



Universiteit
Leiden
The Netherlands

Dancing with the stars

Albert, J.G.

Citation

Albert, J. G. (2020, October 28). *Dancing with the stars*. Retrieved from <https://hdl.handle.net/1887/137988>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/137988>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/137988> holds various files of this Leiden University dissertation.

Author: Albert, J.G.

Title: Dancing with the stars

Issue Date: 2020-10-28



QUESTIONS TWO

A PROBABILISTIC APPROACH
TO DIRECTION-DEPENDENT
IONOSPHERIC
CALIBRATION

J.G. ALBERT
M.S.S.L. OEI
R.J. VAN WEEEREN
H.T. INTEMA
AND
H.J.A. RÖTTGERING

A probabilistic approach to direction-dependent ionospheric calibration

J. G. ALBERT, M. S. S. L. OEI, R. J. VAN WEEREN, H. T. INTEMA, AND H. J. A. RÖTTGERING

Published in A&A

Received 12 April 2019 / Accepted 10 October 2019

Calibrating for direction-dependent ionospheric distortions in visibility data is one of the main technical challenges that must be overcome to advance low-frequency radio astronomy. In this paper, we propose a novel probabilistic, tomographic approach that utilises Gaussian processes to calibrate direction-dependent ionospheric phase distortions in low-frequency interferometric data. We suggest that the ionospheric free electron density can be modelled to good approximation by a Gaussian process restricted to a thick single layer, and show that under this assumption the differential total electron content must also be a Gaussian process. We perform a comparison with a number of other widely successful Gaussian processes on simulated differential total electron contents over a wide range of experimental conditions, and find that, in all experimental conditions, our model is better able to represent observed data and generalise to unseen data. The mean equivalent source shift imposed by our predictive errors are half as large as those of the best competitor model. We find that it is possible to partially constrain the hyperparameters of the ionosphere from sparse-and-noisy observed data. Our model provides an alternative explanation for observed phase structure functions deviating from Kolmogorov's five-thirds turbulence, turnover at high baselines, and diffractive scale anisotropy. We show that our model performs tomography of the free electron density both implicitly and cheaply. Moreover, we find that even a fast, low-resolution approximation of our model yields better results than the best alternative Gaussian process, implying that the geometric coupling between directions and antennae is a powerful prior that should not be ignored.

Acknowledgements. J. G. A. and H. T. I. acknowledge funding by NWO under 'Nationale Roadmap Grootchalige Onderzoeksfaciliteiten', as this research is part of the NL SKA roadmap project. J. G. A. and H. J. A. R. acknowledge support from the ERC Advanced Investigator programme NewClusters 321271. R. J. vW. and M. S. S. L. O. acknowledge support of the VIDI research programme with project number 639.042.729, which is financed by the Netherlands Organisation for Scientific Research (NWO). M. S. S. L. O. thanks Jesse van Oostrum for helpful discussions.

If we know that our individual errors and fluctuations follow the magic bell-shaped curve exactly then the resulting estimates are known to have almost all the nice properties that people have been able to think of.

JOHN W. TUKEY, 1965

2.1 Introduction

Since the dawn of low-frequency radio astronomy, the ionosphere has been a confounding factor in the interpretation of radio data. This is because the ionosphere has a spatially and temporally varying refractive index, which perturbs the radio-frequency radiation that passes through it. This effect becomes more severe at lower frequencies; see [e.g. de Gasperin et al., 2018]. The functional relation between the sky brightness distribution – the image – and interferometric observables – the visibilities – is given by the radio interferometry measurement equation [RIME; Hamaker et al., 1996], which models the propagation of radiation along geodesics from source to observer as an ordered set of linear transformations [Jones, 1941].

A mild ionosphere will act as a weak-scattering layer resulting in a perturbed inferred sky brightness distribution, analogous to the phenomenon of seeing in optical astronomy [Wolf, 1969]. Furthermore, the perturbation of a geodesic coming from a bright source will deteriorate the image quality far more than geodesics coming from faint sources. Therefore, the image-domain effects of the ionosphere can be dependent on the distribution of bright sources on the celestial sphere, that is they can be heteroscedastic. This severely impacts experiments which require sensitivity to faint structures in radio images. Such studies include the search for the epoch of reionisation [e.g. Patil et al., 2017], probes of the morphology of extended galaxy clusters [e.g. van Weeren et al., 2019], efforts to detect the synchrotron cosmic web [e.g. Vernstrom et al., 2017], and analyses of weak gravitational lensing in the radio domain [e.g. Harrison et al., 2016]. Importantly, these studies were among the motivations for building the next generation of low-frequency radio telescopes like the Low Frequency Array (LOFAR), Murchison Widefield Array (MWA), and the future Square Kilometre Array (SKA). Therefore, it is of great relevance to properly calibrate the ionosphere.

Efforts to calibrate interferometric visibilities have evolved over the years from single-direction, narrow-band, narrow-field-of-view techniques [Cohen, 1973], to more advanced multi-directional, wide-band, wide-field methods [e.g. Kazemi et al., 2011, van Weeren et al., 2016, Tasse et al., 2018]. The principle underlying these calibration schemes is that if you start with a rough initial model of the true sky brightness distribution, then you can calibrate against this model and generate an improved sky brightness model. One can then repeat this process for iterative improvement. Among the direction-dependent calibration techniques the most relevant for this paper is facet-based calibration, which applies the single-direction method to piece-wise independent patches of sky called facets. This scheme is possible if there are enough compact bright sources – calibrators – and if sufficient computational resources are available. Ultimately, there are a finite number of calibrators in a field of view and additional techniques must be considered to calibrate all the geodesics involved in the RIME. We note that there are other schemes for ionosphere calibration that do not apply the

facet-based approach, such as image domain warping [Hurley-Walker et al., 2017].

There are two different approaches for calibrating all geodesics involved in the RIME. The first approach is to model the interferometric visibilities from first principles and then solve the joint calibration-and-imaging inversion problem. This perspective is the most fundamental; however, applications [e.g. Bouman et al., 2016] of this type are very rare and often restricted to small data volumes due to exploding computational complexity. However, we argue that investing research capital – in small teams to minimize risk – could be fruitful and disrupt the status quo [Wu et al., 2019]. The second approach is to treat the piecewise independent calibration solutions as data and predict calibration solutions for missing geodesics [e.g. Intema et al., 2009, van Weeren et al., 2016, Tasse et al., 2018]. In this paper, we consider an inference problem of the second kind.

In order to perform inference for the calibration along missing geodesics, a prior must be placed on the model. One often-used prior is that the Jones operators are constant over some solution interval. For example, in facet-based calibration the implicit prior is that two geodesics passing through the same facet and originating from the same antenna have the same calibration – which can be thought of a nearest-neighbour interpolation. One often-neglected prior is the 3D correlation structure of the refractive index of the ionosphere. An intuitive motivation for considering this type of prior is as follows: The ionosphere has some intrinsic 3D correlation structure, and since cosmic radio emission propagates as spatially coherent waves. It follows that the correlation structure of the ionosphere should be present in ground-based measurements of the electric field correlation – the visibilities. The scope of this paper is therefore to build the mathematical prior corresponding to the above intuition.

We arrange this paper by first reviewing some properties of the ionosphere and its relation to interferometric visibilities via differential total electron content in Section 2.2. In Section 2.3, we then introduce a flexible model for the free electron density based on a Gaussian process restricted to a layer. We derive the general relation between the probability measure for free electron density and differential total electron content, and use this to form a strong prior for differential total electron content along missing geodesics. In Section 2.4 we describe a numerical experiment wherein we test our model against other widely successful Gaussian-process models readily available in the literature. In Section 2.5 we show that our prior outperforms the other widely successful priors in all noise regimes and levels of data sparsity. Furthermore, we show that we are able to hierarchically learn the prior from data. In Section 2.6 we provide a justification for the assumptions of the model, and show the equivalence with tomographic inference.

2.2 Ionospheric effects on interferometric visibilities

The telluric ionosphere is formed by the geomagnetic field and a turbulent low-density plasma of various ion species, with bulk flows driven by extreme ultraviolet solar radiation [Kivelson and Russell, 1995]. Spatial irregularities in the free electron density (FED) n_e and magnetic field \mathbf{B} of the ionosphere give rise to a variable refractive index n , described by the Appleton-Hartree equation [Cargill, 2007] – here given in a Taylor series expansion to order $O(\nu^{-5})$:

$$n(\mathbf{x}) \approx 1 - \frac{\nu_p^2(\mathbf{x})}{2\nu^2} \pm \frac{\nu_H(\mathbf{x})\nu_p^2(\mathbf{x})}{2\nu^3} - \frac{\nu_p^4(\mathbf{x}) - 4\nu_H^2(\mathbf{x})\nu_p^2(\mathbf{x})}{8\nu^4}. \quad (2.1)$$

Here $\nu_p(\mathbf{x}) = \left(\frac{n_e(\mathbf{x})q^2}{4\pi^2\epsilon_0 m} \right)^{1/2}$ is the plasma frequency, $\nu_H(\mathbf{x}) = \frac{B(\mathbf{x})q}{2\pi m}$ is the gyro frequency, ν is the frequency of radiation, q is the elementary charge, ϵ_0 is the vacuum permittivity, and m is the effective electron mass. This form of the Appleton-Hartree equation assumes that the ionospheric plasma is cold and collisionless, that the magnetic field is parallel to the radiation wavevector, and that $\nu \gg \max\{\nu_p, \nu_H\}$. The plus symbol corresponds to the left-handed circularly polarised mode of propagation, and the minus symbol corresponds to the right-handed equivalent. Going forward, we will only consider up to second-order effects, and therefore neglect all effects of polarisation in forthcoming analyses.

