Spatial Warped Gaussian Processes: Estimation and Efficient Field Reconstruction

Gareth W. Peters [1], Ido Nevat[2], Sai Ganesh Nagarajan[3], Tomoko Matsui[4]

[1] Department of Statistics and Applied Probability, University of California Santa Barbara, USA

[2] TUMCREATE, Singapore

[3] Singapore University of Technology and Design (SUTD), Singapore

[4] Department of Statistical Modeling, The Institute of Statistical Mathematics (ISM), Tokyo, Japan

Abstract

A class of models for non-Gaussian spatial random fields is explored for spatial field reconstruction in environmental and sensor network monitoring. The family of models explored utilises a class of transformation functions known as the Tukey g-and-h transformations to create a family of warped spatial Gaussian process models which can support various desirable features such as flexible marginal distributions, which can be skewed, leptokurtic and/or heavy-tailed. The resulting model is widely applicable in a range of spatial field reconstruction applications. To utilise the model in applications in practice, it is important to carefully characterise the statistical properties of the Tukey g-and-h random fields. In this work we both study the properties of the resulting warped Gaussian processes as well as using the characterising statistical properties of the warped processes to obtain flexible spatial field reconstructions. In this regard we derive five different estimators for various important quantities often considered in spatial field reconstruction problems. These include the multi-point Minimum Mean Squared Error (MMSE) estimators; the multiple point Maximum A-Posteriori (MAP) estimators; an efficient class of multiple-point linear estimators based on the Spatial-Best Linear Unbiased (S-BLUE) estimators; and two multi-point threshold exceedance based estimators, namely the Spatial Regional and Level Exceedance estimators. Simulation results and real data examples show the benefits of using the Tukey g-and-h transformation as opposed to standard Gaussian spatial random fields in a real data application for environmental monitoring.

Keywords: Random fields, warped Gaussian Process, Spatial field reconstruction

I. Introduction

Spatial monitoring of physical phenomena has been at the center of many different spatial statistics and signal processing innovations in the last few decades. Specific applications in which environmental processes are monitored and spatial field reconstruction is performed include environmental monitoring [1] and weather

monitoring [2], [2]–[4], natural phenomena such as temperature [5], precipitation [6], wind speed [7]–[10] and pollution [11], to name a few.

In this work we explore warping maps of Gaussian process models such as those developed in [7] for spatial field reconstruction. It is most commonly the case that spatial field modelling is undertaken with a variation of a Gaussian Process (GP) model due to its direct interpretation and mathematical tractability [12]. GP models are readily characterized by a mean structure and a valid covariance function [12]. Such models have been applied in many different domains for spatial and temporal modelling: [13]–[15] and the classic texts on such Gaussian process models for spatial statistics [16], [17]. The warping extensions studied in this work are designed to specifically accommodate a flexible range of non-Gaussian skew and kurtosis attributes in order to demonstrate how to undertake accurate and efficient spatial field reconstruction in a more general class of spatial process models, whilst still maintaining much of the practical interpretation offered by Gaussian process modelling settings.

The significance of adding such warping structure arises since increasingly spatial modellers and data scientists are observing that many types of physical phenomena may have finite dimensional distributions which are not adequately captured by symmetric light tailed Gaussian distributions as is mandated by such classical GP spatial field modelling approaches. The reality is that the skew and kurtosis can play a significant role in spatial field modelling, see discussion in [4], [18] and [19]. As noted in these works, there are many real world settings where the physical process being modelled in the spatial field reconstruction may exhibit features which cannot be captured accurately via GP modelling, as the process may display spatial features at different locations that include excess skewness (asymmetry about the mode) , kurtosis and perhaps heavy tails. Hence, the Gaussian process will not provide such flexible features if used as the basis upon which to undertake development of estimators for spatial field reconstruction, therefore we wish to extend this class of models. This will in turn influence the accuracy of any spatial field reconstruction estimator built from such statistical model assumptions.

Therefore, when one finds a setting in which Gaussian process structure will not adequately capture the spatial field statistical structure, there are numerous approaches that one may proceed with when seeking to perform spatial field reconstruction. The first involves seeking a transformation that will induce properties in the transformed observations of the process more akin to a GP model, see examples in [20], [21] for a detailed overview of the subject. Examples include the log-transform, proposed by Galton, which resulted in the nowadays well-known log-normal distribution [22]. The first systematic attempt to fitting asymmetric distributions to non-normal data was made by Edgeworth [23]. Karl Pearson defined several probability distributions as solutions to a particular differential equation, which formed the basis for the Pearson family of distributions [24]. Further development was the *two-piece distributions*, proposed by Fechner by binding together two halves of normal curves, each having a different scaling [25]. A more recent work which had a great impact was by Azzalini [26] who provided a rigorous treatment to the analysis of Skew-Normal distributions.

An alternative, recent class of approaches in statistical machine learning literature has involved the converse of these methods, where one rather warps or "transforms" a GP model to produce additional structure such as skewness or kurtosis and then works with estimation and spatial field reconstruction of the resultant transformed process. This method has two components the specification of the sufficient statistical functions for the input GP model and the specification of the functional characteristic of the transformation map.

In this work, we explore the second approach based on warping a GP model. One may consider our approach as exploring a generalised class of processes that act as extensions of the Gaussian process family to a much larger family of processes that incorporate parameterised features of skewness and kurtosis that arises directly from the class of warping functions adopted. In this family of models we develop, just like the Gaussian process case, we wish to be able to state useful information about the resulting processes finite dimensional distributions: i.e. the marginal and conditional distributions of the process at different locations in space. This can be done by modifying an existing known parametric distribution (Gaussian) by applying a transformation of the random variables pointwise. This concept was first proposed by Tukey in 1977, and further developed in [27]–[29] and in the time series setting by [30]. In this paper we demonstrate that when one selects the family of Tukey G-and-H model distortion functions, one can derive tractable solutions to spatial field reconstruction estimators that will out-perform classical Gaussian process solutions when spatial processes exhibit excess spatial skewness and spatial kurtosis features.

A. *Contributions and Outline*

The main contributions developed in this work are twofold: formulating a class of spatial warped Gaussian process model based on Tukey g-and-h warping functions. This involves deriving important key properties of the Tukey g-and-h transformation and resulting family of spatial warped Gaussian processes (Sections II- III); and using those results to develop new estimation algorithms for efficient spatial field reconstructions (Section IV). Our contributions may be summarised as follows:

1) In Section II we derive mathematical properties of the proposed warped GP models warping function known as the Tukey g-and-h transformation function. This is actually a class of warping functions with sub-classes related to skewness and kurtosis transformations. Since this transformation is not widely known, it is valuable for practitioners to see the basic properties of this class of functions in order to better understand the resulting properties of the warped GP that arise when using this transform. In this regard we explore the transformation and its limiting behaviours such as in (Proposition 2), the existence of the inverse transform (Proposition 1), and its derivatives (Lemma 1). This set of results is instrumental for the analysis of the properties of the Tukey g-and-h process and for the construction of spatial field approximations that form the key application in this paper.

2) In Section III we derive important statistical properties of the warped GP process which include properties of the finite dimensional distributions such as raw and central moments, the multivariate moments (Corollary 1), the population mean and variance (Corollaries 3-4), its higher order multivariate cross moments of any order (Proposition 3), and the index of regular variation (Lemma 2), to determine conditions in terms of the parameters of the g-and-h transform for the existence of all higher order mixed moments. In particular the cross moments are directly relevant to the construction of efficient spatial field reconstructions for multiple locations in space that are performed in the application section. The tail behaviour of the finite dimensional distributions is captured by the tail index analysis and is relevant to understand models with excess kurtosis relative to a GP model, which is induced by our class of warping function.

3) In Section IV we derive the spatial field reconstruction estimators of the Tukey g-and-h random field under five different metrics, most of which are based on the predictive posterior density (Lemma 4).

4) In Section V a detailed set of numerical experiments are provided for both synthetic case studies developed to demonstrate key properties of the spatial field reconstruction methods and their accuracy and performance relative to a baseline GP model. Then we also developed a sequence of real data experiments based on data obtained from the National Climatic Data Center (NCDC) on the spatial observation from ground monitoring stations of precipitation. In particular we took one month of hourly precipitation measurements to fit our warped GP model and to compare the spatial field reconstruction from the warped model discussed in this manuscript compared to a baseline GP model.

II. CONSTRUCTING A FLEXIBLE WARPING FUNCTION FOR SPATIAL GAUSSIAN PROCESSES

Many families of warping function can be developed for extending the class of GP spatial process models. In this work we particularly focus on the study of the family of Tukey G-and-H warping functions. When developing the new warped GP class of flexible models, the following list of desirable properties were considered desirable for the resulting warped GP model to posses (see [20]):

1) Parsimonious and finite number of well interpreted parameters;

2) Well-separated roles for skewness, kurtosis and the tail behaviour parameters;

3) Accurate and computationally efficient parameter estimation and tractability;

4) Derivable and interpretable stochastic properties;

5) Flexible structures in the marginal, conditional and joint density shapes;

6) Good inferential properties;

7) Exact and computationally efficient data generating mechanisms.

Some of these aforementioned properties have been studied before in the context of the class of warping functions we consider, namely the Tukey g-and-h transformed random Gaussian process fields, see for example [30]–[35] and the textbook overview in [36].

In this paper we re-explore these properties and extend them for the purposes of spatial field reconstruction. Therefore, in this section, we first define the fundamental building blocks for deriving the non-Gaussian warped GP fields. As such, we define the basic set of non-linear transformations of univariate and multivariate Gaussian random variables and the resulting Tukey g-and-h univariate and multivariate distributions. We also derive some fundamental properties such as the existence of the inverse transform as well as the derivatives of the distribution that will be important for future results.

Definition 1 (Tukey Transformation [37]).

Consider a Gaussian random variable $W \sim \mathcal{N}(\mu, \sigma)$ and transformation $Y := \mathcal{T}(W; \zeta)$ with parameter vector $\zeta := [m_0, m_1, \mu, \sigma^2, \eta, \theta]$, where $m_0 \in \mathbb{R}$, $m_1 \in \mathbb{R}^+$, $\eta \in \mathbb{R}^d$ and $\theta \in \mathbb{R}$. Then the resultant transformed random variable Y will be from a Tukey law if the corresponding transformation $\mathcal{T}(W; \zeta)$ is given by

$$Y := \mathcal{T}(W; \zeta) = m_0 + m_1 (\mathcal{F}(W; \eta))^\theta W, \tag{1}$$

where the transformation $\mathcal{F}(W; \eta)$ is positive, symmetric, and strictly monotonically increasing for positive values of $W \geq 0$.

We note that as a default we will consider the base transformed random variable W to be Gaussian with $\mu = 0$ and $\sigma = 1$ unless otherwise stated. Furthermore, in general the transformation $\mathcal{F}(W; \eta)$ can have many forms, see discussions in [30]. In this manuscript we concentrate on two specific types of transformation directly related to inducing skewness in the warped GP model or inducing kurtosis in the warped GP model. Note that standard Gaussian processes can have flexible covariance functions (second order flexibility) but there is no flexibility in the higher order structure relating to third co-moments (co-skewness) and fourth co-moments (co-kurtosis) that can often be observed in real spatial field observation data. Therefore, the use of the Tukey sub-families of warping functions that specifically produce parametersization of such structure in the warped GP model is of practical relevance in flexible reconstruction methods for spatial fields.

Definition 2 (Tukey Family of Transformations).

The Tukey g (skewness) and Tukey h (kurtosis) transformations form two classes of sub-transformations in the Tukey family which are characterised by warping functions given by:

1) ***Skew transform:***

$$\mathcal{G}(W; g) = \frac{\exp(gW) - 1}{gW}. \tag{2}$$

2) ***Kurtosis transform:***

$$\mathcal{H}(W; h) = \exp\left(\frac{hW^2}{2}\right), \tag{3}$$

3) **Skew-Kurtosis product transform:**

$$\mathcal{F}(W;\eta) = \mathcal{G}(W;g)\,\mathcal{H}(W;h),\tag{4}$$

for $g \in \mathbb{R} - \{0\}, h \in \mathbb{R}^+$ and $\eta = [g,\ h]$.

These set of Tukey transformations modify the tail behaviour and the skewness relative to the base distribution, in this case the GP model which has finite dimensional Gaussian distributions point-wise in space. As such, the warped GP model will be able to produce a class of processes which point-wise in space will have distributions that are less symmetric around the mode with skew that can be left or right depending on the sign of g in the warping transform $\mathcal{G}$ and heavier tailed than a Gaussian probability distribution via leptokurtic transform $\mathcal{H}$ and the strength of kurtosis relative to the Gaussian at each spatial point is controlled by parameter h. The Tukey g-and-h transformation is known to encompass a large range of distributions and this is further explained in [29] and [30]. This is one of the most desirable properties of this transformation.

Before proceeding in using this family of warping functions to define a warped GP spatial process, it will be useful to provide a few basic properties of this class of transform, that whilst simple mathematically to verify are highly useful in developing further the specification of the properties of the resulting warped processes.

When working with warped GP's it is valuable to be able to evaluate the transformed finite dimensional density at a single location in space point-wise when performing tasks such as hyper-parameter estimation via the likelihood and also for spatial field reconstruction. Therefore, it is of relevance to ask if the warping function is invertible as such a property leads one to obtaining a tractable expression for the finite dimensional distributions. We show in Proposition 1 the conditions under which the inverse of the warping function exists. Consequently, we can be sure that should we satisfy these conditions for the warping function then in the case of the Tukey g-and-h transformation, it does accommodates a closed form distribution representation which will be tractable up to performing a simple univariate root search as we will see later in Proposition 5.

Proposition 1 (Inverse of the Tukey g-and-h transformation)**.**

The inverse of the Tukey g-and-h transformation $\mathcal{T}_{gh}(W;\zeta_{\mathbf{gh}})$ exists if $g \in \mathbb{R} - \{0\}$ and $h \in \mathbb{R}^+$

Proof. See Appendix A   □

Remark 1. *The result in Proposition 1 affirms the existence of the inverse of this transformation but does not provide an analytical expression for the inverse. In practice, one needs to find the inverse using simple univariate root search computational methods in order to facilitate evaluation of the warped distributions or densities point-wise. The existence of the inverse is still critical to verify for the warping transform of the Gaussian process as it will be used to facilitate expression of the spatial processes finite dimensional distributions which will in turn be used for specification of the likelihood for use in calibration and spatial field estimation and reconstruction.*

6

It will also be useful to provide simple characterisations of the derivatives of the warping functions up to third order, as these will be used numerous times in the warped GP spatial field reconstruction estimation frameworks to capture the moments, mode and curvature approximations (covariance) point-wise. The derivatives of the Tukey g-and-h transform are given in Lemma 1 by simple algebra.

Lemma 1 (Derivatives of Tukey g-and-h transform).

