTFM has two-sided filters as defined in (11), however, it is desirable to have one-sided filters in the time dimension. This can be obtained by imposing the following

ASSUMPTION 6.1. For any $\boldsymbol{\varsigma} \in \mathbb{Z}^3$ and $\ell \in \mathbb{N}$: (i) $\chi_{\ell\boldsymbol{\varsigma}} = \underline{\boldsymbol{c}}_\ell(L)\boldsymbol{v}_{\boldsymbol{\varsigma}} = \sum_{\kappa_1,\kappa_2 \in \mathbb{Z}} \sum_{\kappa_3=0}^{\infty} \sum_{j=1}^{q} \mathrm{c}_{\ell j,\boldsymbol{\kappa}} v_{j,\boldsymbol{\varsigma}-\boldsymbol{\kappa}}$, where $\{\boldsymbol{v}_{\boldsymbol{\varsigma}} = (v_{1\boldsymbol{\varsigma}} \cdots v_{q\boldsymbol{\varsigma}})^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ is an i.i.d. q-dimensional zero-mean orthonormal rf; (ii) $\xi_{\ell\boldsymbol{\varsigma}} = \underline{\boldsymbol{\beta}}_\ell(L)\boldsymbol{\varepsilon}_{\boldsymbol{\varsigma}} = \sum_{\kappa_1,\kappa_2} \sum_{\kappa_3=0}^{\infty} \sum_{j=1}^{\infty} \beta_{\ell j,\boldsymbol{k}} \varepsilon_{j,\boldsymbol{\varsigma}-\boldsymbol{k}}$, where $\{\boldsymbol{\varepsilon}_{\boldsymbol{\varsigma}} = (\varepsilon_{1\boldsymbol{\varsigma}}\ \varepsilon_{2\boldsymbol{\varsigma}}\ \cdots)^\top, \boldsymbol{\varsigma} \in \mathbb{Z}^3\}$ is an i.i.d. infinite-dimensional zero-mean orthonormal rf; (iii) For any $\boldsymbol{\varsigma}' \in \mathbb{Z}^3$, any $j = 1, \dots, q$, and any $i \in \mathbb{N}$, $\mathrm{Cov}(v_{j\boldsymbol{\varsigma}}, \varepsilon_{i\boldsymbol{\varsigma}'}) = 0$.

The existence of one-sided time representations in parts (i) and (ii) is a mild one. For the idiosyncratic component, our requirement is for the Wold representation to exist also for an infinite-dimensional process. For the common component, which is singular, the existence of the assumed one-sided representation has been investigated by Forni et al. (2015) in the pure time series case (see also Remark 6.3 below). Notice also that Hallin and Lippi (2013) derived an analogous of our Theorem 4.1, where only one-sided filters are used. Such approach, however, does not ensure the existence of a q-dimensional white noise rf driving the common component, and its existence is instead assumed. For the common component the two-sided representation in space is implied by the q-GSTRF in Definition 4.1, and for the idiosyncratic component we make an analogous assumption but based on an infinite-dimensional white noise rf.

By means of parts (i) and (ii) we also strengthen the conditions on the rf $\boldsymbol{v}$ and $\boldsymbol{\varepsilon}$ which are now independent along the spatio-temporal dimensions. Note that the independence assumption could be relaxed. For example we could just assume $\boldsymbol{v}$ and $\boldsymbol{\varepsilon}_n$ to be martingale differences in the time dimension so to allow for conditional heteroskedasticity in time (see, e.g., Barigozzi, Cho and Owens, 2023).

Part (iii) implies orthgonality of common and idiosyncratic components at all leads and lags consistently with the GSTFM in Definition 4.1.

REMARK 6.3. If for any fixed $n \in \mathbb{N}$ the n-dimensional vector of common components has a spectral density matrix $\boldsymbol{\Sigma}_n^{\chi}(\boldsymbol{\theta})$ which is a rational function of θ_3, then, from Rozanov (1967, Ch.1, Section 10) it follows that, for all $\ell \leq n$ and $\boldsymbol{\varsigma} \in \mathbb{Z}^3$,

$$\chi_{\ell\boldsymbol{\varsigma}} = \sum_{j=1}^{q} \frac{\underline{a}_{\ell j}(L_1, L_2, L_3)}{\underline{d}_{\ell j}(L_3)} v_{j\boldsymbol{\varsigma}} = \sum_{j=1}^{q} \sum_{\kappa_1,\kappa_2 \in \mathbb{Z}} \sum_{\kappa_3=0}^{p_1} \mathrm{a}_{\ell j,(\kappa_1\ \kappa_2\ \kappa_3)} L_1^{\kappa_1} L_2^{\kappa_2} L_3^{\kappa_3} \left[\sum_{h_3=0}^{p_2} \mathrm{d}_{\ell j,h_3} L_3^{h_3} \right]^{-1} v_{j\boldsymbol{\varsigma}},$$

for some finite positive integers p_1 and p_2, which, without loss of generality we can assume to be independent of ℓ. Moreover, $d_{\ell j}(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$, and $a_{\ell j}(z_1, z_2, z_3) \neq 0$ for all $z_3 \in \mathbb{C}$ such that $|z_3| < 1$. By defining $f_{\ell j}(\theta_3) = [d_{\ell j}(\theta_3)]^{-1} = \sum_{\kappa_3=0}^{\infty} \mathrm{f}_{\ell j,\kappa_3} e^{-i\langle \kappa_3, \theta_3 \rangle}$, it follows that