In the regime where refractive index variation over one wavelength is small, we can ignore diffraction and interference, or equivalently think about *wave* propagation as *ray* propagation [e.g. Koopmans, 2010]. This approximation is known as the Jeffreys-Wentzel-Kramers-Brillouin approximation [Jeffreys, 1925], which is equivalent to treating this as a scattering problem, and assuming that the scattered wave amplitude is much smaller than the incident wave amplitude – the weak scattering limit [e.g. Yeh, 1962, Wolf, 1969]. Light passing through a varying refractive index n will accumulate a wavefront phase proportional to the path length of the geodesic traversed. Let $\mathcal{R}_x^{\hat{\mathbf{k}}}$ be a functional of n , so that the geodesic $\mathcal{R}_x^{\hat{\mathbf{k}}}[n]: [0, \infty) \rightarrow \mathbb{R}^3$ maps from some parameter s to points along it. The geodesic connects an Earth-based spatial location \mathbf{x} to a direction on the celestial sphere, indicated by unit vector $\hat{\mathbf{k}}$. The accumulated wavefront phase along the path is then given by

$$\phi_x^{\hat{\mathbf{k}}} = \frac{2\pi\nu}{c} \int_0^\infty n\left(\mathcal{R}_x^{\hat{\mathbf{k}}}[n](s)\right) - 1 \, ds, \quad (2.2)$$

where c is the speed of light *in vacuo*. Hamilton's principle of least-action states that geodesics are defined by paths that extremise the total variation of Eq. 2.2.

By substituting Eq. 2.1 into Eq. 2.2, and by considering terms up to second order in ν^{-1} only, we find that the phase deviation induced by the ionosphere is proportional to the integral of the FED along the geodesic, $\phi_x^{\hat{\mathbf{k}}} \approx \frac{-q^2}{4\pi\epsilon_0 mc\nu} \tau_x^{\hat{\mathbf{k}}}$, where,

$$\tau_x^{\hat{\mathbf{k}}} \triangleq \int_0^\infty n_e\left(\mathcal{R}_x^{\hat{\mathbf{k}}}[n](s)\right) ds. \quad (2.3)$$

Equation 2.3 defines the total electron content (TEC).

In radio interferometry, the RIME states that the visibilities, being a measure of coherence, are insensitive to unitary transformations of the electric field associated with an electromagnetic wave. Thus, the phase deviation associated with a geodesic is a relative quantity, usually referenced to the phase deviation from another fixed parallel geodesic – the origin of which is called the reference antenna. Going forward we use Latin subscripts to specify geodesics with origins at an antenna location; for example $\mathcal{R}_i^{\hat{\mathbf{k}}}[n]$ is used as shorthand for $\mathcal{R}_{x_i}^{\hat{\mathbf{k}}}[n]$. Correspondingly, we introduce the notion of differential total electron content (ΔTEC),

$$\tau_{ij}^{\hat{\mathbf{k}}} \triangleq \tau_i^{\hat{\mathbf{k}}} - \tau_j^{\hat{\mathbf{k}}}, \quad (2.4)$$

which is the TEC of $\mathcal{R}_i^{\hat{\mathbf{k}}}[n]$ relative to $\mathcal{R}_j^{\hat{\mathbf{k}}}[n]$.

2.3 Probabilistic relation between FED and Δ TEC: Gaussian process layer model

In this section we derive the probability distribution of Δ TEC given a specific probability distribution for FED. It helps to first introduce the concept of the ray integral (RI) and the corresponding differenced ray integral (DRI). The RI is defined by the linear operator $G_i^{\mathbf{k}}: \mathcal{V} \rightarrow \mathbb{R}$ mapping from the space of all scalar-valued functions over \mathbb{R}^3 to a scalar value according to,

$$G_i^{\mathbf{k}} f \triangleq \int_0^\infty f(\mathcal{R}_i^{\mathbf{k}}[n](s)) ds, \quad (2.5)$$

where $f \in \mathcal{V} = \{g \mid \int_{\mathbb{R}^3} g^2(\mathbf{x}) d\mathbf{x} < \infty\}$. Thus, an RI simply integrates a scalar field along a geodesic. The DRI $\Delta_{ij}^{\mathbf{k}}: \mathcal{V} \rightarrow \mathbb{R}$ for a scalar field f is straightforwardly defined by

$$\Delta_{ij}^{\mathbf{k}} f \triangleq (G_i^{\mathbf{k}} - G_j^{\mathbf{k}}) f. \quad (2.6)$$

Both the RI and DRI are linear operators in the usual sense. Using Eqs. 2.3 up to 2.6, we see that

$$\tau_{ij}^{\mathbf{k}} = \Delta_{ij}^{\mathbf{k}} n_e. \quad (2.7)$$

Let us now specify that the FED is a Gaussian process (GP) restricted to (and indexed by) the set of spatial locations $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^3 \mid (\mathbf{x} - \mathbf{x}_0) \cdot \hat{\mathbf{z}} \in [a - b/2, a + b/2]\}$. This defines a layer of thickness b at height a above some reference point \mathbf{x}_0 (see **Figure 2.1**). Within this layer the FED is realised from,

$$n_e \sim \mathcal{N}[\mu, K], \quad (2.8)$$

where $\mu: \mathcal{X} \rightarrow \mathbb{R}_{>0}$ is the mean function, and $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is the covariance kernel function. In other words, the ionospheric FED is regarded to be a uncountable infinite set of random variables (RVs) indexed by spatial locations in \mathcal{X} , such that for any finite subset of such locations the corresponding FEDs have a multivariate normal distribution.

In order to extend the scalar field n_e to all of \mathbb{R}^3 , so that we may apply the operator in Eq. 2.6 to FED, we impose that for all $\mathbf{x} \in \mathbb{R}^3 \setminus \mathcal{X}: n_e(\mathbf{x}) = 0$. This simply means that we take electron density to be zero outside the layer, and makes $G_i^{\mathbf{k}}$ well-defined. To further simplify the model, we assume that the mean FED in the layer is constant; that is, for all $\mathbf{x} \in \mathcal{X}: \mu(\mathbf{x}) = \bar{n}_e$.

One immediate question that arises pertains to the reasoning behind using a GP to model the FED in the ionosphere. Currently, there is no adequate probabilistic description of the ionosphere that is valid for all times and at the spatial scales that we require. The state-of-the-art characterisation of the ionosphere at the latitude and scales we are concerned with are measurements of the phase structure function, a second-order statistic [Mevius et al., 2016]. It is well known that second-order statistics alone do not determine a distribution. In general, all moments are required to characterise a distribution, with a determinancy criterion known as Carleman's condition. Furthermore, the ionosphere is highly dynamic and

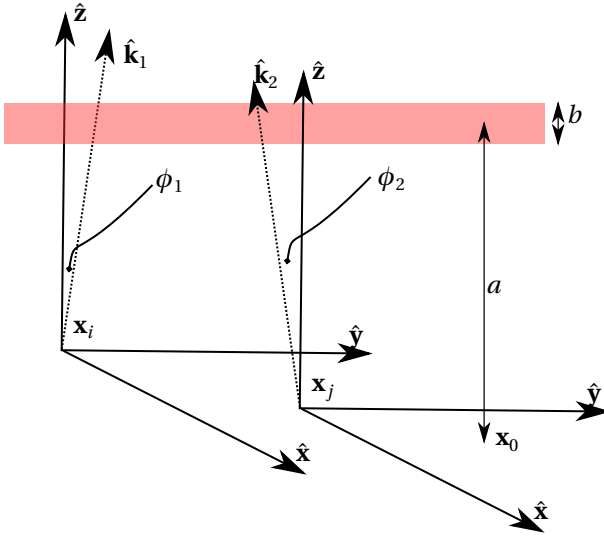


Figure 2.1: Geometry of the toy model. The ionosphere is a layer of thickness b at height a above a reference location \mathbf{x}_0 . In general, ΔTEC is the TEC along one geodesic minus the TEC along another parallel geodesic. Usually, these geodesics are originating at antennae i and j (locations \mathbf{x}_i and \mathbf{x}_j), and pointing in directions $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_2$, respectively. One common choice is to have a fixed reference antenna for all ΔTEC measurements. The corresponding zenith angles are ϕ_1 and ϕ_2 .

displays a multitude of behaviours. Jordan et al. [2017] observed four distinct behaviours of the ionosphere above the MWA. It is likely that there are innumerable states of the ionosphere.

Due to the above issue, it is not our intent to precisely model the ionosphere. We rather seek to describe it with a flexible and powerful probabilistic framework. Gaussian processes have several attractive properties, such as the fact that they are highly expressive, easy to interpret, and (in some cases) allow closed-form analytic integration over hypotheses [Rasmussen and Williams, 2006a].