The first three derivatives of the Tukey g-h transform are given by:

$$\mathcal{T}_{gh}'\left(W;\zeta_{\mathbf{gh}}\right) = hW\mathcal{T}_{gh}\left(W;\zeta_{\mathbf{gh}}\right) + \exp\left(gW + \frac{hW^2}{2}\right), \tag{5}$$

$$\mathcal{T}_{gh}''\left(W;\zeta_{\mathbf{gh}}\right) = h\mathcal{T}_{gh}\left(W;\zeta_{\mathbf{gh}}\right) + hW\mathcal{T}_{gh}'\left(W;\zeta_{\mathbf{gh}}\right) + \exp\left(gW + \frac{hW^2}{2}\right)(hW + g), \tag{6}$$

$$\mathcal{T}_{gh}'''\left(W;\zeta_{\mathbf{gh}}\right) = hW\mathcal{T}_{gh}''\left(W;\zeta_{\mathbf{gh}}\right) + 2h\mathcal{T}_{gh}'\left(W;\zeta_{\mathbf{gh}}\right) + \exp\left(gW + \frac{hW^2}{2}\right)\left((hW + g)^2 + h\right). \tag{7}$$

A. Tukey Warped Spatial Gaussian Process

One may now proceed to developing a characterisation of a warped GP spatial process based around the Tukey family of skewness and kurtosis transformations. We begin by providing a definition of the Gaussian Process (GP) that will set the notation we will use to specify the spatial trend and spatial covariance operator for the input GP spatial model. This GP model will serve as our base distribution (process) to generate non-Gaussian fields i.e. warped GP processes. Then we will show how one can construct the warped GP model based on the Tukey g-and-h random field transform.

Definition 3. *Gaussian Process (GP) [12]:*

Denote by $\mathcal{W}(\mathbf{x}) : \mathcal{X} \mapsto \mathbb{R}$ a stochastic process parameterized by $\{\mathbf{x}\} \in \mathcal{X}$, where $\mathcal{X} \subseteq \mathcal{R}^d$. Then, the random function $\mathcal{W}(\mathbf{x})$ is a Gaussian process if all its finite dimensional distributions are Gaussian, where for any $m \in \mathbb{N}$, the random vector $(W(\mathbf{x}_1), W(\mathbf{x}_2), \ldots, W(\mathbf{x}_m))$ is jointly normally distributed [12].

We can therefore interpret a GP as formally defined by the following class of random functions:

$$\mathcal{W} := \{ \mathcal{W}(\cdot) : \mathcal{X} \mapsto \mathbb{R}$$

$$\text{s.t. } \mathcal{W}(\cdot) \sim \mathcal{GP}\left(\mu\left(\cdot;\boldsymbol{\theta}\right), \mathcal{C}\left(\cdot,\cdot;\boldsymbol{\Psi}\right)\right), \text{ with}$$

$$\mu\left(\cdot;\boldsymbol{\theta}\right) := \mathbb{E}\left[\mathcal{W}(\cdot)\right] : \mathcal{X} \mapsto \mathbb{R},$$

$$\mathcal{C}\left(\cdot,\cdot;\boldsymbol{\Psi}\right) := \mathbb{E}\left[(\mathcal{W}(\cdot) - \mu\left(\cdot;\boldsymbol{\theta}\right))(\mathcal{W}(\cdot) - \mu\left(\cdot;\boldsymbol{\theta}\right))\right] : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+ \}.$$

At each point the function valued process is characterised by the mean function $\mu(\cdot;\boldsymbol{\theta})$, parameterized by $\boldsymbol{\theta}$, and the spatial dependence between any two points is given by the covariance function (Mercer kernel) $\mathcal{C}(\cdot,\cdot;\boldsymbol{\Psi})$, parameterized by $\boldsymbol{\Psi}$, see detailed discussion in [12].

It will be useful to make the following notational definitions for the prior mean vector, cross correlation vector and auto-correlation matrix, respectively:

$$[\mu(\mathbf{x}_1; \boldsymbol{\theta}), \mu(\mathbf{x}_2; \boldsymbol{\theta}), \ldots, \mu(\mathbf{x}_m; \boldsymbol{\theta})] = [a_1, a_2, \ldots, a_m]$$

$$\mathbf{k}(\mathbf{x}_*, \mathbf{x}_{1:m}) := [\mathcal{C}(\mathbf{x}_*, \mathbf{x}_1), \mathcal{C}(\mathbf{x}_*, \mathbf{x}_2), \ldots, \mathcal{C}(\mathbf{x}_*, \mathbf{x}_m)] \in \mathbb{R}^{1 \times m}$$

$$\mathbf{K}(\mathbf{x}_{1:m}, \mathbf{x}_{1:m}) := \begin{bmatrix} \mathcal{C}(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \mathcal{C}(\mathbf{x}_1, \mathbf{x}_m) \\ \vdots & \ddots & \vdots \\ \mathcal{C}(\mathbf{x}_m, \mathbf{x}_1) & \cdots & \mathcal{C}(\mathbf{x}_m, \mathbf{x}_m) \end{bmatrix} \in \mathcal{X}^+(\mathbb{R}^m),$$

with $\mathcal{X}^+(\mathbb{R}^m)$ is the manifold of symmetric positive definite matrices. $\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j)$ is the covariance function. We now provide a formal definition of the Tukey g-and-h process.

Definition 4. *Spatial Tukey Warped Gaussian Process:*

Denote by $\mathcal{Y}(\mathbf{x}) : \mathcal{X} \mapsto \mathbb{R}$ a warped spatial stochastic process parameterized by points in space $\{\mathbf{x}\} \in \mathcal{X} \subseteq \mathbb{R}^2$. Then, $\mathcal{Y}(\mathbf{x})$ is a Tukey warped Gaussian Process if all its m-point finite dimensional distributions, for any $m \in \mathbb{N}$, are jointly distributed according to the random vector obtained by an elementwise Tukey transformation of the underlying spatial Gaussian process on points $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ with $\mathbf{x}_i \in \mathcal{X}$, given by $(Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_m)) := (\mathcal{T}(W(\mathbf{x}_1); \zeta), \ldots, \mathcal{T}(W(\mathbf{x}_m); \zeta))$ for any $m \in \mathbb{N}$, where $(W(\mathbf{x}_1), \ldots, W(\mathbf{x}_m))$ are jointly normally distributed from a spatial Gaussian Process $\mathcal{W}$.

III. Properties of the Spatial Tukey Warped Gaussian Process

Having developed a basic definition of the warped GP model based on a Tukey family of warping functions, we can now study the basic properties of this spatial process by first exploring its finite dimensional distributions. Such basic finite dimensional multiple point characterisations is useful for preparing spatial field reconstruction estimators for these warped GP models. This starts with the model for a single point in space (univariate marginal warped GP model in Section III-A), followed by specification of the properties of multiple-points in space (multivariate finite dimensional marginal warped GP model in Section III-C) .

A. Spatial Tukey Warped Gaussian Characterisation at a Single Point In Space

At any point in space $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^2$ one may consider the properties of the single location warped GP model and its properties. In this case, at one location we may consider as input a standard GP with zero spatial mean and spatial covariance function k giving a Gaussian random variable $W(\mathbf{x}) \sim \mathcal{N}(\mu, \mathcal{C}(\mathbf{x}, \mathbf{x}))$. Then the baseline reference Tukey g-and-h warped random variable is obtained by setting $\zeta_{\mathbf{gh}} := [m_0, m_1, \mu, \mathcal{C}, g, h, \theta = 1]$ where $g \in \mathbb{R} - \{0\}$ and $h \in \mathbb{R}^+$ to give:

$$Y(\mathbf{x}) := \mathcal{T}_{gh}(W(\mathbf{x}); \zeta_{\mathbf{gh}}) = m_0 + m_1 \frac{\exp(gW(\mathbf{x})) - 1}{g} \exp\left(\frac{hW(\mathbf{x})^2}{2}\right). \tag{8}$$

8

In the next proposition, we demonstrate that the warped Gaussian model will include the classical Gaussian model as a limiting case. This result is most readily proven for the setting of a single point in space, where we show that $Y(\mathbf{x})$ has a Gaussian law under some special specification of the warping function or its parameters.

Proposition 2 (Limiting case: Gaussian distribution as a special case)**.**

  The random variable $Y(\mathbf{x}) := \mathcal{T}_{gh}\left(W(\mathbf{x}); \zeta_{\mathbf{gh}}\right)$, *where* $\zeta_{\mathbf{gh}} := [m_0, m_1, \mu, \mathcal{C}, g, h, \theta = 1]$, *in the limit* $g \to 0$ *and* $h \to 0$, *reduces to*

$$Y(\mathbf{x}) \sim \mathcal{N}\left(m_0 + \mu, m_1^2 \mathcal{C}(\mathbf{x}, \mathbf{x})\right). \tag{9}$$

By setting $m_0 = 0, m_1 = 1$ *we recover* $Y(\mathbf{x}) \sim \mathcal{N}\left(\mu, \mathcal{C}(\mathbf{x}, \mathbf{x})\right)$, *i.e.,* $Y(\mathbf{x}) \overset{d}{=} W(\mathbf{x})$.

Proof. See Appendix B   □

Remark 2. *The parameter* $g \in \mathbb{R}$ *modifies the skewness relative the base distribution, in this case the Gaussian. A positive value of g indicates a positive skew in the transformed distribution and a negative value of g indicates a negative skew. The parameter h indicates a transformation in the kurtosis relative to the base distribution and to guarantee existence of moments, one may consider a restricted range such as* $0 \leq h \leq 1$.

We can also easily characterise the tail behaviour of the propose g-and-h warping transform to demonstrate how the h parameter in the $\mathcal{H}$ transform is able to produce leptokurtic behaviour in the resulting warped Gaussian density. One way to understand this is to demonstrate that the resulting warped distribution function has a property of regular variation, in other words it can be expressed as a power law tail decay and a slowly varying function.

Therefore, we briefly study regular variation of a distribution of the Tukey g-and-h distribution, which is a measure of the heavyness of tails of a distribution. Later, we connect this to existence of the nth moment . The index of the regular variation for the Tukey g-and-h distribution has been studied in [30], [38], we summarise this result below for a generic base Gaussian distribution with $W \sim \mathcal{N}\left(\mu, \mathcal{C}(\mathbf{x}, \mathbf{x})\right)$. It is important to understand under what conditions one has finite first and second order moments for a model with a specific kurtosis transform if one is to develop classes of linear spatial field reconstruction based on first and second moment local approximations. An advantage of the Tukey g-and-h warping function is that one can set conditions of the warping functions through simple parameter space restrictions to ensure existence of appropriate moments for use in spatial field reconstructions. The following result in Lemma 2 is meaningful to accommodate this analysis.

Lemma 2. *Consider a location* $\mathbf{x} \in \mathcal{X}$ *and input GP random variable* $W(\mathbf{x}) \sim \mathcal{N}\left(\mu, \mathcal{C}(\mathbf{x}, \mathbf{x})\right)$ *then the Tukey g-and-h random variable* $Y(\mathbf{x}) := \mathcal{T}_{gh}\left(W(\mathbf{x}); \zeta_{\mathbf{gh}}\right)$ *with parameters* $\zeta_{\mathbf{gh}} := [m_0 = 0, m_1 = 1, \mu, \mathcal{C}(\mathbf{x}, \mathbf{x}), g, h, \theta = 1]$ *has an index of regular variation for its survival function* $\overline{F}_{Y(\mathbf{x})}(y) := 1 - F_{Y(\mathbf{x})}(y)$ *given by:*

$$\lambda = \frac{1}{h\,\mathcal{C}(\mathbf{x}, \mathbf{x})}. \tag{10}$$

Proof. See Appendix C   □

B. Spatial Tukey Warped Gaussian Characterisation at Multiple Points In Space

One then readily extends the model development from the univariate Tukey g-and-h random variable that characterises the warped GP model at a single location $\mathbf{x} \in \mathcal{X}$ to a multiple-location specification at points $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathcal{X}$ which requires specification of a multivariate Tukey g-and-h random vector $\boldsymbol{Y} = [Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_m)]$, as for instance specified in constructions in [4], [29], [39] is obtained by applying an element-wise Tukey g-and-h transformation to a Multivariate Normal distribution, $\boldsymbol{W}_{1:m} \sim \mathcal{MVN}(\mathbf{w}; \boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ such that $\Sigma_{ij} = \mathcal{C}(\mathbf{x}_i, \mathbf{x}_j)$ based on the GP covariance kernel k at two points in space $\mathbf{x}_i, \mathbf{x}_j$. This gives for the j-th location $\mathbf{x}_j$ the transformed random variable

$$Y_j := \mathcal{T}_{gh}(W_j; \zeta_{gh}), \quad j \in \{1, \cdots, m\}. \tag{11}$$

In understanding the basic properties of the resulting multivariate warped random vector for the Tukey g-and-h model at multiple spatial locations, we can study the exponential moments (Lemma 3) which will be instrumental in deriving many of the properties of the warped GP models spatial field reconstruction estimators.

Although the Tukey g-and-h multivariate random vector and the Tukey g-and-h processes have been defined previously in aforementioned works. They have not been carefully studied from the perspective of deriving closed form results for their higher order cross moments or spatial moments or any order, including cross covariance, cross skewness and cross kurtosis. In this section, our goal is therefore to derive certain desirable properties which help to characterize the Tukey g-and-h processes further than just their definition.

To achieve these spatial cross moment characterisations it will be useful to first state a result for the quadratic exponential moments of a multivariate normal random vector. This will form of moment structure will arise naturally when expressing the cross moments of the class of warped GP model developed under Tukey transforms. Furthermore, these closed from result will be useful for developing computationally efficient and interpretable spatial field reconstructions based on a spatial Tukey warped GP model class.

Lemma 3 (Exponential Moments of a Gaussian Random Vector). *Consider a m-variate random vector $\boldsymbol{W} := [W_1, \ldots, W_m]^T$ distributed as $\boldsymbol{W} \sim \mathcal{MVN}(\mathbf{w}; \mu, \boldsymbol{\Sigma})$, with diagonal matrix $\boldsymbol{\Sigma}_v$ such that $[\Sigma_v]_{i,i} = v_i, i \in \{1, \ldots, m\}$ and $\mathbf{u} \in \mathbb{R}^m$, then the quadratic exponential moments of $\boldsymbol{W}$ are given by:*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\boldsymbol{W}^T\boldsymbol{\Sigma}_v\boldsymbol{W} + \mathbf{u}^T\boldsymbol{W}\right)\right] = \frac{|\widetilde{\boldsymbol{\Sigma}}|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left(\mu^T\boldsymbol{\Sigma}^{-1}\mu - \widetilde{\mu}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mu}\right)\right), \tag{12}$$

where,

$$\widetilde{\boldsymbol{\Sigma}} = \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma_v}\right)^{-1}, \tag{13}$$

$$\widetilde{\mu} = \widetilde{\boldsymbol{\Sigma}}\left(\mathbf{u} + \boldsymbol{\Sigma}^{-1}\mu\right). \tag{14}$$

Proof. See Appendix D   □

This result can now be developed into a core component to characterising raw cross moments and cross central moments for the warped GP model. We again consider m-points in space in a vector $\mathbf{x}_{1:m} := [x_1, \ldots, x_m]$. Let us first describe a generic equation for an m-variate raw cross moment of a spatial Tukey warped GP model.