$$\chi_{\ell\boldsymbol{\varsigma}} = \sum_{j=1}^{q} \sum_{\kappa_1,\kappa_2 \in \mathbb{Z}} \sum_{\kappa_3=0}^{\infty} \sum_{m_3=0}^{p_1} \mathrm{a}_{\ell j,(\kappa_1\ \kappa_2\ m_3)} \mathrm{f}_{\ell j,\kappa_3 - m_3} L_1^{\kappa_1} L_2^{\kappa_2} L_3^{\kappa_3} v_{j\boldsymbol{\varsigma}},$$

which, by setting $\mathrm{c}_{\ell j,(\kappa_1\ \kappa_2\ \kappa_3)} = \sum_{m_3=0}^{p_1} \mathrm{a}_{\ell j,(\kappa_1\ \kappa_2\ m_3)} \mathrm{f}_{\ell j,\kappa_3 - m_3}$, coincides with Assumption 6.1(i). Thus, for all $n \in \mathbb{N}$ and all $\boldsymbol{\varsigma} \in \mathbb{Z}$, $\boldsymbol{v}_{\boldsymbol{\varsigma}} \in \overline{\mathrm{span}}(\boldsymbol{\chi}_{n\boldsymbol{\varsigma}-\boldsymbol{\kappa}}, \boldsymbol{\kappa} = (\kappa_1\ \kappa_2\ \kappa_3)^{\top}, \kappa_1, \kappa_2 \in \mathbb{Z}, \kappa_3 \geq 0)$, i.e., $\boldsymbol{v}$ is fundamental for $\boldsymbol{\chi}_n$. The generalization of this reasoning to the infinite-dimensional process $\boldsymbol{\chi}$ is considered in Forni et al. (2015, Lemma 1 and 2) in the case of pure time series, where it is shown that, under rationality of the spectral density, then fundamentalness of $\boldsymbol{v}$ is always true for any $n > q$ generically, i.e., for any value of the coefficients $c_{\ell j,\boldsymbol{\kappa}}$ such that Assumption 6.1(i) holds with the exception of a zero-measure set (see also Anderson and Deistler, 2008).

The coefficients of the representations in Assumption 6.1 are characterized by

ASSUMPTION 6.2. For all $\ell \in \mathbb{N}$, $j = 1, \ldots, q$, and $\boldsymbol{\kappa} = (\kappa_1\ \kappa_2\ \kappa_3)^{\top} \in \mathbb{Z}^2 \times \mathbb{N}_0$: (i) $|c_{\ell j,\boldsymbol{\kappa}}| \leq A_{\ell j}^{\chi} \rho_1^{\chi|\kappa_1|} \rho_2^{\chi|\kappa_2|} \rho_3^{\chi\kappa_3}$, for some finite $\rho_1^{\chi}, \rho_2^{\chi}, \rho_3^{\chi} \in (0,1)$ independent of ℓ, j, and $\boldsymbol{\kappa}$, and some finite $A_{\ell j}^{\chi} > 0$ independent of $\boldsymbol{\kappa}$ and such that $\sum_{j=1}^{q} A_{\ell j}^{\chi} \leq A^{\chi}$, for some finite $A^{\chi} > 0$ independent of ℓ; (ii) $|\beta_{\ell j,\boldsymbol{\kappa}}| \leq A_{\ell j}^{\xi} \rho_1^{\xi|\kappa_1|} \rho_2^{\xi|\kappa_2|} \rho_3^{\xi\kappa_3}$, for some finite $\rho_1^{\xi}, \rho_2^{\xi}, \rho_3^{\xi} \in (0,1)$ independent of ℓ, j, and $\boldsymbol{\kappa}$, and some finite $A_{\ell j}^{\xi} > 0$ independent of $\boldsymbol{\kappa}$ and such that $\sum_{j=1}^{\infty} A_{\ell j}^{\xi} \leq A^{\xi}$ and $\sum_{\ell=1}^{\infty} A_{\ell j}^{\xi} \leq A^{\xi}$, for some finite $A^{\xi} > 0$ independent of ℓ and j.

This assumption implies square-summability of the coefficients of the filters, which for the common component is a sufficient condition for (ii) in Definition 4.1 to hold, see Remark 4.2. This assumption has three other important implications. First, part (ii) implies that the largest spatio-temporal dynamic eigenvalue of $\boldsymbol{\xi}_n$ satisfies (see Proposition E.1 in Appendix E)

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \lim_{n\to\infty} \lambda^{\xi}_{n1}(\boldsymbol{\theta}) \leq C, \tag{20}$$

for some finite $C > 0$. Hence, according to (i) in Theorem 4.1, $\boldsymbol{\xi}_n$ is effectively an idiosyncratic component. Second, the generic (i, j)th element of the spatio-temporal spectral density matrix has first derivatives with respect to the components of $\boldsymbol{\theta}$ which are uniformly bounded over all i, j and $\boldsymbol{\theta}$ (see Forni et al., 2017, Proposition 2, for the pure time series case). Third, in part (i) we do not require summability of the coefficients along the rows, so that the spatio-temporal dynamic eigenvalues of $\boldsymbol{\chi}_n$ can be diverging with n. Divergence of those eigenvalues is made formal by means of the following assumption which strengthens (ii) in Theorem 4.1:

ASSUMPTION 6.3. For all $j = 1, \ldots, q-1$ there exist continuous functions $\boldsymbol{\theta} \mapsto \widetilde{\omega}_j(\boldsymbol{\theta})$ and $\boldsymbol{\theta} \mapsto \underset{\sim}{\omega}_j(\boldsymbol{\theta})$ such that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ $0 < \underset{\sim}{\omega}_{j+1}(\boldsymbol{\theta}) \leq \lim_{n\to\infty} \frac{\lambda^{\chi}_{n,j+1}(\boldsymbol{\theta})}{n} \leq \widetilde{\omega}_{j+1}(\boldsymbol{\theta}) \leq \underset{\sim}{\omega}_j(\boldsymbol{\theta}) \leq \lim_{n\to\infty} \frac{\lambda^{\chi}_{nj}(\boldsymbol{\theta})}{n} \leq \widetilde{\omega}_j(\boldsymbol{\theta}) < \infty$.