However, a Gaussian distribution assigns a non-zero probability density to negative values, which is unphysical. One might instead consider the FED to be a log-GP, $n_e(\mathbf{x}) = \bar{n}_e \exp \rho(\mathbf{x})$, where the dimensionless quantity $\rho(\mathbf{x})$ is a Gaussian process. In the limit $\rho(\mathbf{x}) \rightarrow 0$, we recover that n_e is itself a GP. This is equivalent to saying that the $\sigma_{n_e}/\bar{n}_e \ll 1$. As explained in Section 2.4, we determine estimates of σ_{n_e} and \bar{n}_e by fitting our models to actual observed calibrator data, the International Reference Ionosphere (IRI), and observations taken from Kivelson and Russell [1995]. This places the ratio at $\sigma_{n_e}/\bar{n}_e \lesssim 0.06$, suggesting that if the FED can be accurately described with a log-GP, then to good approximation it can also be described with a GP.

We now impose that the geodesics are straight rays, a simplification valid in the weak-scattering limit considered here. The geodesics therefore become $\mathcal{R}_{\mathbf{x}}^{\hat{\mathbf{k}}}[n](s) = \mathbf{x} + s\hat{\mathbf{k}}$. In practice, strong scattering due to small-scale refractive index variations in the ionosphere is negligible at frequencies far above the plasma frequency when the ionosphere is well-

behaved, which is about 90% of the time [Vedantham and Koopmans, 2015]. For frequencies $\lesssim 50$ MHz however, this simplification becomes problematic. Under the straight-ray assumption, Equation 2.7 becomes

$$\tau_{ij}^{\hat{\mathbf{k}}} = \int_{s_i^{\hat{\mathbf{k}}^-}}^{s_i^{\hat{\mathbf{k}}^+}} n_e(\mathbf{x}_i + s\hat{\mathbf{k}}) ds - \int_{s_j^{\hat{\mathbf{k}}^-}}^{s_j^{\hat{\mathbf{k}}^+}} n_e(\mathbf{x}_j + s'\hat{\mathbf{k}}) ds'. \quad (2.9)$$

Here, the integration limits come from the extension of the FED to spatial locations outside the index-set \mathcal{X} , and are given by

$$s_i^{\hat{\mathbf{k}}^\pm} = \left(a \pm \frac{b}{2} - (\mathbf{x}_i - \mathbf{x}_0) \cdot \hat{\mathbf{z}} \right) \sec \phi, \quad (2.10)$$

where $\sec \phi = (\hat{\mathbf{k}} \cdot \hat{\mathbf{z}})^{-1}$ denotes the secant of the zenith angle. It is convenient to collocate the reference point \mathbf{x}_0 with one of the antenna locations, and then to also specify this antenna as the reference antenna, i.e. the origin of all reference geodesics. When this choice is made, ΔTEC becomes $\tau_{i0}^{\hat{\mathbf{k}}}$.

Equation 2.7 shows directly that if n_e is a GP, then so is ΔTEC . This can be understood by viewing the RI as the limit of a Riemann sum. We reiterate that every univariate marginal of a multivariate Gaussian is also Gaussian, and that every finite linear combination of Gaussian RVs is again Gaussian. Taking the Riemann sum to the infinitesimal limit preserves this property. Since the DRI is a linear combination of two RIs, the result follows [e.g. Jidling et al., 2018].

The index-set for the ΔTEC GP is the product space of all possible antenna locations and vectors on the unit 2-sphere, $\mathcal{S} = \{(\mathbf{x}, \hat{\mathbf{k}}) \mid \mathbf{x} \in \mathbb{R}^3, \hat{\mathbf{k}} \in \mathbb{S}^2\}$. This is analogous to saying that the coordinates of the ΔTEC GP are a tuple of antenna location and calibration direction. Thus, given any $\mathbf{y} = (\mathbf{x}, \hat{\mathbf{k}}) \in \mathcal{S}$, the ΔTEC is denoted by $\tau_{\mathbf{x}_0}^{\hat{\mathbf{k}}}$. Because ΔTEC is a GP, its distribution is completely specified by its first two moments.

Since we assume a flat layer geometry, the intersections of two parallel rays with the ionosphere layer have equal lengths of $b \sec \phi$. This results in the mean TEC of two parallel rays being equal, and thus the first moment of ΔTEC is,

$$m_{\Delta\text{TEC}}(\mathbf{y}) = 0, \quad (2.11)$$

where $\mathbf{y} = (\mathbf{x}_i, \hat{\mathbf{k}}) \in \mathcal{S}$. It is important to note that this is not a trivial result. Indeed, a more realistic but slightly more complicated ionosphere layer model would assume the layer follows the curvature of the Earth. In this case, the intersections of two parallel rays with the ionosphere layer have unequal lengths, and the first moment of ΔTEC would depend on the layer geometry and \bar{n}_e .

We now derive the second central moment between two ΔTEC along two different geodesics, as visualised in **Figure 2.1**.

$$K_{\Delta\text{TEC}}(\mathbf{y}, \mathbf{y}') = \mathbb{E} \left[\tau_{i0}^{\hat{\mathbf{k}}} \tau_{j0}^{\hat{\mathbf{k}'}} \right] \quad (2.12)$$

$$= \mathbb{E} \left[(G_i^{\hat{\mathbf{k}}} n_e - G_0^{\hat{\mathbf{k}}} n_e) (G_j^{\hat{\mathbf{k}'}} n_e - G_0^{\hat{\mathbf{k}'}} n_e) \right] \quad (2.13)$$

$$= I_{ij}^{\hat{\mathbf{k}}\hat{\mathbf{k}'}} + I_{00}^{\hat{\mathbf{k}}\hat{\mathbf{k}'}} - I_{i0}^{\hat{\mathbf{k}}\hat{\mathbf{k}'}} - I_{0j}^{\hat{\mathbf{k}}\hat{\mathbf{k}'}} , \quad (2.14)$$

where $\mathbf{y} = (\mathbf{x}_i, \hat{\mathbf{k}}) \in \mathcal{S}$ and $\mathbf{y}' = (\mathbf{x}_j, \hat{\mathbf{k}}') \in \mathcal{S}$ and,

$$I_{ij}^{\hat{\mathbf{k}}\hat{\mathbf{k}}'} = \int_{s_i^{\hat{\mathbf{k}}^-}}^{s_i^{\hat{\mathbf{k}}^+}} \int_{s_j^{\hat{\mathbf{k}}'^-}}^{s_j^{\hat{\mathbf{k}}'^+}} K(\mathbf{x}_i + s\hat{\mathbf{k}}, \mathbf{x}_j + s'\hat{\mathbf{k}}') ds ds'. \quad (2.15)$$

We now see that the GP for ΔTEC is zero-mean with a kernel that depends on the kernel of the FED and layer geometry. The layer geometry of the ionosphere enters through the integration limits of Eq. 2.15. Most notably, the physical kernel is non-stationary even if the FED kernel is. Non-stationarity means that the ΔTEC model is not statistically homogeneous, a fact that is well known since antennae near the reference antenna typically have small ionospheric phase corrections. We henceforth refer to Eq. 2.14 as the physical kernel, or our kernel.

Related work. Modelling the ionosphere with a layer has been used in the past. Yeh [1962] performed analysis of transverse spatial covariances of wavefronts [e.g. Wilcox, 1962, Keller et al., 1964] passing through the ionosphere. Their layer model was motivated by the observation of scintillation of radio waves from satellites [Yeh and Swenson, 1959]. One of their results is a simplified variance function, which can be related to the phase structure functions in Section 2.6.4. In van der Tol [2009], a theoretical treatment of ionospheric calibration using a layered ionosphere with Kolmogorov turbulence is done. More recently, Arora et al. [2016] attempted to model a variable-height ionosphere layer above the MWA using GPS measurements for the purpose of modelling a TEC gradient; however unfortunately they concluded that the GPS station array of the MWA is not dense enough to constrain their model.

2.4 Method

In order to investigate the efficacy of the physical kernel for the purpose of modelling ΔTEC we devise a simulation-based experiment. Firstly, we define several observational setups covering a range of calibration pierce-point sparsity and calibration signal-to-noise ratios. A high signal-to-noise-ratio calibration corresponds to better determination of ΔTEC from gains in a real calibration program. Secondly, we characterise two ionosphere varieties as introduced in Section 2.3. Each ionosphere variety is defined by its layer height and thickness, and GP parameters. For each pair of observational setup and ionosphere variety we realise FED along each geodesic and numerically evaluate Eq. 2.7 thereby producing ΔTEC . We then add an amount of white noise to ΔTEC which mimics the uncertainty in a real calibration program with a given calibration signal-to-noise ratio. Finally, we compare the performance of our kernel against several other common kernels used in machine learning on the problem of Gaussian process regression, known as Kriging. In order to do this, we generate ΔTEC for extra geodesics and place them in a held-out dataset. This held-out dataset is used for validation of the predictive performance to new geodesics given the observed ΔTEC . We refer to the other kernels, which we compare our kernel to, as the competitor kernels, and the models that they induce, as the competitor models.

Table 2.1: Summary of the parameters of the simulated ionospheres.