Proposition 3 (Spatial Tukey Warped GP Raw Moments for m Points in Space)**.**

Consider the m-point spatial Tukey warped GP model with random vector $\mathbf{Y} = [Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_m)]$. Assume w.l.o.g. a standardised transform with location and scale parameters given by $m_0 = 0$, $m_1 = 1$ pointwise and $\zeta_{\mathbf{gh}} := [0, 1, g, h]$ selected such that existence of finite cross moments of orders $(n_1, \ldots, n_m)$ for $n_i \in \mathbb{N}$ holds. Then the raw spatial m-point cross moments of orders $(n_1, \ldots, n_m)$ are given by:

$$
\begin{aligned}
\check{\mu}_{gh}^{(n_1,\ldots,n_m)} : &= \mathbb{E}\left[\prod_{j=1}^m (Y(\mathbf{x}_j))^{n_j}\right] \\
&= \mathbb{E}\left[\prod_{j=1}^m (\mathcal{T}_{gh}(W(\mathbf{x}_j);\zeta_{\mathbf{gh}}))^{n_j}\right] \\
&= \frac{1}{g^{(n_1+n_2+\ldots+n_m)}} \sum_{i_1=0}^{n_1} \cdots \sum_{i_m=0}^{n_m} (-1)^{(i_1+i_2+\ldots+i_m)} \frac{|\widetilde{\boldsymbol{\Sigma}}|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\mu^T \boldsymbol{\Sigma}^{-1}\mu - \left(\widetilde{\mu}^{(i)}\right)^T \widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mu}^{(i)}\right)\right),
\end{aligned}
$$

(15)

where

$$
\widetilde{\boldsymbol{\Sigma}} = \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_v\right)^{-1} \tag{16}
$$

$$
\widetilde{\mu}^{(\mathbf{i})} = \widetilde{\boldsymbol{\Sigma}}\left(\mathbf{u}^{(i)} + \boldsymbol{\Sigma}^{-1}\mu\right), \tag{17}
$$

$$
\mathbf{u}^{(i)} = g\left[(n_1 - i_1), (n_2 - i_2), \ldots, (n_m - i_m)\right]^T \tag{18}
$$

and $\boldsymbol{\Sigma}_v$ is a diagonal matrix with known elements $[\Sigma_v]_{i,i} = hn_i, i \in \{1, \ldots, m\}$.

Proof. See Appendix E   □

This general result is then instrumental in producing practical examples of raw m-location cross moments of the spatial Tukey warped GP model of orders $(n_1, \ldots, n_m)$ denoted by $\check{\mu}_{gh}^{(n_1,\ldots,n_m)}$ as well as the centralised cross moments denoted by $\mu_{gh}^{(n_1,\ldots,n_m)}$ obtained by this same identity coupled with a binomial sum expansion to accommodate the centering transform at each of the locations x_1 to x_m.

This result generalises the previous results in the literature for moments of a Tukey g-and-h process which previously only explored pairwise ($m = 2$ points) second and first order moments, where as this result obviously provides generalised $m \geq 1$ point cross moment expressions of any orders. This is a significant extension in characterising this family of warped GP models, which no-longer have sufficient statistics (functions) given solely by mean and covariance functions.

C. Practical Cross Moment Special Cases for Spatial Field Reconstruction of Tukey Warped GP Models

The result in Proposition 3 whilst very general can be used to readily express a few core identities of relevance to spatial field reconstruction estimators at multiple locations in space, when one assumes to have observations consistent with a class of spatial Tukey warped GP models.

In particular we will state the moments and conditions for their existence in terms of the Tukey h parameter in the warping function $\mathcal{H}$ as well as the two-point cross moments.

Corollary 1 (Single Location Central Moments of Spatial Tukey Warped Gaussian Process)**.**

The n_i-th central moment at location $\mathbf{x}_i$ is given by:

$$\mu_{gh}^{(0,\ldots,0,n_i,0,\ldots,0)} = \frac{1}{g^{n_i}\sqrt{1 - n_i h \mathcal{C}\left(\mathbf{x}_i, \mathbf{x}_i\right)}} \sum_{i=0}^{n_i}(-1)^i \binom{n_i}{i} \exp\left(\frac{n_i h a_i^2 + 2 a_i g\left(n_i - i\right) + \left(n_i - i\right)^2 g^2 \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i)}{2\left(1 - n_i h \mathcal{C}\left(\mathbf{x}_i, \mathbf{x}_i\right)\right)}\right) \tag{19}$$

We state in Proposition 4 below are result that allows one to connect the idea of regular variation to the existence of moments a certain power. We understand that as the tails of the distribution become heavier (as measured by the index of regular variation), the integrals involved in the calculation of moments may cease to converge to a finite value. We formally state the relationship between the moment existence and the regular variation index as follows.

Proposition 4 (Existence of Moments of Single Location Spatial Tukey Warped GP)**.** *For the n_i-th moment to exist in Corollary 1 we have to ensure that the following inequality holds:*

$$h < \frac{1}{n_i \mathcal{C}\left(\mathbf{x}_i, \mathbf{x}_i\right)} \tag{20}$$

Proof. The existence proof is based on the fact that the for the n-th moment to exist for a random variable, the index of regular variation of the random variable must be strictly greater than n. Thus from Lemma 2, we get:

$$\lambda > n,$$
$$\implies \frac{1}{h\sigma^2} > n, \tag{21}$$
$$\implies h < \frac{1}{n\sigma^2}.$$

where $\sigma^2 = \mathcal{C}\left(\mathbf{x}_i, \mathbf{x}_i\right)$, for the Tukey g-and-h process at the location $\mathbf{x}_i$. □

Next we can express as a special case for two points $\mathbf{x}_i, \mathbf{x}_j$ in space $(m = 2)$ the cross moments of order (n_i, n_j) as follows.

Corollary 2 (Two Location Cross Central Moments of Spatial Tukey Warped Gaussian Process)**.**

The n_i-th and the n_j-th order cross moment between two locations $\mathbf{x}_i$ and $\mathbf{x}_j$ is given by:

$$
\mu_{gh}^{(0,\ldots,0,n_i,0,\ldots,0,n_j,0,\ldots,0)} = \frac{1}{g^{n_i+n_j}\sqrt{D_\Sigma}} \sum_{i_i=0}^{n_i} \sum_{i_j=0}^{n_j} (-1)^{i_i+i_j} \binom{n_i}{i_i}\binom{n_j}{i_j} \exp\left(\frac{\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_j\right)\left(\left(\widetilde{\mu}_1^{(i)}\right)^2 - |\mathbf{\Sigma}|^2 D_\Sigma^2 a_i^2\right)}{|\mathbf{\Sigma}|^3 D_\Sigma} \right)
$$

$$
\times \exp\left(\frac{\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)\left(\left(\widetilde{\mu}_2^{(i)}\right)^2 - |\mathbf{\Sigma}|^2 D_\Sigma^2 a_j^2\right) - \left(\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_j\right)+\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_i\right)\right)\left(\widetilde{\mu}_1^{(i)}\widetilde{\mu}_2^{(i)} - |\mathbf{\Sigma}|^2 D_\Sigma^2 a_i a_j\right)}{|\mathbf{\Sigma}|^3 D_\Sigma} \right),
$$

(22)

where

$$
|\mathbf{\Sigma}| = \mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_j\right) - \mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_j\right)\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_i\right),
$$

$$
\widetilde{\mu}_1^{(i)} = |\mathbf{\Sigma}|\left(a_i\left(1 - hn_j\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_j\right)\right) + a_j hn_j\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_j\right) + g\left(\left(\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right) - hn_j|\mathbf{\Sigma}|\right)\left(n_i-i_i\right) + \mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_j\right)\left(n_j-i_j\right)\right)\right),
$$

$$
\widetilde{\mu}_2^{(i)} = |\mathbf{\Sigma}|\left(a_j\left(1 - hn_i\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)\right) + a_i hn_i\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_i\right) + g\left(\left(\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_j\right) - hn_i|\mathbf{\Sigma}|\right)\left(n_j-i_j\right) + \mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_i\right)\left(n_i-i_i\right)\right)\right),
$$

$$
D_\Sigma = h^2 n_i n_j |\mathbf{\Sigma}| - h\left(n_i\mathcal{C}(\mathbf{x}_i,\mathbf{x}_i) + n_j\mathcal{C}\left(\mathbf{x}_j,\mathbf{x}_j\right)\right) + 1.
$$

(23)

Proof. This is direct application for points $\mathbf{x}_i, \mathbf{x}_j$ of Proposition 3. □

We note that with these results one may simply recover the mean at location $\mathbf{x}_i$ with $n_i = 1$ giving

$$
\mathbb{E}\left[Y\left(\mathbf{x}_i\right)\right] = \mu_{gh}^{(0,\ldots,0,1,0,\ldots,0)} = \frac{1}{g\sqrt{1-h\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)}}\left[\exp\left(\frac{ha_i^2 + g^2\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right) + 2a_i g}{2(1-h\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right))}\right) - \exp\left(\frac{ha_i^2}{2(1-h\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right))}\right)\right].
$$

(24)

and the variance at location $\mathbf{x}_i$ with $n_i = 2$ as follows:

$$
\mathbb{V}ar\left[Y\left(\mathbf{x}_i\right)\right] = \kappa_{gh}^{(0,\ldots,0,2,0\ldots,0)} = \frac{1}{g^2\sqrt{1-2h\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)}}\left(\exp\left(\frac{ha_i^2 + 2a_i g + 2g^2\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)}{1-2h\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)}\right)\right.
$$

$$
\left. - 2\exp\left(\frac{2ha_i^2 + 2a_i g + g^2\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)}{2\left(1-2h\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)\right)}\right) + \exp\left(\frac{ha_i^2}{1-2h\mathcal{C}\left(\mathbf{x}_i,\mathbf{x}_i\right)}\right)\right)
$$

$$
- \left(\mu_{gh}^{(0,\ldots,0,1,0,\ldots,0)}\right)^2.
$$

(25)

It will also be practically useful for the spatial field reconstructions to state the cross spatial covariance between the spatial field at new location $\mathbf{x}_*$ and existing observed points $\mathbf{x}_{1:N}$. These cross spatial covariance results will be critical for derivation of the multiple point Spatial-Best Linear Unbiased Estimator class we term the (S-BLUE) spatial field reconstruction. Consider the joint description of a location $\mathbf{x}_*$ and the locations of the observations $\mathbf{Y}_{1:N}$ by the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, \mathbf{x}_*\}$. We also define $\kappa_{gh}^{(0,\ldots,0,1,0,\ldots,0,1,0,\ldots,0)} = \mu_{gh}^{(0,\ldots,0,1,0,\ldots,0,1,0,\ldots,0)} - $

$\mu_{gh}^{(0,...,0,1,0,...,0,0,0,...,0)}\mu_{gh}^{(0,...,0,0,0,...,0,1,0,...,0)}$ as the covariance at any two locations $\mathbf{x}_i$ and $\mathbf{x}_j$. Then we obtain the following quantities using the aforementioned derivations applied to obtain these cross moments:

$$\mu_{p_*} = a_* + \mathbf{k}\left(\mathbf{x}_*, \mathbf{x}_{1:N}\right)\mathbf{K}^{-1}\left(\mathbf{x}_{1:N}, \mathbf{x}_{1:N}\right)\left(\mathcal{T}_{gh}^{-1}\left(\mathbf{Y}_{1:N} - \mathbb{E}\left[\mathbf{Y}_{1:N}\right]\right)\right)$$

$$\mathbb{C}ov\left(\mathbf{Y}_*, \mathbf{Y}_{1:N}\right) := \left[\kappa_{gh}^{(1,0,...,0,1)}, \kappa_{gh}^{(0,1,...,0,1)}, \ldots, \kappa_{gh}^{(0,0,...,1,1)}\right] \in \mathbb{R}^{1 \times N}$$

$$\mathbb{C}ov\left(\mathbf{Y}_{1:N}, \mathbf{Y}_{1:N}\right) := \begin{bmatrix} \kappa_{gh}^{(1,0,...,0,0)} & \cdots & \kappa_{gh}^{(1,0,...,1,0)} \\ \vdots & \ddots & \vdots \\ \kappa_{gh}^{(1,0,...,1,0)} & \cdots & \kappa_{gh}^{(0,0,...,1,0)} \end{bmatrix} \in \mathcal{X}^+\left(\mathbb{R}^N\right),$$

Having derived these properties of the Tukey g-and-h random field we are now ready to derive the results for various field reconstruction metrics.

IV. Field Reconstruction for Spatial Tukey Warped Gaussian Processes

In many real world settings for spatial field modelling and analysis the spatial fields finite dimensional distributions or process may not be used explicitly in modelling and instead a selection of relevant point estimators will be used to characterise a statistical summary of the warped Gaussian process, in our case the Tukey g-and-h warped spatial processes. In this section we propose and solve for a variety of practically relevant spatial field point estimators that is often called in spatial statistics "spatial field reconstruction" or "spatial field approximation/estimation" [7], [8]. In our case, we tailored the spatial field reconstructions to three classes of estimator with varying efficacy and computational cost in construction.

We have formulated this spatial field reconstruction context in terms of classical risk minimization through a loss function formulating the class of estimator which is resolved with respect to the spatial process, namely the warped Gaussian process model. We are assuming the warped Gaussian process model is calibrated and here the exercise is not in calibration of this models hyper parameters for the spatial covariance kernel and warping skew and kurtosis parameters, namely g and h respectively, but rather in regards to approximating efficiently point estimators for the spatial field at any collection of target locations.

In this regard we derive five different estimators for various important quantities often considered in spatial field reconstruction problems. The multi-point Minimum Mean Squared Error (MMSE) estimators; the multiple point Maximum A-Posteriori (MAP) estimators; an efficient class of multiple-point linear estimators based on the Spatial-Best Linear Unbiased (S-BLUE) estimators; and two multi-point threshold exceedance based estimators, namely the Spatial Regional and Level Exceedance estimators.

Throughout the remainder of the manuscript, in order to simplify the notation we will denote process $Y(\mathbf{x}_i)$ at location $\mathbf{x}_i$ simply by Y_i. To perform spatial field reconstruction under the assumption that the data generating process for the spatial observations is a spatial Tukey warped Gaussian Process, one may approach the problem under a Bayesian paradigm. Then, it would be natural to base field reconstruction estimators at an arbitrary location

$\mathbf{x}_*$, at which one has not observed the process, via the predictive posterior density at any location in space, $\mathbf{x}_* \in \mathcal{X}$, denoted $p\left(Y_*|\mathbf{Y}_{1:N}\right)$, where $\mathbf{Y}_{1:N}$ is the measurements obtained from the sensors that observe the process at N locations $\mathbf{x}_{1:N} \in \mathcal{X}$. This approach is widely used and is the standard in many applications in sensor networks [1]–[4], [6]–[10]

To proceed with development of the conditional probability of prediction of the distribution of the spatial field at unobserved location $\mathbf{x}_*$ we need to specify the warped spatial GP's density, given in Lemma 4 which shows how to predict the spatial field at the new unobserved location $\mathbf{x}_*$ conditional on the observations of the warped GP spatial field $\mathbf{Y}_{1:N}$ at observed locations $\mathbf{x}_{1:N}$. Due to the structure of the warping spatial process developed, we are able to analytically express the first two moments for the predictive condition mean and covariance in close form. This is critical for efficient development of point estimators for the spatial field reconstructions that will be developed in following sections.