The requirement of linearly diverging eigenvalues is standard in the factor model literature. It implies that here we are dealing only with factors which are pervasive for the whole cross-section, which in turn implies that the ordering of the cross-sectional units is irrelevant for estimation. We also stress the fact that we do not require distinct eigenvalues. We just need the model to be identified and this is always the case as long as we have an eigengap, diverging with n, between the qth and the $(q+1)$th dynamic spatio-temporal eigenvalues, as guaranteed by Lemma E.5 in Appendix E.

We do not make any distributional assumption and we require only the following moment conditions

ASSUMPTION 6.4. For all $j = 1, \ldots, q$ and $\ell \in \mathbb{N}$, $\max\{\mathrm{E}(|v_{h\varsigma}|^p), \mathrm{E}(|\varepsilon_{j\varsigma}|^p)\} \leq \bar{A}$, for some $p \geq 4$ and $\bar{A} > 0$ independent of j and ℓ.

Two technical assumptions are also required. First, we characterize the kernel functions and bandwidths needed to estimate the spectral density matrix and the truncation levels in (18) by the following

ASSUMPTION 6.5. (i) For any $l = 1, 2, 3$, the kernel functions $K_l : [-1, 1] \to \mathbb{R}^+$ are symmetric and bounded, and such that (a) $K_l(0) = 1$; (b) for some $\vartheta_l > 0$, $|K_l(u) - 1| = O(|u|^{\vartheta_l})$ as $u \to 0$; (c) $\int_{\mathbb{R}} K_l^2(u)\mathrm{d}u < \infty$; (d) $\sum_{h_1\in\mathbb{Z}} \sup_{|h_1-h_2|\leq 1} |K_l(h_1 u) - K_l(h_2 u)| = O(1)$ as $u \to 0$. (ii) The bandwidths are such that $c_1 S_1^{b_1} < B_{S_1} < c_2 S_1^{b_2}$, $c_1^* S_2^{b_1^*} < B_{S_2} < c_2^* S_2^{b_2^*}$, and $c_1^{**} T^{b_1^{**}} < B_T < c_2^{**} T^{b_2^{**}}$, for some $c_1, c_2, c_1^*, c_2^*, c_1^{**}, c_2^{**} > 0$ and $0 < b_1 < b_2 < 1$, $0 < b_1^* < b_2^* < 1$, $0 < b_1^{**} < b_2^{**} < 1$. (iii) $d_1 S_1^{p_1} < M_{S_1} < d_2 S_1^{p_2}$, $d_1^* S_2^{p_1^*} < M_{S_2} < d_2^* S_2^{p_2^*}$, and $d_1^{**} T^{p_1^{**}} < M_T < d_2^{**} T^{p_2^{**}}$, for some $d_1, d_2, d_1^*, d_2^*, d_1^{**}, d_2^{**} > 0$ and $0 < p_1 < p_2 < 1$, $0 < p_1^* < p_2^* < 1$, $0 < p_1^{**} < p_2^{**} < 1$.

Part (i) and (ii) are standard. Part (iii) controls the truncation of the linear filter defined in (17) and (18).

Second, we assume that the effect of the linear spatio-temporal filters $\underline{\boldsymbol{K}}^{\chi}_{n\ell}(L)$, as defined in (15), decreases geometrically.

ASSUMPTION 6.6. For any $\ell = 1, \dots, n$, let $\underline{\boldsymbol{K}}^{\chi}_{n\ell}(L) = \sum_{(\kappa_1\ \kappa_2\ \kappa_3)^\top \in \mathbb{Z}^3} \boldsymbol{K}^{\chi}_{n\ell,(\kappa_1\ \kappa_2\ \kappa_3)} L_1^{\kappa_1} L_2^{\kappa_2} L_3^{\kappa_3}$, then, $\|\boldsymbol{K}^{\chi}_{n\ell,(\kappa_1\ \kappa_2\ \kappa_3)}\| \leq C_0(1+\varepsilon_1)^{-|\kappa_1|}(1+\varepsilon_2)^{-|\kappa_2|}(1+\varepsilon_3)^{-|\kappa_3|}\|\boldsymbol{K}^{\chi}_{n\ell,(0\ 0\ 0)}\|$, for some finite $C_0, \epsilon_1, \epsilon_2, \epsilon_3 > 0$ independent of ℓ.

6.3. *Asymptotic results.* To study the asymptotic properties of the estimated spectral density matrix, we generalize to the case of spatio-temporal rf the approaches by Wu and Zaffaroni (2018) and Zhang and Wu (2021) for time series and by Deb, Pourahmadi and Wu (2017) for purely spatial models, which in turn are all are based on the notion of functional dependence originally proposed by Wu (2005) in a univariate time series context. The resulting estimation theory is available in Appendix D and represents a novel contribution to the literature on the inference for spatio-temporal rf.