Variety	a (km)	b (km)	K_{FED}	σ_{n_e} (m^{-3})	HPD (km)
dawn	250	100	M32	$6 \cdot 10^9$	15
dusk	350	200	EQ	$3 \cdot 10^9$	15

2.4.1 Data generation

For all simulations, we have chosen the core and remote station configuration of LOFAR [van Haarlem et al., 2013], which is a state-of-the-art low-frequency radio array centred in the Netherlands and spread across Europe. The core and remote stations of LOFAR are located within the Netherlands with maximal baselines of 70 km, and we term this array the Dutch LOFAR configuration. We thinned out the array such that no antenna is within 150 m of another. We made this cutoff to reduce the data size because nearby antennae add little new information and inevitably raise computational cost. For example, antennae like CS001HBA0 and CS001HBA1 are so close that their joint inclusion was considered redundant.

We consider several different experimental conditions, with a particular choice denoted by η , under which we compare our model to competitors. We consider five levels of pierce-point sparsity: {10, 20, 30, 40, 50} directions per field of view (12.6 deg^2). For a given choice of pierce-point sparsity we place twice as many directions along a Fibonacci spiral – scaled to be contained within the field of view – and randomly select half of the points to be in the observed dataset and the other half to be in the held-out dataset. The Fibonacci spiral is slightly overdense in the centre of the field of view, which mimics selecting bright calibrators from a primary-beam uncorrected radio source model. We consider a range of calibration signal-to-noise ratios, which correspond to Gaussian uncertainties of ΔTEC that would be inferred from antenna-based gains in a real calibration program. We therefore consider 11 uncertainty levels on a logarithmic scale from 0.1 to 10 mTECU. A typical state-of-the-art Dutch LOFAR-HBA (high-band antennae) direction-dependent calibration is able to produce on the order of 30 calibration directions [Shimwell et al., 2019], based on the number of bright sources in the field of view, and produce ΔTEC with an uncertainty of approximately 1 mTECU; these levels of sparsity and noise probe above and below nominal LOFAR-HBA observing conditions.

We define an ionosphere variety as an ionosphere layer model with a particular choice of height a , thickness b , mean electron density \bar{n}_e , and FED kernel K_{FED} with associated hyperparameters, namely length-scale and variance. As mentioned in Section 2.3, due to the innumerable states of the ionosphere our intent is not to exactly simulate the ionosphere, but rather to derive a flexible model. Therefore, to illustrate the flexibility of our model, we have chosen to experiment with two very different ionosphere varieties which we designate the *dawn* and *dusk* ionosphere varieties. These ionosphere varieties are summarised in **Table 2.1**. In Section 2.6.4 we show that these ionosphere varieties predict phase structure functions which are indistinguishable from real observations. In order to select the layer height and thickness parameters for the dawn and dusk varieties we took height profiles from the International Reference Ionosphere [IRI; Bilitza and Reinisch, 2008] model.

In order to choose the FED GP kernels and hyperparameters we note that it has been suggested that scintillation is more pronounced during mornings, due to increased FED variation [e.g. Spoelstra, 1983]; therefore we chose a rough FED kernel for our dawn simulation.

Roughness corresponds to how much spectral power is placed on the shorter length-scales, and also relates to how differentiable realisations from the process are; e.g. see **Figure 2.2**. For the dawn ionosphere we choose the Matérn-3/2 (M32) kernel,

$$K_{\text{M32}}(\mathbf{x}, \mathbf{x}') = \sigma_{n_e}^2 \left(1 + \frac{\sqrt{3}}{l_{\text{M32}}} |\mathbf{x} - \mathbf{x}'| \right) \exp \left[\frac{-\sqrt{3}}{l_{\text{M32}}} |\mathbf{x} - \mathbf{x}'| \right], \quad (2.16)$$

which produces realisations that are only once differentiable and therefore rough. For the dusk ionosphere we choose the exponentiated quadratic (EQ) kernel,

$$K_{\text{EQ}}(\mathbf{x}, \mathbf{x}') = \sigma_{n_e}^2 \exp \left[\frac{-|\mathbf{x} - \mathbf{x}'|^2}{2l_{\text{EQ}}^2} \right], \quad (2.17)$$

which produces realisations that are infinitely differentiable and smooth.

Both kernels have two hyperparameters, variance $\sigma_{n_e}^2$ and length-scale l . In order to estimate the FED variation, σ_{n_e} , we used observations from Kivelson and Russell [1995] that TEC measurements are typically on the order of 10 TECU, with variations of about 0.1 TECU. Following the observation that the dawn typically exhibits more scintillation we choose a twice higher σ_{n_e} for our dawn simulation. In addition to the length-scale we consider the half-peak distance (HPD) h , which corresponds to the distance at which the kernel reaches half of its maximum. This parameter has a consistent meaning across all monotonically decreasing isotropic kernels, whereas the meaning of l depends on the kernel. It is related to h by $h \approx 1.177l_{\text{EQ}}$ for the EQ and $h \approx 0.969l_{\text{M32}}$ for the M32 kernel. The length-scales were chosen by simulating a set of ionospheres with different length-scales and choosing the length-scale that resulted in ΔTEC screens that are visually similar to typical Dutch LOFAR-HBA calibration data. For a given ionosphere variety, we numerically integrate the FED realised from the corresponding GP along the rays in order to compute TEC. From TEC we compute DTEC by taking the difference with the reference antenna TEC. We note that this requires a much higher relative precision in the absolute TEC calculations, since TEC is typically two orders of magnitude large than DTEC. Due to computational limits, we only realise one simulation per experimental condition – that is, we do not average over multiple realisations per experimental condition – however given the large number of experimental conditions there is enough variation to robustly perform a comparative analysis.

2.4.2 Competitor models

For the comparison with competitor models, we compare the physical kernel with: exponential quadratic (EQ), Matérn-5/2 (M52), Matérn-3/2 (M32), and Matérn-1/2 (M12) [Rasmussen and Williams, 2006a]. The EQ and M32 kernels have already been introduced as FED kernels. The M52 and M12 are very similar except for having different roughness properties. Each of these kernels results in a model that spatially smooths – and therefore interpolates – the observed data, but involves a different assumption on the underlying roughness of the function. In order to use these kernels to model ΔTEC , we give each subspace of \mathcal{S} its own kernel and take the product. For example, if K_C is the competitor kernel type, and $(\mathbf{x}, \hat{\mathbf{k}}), (\mathbf{x}', \hat{\mathbf{k}}') \in \mathcal{S}$, then we form the kernel $K_C((\mathbf{x}, \hat{\mathbf{k}}), (\mathbf{x}', \hat{\mathbf{k}}')) = K_C^1(\mathbf{x}, \mathbf{x}')K_C^2(\hat{\mathbf{k}}, \hat{\mathbf{k}}')$ thereby giving each subspace of the index set, \mathcal{S} , its own kernel with associated hyperparameters.

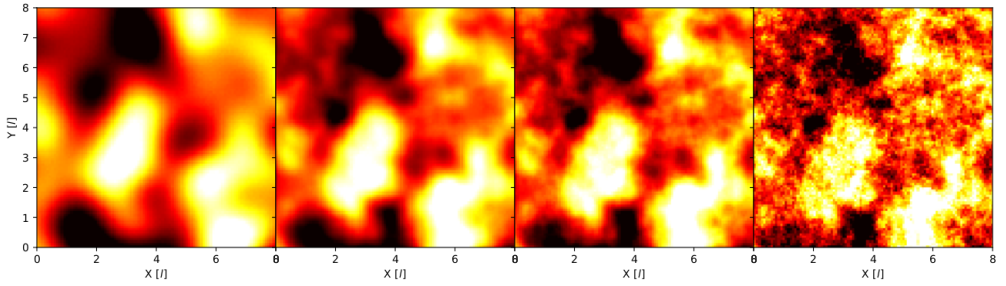


Figure 2.2: Example realisations from exponential quadratic, Matérn-5/2, Matérn-3/2, and Matérn-1/2 kernels. The same HPD was used in all kernels, however the smoothness of the resulting process realisation is different for each.

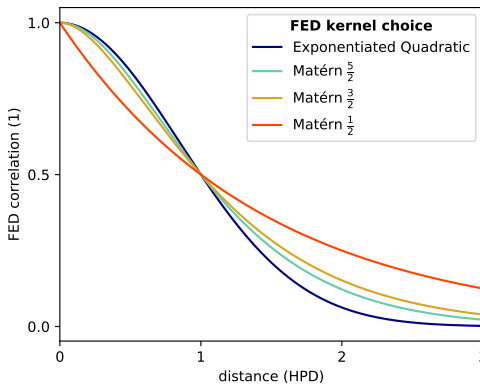


Figure 2.3: Shape of several kernels as a function of separation in units of the HPD of the kernel.

Figure 2.3 shows each kernel profile with the same HPD and Figure 2.2 shows example realisations from the same kernels. It can be visually verified that the M32 kernel has more small-scale variation than the EQ kernel, while maintaining similar large-scale correlation features.

We note that evaluation of the physical kernel requires that a double integral be performed, which can be done in several ways [e.g. Hendriks et al., 2018]. In our experiments we tried both explicit adaptive step-size Runge-Katta quadrature, and two-dimensional trapezoid quadrature. We found via experimentation that we could simply use the trapezoid quadrature with each abscissa partitioned into four equal intervals without loss of effectiveness. However, we chose to use seven partitions. We discuss this choice in Section 2.6.5.