Lemma 4 (Predictive Posterior Density at Location $\mathbf{x}_*$)**.**

The predictive posterior density $p\left(Y_*|\mathbf{Y}_{1:N}, \mathbf{x}_{l:N}, \mathbf{x}_*\right)$ is given by:

$$Y_*|\mathbf{Y}_{1:N}, \mathbf{x}_{l:N}, \mathbf{x}_* \sim \mathcal{GH}\left(y_*; \zeta_{gh}\right), \tag{26}$$

where $\zeta_{gh} := \left[m_0 = 0, m_1 = 1, \mu_{p_*}, \sigma_{p_*}^2, g, h, \theta = 1\right]$ and

$$
\begin{aligned}
\mu_{p_*} &= a_* + \mathbf{k}\left(\mathbf{x}_*, \mathbf{x}_{l:N}\right) \mathbf{K}^{-1}\left(\mathbf{x}_{l:N}, \mathbf{x}_{l:N}\right)\left(\mathcal{T}_{gh}^{-1}\left(\mathbf{Y}_{1:N} - \mathbb{E}\left[\mathbf{Y}_{1:N}\right]\right)\right), \\
\sigma_{p_*}^2 &= k\left(\mathbf{x}_*, \mathbf{x}_*\right) - \mathbf{k}\left(\mathbf{x}_*, \mathbf{x}_{l:N}\right) \mathbf{K}^{-1}\left(\mathbf{x}_{l:N}, \mathbf{x}_{l:N}\right) \mathbf{k}\left(\mathbf{x}_{l:N}, \mathbf{x}_*\right).
\end{aligned}
\tag{27}
$$

Note that $\mathbb{E}\left[\mathbf{Y}_{1:N}\right]$ here denotes the prior mean of all the observations $\mathbf{Y}_{1:N}$ and $\mathcal{T}_{gh}^{-1}\left(.\right)$ denotes a pointwise application of the inverse.

Proof. A derivation of this conditional predictive density for a general g-and-h density with Guassian input process is given in [29]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Furthermore, we may derive based on this predictive density the first two predictive conditional moments which we state in the following corollaries describing the posterior mean and posterior variance of the Tukey g-h process at some arbitrary location $\mathbf{x}_*$. There are obtained explicitly using results derived in Proposition 3 and the distribution characteristics specified in Section III.

Corollary 3 (Posterior Predictive Mean at an Arbitrary Location $\mathbf{x}_*$)**.**

$$\mathbb{E}\left[Y\left(\mathbf{x}_*\right)|\mathbf{Y}_{1:N}, \mathbf{x}_{l:N}, \mathbf{x}_*\right] = \frac{1}{g\sqrt{1 - h\sigma_{p_*}^2}}\left[\exp\left(\frac{h\mu_{p_*}^2 + g^2\sigma_{p_*}^2 + 2\mu_{p_*}g}{2(1 - h\sigma_{p_*}^2)}\right) - \exp\left(\frac{h\mu_{p_*}^2}{2(1 - h\sigma_{p_*}^2)}\right)\right]. \tag{28}$$

Corollary 4 (Posterior Predictive Variance at an Arbitrary Location $\mathbf{x}_*$)**.**

$$\mathbb{V}ar\left[Y\left(\mathbf{x}_*\right)|\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right] = \frac{1}{g^2\sqrt{1-2h\sigma_{p_*}^2}}\left(\exp\left(\frac{h\mu_{p_*}^2 + 2\mu_{p_*}g + 2g^2\sigma_{p_*}^2}{1-2h\sigma_{p_*}^2}\right)\right.$$

$$\left. - 2\exp\left(\frac{2h\mu_{p_*}^2 + 2\mu_{p_*}g + g^2\sigma_{p_*}^2}{2\left(1-2h\sigma_{p_*}^2\right)}\right) + \exp\left(\frac{h\mu_{p_*}^2}{1-2h\sigma_{p_*}^2}\right)\right) \quad (29)$$

$$- \left(\mathbb{E}\left[Y\left(\mathbf{x}_*\right)|\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right]\right)^2 .$$

A. Spatial Field Reconstruction Risk Functionals and Loss Functions

In developing estimators formally for the spatial field reconstruction we introduce a set of different loss functions that will characterise the risk functionals to be minimized with respect to the warped GP model in order to represent the spatial field reconstructions of different estimator types. We have developed five different spatial field reconstruction estimators that we formalise below based on the conditional predictive posterior distribution.

Based on the predictive posterior density, various point estimators, like the MMSE and the MAP estimators can be derived. These estimators provide a pointwise estimator of the intensity of the spatial filed, $\widehat{Y}_*$ at location $\mathbf{x}_*$. This enables us to reconstruct the whole spatial field by evaluating $\widehat{Y}_*$ on a fine grid of points.

We first define a set of objective functions that covers different aspects of spatial regression and hence we calculate different estimators, each one having its special characteristic. We then derive explicitly each of the estimators in closed form expressions.

1) Minimum Mean Squared Estimator (MMSE) estimator

The MMSE estimator minimises the MSE by utilising a *squared loss function*:

$$L\left(Y_*,\widehat{Y}_*\right) := \left(Y_* - \widehat{Y}_*\right)^2 . \quad (30)$$

Then, the MMSE spatial field reconstruction estimator is given by

$$\widehat{Y}_*^{\text{MMSE}} = \arg\min_y \mathbb{E}\left[L\left(Y_*,\widehat{Y}_*\right)|\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right]$$

$$= \arg\min_y \mathbb{E}\left[\left(Y_* - \widehat{Y}_*\right)^2 |\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right] \quad (31)$$

$$= \mathbb{E}\left[Y_*|\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right]$$

$$= \int_{-\infty}^{\infty} Y_* p\left(Y_*|\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right)\mathrm{d}\,Y_* .$$

The conditional expectation is the the estimate that minimises the conditional MMSE and the minimum value is given by the conditional variance shown below:

$$\sigma_{Y*|\mathbf{Y}_{1:N}}^2 = \mathbb{E}\left[\left(Y_* - \mathbb{E}\left[Y_*|\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right]\right)^2 |\mathbf{Y}_{1:N},\mathbf{x}_{1:N},\mathbf{x}_*\right] . \quad (32)$$

2) Maximum A-Posteriori (MAP) estimator:

The MAP estimator is the mode of the predictive posterior density and utilizes the $0/1$ *loss function*:

$$L\left(Y_*, \widehat{Y}_*\right) := \begin{cases} 0 & \text{, If } \quad Y_* = \widehat{Y}_* \\ 1 & \text{, Otherwise} \end{cases} \tag{33}$$

Then, the MAP spatial field reconstruction estimator is given by

$$\widehat{Y}_*^{\text{MAP}} = \arg\max_{y^*} p\left(Y_* | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_*\right). \tag{34}$$

3) Spatial-Best Linear Unbiased Estimator (S-BLUE):

The S-BLUE utilises the same loss function as the MMSE estimator, but is restricted to the linear family of estimators, given by:

$$\widehat{Y}_*^{\text{S-BLUE}} = \arg\min_{a_*, \mathbf{B}_*} \mathbb{E}\left[\left(Y_* - a_* + \mathbf{B}_*^T \mathbf{Y}_{1:N}\right)^2 | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_*\right], \tag{35}$$

where $a_* \in \mathbb{R}$ and $\mathbf{B}_*^T \in \mathbb{R}^{1 \times N}$

4) Spatial Regional and Level Exceedance estimators:

There are a few ways of characterizing the spatial exceedance of these processes, either through the region or the location of the spatial exceedance given a user defined threshold and the tolerance. The other characteristic is the level at which the random process exceeds the given threshold at a given location. We define these characterizations mathematically as follows:

a) Spatial-Regional Right Tail Exceedance (S-RTE):

$$D_{\mathbf{x}}\left(\mathbf{x}_*; \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, T, \alpha\right) = \{\mathbf{x}_* : \mathrm{P}\left(Y_* > T | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_*\right) \geq 1 - \alpha\}, \tag{36}$$

b) Spatial-Regional Left Tail Exceedance (S-LTE):

$$D_{\mathbf{x}}\left(\mathbf{x}_*; \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, T, \alpha\right) = \{\mathbf{x}_* : \mathrm{P}\left(Y_* < T | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_*\right) \geq 1 - \alpha\}, \tag{37}$$

where T is the pre-defined threshold and α is the user defined tolerance. Here we are interested in the locations at which the above exceedance is observed. The function $D_{\mathbf{x}}\left(\mathbf{x}_*; \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, T, \alpha\right) : \mathcal{X} \mapsto \{0, 1\}$, i.e., it is a binary map in space indicating the locations $\mathbf{x}_*$ at which the process exceeds a pre-defined threshold with a specified probability α. These locations may form points, lines or surfaces in $\mathcal{X} \subseteq \mathbb{R}^2$. It is useful to define exceedances on both tails due to the presence of skewness and kurtosis which affects the tail behaviour asymmetrically.

c) Spatial-Level Exceedance (S-LE):

$$D_\alpha = \inf \left\{ \alpha : \mathrm{P}\left(Y_* > T | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_* \right) \geq 1 - \alpha \right\}. \tag{38}$$

Here we are interested to find the minimum quantile level α at which the process exceeds the given threshold T at a given location $\mathbf{x}_*$. The function $D_\alpha \left(\mathbf{x}_*; \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, T \right) : \mathcal{X} \mapsto (0,1)$, i.e.,

Now that we have specified various metrics by which the field reconstruction can be carried out, we detail each of the estimators.

B. Spatial Field Reconstruction Estimator Derivations

All the aforementioned objective functions except for the S-BLUE require the knowledge of the predictive posterior distribution conditioned on the observations and the sensor locations, given in Lemma 4. We begin by deriving the classical predictive spatial field estimator based on the minimum mean square error for the warped spatial Gaussian process.

Lemma 5 (MMSE spatial field reconstruction estimator [29]).

The MMSE spatial field reconstruction function is given by the mean of the predictive posterior which is of the form similar to the one described in Corollary 3 and is given by:

$$\widehat{Y}_{*|Y_{1:N}}^{MMSE} = \frac{1}{g\sqrt{1 - h\sigma_{p_*}^2}} \left[\exp\left(\frac{h\mu_{p_*}^2 + g^2\sigma_{p_*}^2 + 2\mu_{p_*}g}{2(1 - h\sigma_{p_*}^2)}\right) - \exp\left(\frac{h\mu_{p_*}^2}{2(1 - h\sigma_{p_*}^2)}\right) \right]. \tag{39}$$

The Mean Squared Error (MSE) is given by the variance of the predictive posterior which is of the form similar to the one described in Corollary 4 and is given by:

$$SMSE\left(\mathbf{x}_*\right) = \sigma_{Y*|\mathbf{Y}_{1:N}}^2 = \frac{1}{g^2\sqrt{1 - 2h\sigma_{p_*}^2}} \left[\exp\left(\frac{h\mu_{p_*}^2 + 2\mu_{p_*}g + 2g^2\sigma_{p_*}^2}{1 - 2h\sigma_{p_*}^2}\right) - 2\exp\left(\frac{2h\mu_{p_*}^2 + 2\mu_{p_*}g + g^2\sigma_{p_*}^2}{2\left(1 - 2h\sigma_{p_*}^2\right)}\right) \right.$$
$$\left. + \exp\left(\frac{h\mu_{p_*}^2}{1 - 2h\sigma_{p_*}^2}\right) \right] - (\widehat{Y}_*^{MMSE})^2. \tag{40}$$

Remark 3. *The expression for the uncertainty associated with the MMSE estimate is shown in Equation 40. We can see that the expression depends on the predictive posterior density parameter μ_{p_*}, which in turn depends on the observations $\mathbf{Y}_{1:N}$ as seen from Lemma 4. This is significantly different from the Gaussian process, where the uncertainty does no depend on the observations and only depends on the covariances between the locations at which the data were observed. This influences applications that involve experimental design problems like sensor placement, sensor selection, etc, in that just knowing the information of the locations alone is no longer sufficient*

for the Tukey g-h process and repeated experiments may need to be performed for evaluating the characteristics of just one location set.

Remark 4. *Note that the existence of the MMSE at a location $\mathbf{x}_*$ is given by, $h < \frac{1}{n\sigma^2_{p_*}}$. We also know that $\sigma^2_{p_*} \leq k(\mathbf{x}_*, \mathbf{x}_*)$. This means that for a given value of the parameter h, there may arise a case when the Tukey g-and-h population mean at a location does not exist but the conditional mean can exist. This is because as we gather observations around the location of interest, we reduce $\sigma^2_{p_*}$ and hence the effect of the kurtosis parameter on the conditional MMSE.*

Proposition 5 (MAP spatial field reconstruction estimator:)**.**

The spatial MAP estimator is obtained from the solution to the following optimisation problem:

$$\widehat{Y}^{MAP}_* = \mathcal{T}_{gh}\left(w^o_*; \zeta_{gh}\right) \tag{41}$$

where w^o_* is the solution to the following optimsation problem.

$$w^o_* = \arg\max_{w_*} \frac{\exp\left(-\frac{(w_* - \mu_{p_*})^2}{2\sigma^2_{p_*}}\right)}{hw_* \frac{\exp(gw_*) - 1}{g} \exp\left(\frac{hw^2_*}{2}\right) + \exp\left(gw_* + \frac{hw^2_*}{2}\right)} \tag{42}$$

Proof. See Appendix F □

Remark 5. *The optimisation problem in specified in Proposition 5, to obtain the MAP estimate of the Tukey g-and-h process at a location $\mathbf{x}_*$ does not admit analytical solutions and hence has to be dealt with numerically. To this end, one can solve this one-dimensional optimisation through standard techniques such as Newton methods adopting a simple gradient descent (since we can obtain the derivatives of this expression at all values of w_*).*

Proposition 6 (MAP spatial variance:)**.**

The variance of the spatial MAP estimator is given by the hessian of the posterior GH distribution at a location $\mathbf{x}_*$, evaluated at the optimum value w^o_* obtained from Proposition 5. It is represented as follows:

$$\sigma^2_{*MAP} = \frac{\left(\mathcal{T}'_{gh}\left(w^o_*\right)\right)^2 f''_W\left(w^o_*\right) + 2f'_W\left(w^o_*\right)\mathcal{T}''_{gh}\left(w^o_*\right)\left(\mathcal{T}''_{gh}\left(w^o_*\right) - \mathcal{T}'_{gh}\left(w^o_*\right)\right) - f_W\left(w^o_*\right)\mathcal{T}'''_{gh}\left(w^o_*\right)\mathcal{T}'_{gh}\left(w^o_*\right)}{\left(\mathcal{T}'_{gh}\left(w^o_*\right)\right)^3} \tag{43}$$

where w^o_* is the mode of the GH distribution at $\mathbf{x}_*$ and $f_W(.)$ is the Gaussian pdf with mean and variance as derived in Lemma 4.

Proof. See Appendix G □

We now state a few corollaries for finding the solution to the optimisation problem posed for finding the MAP estimate at any arbitrary location. These are the cases when the parameters $g \to 0$ and $h \to 0$.