Letting $\widehat{\sigma}^x_{ij}(\boldsymbol{\theta})$ be the (i,j)-th entry of the estimator $\widehat{\boldsymbol{\Sigma}}^x_n(\boldsymbol{\theta})$, defined in (16), we prove the following

THEOREM 6.1. *Let Assumptions 3.1, 3.2, 6.1, 6.2, 6.4, and 6.5 hold. Define*

$$\alpha_{S_1,S_2,T} = \max\left\{(\log B_{S_1} \log B_{S_2} \log B_T)^2 \frac{B_{S_1} B_{S_2} B_T}{S_1 S_2 T}, \frac{1}{B_{S_1}^{2\vartheta_1}}, \frac{1}{B_{S_2}^{2\vartheta_2}}, \frac{1}{B_T^{2\vartheta_3}}\right\},$$

where ϑ_1, ϑ_2, and ϑ_3 are defined in Assumption 6.5. Then, there exists a finite $C > 0$ independent of n, S_1, S_2 and T, such that

$$\max_{1\leq i,j\leq n} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathrm{E}\left|\widehat{\sigma}^x_{ij}(\boldsymbol{\theta}) - \sigma^x_{ij}(\boldsymbol{\theta})\right|^2 \leq C\alpha_{S_1,S_2,T}.$$

Our results are nonstandard in the literature on spatial statistics: we do not need to choose between in-fill or long-span asymptotic regime and we simply require that both S_1 and S_2 diverge, so $S \to \infty$. With this regard, we emphasize that our estimator of the spectral density matrix entries as in (16) bears some similarities with the tapered estimator of the Fourier transform of the covariance matrix of a spatial rf on a lattice proposed by Dahlhaus and Künsch (1987). Differently from their method, in our approach we replace data tapers with kernels. This yields a two-fold advantage: first, it allows to control for the estimation bias of $\sigma^x_{ij}(\boldsymbol{\theta})$, taking care of the boundary effects; second, it offers the possibility of using the mentioned flexible asymptotic regime. We refer to El Machkouri, Volnỳ and Wu (2013) for a related discussion; see also Deb, Pourahmadi and Wu (2017) for similar comments.

REMARK 6.4. The rate in Theorem 6.1 depends on the kernel smoothness ϑ_l, $l = 1, 2, 3$ and the bandwidths B_{S_1}, B_{S_2}, and B_T (see Assumption 6.5). Typically the same kernel is used in all dimensions, so we can assume $\vartheta_l = \vartheta_o$ for all $l = 1, 2, 3$. In our applications we used the Epanechnikov kernel for which $\vartheta_o = 2$. We consider some special cases. First, consider the case in which $S_1 \asymp S_2 \asymp T$, then, up to logarithmic terms, the optimal bandwidths are such that $B_{S_1} \asymp S_1^{3/7}$, $B_{S_2} \asymp S_2^{3/7}$, and $B_T \asymp T^{3/7}$. This implies that the optimal rate of consistency for our estimator of the spectral density matrix is $S_1^{6/7} = S_2^{6/7} = T^{6/7}$.

Second, consider the case in which $S_1 \asymp S_2$, and, given that each bandwidth must satisfy Assumption 6.5, we have two sub-cases. Either, we can set $B_{S_1} \asymp S_1^{1/3}$, $B_{S_2} \asymp S_2^{1/3}$, and $B_T \asymp T^{1-\eta}$, for some small $\eta > 0$, which would entail a consistency rate $\min(S_1^{2/3}T^{\eta/2}, S_1^{2/3}, T^{2-2\eta}) = \min(S_2^{2/3}T^{\eta/2}, S_2^{2/3}, T^{2-2\eta})$, which for large T is approximately $S_1^{2/3} = S_2^{2/3}$. Or we can set $B_{S_1} \asymp S_1^{1-\eta}$, $B_{S_2} \asymp S_2^{1-\eta}$, for some small $\eta > 0$, and $B_T \asymp T^{1/5}$, which would entail a consistency rate $\min(S_1^{\eta}T^{2/5}, S_1^{2-2\eta}, T^{2/5}) = \min(S_2^{\eta}T^{2/5}, S_2^{2-2\eta}, T^{2/5})$, which for large S_1 and S_2 is approximately $T^{2/5}$ as in a pure

time series model (Barigozzi and Farnè, 2022). In both cases, we could let η grow but at the price of slower consistency rates along the other dimensions. The case of spatial bandwidths depending also on T and/or time bandwidth depending also on S_1, S_2 is ruled out by our Assumption 6.5.

We then prove consistency of the common component estimator $\widehat{\chi}^{(n)}_{\ell\boldsymbol{\varsigma}}$ defined in (19)

THEOREM 6.2. *Let Assumptions 3.1, 3.2 and 6.1-6.6 hold. Recall the definition of $\alpha_{S_1,S_2,T}$ in Theorem 6.1. Then, there exists finite $C, C^*, \widetilde{C} > 0$ independent of n, S_1, S_2, and T, such that, for n large enough,*

(i) *for any $\boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top$ with $s_1 = 1, \dots, S_1$, $s_2 = 1, \dots, S_2$, and $t = 1, \dots, T$,*

$$\max_{1\le\ell\le n} \mathrm{E}\left|\widehat{\chi}^{(n)}_{\ell\boldsymbol{\varsigma}} - \chi_{\ell\boldsymbol{\varsigma}}\right| - b_{\boldsymbol{\varsigma}} \le C \max(n^{-1/2}, \alpha^{1/2}_{S_1,S_2,T}) M_{S_1} M_{S_2} M_T,$$

where $b_{\boldsymbol{\varsigma}} = C^*(1+\varepsilon_1)^{-\kappa_1^*(s_1)}(1+\varepsilon_2)^{-\kappa_2^*(s_2)}(1+\varepsilon_3)^{-\kappa_3^*(t)}$, with $\kappa_1^*(s_1) = \min\{|\underline{\kappa}_1(s_1) - 1|, \overline{\kappa}_1(s_1) + 1\}$, $\kappa_2^*(s_2) = \min\{|\underline{\kappa}_2(s_2) - 1|, \overline{\kappa}_2(s_2) + 1\}$ and $\kappa_3^*(t) = \min\{|\underline{\kappa}_3(t) - 1|, \overline{\kappa}_3(t) + 1\}$, and where M_{S_1}, M_{S_2}, and M_T are defined in Assumption 6.5, $\varepsilon_1, \varepsilon_2$, and ε_3 are defined in Assumption 6.6, and $\underline{\kappa}_1(s_1), \overline{\kappa}_1(s_1), \underline{\kappa}_2(s_2), \overline{\kappa}_2(s_2), \underline{\kappa}_3(t)$, and $\overline{\kappa}_3(t)$ are defined in (18).