However, we chose to use seven partitions. We discuss this choice in Section 2.6.5.

2.4.3 Model comparison

For model comparison, we investigate two key aspects of each model: the ability to accurately model observed ΔTEC , and the ability to accurately infer the held-out ΔTEC . In the language of the machine-learning community these are often referred to as minimising the data loss and the generalisation error, respectively. We also investigate the ability to learn the hyperparameters of the physical kernel from sparse data. Finding that the physical model accurately models both observed and held-out ΔTEC , while also being able to learn the hyper parameters, would be a positive outcome.

To measure how well a model represents the observed data, given a particular choice of kernel K and hyperparameters, we compute the log-probability of the observed (LPO)

ΔTEC data – Bayesian evidence – which gives a measure of how well a GP fits the data with intrinsically penalised model complexity. If we have data measured at $\mathbf{X} \in \mathcal{S}$ according to $\tau_{\text{obs}} = \tau(\mathbf{X}) + \epsilon$ where $\epsilon \sim \mathcal{N}[0, \sigma^2]$ and $\tau(\mathbf{X}) \sim \mathcal{N}[0, K(\mathbf{X}, \mathbf{X})]$ then the LPO is,

$$\log P_K(\tau_{\text{obs}}) = \log \mathcal{N}[0, \mathbf{B}], \quad (2.18)$$

where $\mathbf{B} = K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}$. To measure how well a model generalises to unseen data, given a particular choice of kernel K , we compute the conditional log-probability of held-out (LPH) data given the observed data. That is, if we have a held-out dataset measured at $\mathbf{X}^* \in \mathcal{S}$ according to $\tau_{\text{obs}}^* = \tau(\mathbf{X}^*) + \epsilon^*$ with $\epsilon^* \sim \mathcal{N}[0, \sigma^2]$ then the LPH conditional on observed τ_{obs} is,

$$\log P_K(\tau_{\text{obs}}^* | \tau_{\text{obs}}) = \log \mathcal{N}[K(\mathbf{X}^*, \mathbf{X})\mathbf{B}^{-1}\tau_{\text{obs}}, \mathbf{B}^* - K(\mathbf{X}^*, \mathbf{X})\mathbf{B}^{-1}K(\mathbf{X}, \mathbf{X}^*)] \quad (2.19)$$

where $\mathbf{B}^* = K(\mathbf{X}^*, \mathbf{X}^*) + \sigma^2 \mathbf{I}$.

In order to make any claims of model superiority, we will define the following two figures of merit (FOMs),

$$\Delta\text{LPO}_C(\eta) \triangleq \frac{P_{\Delta\text{TEC}}(\tau_{\text{obs}} | \eta)}{P_C(\tau_{\text{obs}} | \eta)}, \quad (2.20)$$

$$\Delta\text{LPH}_C(\eta) \triangleq \frac{P_{\Delta\text{TEC}}(\tau_{\text{obs}}^* | \tau_{\text{obs}}, \eta)}{P_C(\tau_{\text{obs}}^* | \tau_{\text{obs}}, \eta)}, \quad (2.21)$$

where $P_{\Delta\text{TEC}}$ is the probability distribution using the physical kernel and P_C is the distribution using a competitor kernel. The variable η represents a particular choice of experimental conditions, for example pierce point sparsity and noise.

These FOMs specify how much more or less probable the physical kernel model is than a competitor for the given choice of experimental conditions, and are therefore useful interpretable numbers capable of discriminating between two models. For example, a $\Delta\text{LPO}_C(\eta)$ value of 1 implies that for the given experimental conditions, η , both models have an equal probability of representing the observed data, and a value of 1.5 would imply that the physical kernel representation is 50% more probable than the competitor kernel. We note that considering the ratio of marginal probabilities is the canonical way of model selection [Rasmussen and Williams, 2006a]. For a rule-of-thumb using these FOMs, we empirically visually find that models produce noticeably better predictions starting at around 1.10 (10%).

For each choice of experimental conditions, η , and kernel model, we first infer the maximum *a posteriori* estimate of the hyperparameters of the kernel by maximising the marginal log-likelihood of the corresponding GP [Rasmussen and Williams, 2006a], which is equivalent to maximising the LPO of that model on the available observed dataset. We maximise the marginal log-likelihood using the variable metric BFGS method, which uses a low-rank approximation to the Hessian to perform gradient-based convex optimisation [Byrd et al., 1995]. We use the GPflow library [Matthews et al., 2017], which simplifies the algorithmic process considerably. On top of this we perform optimisation from multiple random initialisations to avoid potential local minima. For the physical kernel this corresponds to learning the layer height a and thickness b , and FED kernel length-scale l , and variance $\sigma_{n_e}^2$, and for the competitor kernels this corresponds to learning a variance and the length-scales for each subspace.

Table 2.2: Average and standard deviation, over all experimental conditions, of the difference between the learned physical hyperparameters and the true hyperparameters.

Variety	a (km)	b (km)	HPD (km)	$b\sigma_{n_e}$ (10^{11}km m^{-3})
dawn	10 ± 10	48 ± 18	4 ± 3	1.9 ± 1.2
dusk	16 ± 9	82 ± 20	1 ± 0.5	2.2 ± 0.3

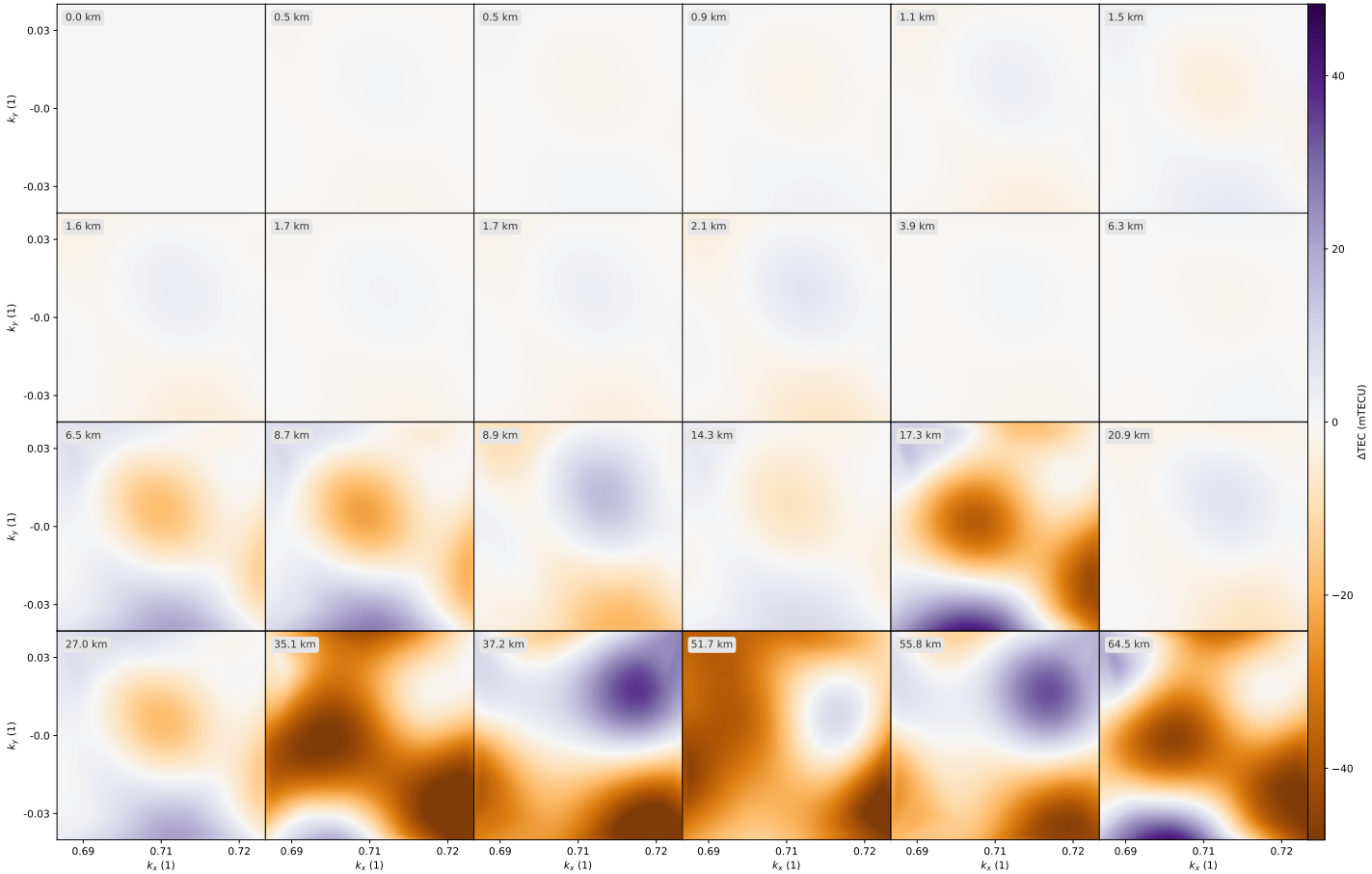


Figure 2.4: Caption continued on next page.

Figure 2.4: Example of antenna-based ΔTEC screens from the dusk ionosphere simulation. Each plot shows the simulated ground truth (noise-free) ΔTEC for each geodesic originating from that station with axes given in direction components k_x and k_y . The inset label gives how far the antenna is from the reference antenna. Antennae further from the reference antenna tend to have a larger magnitude ΔTEC as expected. Each plot box bounds a circular 12.6 deg^2 field of view.