Proposition 7 (S-BLUE spatial field reconstruction estimator:)**.**

The Spatial-Best Linear Unbiased Estimator (S-BLUE) at any location $\mathbf{x}_*$ *is given by:*

$$\widehat{Y}_*^{\text{S-BLUE}} = \widehat{a}_* + \widehat{\mathbf{B}}_*^T \left(\mathbf{Y}_{1:N} - \mathbb{E}\left[\mathbf{Y}_{1:N}\right]\right) \tag{44}$$

where $\widehat{a}_* = \mu_{p_*}$ *and* $\widehat{\mathbf{B}} = \mathbb{C}ov^{-1}\left(\mathbf{Y}_{1:N}, \mathbf{Y}_{1:N}\right)\mathbb{C}ov\left(Y_*, \mathbf{Y}_{1:N}\right)$. *In addition,* $\mathbb{E}\left[\mathbf{Y}_{1:N}\right]$ *is just the prior mean of the observations* $\mathbf{Y}_{1:N}$.

The exceedance problem of the Tukey g-and-h process as described in the objective functions, given by Equations 36 and 37 are reduced to the problem of obtaining the multivariate quantile level sets of the process at d locations. First we look at univariate description of the quantile function at just one location. Then we will see that the existence of a tractable quantile Tukey g-and-h function enables us to reduce the exceedance problems to the description of these quantile function along with the user given parameters.

Proposition 8 (Spatial region of right tail exceedance estimator)**.**

The region of right tail exceedance described in Equation 36 is reduced to finding the quantile level sets at the user given tolerance to see if it exceeds the given threshold. This is a binary estimate at location $\mathbf{x}_*$, *that is characterized by:*

$$\widehat{Y}_*^{\text{S-RTE}} = \begin{cases} 1, & \text{if } q_*\left(\alpha\right) > \mathcal{T}_{gh}^{-1}\left(T; \zeta_{gh}\right), \\ 0, & \text{Otherwise} \end{cases} \tag{45}$$

where $q_*\left(\alpha\right) = \mu_{p_*} + \sigma_{p_*}\Phi^{-1}\left(\alpha\right)$ *is the quantile as seen from Lemma 6. The region is created by accumulating all the locations that produce 1 in the binary estimate.*

Remark 6. *We note that derivation of the left tail exceedance spatial field reconstruction estimator follows analogously the result obtained in Proposition 8 with a reversal of the inequality and a quantile level* $(1-\alpha) \in [0, 1]$.

Proposition 9 (Spatial level of exceedance estimator)**.**

The region of exceedance described in Equation 38 is reduced to finding the quantile level sets at the user given tolerance to see if it exceeds the given threshold. This is a binary estimate at location $\mathbf{x}_*$, *that is characterized by:*

$$\widehat{Y}_*^{\text{S-LE}} = \inf\left\{\alpha : w_*\left(\alpha\right) > \mathcal{T}_{gh}^{-1}\left(T; \zeta_{gh}\right)\right\} \tag{46}$$

where $w_*\left(\alpha\right) = \mu_{p_*} + \sigma_{p_*}\Phi^{-1}\left(\alpha\right))$ *as seen in Lemma 6.*

Note that in Propositions 8 and 9, we used $\mathbf{x}_*$ to denote the unobserved spatial field location. In these cases the quantile function is essentially the conditional quantile function from the conditional predictive posterior distribution obtained in Lemma 4.

V. EXPERIMENTAL RESULTS: WARPED GAUSSIAN PROCESS SPATIAL FIELD RECONSTRUCTIONS

In this section, we look at results concerning spatial field reconstruction based on a warped Gaussian process model of the Tukey g-and-h family of spatial processes. We consider both central estimators of spatial field as well as exceedance spatial maps. We compare the warped Gaussian process reconstructions of the Tukey g-and-h processes to those one would obtain had one miss-specified the model and applied a classical standard Gaussian process field reconstruction. It is of interest to assess the ability of the Tukey g-and-h process to detect attributes of spatial skewness and kurtosis, compared to the Gaussian process in a range of different settings. We consider both synthetic as well as real data studies to illustrate the Tukey g-and-h model developed.

A. Simulation Setup

In these case studies we model the physical phenomenon or spatial process that is partially observed by a Tukey transformation of an underlying Gaussian process over a spatial grid where $x \in [-5, 5]$ and $y \in [-5, 5]$. We work with the well known squared exponential kernel with a known unit variance and known length-scale. We also subtract the mean to obtain a zero mean GP. Furthermore, the parameters g, h are assumed to be known and are set to different values within the allowable range, to test various cases. This allows us to demonstrate on a synthetic case study the importance of developing spatial field reconstructions that accommodate warped spatial GP structures for spatial skewness and kurtosis by comparing the loss in accuracy if one applies just a standard GP model.

Moreover, we have N sensors with known locations, spread uniformly in over the spatial grid region of interest. We vary N to study the effect of the number of sensors on different regression estimators. To this end, we use the normalised spatial mean squared error (N-MSE) to characterize the performance of the different estimators and we define it as follows:

$$\text{N-MSE} = \frac{1}{J \times T} \sum_{j=1}^{J} \sum_{i=1}^{T} \frac{\left(\widehat{Y}\left(x_*^{(ij)} \right) - Y\left(x^{(ij)} \right) \right)^2}{maxY - minY}$$

where J is the number of spatial locations to regress and T is the number of trials of the Tukey g-and-h process simulation. In addition, $maxY = \max_{i,j} x^{(ij)}$ is the maximum value of the simulated process across all locations and trials and similarly $minY = \min_{i,j} x^{(ij)}$ is the minimum value of the simulated process.

We outline the following test scenarios studied in this setting.

1) **S**cenario 1 : We compare the performance of the different estimators such as, MMSE (Equation 39), S-MAP (Equation 41), S-BLUE (Equation 35) to the Gaussian Process estimate in various scenarios such as, small g and h with low spatial correlation, high spatial correlation and other such combinations of g, h and l.

2) **S**cenario 2 : We characterize the exceedance estimation ability of the Tukey g-and-h process as shown in Proposition 8 (g,h), through spatial ROC curves compared to a Gaussian process.

3) **S**cenario 3: We adopt real data case studies for the US and we apply our estimators on the noisy observed US precipitation data sets [40]. Wer consider a spatial field reconstruction that involves latitudes between 34 to 43 degress and longitudes between -110 to -100 degrees and we compare the different spatial field reconstruction estimators and their performance on real data versus a classical Gaussian process model.

Before we look at the simulation results, we show the actual differences in the distribution of the Tukey g-and-h finite dimensional distribution with the parameters that are used in the simulation case studies in order to compare to the Gaussian distribution to demonstrate clearly that at each spatial field location one would expect that skewness and kurtosis will be meaningful to consider in constructing the spatial field reconstruction estimators.
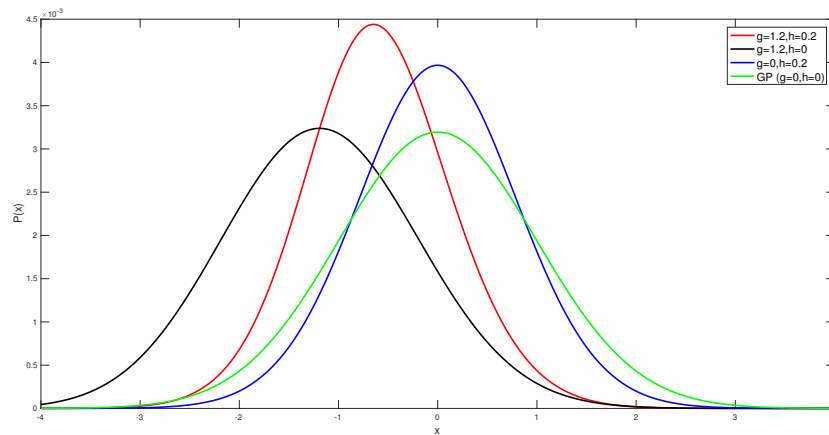

Fig. 1: Compare standard Gaussian distribution to the Tukey g-and-h distributions with values from $\{g = 1.2, h = 0.2, \mu = 0, \sigma^2 = 1\}$.

B. Scenario 1: Spatial Field Reconstruction on Synthetic Data

In this section we present the estimation performance of the Spatial-BLUE estimator, Gaussian Process estimator, Spatial-MAP estimator and the Spatial MMSE estimator. The results here are shown for a synthetic data set constructed with given parameter values for g, h and l, using a squared exponential kernel. The study utilised 50 realisations of the Tukey g-and-h random field and the average spatial Mean Squared Error reported below is the average across the 50 realisations and all the locations in space where the field was reconstructed. To study the

variation across the number of sensors, we used $N = \{9, 16, 25, 36, 49, 64, 81\}$ and spread it uniformly across the field.
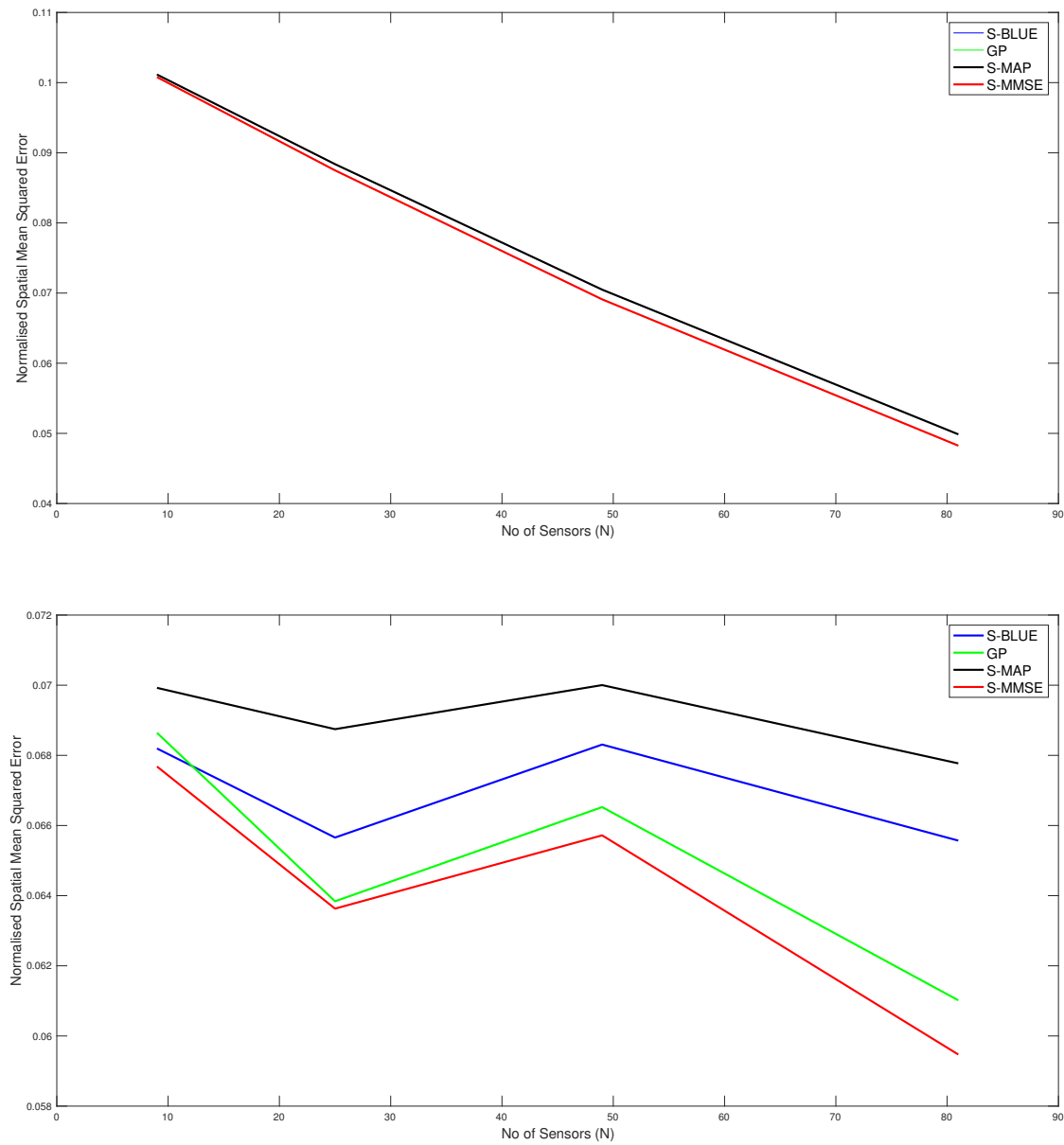

Fig. 2: Plots of comparison of the N-MSE as a function of number of sensors (N). Top Subplot: $g = 0.001$, $h = 0.001$, $l = 0.5$. Bottom Subplot: $g = 1.2$, $h = 0.2$, $l = 0.5$

As we can see in the first top sub-figure in Figure 2, this is the situation in which the true spatial field process is very close to Gaussian as the $g \approx h \approx 0$ settings was used and therefore with both l values only two curves are seen for the different spatial field reconstructions and they are basically almost indistinguishable. The difference

23

reflects the very slight skew and kurtosis still present in the target warped spatial Gaussian process since g and h were not set exactly 0. Basically in this scenario we see that the curve associated with the S-BLUE and the S-MAP coincide, as is the case with the S-MMSE and the GP curves. Furthermore all the estimators perform better as the number of sensors increase. Overall the difference in the reconstruction accuracy is negligible between the four estimators as is expected since the values of g and h are very close to 0 as is to be expected. This is simply a case study to illustrate that our spatial field reconstructions for the warped spatial GP model can recover a classica GP reconstruction if the target spatial field truly has this characterisation.

In the Bottom subplot of Figure 2, we see that with some significant increase in g and h values that adds to the warped spatial GP skewness and kurtosis, then in this case the standard GP spatial field reconstructions degrade in accuracy. The classical GP estimate is unable to account for the significant non-linearities in the transform that induce the skewness and kurtosis in the spatial field, while the other estimators that we have derived for spatial field reconstruction specifically in the setting of warped spatial GP significantly outperform the classical GP reconstructions of comparitative loss functional type, especially as the number of sensors increase. We also observe that the linearisation of the spatial field reconstruction provides an efficient estimator computationally compared to the MMSE estimator but the cost of this computational efficiency is a loss in accuracy due to the attemtp to linearise locally the highly non-linear spatial field warping transforms. We see that it becomes particularly effective to use an efficient linear S-BLUE estimator in settings in which observation coverage is sparse and sensors are placed sparsely in the spatial region of interest. In these cases we show that the S-BLUE estimator performs comparatively with the more computationally expensive MMSE estimators.

C. Scenario 2: Spatial Exceedance on Synthetic Data

In the exceedance examples, we characterize the spatial region of exceedance estimator (Proposition 8) and its left tail equivalent, and the region of exceedanc (Proposition 9). We treat these problems as a detection problem, since we obtain a binary estimate at each location. We characterize such performance via plots of the Receiver Operating Characteristic (ROC) curves associated with different values of N. In these illustrations we use $\alpha = 0.1$ and the threshold λ is varied from the minimum value of the realisations of the Tukey g-and-h field to the maximum value of the realisations. Once the predictive distribution is obtained at the target locations, we calculate the binary estimate of exceedance at each target location and finally to obtain the required probabilities of detection and false alarms, we collate the proportion of false alarms and correct detections at all locations and get the probabilities of false alarm and detection.