(ii) *for any $\boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top$ with $s_1 = M_{S_1}, \dots, S_1 - M_{S_1}$ and $s_2 = M_{S_2}, \dots, S_2 - M_{S_2}$ and $t = M_T, \dots, T - M_T$,*

$$\max_{1\le\ell\le n} \mathrm{E}\left|\widehat{\chi}^{(n)}_{\ell\boldsymbol{\varsigma}} - \chi_{\ell\boldsymbol{\varsigma}}\right| \le \widetilde{C} \max(n^{-1/2}, \alpha^{1/2}_{S_1,S_2,T}) M_{S_1} M_{S_2} M_T.$$

We derive L_1 bounds which are more general than those derived for the GDFM in the purely time series case. If the number of lags M_{S_1}, M_{S_2}, and M_T used in (17) is not too large and n, S_1, S_2 and T all diverge to infinity, then, since Theorem 6.2 directly implies convergence in probability, consistency of $\widehat{\chi}^{(n)}_{\ell\boldsymbol{\varsigma}}$ readily follows. As it is clear from Theorem 5.1, we need a large n to disentangle the common and the idiosyncratic components, while from Theorem 6.1 we see that we need large S_1, S_2, and T to consistently estimate the spectral density matrix of the observed rf. We remark that part (i) characterizes the bias due to boundary effects and it yields a rate of convergence when the spatial locations and the time are close to the boundaries: to the best of our knowledge, this inference aspect has been neglected in the literature on factors models.

REMARK 6.5. The consistency rate depends on the truncation level we choose when applying the two-sided filter in (17). When considering the same setting as in Remark 6.4 where $S_1 \asymp S_2 \asymp T$, the consistency rate for the estimate spectral density is $T^{6/7}$, and assuming $M_{S_1} = M_{S_2} = M_T = M$, we need $M = o(T^{2/7})$.

7. Determining the number of factors. An essential aspect for the implementation of the GSTFM is the correct identification of the number of factors q. Theorem 4.1 provides a rough guideline for this: intuitively, one should choose the value of q such that the q-th dynamic eigenvalue should be "sufficiently large" while the $q+1$-th one should not be "small". To provide a more precise selection procedure, we define an information criterion (IC) that enables us to estimate q consistently. To this end, we propose the use of a criterion which is based on the eigenvalues, $\widehat{\lambda}^x_{nj}(\boldsymbol{\theta})$, $j = 1, \dots, n$, of $\widehat{\boldsymbol{\Sigma}}^x_n(\boldsymbol{\theta})$.

Letting $p(n, S_1, S_2, T)$ denote a penalty depending on both n and on S_1, S_2, and T, we consider the information criterion

$$\widehat{\mathrm{IC}}^{(n)}(k) = \log\left[\frac{1}{n}\sum_{j=k+1}^{n}\frac{1}{8\pi^3}\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\widehat{\lambda}^x_{nj}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\right] + k\,p(n, S_1, S_2, T),$$

and we define the estimator of the number of factors

$$\widehat{q}^{(n)} = \arg\min_{0\le k\le q_{\max}} \widehat{\mathrm{IC}}^{(n)}(k), \tag{21}$$

for some a priori chosen maximum number of factors $q_{\max}$. We assume the following standard divergence rate of the penalty

ASSUMPTION 7.1. As $n, S_1, S_2, T \to \infty$, $p(n, S_1, S_2, T) \to 0$ and

$$\min\left\{n, \frac{1}{\log B_{S_1}\log B_{S_2}\log B_T}\sqrt{\frac{S_1S_2T}{B_{S_1}B_{S_2}B_T}}, B_{S_1}^{\vartheta_1}, B_{S_2}^{\vartheta_2}, B_T^{\vartheta_3}\right\} p(n, S_1, S_2, T) \to \infty.$$

Finally, we establish consistency of $\widehat{q}^{(n)}$

THEOREM 7.1. *Let Assumptions 3.1, 3.2, 6.1-6.5, and 7.1 hold. Then, as* $n, S_1, S_2, T \to \infty$, $\mathrm{P}(\widehat{q}^{(n)} = q) \to 1$.

8. Monte Carlo experiments. Before delving into numerical studies, we summarize the estimation procedure in the following

Algorithm 1: Algorithm for estimating the common component.

Input: data $\{x_{\ell\boldsymbol{\varsigma}},\ \ell = 1,\dots,n, \boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top, s_1 = 1,\dots,S_1, s_2 = 1,\dots,S_2, t = 1,\dots,T\}$; estimated number of factors $\widehat{q}^{(n)}$ (see Algorithm 1 in Appendix G);
kernel functions $K_1(\cdot)$, $K_2(\cdot)$, and $K_3(\cdot)$;
bandwidths integers B_{S_1}, B_{S_2}, and B_T;
truncation integers M_{S_1}, M_{S_2}, and M_T.

Output: $\{\widehat{\chi}^{(n)}_{\ell\boldsymbol{\varsigma}}, \ell = 1,\dots,n,$
$\boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top, s_1 = 1,\dots,S_1, s_2 = 1,\dots,S_2, t = 1,\dots,T\}$.