2.5 Results

In **Table 2.2** we report the average and standard deviation, over all experimental conditions, of the difference between the learned physical hyperparameters and the true hyperparameters, which we term the discrepancy. The optimisation converged in all cases. We observe that for both ionosphere varieties the discrepancy of a is on the order of a $\sim 10 \text{ km}$, or a few percent, implying that a can be learned from data. The discrepancy of HPD, is on the order of 1 km , or around 10%, implying the spectral shape information of the FED can be constrained from data. We observe that the discrepancy of layer thickness, b , is large and on the order of 50%. One reason for this is because Eq. 2.15 will scale to first order with b – which is degenerate with the function of σ_{n_e} – and the only way to break the degeneracy is to have enough variation in the secant of the zenith angle. In a sparse and noisy observation of ΔTEC , the secant variation is poor and it is expected that this degeneracy exists. Therefore we also show the product $b\sigma_{n_e}$, and we see that this compound value discrepancy is smaller by approximately 35%.

In **Table 2.3** we summarise the performance of the physical kernel against each competitor kernel. We display the mean of $\Delta\text{LPO}_C(\eta)$, and $\Delta\text{LPH}_C(\eta)$ over all experimental conditions, as well as their values at the nominal experimental conditions of 30 directions per 12.6 deg^2 , and ΔTEC noise of 1 mTECU , which is indicated with η_{nom} . We use bold font in **Table 2.3** to indicate the best competitor model.

We first consider the ability of each model to represent the observed data. For the dawn ionosphere, the M52 competitor kernel has the best (lowest) $\langle\Delta\text{LPO}_C\rangle_\eta = 1.55$ and $\Delta\text{LPO}_C^{\eta_{\text{nom}}} = 1.46$, implying that the M52 kernel model is 55% and 46% less probable than the physical kernel model on average over all experimental conditions, and at nominal conditions, respectively. We note that the M32 kernel produced similar results. For the dusk ionosphere, the EQ kernel model is likewise the best among all competitors, being only 73% and 54% less probable than the physical kernel model on average over all experimental conditions, and at nominal conditions, respectively. In all experimental conditions, the physical model provides a significantly more probable explanation of the observed data.

We now consider the ability of each model to infer the held-out data. For the dawn ionosphere, the M52 competitor kernel has the best (lowest) $\langle\Delta\text{LPH}_C\rangle_\eta = 1.49$ and $\Delta\text{LPH}_C^{\eta_{\text{nom}}} = 1.31$, implying that the M52 kernel prediction is 49% and 31% less probable than the physical kernel model on average over all experimental conditions, and at nominal conditions, respectively. We note that the M32 kernel produced similar results. For the dusk ionosphere, the EQ kernel model is likewise the best among all competitors, with predictions only 16% and 12% less probable than the physical kernel model on average over all experimental conditions, and at nominal conditions, respectively. In the case of the rougher dawn ionosphere, the physical model provides a significantly more probable prediction of the held-out data in all

Table 2.3: Shows the probability ratio FOMs (see text) averaged over experimental conditions and at nominal conditions. Larger values indicate that the physical model is more probable. Bold face indicates the best performing competitor model (lower number).

	$\langle \Delta LPO_C \rangle_\eta$	$\Delta LPO_C^{\eta_{nom}}$	$\langle \Delta LPH_C \rangle_\eta$	$\Delta LPH_C^{\eta_{nom}}$
dawn				
M12	1.86	1.79	1.82	1.61
M32	1.56	1.49	1.50	1.33
M52	1.55	1.46	1.49	1.31
EQ	1.63	1.48	1.84	1.35
dusk				
M12	2.72	2.19	2.24	1.73
M32	1.96	1.69	1.50	1.29
M52	1.82	1.60	1.33	1.20
EQ	1.73	1.54	1.16	1.12

experimental conditions. However, for the smoother dusk ionosphere at nominal conditions, the physical model is only 12% more probable than the EQ kernel model, which is not very significant.

Figure 2.7 shows a visual comparison of the predictive distributions of the physical and best competitor kernel for the dawn ionosphere, for nominal and sparse-and-noisy conditions, for a subset of antennae over the field of view. In the first row we show the ground truth and observed data. In the second and third rows we plot the mean of the predictive distribution with uncertainty contours of the physical and best competitor models, respectively. At nominal conditions, the predictive means of the best competitor and physical models both visually appear to follow the shape of the ground truth. However, for the sparse-and-noisy condition, only the physical model predictive mean visually follows the shape of the ground truth. The uncertainty contours of the physical model vary in height slowly over the field of view, and are on the order of 0.5–1 mTECU. The uncertainty contours for the physical model indicate that we can trust the predictions near the edges of the field of view. In comparison, the uncertainty contours of the best competitor model steeply grow in regions without calibrators, and are on the order of 2–10 mTECU, indicating that only predictions in densely sampled regions should be trusted.

The last two rows show the residuals between the posterior means and the ground truth for the physical and best competitor models respectively. From this we can see that even when the best-competitor predictive mean visually appears to follow the ground truth the residuals are larger in magnitude than those of the physical models.

In order to quantify the effect of the residuals, a Δ TEC error, $\delta\tau$, can be conveniently represented by the equivalent source shift for a source at zenith on a baseline of r ,

$$\delta l \approx \frac{q^2}{\epsilon_0 m_e \nu^2 r} \delta\tau \quad (2.22)$$

$$\approx 1.16'' \left(\frac{r}{10\text{km}} \right)^{-1} \left(\frac{\nu}{150\text{MHz}} \right)^{-2} \left(\frac{\delta\tau}{\text{mTECU}} \right). \quad (2.23)$$

Figure 2.5 shows the mean linear regression of the absolute equivalent source shift of the

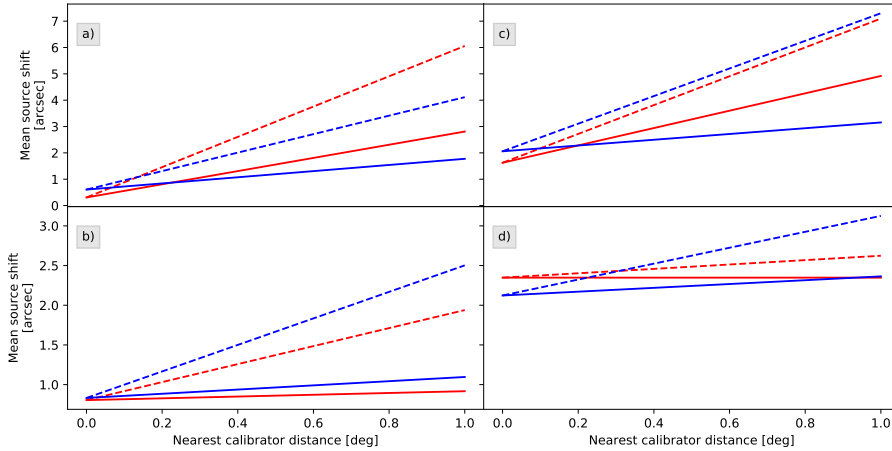


Figure 2.5: Mean equivalent source shift as a function of angular distance from the nearest calibrator caused by inference errors from the ground truth for **a**) remote stations (RS; > 3 km from the reference antenna) at nominal conditions (30 calibrators for 12.6deg^2 and 1 mTECU noise), **b**) core stations (CS; < 2 km) at nominal conditions, **c**) RS with sparse-and-noisy conditions (10 calibrators for 12.6deg^2 and 2.6 mTECU noise), and **d**) CS with sparse-and-noisy conditions. The *dash line styles* are the best competitor models (see text), the *solid line styles* are the physical model. The red lines are dawn ionospheres, and the blue lines are dusk ionospheres.

residuals for each point in the held-out data set, for nominal (left) and sparse-and-noisy (right) conditions, at 150 MHz on a baseline of 10 km, as a function of the nearest calibrator. For visual clarity we have not plotted confidence intervals, however we note that for nominal conditions the 1σ confidence width is about $2''$ and for the sparse-and-noisy conditions it is about $4''$. Because there are few nearest-calibrator distances exceeding 1 degree at nominal conditions, we only perform a linear regression out to 1 degree.

The upper row shows the source shift for the remote stations (RS) residuals, which are generally much larger than the source shifts for core stations (CS) in the bottom row, since the CS antennae are much closer to the reference antenna and have smaller ΔTEC variance. We observe that the physical model (dashed line styles) generally has a shallower slope than the best competitor model (solid line styles). Indeed, for the CS antennae the physical model source shift is almost independent of distance from a calibrator. The offset from zero at 0 degrees of separation comes from the fact that the predictive variance cannot be less than the variance of the observations; see the definition of \mathbf{B}^* in Eq. 2.19. At 1 degree of separation, the physical model mean equivalent source shift is approximately half of that of the best competitor model. At 0 degrees of separation, the mean source shift is the same for both models as expected.