From the ROC curves in figure 3 we see that the Tukey g-and-h distribution in different cases outperforms the Gaussian process in all the scenarios created, with closest performance corresponding to cases when the generated synthetic data is from settings of g and h parameters in which the data generating mechanism is closest to no skewness or kurtosis effects.

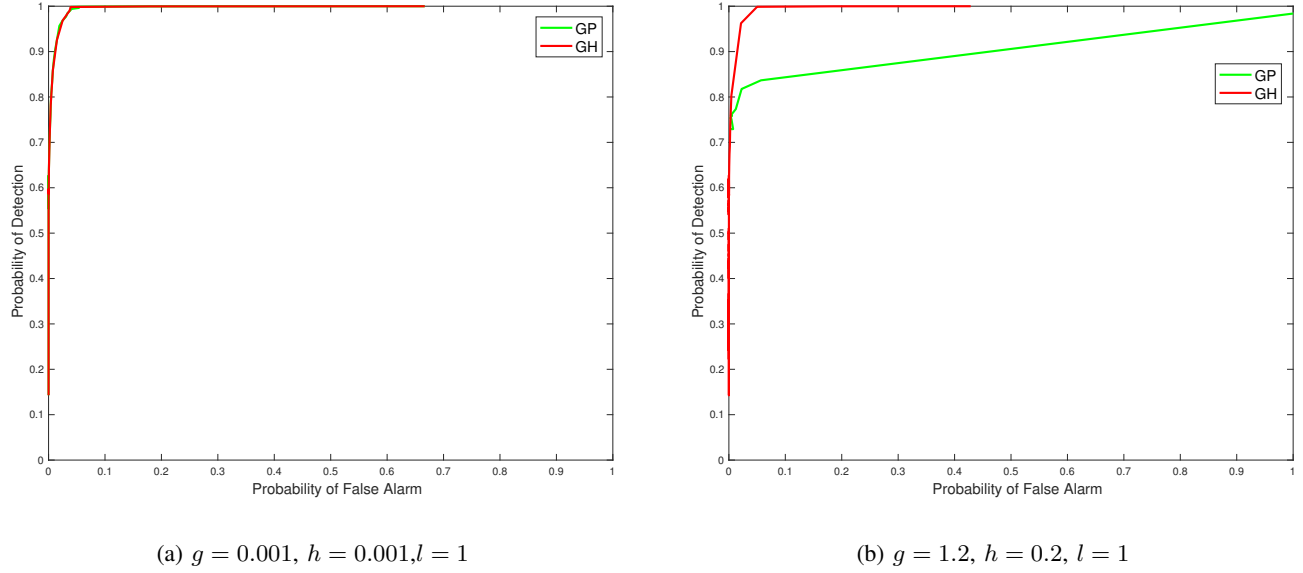(a) $g = 0.001$, $h = 0.001, l = 1$ (b) $g = 1.2$, $h = 0.2$, $l = 1$

Fig. 3: Plots of ROC curves comparing the Tukey g-and-h distribution to the Gaussian process for different values of g and h.

Now the figure 4 contains the ROC curve corresponding to $g = 1.2, h = 0.001$ and $l = 1$. This is the case with high positive skew and almost zero kurtosis. We notice that the false alarm rate for the quantile exceedance is extremely low and specifically we defined the exceedance as the right tail exceedance. This may be due to the fact that the estimation of the Tukey g-and-h posterior at all the target locations depend on the data from the sensors which tend to be considerably positively skewed ($g = 1.2$), without heavy tails. Since, the GP estimation involves the actual observations (without any transformation), the predictive distribution tends to have a mean with a high positive value and since there is the asymmetry and the lack of heavy tails, the GP also exhibits reasonably low false alarm rate for the right tail exceedance that is comparable to the Tukey g-and-h process. However, for the same case if we calculate the left tail exceedance as characterized in Proposition 9, we notice that these ROC curves exhibit a higher range of false alarms and detection rate and we obtain the following figure.

In addition, we observed that when we considered the case $g = -1.2$, $h = 0.001$, $l = 1$ which is negatively skewed with 0 kurtosis, we noticed that the above trend is reversed, i.e, the right tail exceedance exhibited a larger false alarm rate and the left tail exceedance exhibited a very small rate of false alarms.

D. Scenario 3: Spatial Field Reconstruction on Real Spatial Sensor Data for Hourly Precipitation

In this section, we show results of spatial field reconstruction on the US precipitation dataset, available at https://www.ngdc.noaa.gov. This shows the hourly precipitation in digital data set DSI-3240, archived at the National Climatic Data Center (NCDC). It is obtained from US National Weather Service (NWS), Federal Aviation
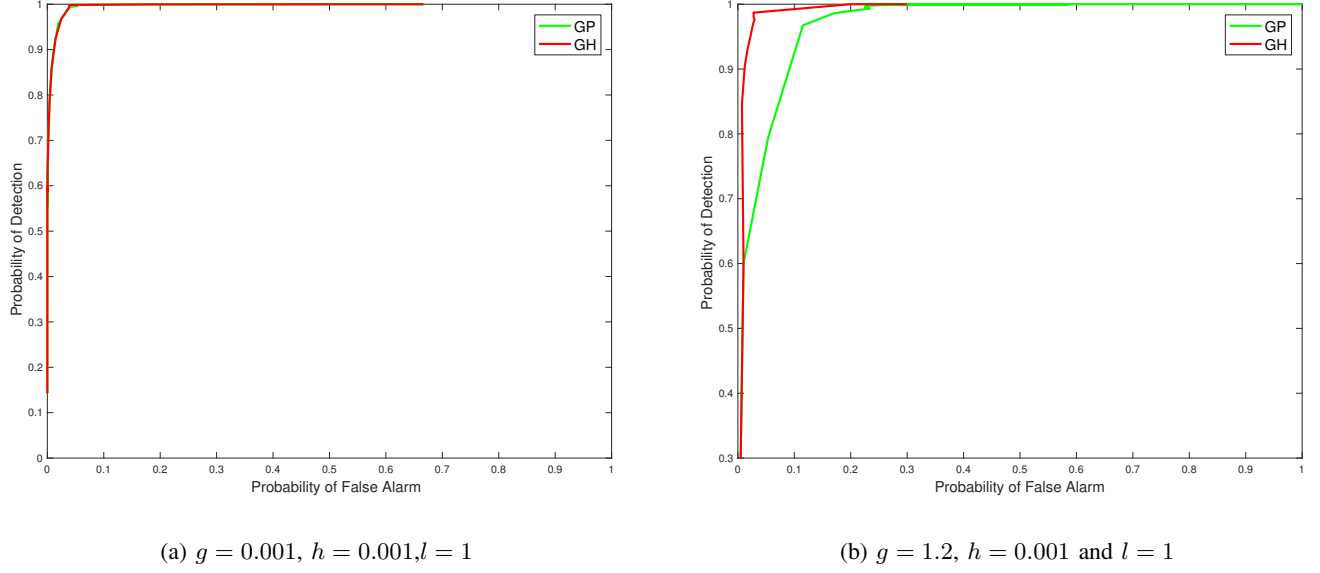
25

(a) $g = 0.001$, $h = 0.001, l = 1$ (b) $g = 1.2$, $h = 0.001$ and $l = 1$

Fig. 4: Plots of ROC curves comparing the Tukey g-and-h distribution to the Gaussian process for different values of g and h.

Administration (FAA), and cooperative observer stations in the United States of America, Puerto Rico, the US Virgin Islands, and various Pacific Islands. We extract data for one month (November 1994) from the sensors within the area shown in the map in Fig 5.

Here, to evaluate the performance of our estimators, we estimate the parameters from the US precipitation dataset corresponding to the data from the above locations. For both the Tukey g-and-h and the GP we use an underlying GP with zero mean and a squared exponential kernel where the length-scale and the variance are estimated appropriately. In addition, The Tukey g-and-h requires estimation of two additional parameters namely the skew parameter g and the kurtosis parameter h. After a maximum likelihood estimation in both cases, we obtain the estimates of the Tukey g-and-h to be $l_{gh} = 0.2$, $\sigma_{gh}^2 = 1$, $g = 0.001$ and $h = 0.2$ and the estimates for the GP are $l_{gp} = 0.1$ and $\sigma_{gp}^2 = 2.5$.

Additionally, we define a normalised version of the median of absolute deviations for the comparison of our estimates on the real data. It is defined as follows:

$$\text{N-MAD} = median_{j \in J, i \in T} \frac{\left| \widehat{Y}\left(x_*^{(ij)} \right) - Y\left(x^{(ij)} \right) \right|}{maxY - minY}$$

where J is the number of spatial locations to regress and T is the number of trials of the Tukey g-and-h process simulation. In addition, $maxY = \max_{i,j} x^{(ij)}$ is the maximum value of the simulated process across all locations and trials and similarly $minY = \min_{i,j} x^{(ij)}$ is the minimum value of the simulated process.
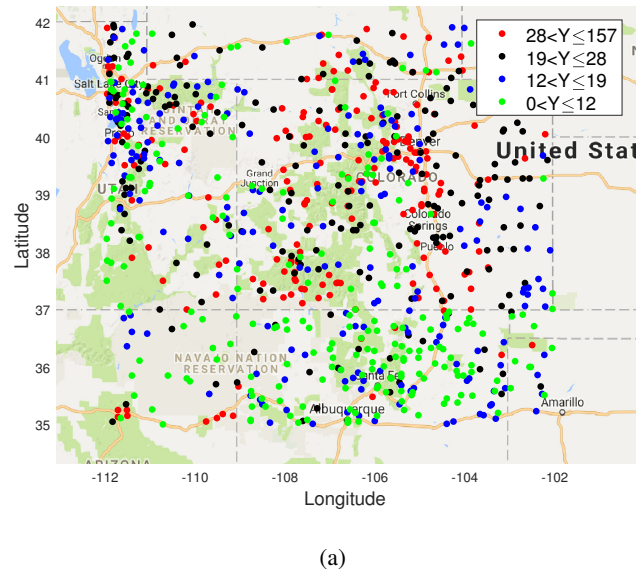
26

(a)

Fig. 5: USA map with sensor locations measuring precipitation in 4 states showing the average precipitation for the month of Nov 1994 at 850 locations. The values (in mm) are colour coded according to the ranges shown.
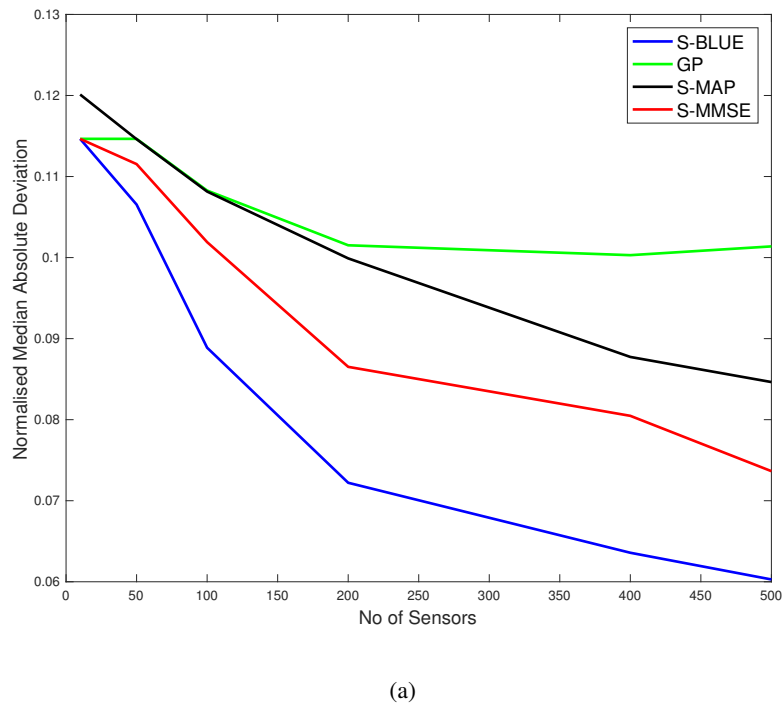

(a)

Fig. 6: The N-MAD comparison between the various Tukey g-and-h estimators and the Gaussian process

We can see from Figure 6, any estimator of the Tukey g-and-h process performs much better than the GP. Particularly in both cases the S-BLUE estimate performs the best in this case and this maybe because in the real data we do not have multiple realisations of the data, thus to get a good estimate we should exploit the spatial information well and the S-BLUE estimate does that precisely, as it requires the spatial Tukey g-and-h covariance, which enables it to produce better estimates and these estimates improve as the number of sensor locations having data increase. Moreover, we notice that the N-MSE of the GP increases as the number of sensors increase and this shows that we are fitting the data with a model that is biased. In Figure 7, we see that the ratio
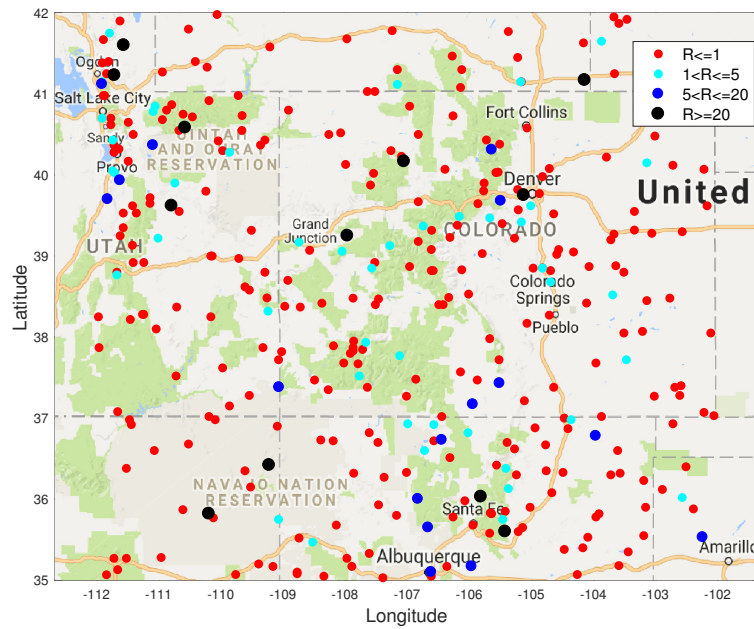
(a)

Fig. 7: The ratio R of the N-MSE of the S-BLUE estimate and the GP estimate at each of the test locations when $N = 500$, i.e, 500 locations were used as the training set.

$R_* = \frac{N-MSE \ of \ the \ BLUE \ estimate \ at \ *}{N-MSE \ of \ the \ GP \ estimate \ at \ *}$, in majority (80 %) of the test locations is less than or equal to 1, which means that the S-BLUE estimator performs better than the GP at those locations. The light blue markers indicated the locations where this ratio is between 1 and 5, which is around 11% of the test data. Similarly the dark blue markers and the black markers represent the small percentage of the locations where the GP performs better than the S-BLUE. We notice that these locations are generally concentrated in the corners of the grid, which might explain the presence of a few outliers as the main factor that affects the estimates is the covariance between the test location and the sensor locations.