1 Compute $\widehat{\boldsymbol{\Sigma}}^x_n(\boldsymbol{\theta}_{\boldsymbol{h}})$ as in (16), with $\boldsymbol{\theta}_{\boldsymbol{h}}$ as in Remark 6.1.

2 Compute the $\widehat{q}^{(n)}$ eigenvectors $\widehat{\boldsymbol{p}}^x_{nj}(\boldsymbol{\theta}_{\boldsymbol{h}}), j = 1,\dots,\widehat{q}$, of $\widehat{\boldsymbol{\Sigma}}^x_n(\boldsymbol{\theta}_{\boldsymbol{h}})$, with $\boldsymbol{\theta}_{\boldsymbol{h}}$ as in Remark 6.1.

3 Compute $\widehat{\boldsymbol{K}}^x_{n\ell}(\boldsymbol{\theta}_{\boldsymbol{h}})$ and $\underline{\widehat{\boldsymbol{K}}}^{x,\boldsymbol{\varsigma}}_{n\ell}(L)$ as in (17), with $\boldsymbol{\theta}_{\boldsymbol{h}}$ as in Remark 6.1.

4 Compute $\widehat{\chi}^{(n)}_{\ell\boldsymbol{\varsigma}} = \underline{\widehat{\boldsymbol{K}}}^{x,\boldsymbol{\varsigma}}_{n\ell}(L)\boldsymbol{x}_{n\boldsymbol{\varsigma}}$ as in (19).

We illustrate how Algorithm 1 works and we provide evidence of our key theoretical results. In Section 2 we already showed the presence of the eigen-gap in finite-samples as predicted by our results in Section 4, further evidence is available in Appendix I; in Section 8.1, we study the performance of the estimator of the common component proposed in Section 6, and we provide a comparison of our GSTFM with the extant GDFM; in Section 8.2, we explain how to select the number of factors following Section 7.

In the whole section we simulate data using $x_{\ell\boldsymbol{\varsigma}} = \chi_{\ell\boldsymbol{\varsigma}} + \xi_{\ell\boldsymbol{\varsigma}}$, for $\ell = 1,\dots,n$, $\boldsymbol{\varsigma} = (s_1\ s_2\ t)^\top$ with $s_1 = 1,\dots,S_1$, $s_2 = 1,\dots,S_2$, and $t = 1,\dots,T$. The case of cross- and serially correlated idiosyncratic components is studied in Appendix I. The idiosyncratic component $\xi_{\ell\boldsymbol{\varsigma}}$ is i.i.d. from a standard normal distribution and the common component $\chi_{\ell\boldsymbol{\varsigma}}$ is generated according to two different mechanisms.

Model (a) is an infinite convolution over the lattice:

$$\chi_{\ell\boldsymbol{\varsigma}} = \sum_{\boldsymbol{\kappa}} \sum_{j=1}^{q} a_{\ell j} b_{\ell j}^{|\kappa_1|+|\kappa_2|+|\kappa_3|} L^{\boldsymbol{\kappa}} u_{j,\boldsymbol{\varsigma}}. \tag{22}$$

Model (b) is a finite convolution over the lattice:

$$\chi_{\ell\boldsymbol{\varsigma}} = \sum_{\boldsymbol{\kappa}=(-1\ -1\ 0)^\top}^{(1\ 1\ 1)^\top} \sum_{j=1}^{q} a_{\ell j} 0.5^{|\kappa_1|+|\kappa_2|+|\kappa_3|} L^{\boldsymbol{\kappa}} u_{j,\boldsymbol{\varsigma}}. \tag{23}$$

We generate $a_{\ell j}$ and $u_{j,\boldsymbol{\varsigma}}$, $j = 1, \dots, q$, from i.i.d. standard normal distributions and $b_{\ell j}$ from i.i.d. uniform distributions on $[0.5, 0.8]$. The Monte Carlo (MC) experiments are repeated $N = 100$ times.

8.1. *The common component.* Section 6 contains the asymptotics of the proposed estimation methods. A practically relevant question is related to the finite-sample behaviour of the proposed estimators. To investigate this aspect, we set $q = 2$ and we study numerically how the mean square error (MSE)

$$E_1 = \frac{1}{nS_1S_2T} \sum_{\ell=1}^{n} \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} \sum_{t=1}^{T} (\widehat{\chi}_{\ell\boldsymbol{\varsigma}}^{(n)} - \chi_{\ell\boldsymbol{\varsigma}})^2$$

and the standardised MSE

$$E_2 = \frac{\sum_{\ell=1}^{n} \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} \sum_{t=1}^{T} (\widehat{\chi}_{\ell\boldsymbol{\varsigma}}^{(n)} - \chi_{\ell\boldsymbol{\varsigma}})^2}{\sum_{\ell=1}^{n} \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} \sum_{t=1}^{T} \chi_{\ell\boldsymbol{\varsigma}}^2}$$

change with n and with the spatio-temporal dimensions S_1, S_2 and T.

In the top panel of Table 1, we display the averaged (over all MC runs) E_1 and E_2 for $n = 20, 40, 60, 80$ and $(S_1, S_2, T) = (20, 20, 20)$. The table clearly shows that the estimation errors decrease as n increases: this illustrates the blessing of dimensionality for the estimation of the common component. Interestingly, we remark that already with $n = 20$, E_1 and E_2 have values that are very similar to the ones obtained for larger sample sizes (e.g. $n = 60$).

In the bottom panel of Table 1 we report the averaged (over all MC runs) values of E_1 and E_2 for $n = 40$ and $(S_1, S_2, T) = (10, 10, 10)d$, with $d = 1, 2, 3, 4$. In line with the theoretical results, the errors decrease as the spatio-temporal dimensions increase.