2.6 Discussion

2.6.1 Model selection bias

Our derived model is a probabilistic model informed by the physics of the problem. We use the same physical model to simulate the data. Therefore it should perform better than any other general-purpose model. The fact that we simulate from the same physical model as used to derive the probabilistic model does not detract from the efficacy of the proposed model to represent the data. Indeed, it should be seen as a reason for preferring physics-based approaches when the physics are rightly known. The Gaussian random field layer model for the ionosphere has been a useful prescription for the ionosphere for a long time [e.g. Yeh and Swenson, 1959].

One type of bias that should be addressed is the fact that we assume we know the FED kernel type of the ionosphere. We do not show, for example, what happens when we assume the wrong FED kernel. However, since we are able to converge on optimal hyper parameters for a given choice of FED kernel, we can therefore imagine performing model selection based on the values of the Bayesian evidence (LPO) for different candidate FED kernels. Thus, we can assume that we could correctly select the right FED kernel in all the experimental conditions that we chose in this work.

2.6.2 Implicit tomography

The results of Section 2.5 indicate that the physical model provides a better explanation of ΔTEC data than any of the competitor models. One might ask how it performs so well. The approach we present is closely linked to tomography, where (possibly non-linear) projections of a physical field are inverted for a scalar field. In a classical tomographic approach, the posterior distribution for the FED given observed ΔTEC data would be inferred and then the predictive ΔTEC would be calculated from the FED, marginalising over all possible FEDs,

$$P(\boldsymbol{\tau} | \boldsymbol{\tau}_{\text{obs}}) = \int_{\mathbf{n}_e} P(\boldsymbol{\tau} | \mathbf{n}_e) P(\mathbf{n}_e | \boldsymbol{\tau}_{\text{obs}}) d\mathbf{n}_e, \quad (2.24)$$

where $\mathbf{n}_e = \{n_e(\mathbf{x}) | \mathbf{x} \in \mathcal{X}\}$ is the set of FEDs over the entire index set \mathcal{X} , $\boldsymbol{\tau} = \{\tau_{\mathbf{x}}^{\hat{\mathbf{k}}} | (\mathbf{x}, \hat{\mathbf{k}}) \in \mathcal{S}_* \subset \mathcal{S}\}$ is the ΔTEC over some subset \mathcal{S}_* of the index set \mathcal{S} , $\boldsymbol{\tau}_{\text{obs}} = \{\tau_{\mathbf{x}}^{\hat{\mathbf{k}}} + \epsilon | (\mathbf{x}, \hat{\mathbf{k}}) \in \mathcal{S}_{\text{obs}} \subset \mathcal{S}\}$ is the observed ΔTEC over a different subset \mathcal{S}_{obs} of \mathcal{S} , and $\epsilon \sim \mathcal{N}[\mathbf{0}, \sigma^2 \mathbf{I}]$.

In our model, the associated equation for $P(\boldsymbol{\tau} | \boldsymbol{\tau}_{\text{obs}})$ is found by conditioning the joint distribution on the observed ΔTEC and then marginalising out FED,

$$P(\boldsymbol{\tau} | \boldsymbol{\tau}_{\text{obs}}) = \int_{\mathbf{n}_e} P(\mathbf{n}_e, \boldsymbol{\tau} | \boldsymbol{\tau}_{\text{obs}}) d\mathbf{n}_e \quad (2.25)$$

$$= \int_{\mathbf{n}_e} P(\mathbf{n}_e | \boldsymbol{\tau}_{\text{obs}}) P(\boldsymbol{\tau} | \mathbf{n}_e, \boldsymbol{\tau}_{\text{obs}}) d\mathbf{n}_e, \quad (2.26)$$

where in the second line we used the product rule of probability distributions [Kolmogorov, 1956]. By working through Eqs. 2.24 and 2.26, we discover that if $P(\boldsymbol{\tau} | \mathbf{n}_e) = P(\boldsymbol{\tau} | \mathbf{n}_e, \boldsymbol{\tau}_{\text{obs}})$ is true, then our method is equivalent to first inferring FED and then using that distribution to

calculate ΔTEC . In Appendix A we prove that the expressions in Eqs. 2.24 and 2.26 are equal due to the linear relation between FED and ΔTEC because the sum of two Gaussian RVs is again Gaussian. Most importantly, this result would not be true if ΔTEC was a non-linear projection of FED.

We refer to this as implicit tomography as opposed to explicit tomography, wherein the FED distribution would be computed first and the ΔTEC computed second [e.g. Jidling et al., 2018]. This explains why our kernel is able to accurately predict ΔTEC in regions without nearby calibrators. The computational savings of our approach is many-fold compared with performing explicit tomography, since the amount of memory that would be required to evaluate the predictive distribution of FED everywhere would be prohibitive. Finally, the use of GPs to model ray integrals of a GP scalar field is used in the seismic physics community for performing tomography of the interior of the Earth.

2.6.3 Temporal differential TEC correlations

One clearly missing aspect is the temporal evolution of the ionosphere. In this work we have considered instantaneous realisations of the FED from a spatial GP; however, the inclusion of time in the FED GP is straightforward in principle. One way to include time is by appending a time dimension to the FED kernel, which would mimic internal (e.g. turbulence-driven) evolution of the FED field. Another possibility is the application of a frozen flow assumption, wherein the ionospheric time evolution is dominated by a wind of constant velocity \mathbf{v} , so that $n_e(\mathbf{x}, t) = n_e^0(\mathbf{x} - \mathbf{v}t)$. Here, n_e^0 represents the FED at time $t = 0$, and n_e is a translation over the array as time progresses. In modelling a real dataset with frozen flow the velocity could be assumed to be *piece-wise* constant in time. We briefly experimented with frozen flow and found hyperparameter optimisation to be sensitive to the initial starting point due to the presence of many local optima far from the ground-truth hyperparameters. We suggest that a different velocity parametrisation might facilitate implementation of the frozen flow approach.

2.6.4 Structure function turnover and anisotropic diffractive scale

The power spectrum is often used to characterise the second-order statistics of a stationary random medium, since according to Bochner's Theorem the power spectrum is uniquely related to the covariance function via a Fourier transform. In 1941, Kolmogorov [translated from Russian in Kolmogorov, 1991] famously postulated that turbulence of incompressible fluids with very large Reynolds numbers displays self-similarity. From this assumption, he used dimensional analysis to show that the necessary power spectrum of self-similar turbulence is a power-law with an exponent of $-5/3$. A convenient related measurable function for the ionosphere is the phase structure function [van der Tol, 2009],

$$D(r) = \langle (\phi_{\nu}(R) - \phi_{\nu}(r+R))^2 \rangle_R \quad (2.27)$$

$$\triangleq \left(\frac{r}{r_{\text{diff}}} \right)^{\beta}, \quad (2.28)$$

where the expectation is locally over locations far from the boundaries of the turbulent medium, which is often characterised by an outer scale. The quantity r_{diff} is referred to as the

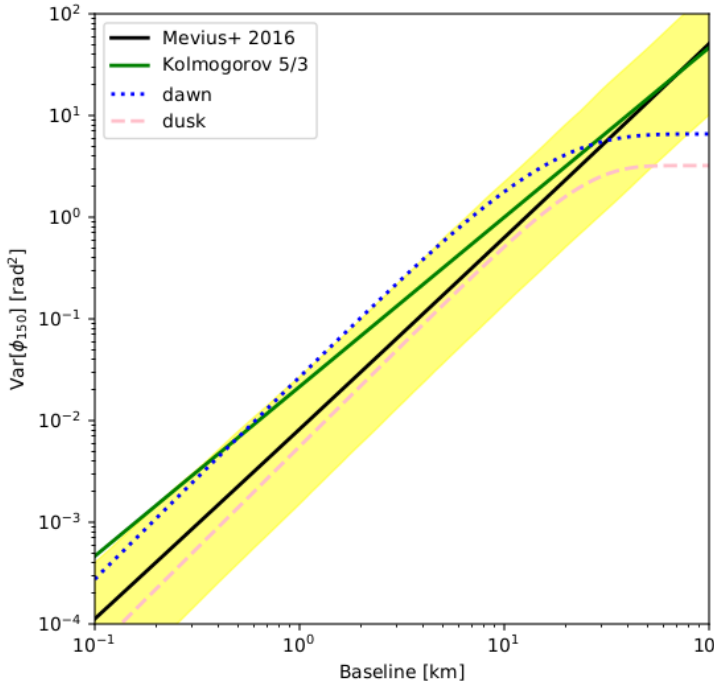


Figure 2.6: Structure functions predicted by our model compared with observations and theory. The dotted and dashed lines show the phase structure function corresponding to the physical kernel, with the dawn and dusk configurations, respectively (see **Figure 2.1**). Along side is the predicted structure function of Kolmogorov turbulence with a diffraction scale of 10 km, and the structure function constrained from observations in Meivius et al. [2016] with 1σ confidence region in yellow. We note that Meivius et al. [2016] observes a turnover, but does not characterise it, and therefore we do not attempt to plot it here.

diffraction scale, and is defined as the length where the structure function is 1 rad^2 . Under Kolmogorov’s theory of 1941, $\beta = 5/3$. Observations from 29 LOFAR pointings constrain β to be 1.89 ± 0.1 , slightly higher than predicted by Kolmogorov’s theory, and the diffraction scale to range from 5 to 30 km [Meivius et al., 2016].