VI. CONCLUSION

We study in detail, various aspects regarding Tukey g-and-h processes, from its construction and derivation of n-variate cross moments, which helped us develop three different estimators for spatial regression, the MMSE, S-BLUE and the S-MAP. We discuss their derivation through the predictive posterior distribution. In addition, we define the right tail and the left tail exceedance estimators using the quantile level sets of these processes. Finally, we test the performance of our estimators on a simulated data setting and the US precipitation dataset. We find that although the Gaussian process is a useful model, we show that it suffers when the data exhibits heavy-tailedness and/or skewness, where the Tukey g-and-h process is better suited and is also practically applicable. The flexibility of the construction and the estimation of the Tukey g-and-h process can enable us to incorporate even more complex spatial and temporal patterns through different kernels much like the Gaussian process whilst enabling a way to account for non-Gaussian behaviour which arises in many real world applications.

REFERENCES

[1] J. K. Hart and K. Martinez, "Environmental sensor networks: A revolution in the earth system science?" *Earth-Science Reviews*, vol. 78, no. 3-4, pp. 177–191, 2006.

[2] S. Rajasegarar, T. C. Havens, S. Karunasekera, C. Leckie, J. C. Bezdek, M. Jamriska, A. Gunatilaka, A. Skvortsov, and M. Palaniswami, "High-resolution monitoring of atmospheric pollutants using a system of low-cost sensors," *IEEE transactions on geoscience and remote sensing*, vol. 52, no. 7, pp. 3823–3832, 2014.

[3] A. Kottas, Z. Wang, and A. Rodríguez, "Spatial modeling for risk assessment of extreme values from environmental time series: A bayesian nonparametric approach," *Environmetrics*, vol. 23, no. 8, pp. 649–662, 2012.

[4] D. Murakami, G. W. Peters, Y. Yamagata, and T. Matsui, "Participatory sensing data tweets for micro-urban real-time resiliency monitoring and risk management," *IEEE Access*, vol. 4, pp. 347–372, 2016.

[5] K. Stahl, R. Moore, J. Floyer, M. Asplin, and I. McKendry, "Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density," *Agricultural and Forest Meteorology*, vol. 139, no. 3, pp. 224–236, 2006.

[6] X. Zhang and R. Srinivasan, "Gis-based spatial precipitation estimation: A comparison of geostatistical approaches1," *JAWRA Journal of the American Water Resources Association*, vol. 45, no. 4, pp. 894–906, 2009.

[7] I. Nevat, G. W. Peters, F. Septier, and T. Matsui, "Estimation of spatially correlated random fields in heterogeneous wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2597–2609, 2015.

[8] I. Nevat, G. W. Peters, and I. B. Collings, "Random field reconstruction with quantization in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 6020–6033, 2013.

[9] I. Nevat, G. W. Peters, F. Septier, and T. Matsui, "Wind storm estimation using a heterogeneous sensor network with high and low resolution sensors," in *Communications (ICC), 2015 IEEE International Conference on*.   IEEE, 2015, pp. 4865–4870.

[10] T. L. Thorarinsdottir and T. Gneiting, "Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 173, no. 2, pp. 371–388, 2010.

[11] H. Rajaona, F. Septier, P. Armand, Y. Delignon, C. Olry, A. Albergel, and J. Moussafir, "An adaptive bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release," *Atmospheric Environment*, vol. 122, pp. 748–762, 2015.

[12] C. E. Rasmussen, "Gaussian processes for machine learning."   MIT Press, 2006.

[13] S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang, "Gaussian predictive process models for large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 4, pp. 825–848, 2008.

[14] R. B. Gramacy and D. W. Apley, "Local gaussian process approximation for large computer experiments," *Journal of Computational and Graphical Statistics*, vol. 24, no. 2, pp. 561–578, 2015.

[15] D. Gu and H. Hu, "Spatial gaussian process regression with mobile sensor networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1279–1290, 2012.

[16] N. A. Cressie, *Statistics for spatial data*.   Wiley Online Library, 1993.

[17] T. Gneiting, M. G. Genton, and P. Guttorp, "Geostatistical space-time models, stationarity, separability, and full symmetry," *Monographs On Statistics and Applied Probability*, vol. 107, p. 151, 2006.

[18] M. G. Genton, *Skew-elliptical distributions and their applications: a journey beyond normality*.   CRC Press, 2004.

[19] F. Tagle, S. Castruccio, and M. G. Genton, "A multi-resolution spatial model for large datasets based on the skew-t distribution," *arXiv preprint arXiv:1712.01992*, 2017.

[20] C. Ley, "Flexible modelling in statistics: past, present and future," *Journal de la Société Française de Statistique*, vol. 156, no. 1, pp. 76–96, 2015.

[21] F. J. Rubio, "Modelling of kurtosis and skewness: Bayesian inference and distribution theory," Ph.D. dissertation, University of Warwick, 2013.

[22] F. Galton, "The geometric mean, in vital and social statistics," *Proceedings of the Royal Society of London*, vol. 29, no. 196-199, pp. 365–367, 1879.

[23] F. Y. Edgeworth, "Xlii. the law of error and the elimination of chance," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 21, no. 131, pp. 308–324, 1886.

[24] K. Pearson, "Contributions to the mathematical theory of evolution," *Philosophical Transactions of the Royal Society of London. A*, vol. 185, pp. 71–110, 1894.

[25] G. T. Fechner, *Kollektivmasslehre*.   Engelmann, 1897.

[26] A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian journal of statistics*, pp. 171–178, 1985.

[27] D. C. Hoaglin, S. C. Peters *et al.*, *Software for exploring distribution shape*.   Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 1979.

[28] D. C. Hoaglin, "Summarizing shape numerically: The g-and-h distributions," *Exploring data tables, trends, and shapes*, pp. 461–513, 1985.

[29] G. Xu and M. G. Genton, "Tukey g-and-h random fields," *Journal of the American Statistical Association*, no. just-accepted, 2016.

[30] G. W. Peters, W. Y. Chen, and R. H. Gerlach, "Estimating quantile families of loss distributions for non-life insurance modelling via l-moments," *Risks*, vol. 4, no. 2, p. 14, 2016.

[31] W. Y. Chen, G. W. Peters, R. H. Gerlach, and S. A. Sisson, "Dynamic quantile function models," *arXiv preprint arXiv:1707.02587*, 2017.

[32] G. W. Peters and S. Sisson, "Bayesian inference, monte carlo sampling and operational risk." *Journal of Operational Risk*, 2006.

[33] C. Field and M. G. Genton, "The multivariate g-and-h distribution," *Technometrics*, vol. 48, no. 1, pp. 104–111, 2006.

[34] G. Xu and M. G. Genton, "Tukey g-and-h random fields," *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 1236–1249, 2017.

[35] J. Jeong, Y. Yan, S. Castruccio, and M. G. Genton, "A stochastic generator of global monthly wind energy with tukey g-and-h autoregressive processes," *arXiv preprint arXiv:1711.03930*, 2017.

[36] M. G. Cruz, G. W. Peters, and P. V. Shevchenko, *Fundamental aspects of operational risk and insurance analytics: A handbook of operational risk*.   John Wiley & Sons, 2014.

[37] J. W. Tukey, "Modern techniques in data analysis," in *Proceedings of the NSF-Sponsored Regional Research Conference*. Southern Massachusetts University, 1977.

[38] M. Degen, P. Embrechts, and D. D. Lambrigger, "The quantitative modeling of operational risk: between g-and-h and evt," *ASTIN Bulletin: The Journal of the IAA*, vol. 37, no. 2, pp. 265–291, 2007.

[39] C. Field and M. G. Genton, "The multivariate g-and-h distribution," *Technometrics*, 2012.

[40] "Us precipitation data web-site," 09 2015. [Online]. Available: http://www.image.ucar.edu/Data/US.monthly.met/

APPENDIX A

PROOF OF PROPOSITION 1

To prove that the inverse of the transformation $\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}})$ exists we prove that it is a bijection.

Firstly, we notice that the transformation is continuous in $\mathbb{R}$, since it is essentially the difference of two exponential functions. The Intermediate Value Theorem states that if a function f is continuous in an interval $[a, b]$ and $\min(f(a), f(b)) < L < \max(f(a), f(b))$, then there exists $c \in [a, b]$ such that $f(c) = L$. We apply a version of this theorem to a function f that is continuous in an open interval (a, b).

Now $\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}}) : \mathbb{R} \mapsto \mathbb{R}$, where this function is continuous in $(-\infty, \infty)$. Let $L_1 := \lim_{W \to -\infty} \mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}}) = -\infty$ and $L_2 := \lim_{W \to \infty} \mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}}) = \infty$. Since $L_1 < L < L_2$ for any arbitrary $L \in (-\infty, \infty)$ and since $\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}})$ is continuous in this interval then there exists a $c \in (-\infty, \infty)$ such that $\mathcal{T}_{gh}(c; \zeta_{\mathbf{gh}}) = L$. Hence this function is surjective.

To prove that it is injective, we use the derivative information presented in Lemma 1. The first derivative of the Tukey g-and-h transform is given by:

$$\mathcal{T}'_{gh}(W; \zeta_{\mathbf{gh}}) = hW\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}}) + \exp\left(gW + \frac{hw^2}{2}\right).$$

We know that the kurtosis parameter is always non-negative and hence we need to find the sign of $W\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}})$, since $\exp\left(gW + \frac{hW^2}{2}\right)$ is always positive.

Consider $M := W\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}})$, we then have that

$$\begin{aligned} M &= W\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}}) \\ &= Wg\frac{\exp(gW) - 1}{g^2}\exp\left(\frac{hW^2}{2}\right). \end{aligned} \tag{47}$$

From the above expression to determine the sign of M, we just have to consider the sign of $Wg(\exp(gW) - 1)$. We know that when $Wg > 0$, $\exp(gW) > 1$ and when $Wg < 0$, $\exp(gW) < 1$, and therefore $M > 0$.

Hence the derivative is positive and this means that the transform $\mathcal{T}_{gh}(W; \zeta_{\mathbf{gh}})$ is strictly increasing and as a result it is injective. Since it is surjective and injective, we show that this Tukey g-and-h transformation has an inverse.

APPENDIX B

PROOF OF PROPOSITION 2

We have the random variable $Y := \mathcal{T}_{gh}\left(W; \zeta_{\mathbf{gh}}\right)$, where $\zeta_{\mathbf{gh}} := [m_0, m_1, \mu, \sigma^2, g, h, \theta = 1]$. Then we have that

$$\mathrm{P}\left(Y \leq y\right) = \mathrm{P}\left(\mathcal{T}_{gh}\left(W; \zeta_{\mathbf{gh}}\right) \leq y\right). \tag{48}$$

The limit of the function $\mathcal{T}_{gh}\left(W; \zeta_{\mathbf{gh}}\right)$ as $g \to 0$ and $h \to 0$ is given by:

$$\begin{aligned} L &= \lim_{(g,h)\to(0,0)} \mathcal{T}_{gh}\left(W; \zeta_{\mathbf{gh}}\right) \\ &= m_0 + m_1 \lim_{g\to 0} \frac{W\left(\exp\left(gW\right) - 1\right)}{gW} \lim_{h\to 0} \exp\left(\frac{hW^2}{2}\right). \end{aligned} \tag{49}$$

Let, $L_1 = \frac{W \exp(gW) - 1}{gW}$ and $L_2 = \exp\left(\frac{hW^2}{2}\right)$. Now the functions are separable in g and h, and the limit exists and can be evaluated separately if and only if the limits in g and h exist independently. We know that:

$$\begin{aligned} L_1 &= \lim_{g\to 0} \frac{W\left(\exp\left(gW\right) - 1\right)}{gW} = W, \\ L_2 &= \lim_{h\to 0} \exp\left(\frac{hW^2}{2}\right) = 1. \end{aligned} \tag{50}$$

Both the limits exist independently in g and h. Hence, $L = m_0 + m_1 L_1 L_2 = m_0 + m_1 W$. Also, we know that $W \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Then $L \sim \mathcal{N}\left(m_0 + \mu, m_1^2 \sigma^2\right)$ Finally,

$$\begin{aligned} \mathrm{P}\left(Y \leq y\right) &= \mathrm{P}\left(L \leq y\right) \\ \implies Y &\overset{d}{=} L \end{aligned} \tag{51}$$

As a corollary, the special case when, $m_0 = 0$ and $m_1 = 1$, we get $Y \overset{d}{=} W$

APPENDIX C

PROOF OF PROPOSITION 2

The index of regular variation for the survival function $\overline{F}_Y(y)$ associated with random variable Y is λ where:

$$\lim_{y\to\infty} \frac{y f(y)}{1 - F(y)} = \lambda \tag{52}$$

For a generic transformation $\mathcal{T}_{gh}\left(W; \zeta_{\mathbf{gh}}\right)$ with parameters $\zeta_{\mathbf{gh}} := [m_0 = 0, m_1 = 1, \mu, \sigma^2, g, h, \theta = 1]$, let $f_W(.)$ be the Gaussian pdf of the underlying distribution and $F_W(.)$ be the Gaussian cdf and let $u = \mathcal{T}_{gh}^{-1}(.)$. Then, to find the limit we do as follows:

$$L = \lim_{y\to\infty} \frac{y f_Y(y)}{1 - F_Y(y)} = \frac{f_W(u)}{u\left(1 - F_W(u)\right)\left(\frac{g}{u\left(1 - \frac{1}{\exp(gu)}\right)} + h\right)} \tag{53}$$

Let $L_1 = \frac{f_W(u)}{u(1 - F_W(u))}$ and $L_2 = \frac{1}{\left(\frac{g}{u\left(1 - \frac{1}{\exp(gu)}\right)} + h\right)}$. We get $L_1 = \frac{1}{\sigma^2}$ by repeatedly applying L-Hospitals rule and $L_2 = \frac{1}{h}$. Hence $L = L_1 L_2 = \frac{1}{h\sigma^2}$.