To elaborate on the motivating example of Section 2, we compare the performance of the GSTFM and the GDFM in terms of estimation accuracy of the common components. We set $n = 30$, $q = 2$ and $(S_1, S_2, T) = (10, 10, 20)$ or $(S_1, S_2, T) = (20, 10, 20)$. In Table 2, we report the average (over all MC runs) values of E_1 and E_2, for the GSTFM and GDFM. The advantage of our approach is evident: the GSTFM produces smaller estimation errors of the common components than the GDFM. We emphasize that E_1 of the GDFM displays a sharp rise as S_1 increases from 10 to 20. This aspect illustrates that there is no blessing of dimensionality for the GDFM if the spatial dependencies are ignored: adding more time series does not yield any accuracy improvement and the results of Forni et al. (2000) do not apply. Indeed, when S_1 increases, stacking the new observations in a vector, as in Section 2, implies that we are dealing with a larger number of spatially dependent variables: the GDFM ignores these spatial dependencies and, as a result, it becomes less reliable in the estimation of the common component, entailing larger values of E_1—incidentally, this point is not detectable looking at E_2 because of its standardisation based on the variance of the true common component.

TABLE 1
E_1 and E_2 of the GSTFM, $q = 2$.

	n			
$(S_1, S_2, T) = (20, 20, 20)$	20	40	60	80
Model (a) in (22)				
E_1	0.389	0.346	0.339	0.331
E_2	0.066	0.060	0.059	0.058
Model (b) in (23)				
E_1	0.251	0.193	0.175	0.164
E_2	0.047	0.036	0.031	0.030
	(S_1, S_2, T)			
$n = 40$	$(10, 10, 10)$	$(20, 20, 20)$	$(30, 30, 30)$	$(40, 40, 40)$
Model (22)				
E_1	0.372	0.289	0.302	0.301
E_2	0.077	0.051	0.050	0.046
Model (23)				
E_1	0.196	0.115	0.146	0.118
E_2	0.036	0.021	0.027	0.021

TABLE 2
E_1 and E_2 of the GSTFM and GDFM, $n = 30$, $S_2 = 10$, $T = 20$, and $q = 2$.

		Model (a) in (22)		Model (b) in (23)	
		GSTFM	GDFM	GSTFM	GDFM
$S_1 = 10$	E_1	0.606	1.632	0.566	2.773
	E_2	0.325	0.807	0.149	0.709
$S_1 = 20$	E_1	0.817	4.202	0.470	4.160
	E_2	0.150	0.747	0.085	0.731

TABLE 3
Under- and over-identification rates for $\widehat{q}_{\widehat{c}}^{(n)}$, with $q = 0, 1, 2, 3$.

		$q = 0$	$q = 1$	$q = 2$	$q = 3$
Model (a) in (22)	Under-identification	0	0	0	0
	Over-identification	0	0.10	0.08	0.04
Model (b) in (23)	Under-identification	0	0	0	0
	Over-identification	0	0	0	0

8.2. *Selection of the number of factors* . We investigate the finite sample performance of the estimator of q defined in Section 7. However, looking at (21), we remark that, if the estimator $\widehat{q}^{(n)}$ is consistent, then the estimator $\widehat{q}_c^{(n)}$ obtained via the penalty $cp(n, S_1, S_2, T)$, $c > 0$, is consistent as well. Hence, in practice, one needs to choose also c to estimate q consistently. The detailed procedure for the automatic selection of the number of factors is summarized in Algorithm 1 in Appendix G. To evaluate the estimation accuracy of Algorithm 1, we set $n = 100$, $(S_1, S_2, T) = (25, 25, 25)$, and $q = 0, 1, 2, 3$ and we run 200 MC replications. Table 3 shows the under- and over-identification proportions for $\widehat{q}_{\widehat{c}}^{(n)}$. The results illustrate good finite-sample performance of the selection procedure of q: for Model (b) in (23), the algorithm identifies q correctly for all replications and for all values of q; for Model (a) in (22), the over-identification rate is not zero for $q = 1, 2, 3$ but it is nevertheless very small.

9. Conclusions and further developments. We develop the theory and provide the complete inference toolkit (estimation of the common component and selection of the number of factors) for the factor analysis of high-dimensional spatio-temporal rf defined on a lattice. Our model accounts for all spatio-temporal common correlations among all components of the rf. We give statistical guarantees of the proposed estimation methods. Our asymptotic theory extends the one available in Forni et al. (2000), whose rates of convergence, which are unavailable in the literature on factor models for time series, can be derived as a special case of our rates in Section 6. Monte Carlo studies illustrate the applicability and the good performance of our GSTFM under many different settings, commonly encountered in data analysis.

We foresee some extensions of our results. For instance, one may define estimators of the common component which involve one-sided filters in time, thus allowing for forecasting. We conjecture that this is possible along the lines of Forni et al. (2005) or Forni et al. (2017). Moreover, one may consider the case of irregularly spaced (in time and in space) data. Given that in this case the frequencies are no longer in $\boldsymbol{\Theta}$ the main challenge is then to introduce a new notion of spectral density and its estimator; see, e.g., Deb, Pourahmadi and Wu (2017) where it is also shown that, in the case of irregularly spaced data on a lattice, the implied asymptotic properties of the related estimator are likely to be affected in a non-trivial way. These extensions are part of our ongoing research.

Funding. The third author was supported by grants WK2040000055 and YD2040002016.

REFERENCES

ALTISSIMO, F., CRISTADORO, R., FORNI, M., LIPPI, M. and VERONESE, G. (2010). New Eurocoin: Tracking economic growth in real time. *The Review of Economics and Statistics* **92** 1024–1034.