In **Figure 2.6** the structure functions of the physical kernel are shown for the dawn and dusk varieties, alongside Kolmogorov’s $\beta = 5/3$ and the Meivius et al. [2016] observations. Though not plotted, Meivius et al. [2016] also find that there is a hint of a turnover in the structure functions they observed, which they suggest might be a result of an outer scale in the context of Kolmogorov turbulence. However, these latter authors conclude that longer baselines are needed to properly confirm the turnover and its nature. The dawn and dusk structure functions are nearly parallel with observations, and have turnovers that result because the FED covariance functions decay to zero monotonically and rapidly beyond the HPD. Interestingly, despite the fact that the FED kernels used for the dawn and dusk ionospheres have different spectral shapes, the structure functions have similar slopes. The difference between the dawn and dusk structure functions can be seen in the curvature of

their turnovers.

Our model provides an explanation for the observed shape of structure functions, which Kolmogorov's theory of 1941 fails to provide, namely the existence of a turnover, and a slope deviating from five-thirds. Specifically, a turnover requires only FED correlations that are stationary, isotropic, and monotonically decreasing (SIMD). Both the dawn and dusk ionosphere varieties experimented with predict slopes compatible with observations. Moreover, as shown in Appendix B, our model in conjunction with the SIMD FED kernel is falsifiable by observing a lack of plateau.

Mevius et al. [2016] also observe anisotropy in the measured r_{diff} as a function of pointing direction, and suggest that it is due to FED structures aligned with magnetic field lines [Loi et al., 2015]. In total, 12 out of 29 (40%) of their observations show anisotropy unaligned with the magnetic field lines of Earth. We propose a complementary explanation for the anisotropy of diffractive scale, without appealing to magnetic field lines. Our model implies that diffractive scale monotonically decreases with zenith angle. This is a result of the non-stationarity of the physical kernel even if the FED is stationary.

2.6.5 Low-accuracy numerical integration

The numerical integration required to compute Eq. 2.14 is performed using the 2D Trapezoid rule. This requires the selection of a number of partitions along the ray. The computational complexity scales quadratically with the number of partitions chosen, and thus a trade-off between accuracy and speed must be chosen. We found the relative error (using the Frobenius norm) to be 80% with two partitions, 20% with three partitions, 10% with four partitions, and 6% with seven partitions. After experimentation it was surprisingly found that two partitions was sufficient to beat all competitor models, and that marginal improvement occurs after five partitions. This suggests that even a low-accuracy approximation of our model encodes enough geometric information to make it a powerful tool in describing the ionosphere. Ultimately, we chose to use four partitions for our trials.

2.7 Conclusion

In this work, we put forth a probabilistic description of antenna-based ionospheric phase distortions, which we call the physical model. We assumed a single weakly scattering ionosphere layer with arbitrary height and thickness, and free electron density (FED) described by a Gaussian process (GP). We argue that modelling the FED with a GP locally about the mean is a strong assumption due to the small ratio of FED variation to mean as evidenced from ionosphere models. We show that under these assumptions the directly observable ΔTEC must also be a GP. We provide a mean and covariance function that are analytically related to the FED GP mean and covariance function, the ionosphere height and thickness, and the geometry of the interferometric array.

In order to validate the efficacy of our model, we simulated two varieties of ionosphere – a dawn (rough FED) and dusk (smooth FED) scenario – and computed the corresponding ΔTEC for the Dutch LOFAR-HBA configuration over a wide range of experimental conditions including nominal and sparse-and-noisy conditions. We compared this physical kernel to other widely successful competitor GP models that might naively be applied to the same

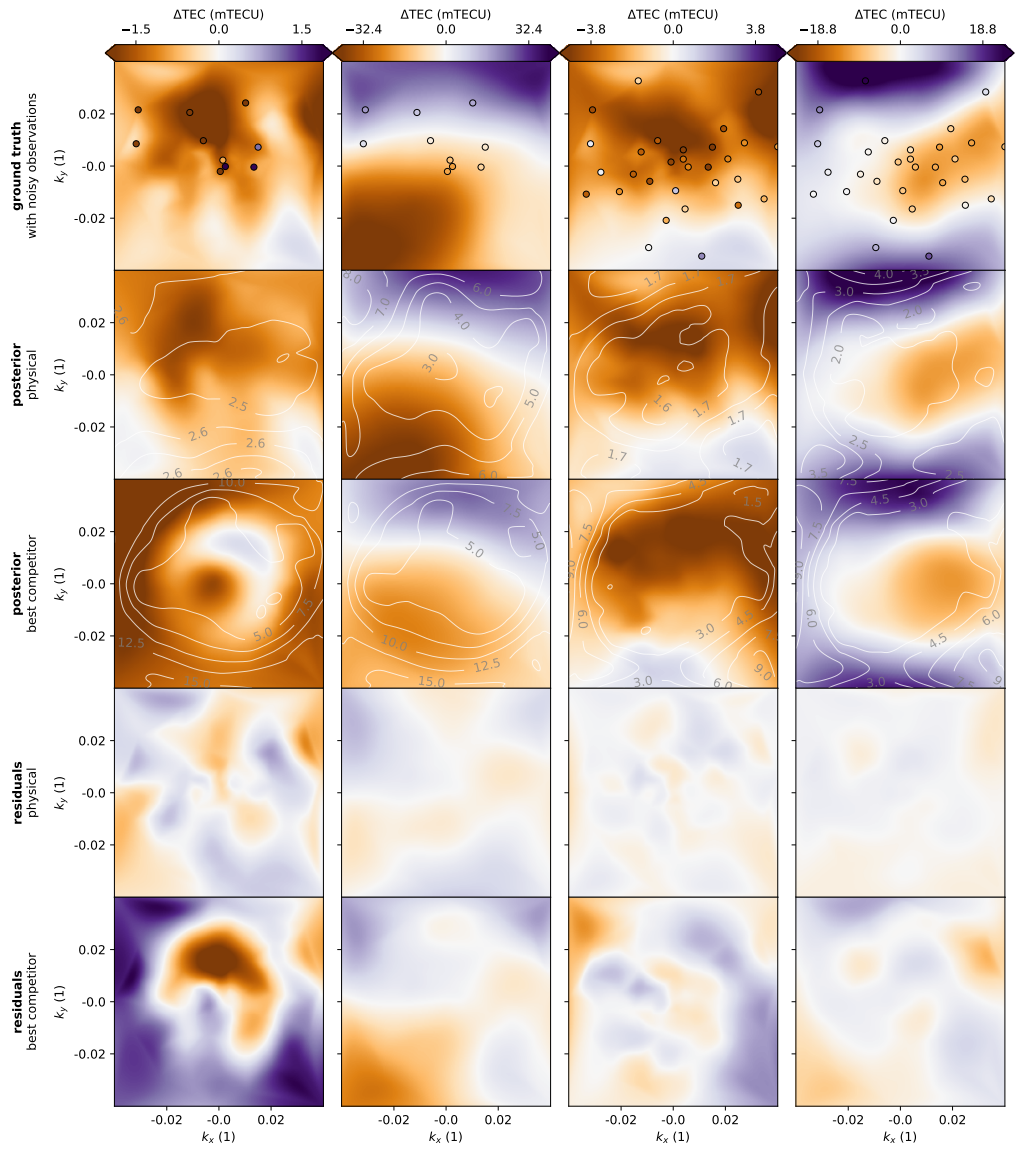


Figure 2.7: Caption continued on next page.

Figure 2.7: Example visual comparison of the predictive performance of our physical model with that of the best competitor model for the dawn ionosphere. *First row* – ground truth ΔTEC overlaid on noisy draws from the ground truth which are the observations; *Second and third rows* – posterior mean with uncertainty contours for the physical model and best competitor model respectively. *Fourth and fifth rows* – residuals between posterior means and ground truth for the physical model and best competitor model respectively. *First two columns*: Results for experimental conditions, (10 directions, 2.5 mTECU noise), for a central antenna (near to reference antenna) and a remote station (far from reference antenna); *Last two columns*: Results for experimental conditions, (30 directions, 1.6 mTECU noise), for a central antenna and a remote station.

problem. Our results show that we are always able to learn the FED GP hyperparameters and layer height – including from sparse-and-noisy ΔTEC data – and that the layer thickness could likely be learned if a height prior was provided. In general, the physical model is better able to represent observed data and generalises better to unseen data.

Visual validation of the predictive distributions of ΔTEC show that the physical model can accurately infer ΔTEC in regions far from the nearest calibrator. Residuals from the physical model (0.5–1 mTECU) are smaller and less correlated than those of competitor models (2–10 mTECU). In terms of mean equivalent source shift resulting from incorrect predictions, the physical model mean equivalent source shift is approximately half of that of the best competitor model. We show that our model performs implicit tomographic inference at low cost, which is because ΔTEC is a linear projection of FED and the FED is a GP. We suggest possible extensions to incorporate time, including frozen flow and appending the FED spectrum with a temporal power spectrum. Our model provides an alternative explanation for the Mevius et al. [2016] observations: phase structure function slope deviating from Kolmogorov’s five-thirds, the turnover on large baselines, and diffractive scale anisotropy.

In the near future, we will apply this model to LOFAR-HBA datasets and perform precise ionospheric calibration for all bright sources in the field of view. It is envisioned that this will lead to clearer views of the sky at the longest wavelengths, empowering a plethora of science goals.