APPENDIX D

PROOF OF LEMMA 3

We begin by writing an explicit expression for the expectation:

$$
\mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}_v\mathbf{w} + \mathbf{u}^T\mathbf{w}\right)\right] = \frac{1}{\sqrt{2\pi}|\boldsymbol{\Sigma}|^{0.5}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(\frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}_v\mathbf{w} + \mathbf{u}^T\mathbf{w}\right)
$$
$$
\times \exp\left(-\frac{1}{2}\left((\mathbf{w}-\mu)^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\mu)\right)\right)\mathbf{dw}.
\tag{54}
$$

Now consider the integrand of the high-dimensional integral shown above. We would like to complete the squares of this term, in order to get an unnormalised multivariate normal pdf, such that we can evaluate this integral with ease. In this regard, we do the following:

$$
\exp\left(\frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}_v\mathbf{w} + \mathbf{u}^T\mathbf{w}\right)\exp\left(-\frac{1}{2}\left((\mathbf{w}-\mu)^T\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\mu)\right)\right)
$$
$$
= \exp\left(-\frac{1}{2}\left(\mathbf{w}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{w} - \mathbf{w}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Sigma}}\left(\mathbf{u}+\boldsymbol{\Sigma}^{-1}\mu\right) - \left(\mathbf{u}^T + \mu^T\boldsymbol{\Sigma}^{-1}\right)\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{w} + \mu^T\boldsymbol{\Sigma}^{-1}\mu\right)\right)
$$
$$
= \exp\left(-\frac{1}{2}\left(\mathbf{w}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{w} - \mathbf{w}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mu} - \widetilde{\mu}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{w} + \mu^T\boldsymbol{\Sigma}^{-1}\mu + \widetilde{\mu}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mu} - \widetilde{\mu}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mu}\right)\right)
$$
$$
= \exp\left(-\frac{1}{2}\left((\mathbf{w}-\widetilde{\mu})^T\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{w}-\widetilde{\mu})\right)\right)\exp\left(-\frac{1}{2}\left(\mu^T\boldsymbol{\Sigma}^{-1}\mu - \widetilde{\mu}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mu}\right)\right).
\tag{55}
$$

Now substituting this results into the integral we obtain that:

$$
\mathbb{E}\left[\exp\left(\frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}_v\mathbf{w} + \mathbf{u}^T\mathbf{w}\right)\right] = \frac{|\widetilde{\boldsymbol{\Sigma}}|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left[\mu^T\boldsymbol{\Sigma}^{-1}\mu - \widetilde{\mu}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\mu}\right]\right),
\tag{56}
$$

where

$$
\widetilde{\boldsymbol{\Sigma}} = \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_v\right)^{-1}
$$
$$
\widetilde{\mu} = \widetilde{\boldsymbol{\Sigma}}\left(\mathbf{u} + \boldsymbol{\Sigma}^{-1}\mu\right).
\tag{57}
$$

APPENDIX E

PROOF OF PROPOSITION 3

We begin by using Equation 3

$$
\begin{aligned}
\mu_{gh}^{(n_1,\ldots,n_m)} &= \mathbb{E}\left[\prod_{j=1}^{m}\left(\mathcal{T}_{gh}\left(W\left(\mathbf{s}_j\right);\zeta\right)\right)^{n_j}\right] \\
&= \frac{1}{g^{(n_1+\ldots+n_m)}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\exp\left(gw_1+\frac{hw_1^2}{2}\right)-\exp\left(\frac{hw_1^2}{2}\right)\right)^{n_1}\cdots \\
&\quad\left(\exp\left(gw_m+\frac{hw_m^2}{2}\right)-\exp\left(\frac{hw_m^2}{2}\right)\right)^{n_m}\mathbf{dw} \\
&= \frac{1}{g^{(n_1+\ldots+n_m)}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\sum_{i_1=0}^{n_1}(-1)^{i_1}\binom{n_1}{i_1}\left(\exp\left(gw_1+\frac{hw_1^2}{2}\right)\right)^{n_1-i_1}\left(\exp\left(\frac{hw_1^2}{2}\right)\right)^{i_1}\cdots \\
&\quad\sum_{i_m=0}^{n_m}(-1)^{i_m}\binom{n_m}{i_m}\left(\exp\left(gw_m+\frac{hw_m^2}{2}\right)\right)^{n_m-i_m}\left(\exp\left(\frac{hw_m^2}{2}\right)\right)^{i_m}\mathbf{dw} \\
&= \frac{1}{g^{(n_1+\ldots+n_m)}}\sum_{i_1=0}^{n_1}\cdots\sum_{i_m=0}^{n_m}(-1)^{(i_1+\ldots+i_m)}\binom{n_1}{i_1}\cdots\binom{n_m}{i_m} \\
&\quad\mathbb{E}\left[\exp\left((n_1-i_1)gw_1+\ldots+(n_m-i_m)gw_m+\frac{n_1h}{2}w_1^2+\ldots+\frac{n_mh}{2}w_m^2\right)\right]
\end{aligned}
\tag{58}
$$

The last line can be obtained from Lemma 3, using the following substitution

$$
\mathbf{u}^{(i)} = [g(n_1-i_1), g(n_2-i_2),\ldots,g(n_m-i_m)]^T
\tag{59}
$$

$$
\mathbf{\Sigma}_v = \begin{pmatrix} hn_1 & 0 & \cdots & 0 \\ 0 & hn_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & hn_m \end{pmatrix}
\tag{60}
$$

Which gives us the following result:

$$
\begin{aligned}
\mu_{gh}^{(n_1,\ldots,n_m)} &= \mathbb{E}\left[\prod_{j=1}^{m}\left(\mathcal{T}_{gh}\left(W\left(\mathbf{s}_j\right);\zeta\right)\right)^{n_j}\right] \\
&= \frac{1}{g^{(n_1+n_2+\ldots+n_m)}}\sum_{i_1=0}^{n_1}\cdots\sum_{i_m=0}^{n_m}(-1)^{(i_1+i_2+\ldots+i_m)}\frac{|\widetilde{\mathbf{\Sigma}}|^{\frac{1}{2}}}{|\mathbf{\Sigma}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left[\mu^T\mathbf{\Sigma}_{\mathbf{s}}^{-1}\mu-\left(\widetilde{\mu}^{(i)}\right)^T\widetilde{\mathbf{\Sigma}}^{-1}\widetilde{\mu}^{(i)}\right]\right)
\end{aligned}
\tag{61}
$$

where

$$
\widetilde{\mathbf{\Sigma}} = \left(\mathbf{\Sigma}^{-1}-\mathbf{\Sigma}_v\right)^{-1}
\tag{62}
$$

$$
\widetilde{\mu}^{(\mathbf{i})} = \widetilde{\mathbf{\Sigma}}\left(\mathbf{u}^{(i)}+\mathbf{\Sigma}^{-1}\mu\right)
\tag{63}
$$

34

APPENDIX F

PROOF OF PROPOSITION 5

$$\widehat{Y}_*^{\text{MAP}} = \arg\max_{y^*} p\left(Y_* | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_*\right)$$

From Lemma 4, we understand that the predictive posterior density $p\left(Y_* | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_*\right)$ is given by:

$$Y_* | \mathbf{Y}_{1:N}, \mathbf{x}_{1:N}, \mathbf{x}_* \sim \mathcal{GH}\left(y_*, \zeta_{gh}\right), \tag{64}$$

where $\zeta_{gh} := \left[\mu_p, \sigma_p^2, g, h, 1\right]$. Now the MAP estimate boils down to finding the maximum of this GH distribution given by the g,h transformation of the underlying normal distribution with mean μ_p and σ_p^2. We denote the Gaussian pdf and cdf by $f_W\left(w\right)$ and $F_W\left(w\right)$ respectively. Also $Y = \mathcal{T}_{gh}\left(W; \zeta_{gh}\right)$, where $\zeta_{gh} := \left[m_0 = 0, m_1 = 1\mu_p, \sigma_p^2, g, h, 1\right]$. Let the transformed pdf and cdf be represented by $f_Y\left(y\right)$ and $F_Y\left(y\right)$. As a shorthand we are going to refer to the transformation as $\mathcal{T}_{gh}\left(w\right)$.

$$
\begin{aligned}
\widehat{Y}_*^{\text{MAP}} &= \arg\max_{y_*} f_{Y_*}\left(y_*\right) \\
&= \arg\max_{y_*} \frac{d}{dy_*} F_{Y_*}\left(y_*\right) \\
&= \arg\max_{y_*} \frac{d}{dy_*} F_{W_*}\left(\mathcal{T}_{gh}^{-1}\left(y_*\right)\right) \\
&= \arg\max_{y_*} f_{W_*}\left(\mathcal{T}_{gh}^{-1}\left(y_*\right)\right) \frac{d}{dy_*} \mathcal{T}_{gh}^{-1}\left(y_*\right)
\end{aligned}
\tag{65}
$$

We know that $\mathcal{T}_{gh}\left(\mathcal{T}_{gh}^{-1}\left(y_*\right)\right) = y_*$. Using this property and differentiating on both sides we get:

$$\frac{d}{dy_*} \mathcal{T}_{gh}^{-1}\left(y_*\right) = \frac{1}{\mathcal{T}_{gh}'\left(\mathcal{T}_{gh}^{-1}\left(y_*\right)\right)}$$

Substituting this back in the Equation 65, we get:

$$\widehat{Y}_*^{\text{MAP}} = \arg\max_{y_*} \frac{f_{W_*}\left(\mathcal{T}_{gh}^{-1}\left(y_*\right)\right)}{\mathcal{T}_{gh}'\left(\mathcal{T}_{gh}^{-1}\left(y_*\right)\right)} \tag{66}$$

$$= \arg\max_{w_*} \frac{\exp\left(\frac{-(w_* - \mu_{p_*})^2}{2\sigma_{p_*}^2}\right)}{hw_* \frac{\exp(gw_*) - 1}{g} \exp\left(\frac{hw_*^2}{2}\right) + \exp\left(gw_* + \frac{hw_*^2}{2}\right)} \tag{67}$$

where $w_* = \mathcal{T}_{gh}^{-1}\left(y_*\right)$, since $\mathcal{T}_{gh}^{-1}$ is a strictly increasing transformation, maximising it in w_* is equivalent to maximising $\frac{f_{W_*}\left(\mathcal{T}_{gh}^{-1}(y_*)\right)}{\mathcal{T}_{gh}'\left(\mathcal{T}_{gh}^{-1}(y_*)\right)}$.

APPENDIX G

PROOF OF PROPOSITION 6

The uncertainty is given by the second derivative of the GH distribution at location $\mathbf{s}_*$ evaluated at the mode of the distribution w_*^o, which is obtained by solving the optimisation problem in Proposition 5. Our goal is to get the following quantity:

$$\sigma_{*MAP}^2 = \frac{d^2}{dw_*^2}\left(\frac{f_W(w_*)}{\mathcal{T}_{gh}'(w_*)}\right)\bigg|_{w_*^o} \tag{68}$$

$$\frac{d}{dw_*}\left(\frac{f_W(w_*)}{\mathcal{T}_{gh}'(w_*)}\right) = \frac{\mathcal{T}_{gh}'(w_*)f_W'(w_*) - f_W(w_*)\mathcal{T}_{gh}''(w_*)}{\left(\mathcal{T}_{gh}'(w_*)\right)^2}$$

$$\frac{d^2}{dw_*^2}\left(\frac{f_W(w_*)}{\mathcal{T}_{gh}'(w_*)}\right) = \frac{d}{dw_*}\left(\frac{f_W'(w_*)}{\mathcal{T}_{gh}'(w_*)}\right) - \frac{d}{dw_*}\left(\frac{f_W(w_*)\mathcal{T}_{gh}''(w_*)}{\left(\mathcal{T}_{gh}'(w_*)\right)^2}\right)$$

$$= \frac{\left(\mathcal{T}_{gh}'(w_*)\right)^3 f_W''(w_*) - 2f_W'(w_*)\mathcal{T}_{gh}''(w_*)\left(\mathcal{T}_{gh}'(w_*)\right)^2 + 2f_W'(w_*)\left(\mathcal{T}_{gh}''(w_*)\right)^2\mathcal{T}_{gh}'(w_*)}{\left(\mathcal{T}_{gh}'(w_*)\right)^4}$$

$$- \frac{f_W'(w_*)\mathcal{T}_{gh}'''(w_*)\left(\mathcal{T}_{gh}'(w_*)\right)^2}{\left(\mathcal{T}_{gh}'(w_*)\right)^4}$$

$$= \frac{\left(\mathcal{T}_{gh}'(w_*)\right)^2 f_W''(w_*) + 2f_W'(w_*)\mathcal{T}_{gh}''(w_*)\left[\mathcal{T}_{gh}''(w_*) - \mathcal{T}_{gh}'(w_*)\right] - f_W(w_*)\mathcal{T}_{gh}'''(w_*)\mathcal{T}_{gh}'(w_*)}{\left(\mathcal{T}_{gh}'(w_*)\right)^3}$$

$$\tag{69}$$

Hence evaluating at w_*^o, we get:

$$\sigma_{*MAP}^2 = \frac{\left(\mathcal{T}_{gh}'(w_*^o)\right)^2 f_W''(w_*^o) + 2f_W'(w_*^o)\mathcal{T}_{gh}''(w_*^o)\left[\mathcal{T}_{gh}''(w_*^o) - \mathcal{T}_{gh}'(w_*^o)\right] - f_W(w_*^o)\mathcal{T}_{gh}'''(w_*^o)\mathcal{T}_{gh}'(w_*^o)}{\left(\mathcal{T}_{gh}'(w_*^o)\right)^3} \tag{70}$$

APPENDIX H

PROOF OF LEMMA 6

Let us consider the Tukey g-h process at a location $\mathbf{x}_k$, then $Y_k := \mathcal{T}_{gh}(W_k; \zeta_{gh})$, where $W_k \sim \mathcal{N}(a_k, \mathcal{C}(\mathbf{x}_k, \mathbf{x}_k))$.

$$F_{Y_k}(y) = \mathrm{P}(Y_k \leq y)$$

$$= \mathrm{P}(\mathcal{T}_{gh}(W_k; \zeta_{gh}) \leq y)$$

$$= \mathrm{P}\left(W_k \leq \mathcal{T}_{gh}^{-1}(y; \zeta_{gh})\right) \tag{71}$$

$$\implies \alpha = F_{W_k}\left(\mathcal{T}_{gh}^{-1}(y; \zeta_{gh})\right)$$

Rewriting the above equation in terms of the quantile function we get:

$$Q_{W_k}(\alpha) = \mathcal{T}_{gh}^{-1}(y; \zeta_{gh})$$

$$y = \mathcal{T}_{gh}(Q_{W_k}(\alpha); \zeta_{gh})$$

$$Q_{Y_k}(\alpha) = \mathcal{T}_{gh}(Q_{W_k}(\alpha); \zeta_{gh}) \tag{72}$$

$$= \mathcal{T}_{gh}\left(a_k + \sqrt{\mathcal{C}(\mathbf{x}_k, \mathbf{x}_k)}\Phi^{-1}(\alpha); \zeta_{gh}\right)$$

$$= \mathcal{T}_{gh}(w_k(\alpha); \zeta_{gh})$$

We can apply the same proof to the Tukey g-h process at location $\mathbf{x}_*$, where we replace $W_* \sim \mathcal{N}\left(\mu_{p_*}, \sigma_{p_*}^2\right)$, by the predictive posterior parameters, which gives us the conditional quantile function.

Lemma 6 (Marginal and Conditional Quantile functions of the Tukey g-and-h process)**.**

1) *Marginal Quantile Function:*

   The univariate marginal quantile function of the Tukey g-and-h process at any arbitrary location $\mathbf{x}_k$ is described as follows:

   $$Q_{Y_k}(\alpha) = \mathcal{T}_{gh}(q_k(\alpha); \zeta_{gh}) \tag{73}$$

   where $q_k(\alpha) = a_k + \sqrt{\mathcal{C}(\mathbf{x}_k, \mathbf{x}_k)}\Phi^{-1}(\alpha)$, and $\Phi^{-1}(\alpha)$ denotes the inverse cdf of the standard normal distribution.

2) *Conditional Quantile Function:*

   The univariate conditional quantile function of the Tukey g-and-h process at any arbitrary location $\mathbf{x}_*$ is described as follows:

   $$Q_{Y_*|Y_{1:N}}(\alpha) = \mathcal{T}_{gh}(q_*(\alpha); \zeta_{gh}) \tag{74}$$

   where $q_*(\alpha) = \mu_{p_*} + \sigma_{p_*}\Phi^{-1}(\alpha)$. Also, $\Phi^{-1}(\alpha)$ is the inverse cdf of the standard normal distribution.