ANDERSON, B. D. O. and DEISTLER, M. (2008). Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement, and System Integration* **1** 284–292.

BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135-171.

BALTAGI, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.

BARIGOZZI, M., CHO, H. and OWENS, D. (2023). FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series. *Journal of Business & Economic Statistics* 1–13.

BARIGOZZI, M. and FARNÈ, M. (2022). An algebraic estimator for large spectral density matrices. *Journal of the American Statistical Association* 1–13.

BODELET, J. and LA VECCHIA, D. (2022). Robust sieve M-estimation with an application to dimensionality reduction. *Electronic Journal of Statistics* **16** 3996–4030.

BRILLINGER, D. R. (1970). The frequency analysis of relations between stationary spatial series. In *Proc. Twelfth Biennial Seminar Canadian Math. Congress (ed. R. Pyke). Canadian Math. Congress, Montreal* 39–81.

BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time series: theory and methods*, Second ed. Springer Science & Business Media.

CHAMBERLAIN, G. and ROTHSCHILD, M. (1983). Arbitrage, factor structure and mean– variance analysis in large asset markets,. *Econometrica* **51** 1305–1324.

CHANG, J., HE, J., YANG, L. and YAO, Q. (2023). Modelling matrix time series via a tensor CP-decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **85** 127–148.

CHEN, R., YANG, D. and ZHANG, C.-H. (2022). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association* **117** 94–116.

CHRISTENSEN, W. F. and AMEMIYA, Y. (2002). Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association* **97** 302–317.

CRESSIE, N. and WIKLE, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

CRISTADORO, R., FORNI, M., REICHLIN, L. and VERONESE, G. (2005). A core inflation indicator for the euro area. *Journal of Money, Credit and Banking* **37** 539–560.

DAHLHAUS, R. and KÜNSCH, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74** 877–882.

DEB, S., POURAHMADI, M. and WU, W. B. (2017). An asymptotic theory for spectral analysis of random fields. *Electronic Journal of Statistics* **11** 4297–4322.

EL MACHKOURI, M., VOLNỲ, D. and WU, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications* **123** 1–14.

FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **75** 603–680.

FORNI, M. and LIPPI, M. (2001). The generalized dynamic factor model: representation theory. *Econometric theory* **17** 1113–1141.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics* **82** 540–554.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American statistical association* **100** 830–840.

FORNI, M., GIANNONE, D., LIPPI, M. and REICHLIN, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory* **25** 1319–1347.

FORNI, M., HALLIN, M., LIPPI, M. and ZAFFARONI, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics* **185** 359–371.

FORNI, M., HALLIN, M., LIPPI, M. and ZAFFARONI, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics* **199** 74–92.

GERSING, P., RUST, C. and DEISTLER, M. (2024). Reconciling the Theory of Factor Sequences arXiv No. 2307.10067.

GEWEKE, J. (1977). The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models 1* (D. J. Aigner and A. S. Goldberger, eds.) 365–383. North-Holland.

HALLIN, M., HÖRMANN, S. and LIPPI, M. (2018). Optimal dimension reduction for high-dimensional and functional time series. *Statistical Inference for Stochastic Processes* **21** 385–398.

HALLIN, M. and LIPPI, M. (2013). Factor models in high-dimensional time series—A time-domain approach. *Stochastic processes and their applications* **123** 2678–2695.

HALLIN, M. and LIŠKA, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* **102** 603–617.

HALLIN, M. and LIŠKA, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics* **163** 29–41.

HALLIN, M. and TRUCÍOS, C. (2021). Forecasting value-at-risk and expected shortfall in large portfolios: A general dynamic factor model approach. *Econometrics and Statistics* **27** 1–15.

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics* **40** 694–726.

LAZAR, N. (2008). *The statistical analysis of functional MRI data*. Springer Science & Business Media.

LEONENKO, N. (1999). *Limit theorems for random fields with singular spectrum* **465**. Springer Science & Business Media.

MANDREKAR, V. S. and REDETT, D. A. (2017). *Weakly Stationary Random Fields, Invariant Subspaces and Applications*. Chapman and Hall/CRC.

MOTTA, G., WU, W. B. and POURAHMADI, M. (2023). $\sqrt{2}$-estimation for smooth eigenvectors of matrix-valued functions. *Biometrika* **110** 1077–1098.

PARK, B. U., MAMMEN, E., HÄRDLE, W. and BORAK, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association* **104** 284–298.

ROYDEN, H. L. and FITZPATRICK, P. (1988). *Real analysis* **32**. Macmillan New York.

ROZANOV, Y. A. (1967). *Stationary random processes*. Holden-Day.

RUDIN, W. (1987). *Real and complex analysis. 1987*. McGraw-Hill.

SARGENT, T. J. and SIMS, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research* **1** 145–168.

STEIN, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

STOCK, J. H. and WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97** 1167–1179.

TRUCÍOS, C., MAZZEU, J. H., HALLIN, M., HOTTA, L. K., VALLS PEREIRA, P. L. and ZEVALLOS, M. (2022). Forecasting conditional covariance matrices in high-dimensional time series: a general dynamic factor approach. *Journal of Business & Economic Statistics* **41** 40–52.

VANMARCKE, E. (2010). *Random fields: analysis and synthesis*. World scientific.

WANG, F. and WALL, M. M. (2003). Generalized common spatial factor model. *Biostatistics* **4** 569–582.

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* 434–449.

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* **102** 14150–14154.

WU, W. B. and ZAFFARONI, P. (2018). Asymptotic theory for spectral density estimates of general multivariate time series. *Econometric Theory* **34** 1–22.

ZHANG, D. and WU, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics* **49** 233–254.