A probabilistic approach to direction-dependent ionospheric calibration

Albert, J.G.; Oei, M.S.S.L.; Weeren, R.J. van; Intema, H.T.; Rottgering, H.J.A.

Astronomy & Astrophysics manuscript no. main
October 11, 2019

A probabilistic approach to direction-dependent ionospheric calibration

J. G. Albert[1], M. S. S. L. Oei[1], R. J. van Weeren[1], H. T. Intema[1,2], and H. J. A. Röttgering[1]

1. Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, the Netherlands
2. International Centre for Radio Astronomy Research – Curtin University, GPO Box U1987, Perth WA 6845, Australia
E-mail: albert@strw.leidenuniv.nl & oei@strw.leidenuniv.nl

Received 12 April 2019 / Accepted 10 October 2019

ABSTRACT

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 the best competitor model's. We find that it is possible to partially constrain the ionosphere's hyperparameters from sparse-and-noisy observed data. Our model provides an alternative explanation for observed phase structure functions deviating from Kolmogorov's 5/3 turbulence, turnover at high baselines, and diffractive scale anisotropy. We show that our model implicitly cheaply performs tomography of the free electron density. Moreover, we find that even a fast, low-resolution approximation to 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.

1. Introduction

Since the dawn of low-frequency radio astronomy, the ionosphere has been a confounding factor in astronomers' interpretations of radio data. This is because the ionosphere has a spatially and temporally varying refractive index, which perturbs 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, i.e. 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 LOFAR, MWA, and the future SKA. Thus 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 estimate of the true sky brightness distribution, you can calibrate based on this image, generate a revised image, and repeat this process iteratively. 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. 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 more 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 piece-wise 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 usually 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. We thus form the scope of this paper as building 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. In Section 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 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 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 6 we provide a justification for the assumptions of the model, and show the equivalence with tomographic inference.

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 & 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 $O(\nu^{-5})$:

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

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, and 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 sign corresponds to the left-handed circularly polarised mode of propagation, and the minus sign 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 $\mathscr{R}_{\mathbf{x}}^{\hat{\mathbf{k}}}$ be a functional of n, so that the geodesic $\mathscr{R}_{\mathbf{x}}^{\hat{\mathbf{k}}}[n] : [0, \infty) \to \mathbb{R}^3$ 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

$$\phi_{\mathbf{x}}^{\hat{\mathbf{k}}} = \frac{2\pi\nu}{c} \int_0^\infty n\left(\mathscr{R}_{\mathbf{x}}^{\hat{\mathbf{k}}}[n](s)\right) - 1 \, \mathrm{d}s, \quad (2)$$

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.

By substituting Eq. 1 into Eq. 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, $\phi_{\mathbf{x}}^{\hat{\mathbf{k}}} \approx \frac{-q^2}{4\pi\epsilon_0 mc\nu} \tau_{\mathbf{x}}^{\hat{\mathbf{k}}}$, where,

$$\tau_{\mathbf{x}}^{\hat{\mathbf{k}}} \triangleq \int_0^\infty n_e\left(\mathscr{R}_{\mathbf{x}}^{\hat{\mathbf{k}}}[n](s)\right) \mathrm{d}s. \quad (3)$$

Equation 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 will use Latin subscripts to specify geodesics with origins at an antenna location; e.g. it shall be understood that $\mathscr{R}_i^{\hat{\mathbf{k}}}[n]$ is short for $\mathscr{R}_{\mathbf{x}_i}^{\hat{\mathbf{k}}}[n]$. Correspondingly, we introduce the notion of differential total electron content (ΔTEC),

$$\tau_{ij}^{\hat{\mathbf{k}}} \triangleq \tau_i^{\hat{\mathbf{k}}} - \tau_j^{\hat{\mathbf{k}}}, \quad (4)$$

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

3. Probabilistic relation between FED and ΔTEC: Gaussian process layer model

In this section we derive the probability distribution of ΔTEC 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 $G_i^{\hat{\mathbf{k}}} : \mathscr{V} \to \mathbb{R}$ mapping from the space of all scalar-valued functions over $\mathbb{R}^3$ to a scalar value according to,

$$G_i^{\hat{\mathbf{k}}} f \triangleq \int_0^\infty f\left(\mathscr{R}_i^{\hat{\mathbf{k}}}[n](s)\right) \mathrm{d}s, \tag{5}$$

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

$$\Delta_{ij}^{\hat{\mathbf{k}}} f \triangleq \left(G_i^{\hat{\mathbf{k}}} - G_j^{\hat{\mathbf{k}}}\right) f. \tag{6}$$

Both the RI and DRI are linear operators in the usual sense. Using Eqs. 3 up to 6, we see that

$$\tau_{ij}^{\hat{\mathbf{k}}} = \Delta_{ij}^{\hat{\mathbf{k}}} n_e. \tag{7}$$

Let us now specify that the FED is a Gaussian process (GP) restricted to (and indexed by) the set of spatial locations $\mathscr{X} = \left\{\mathbf{x} \in \mathbb{R}^3 \mid (\mathbf{x} - \mathbf{x}_0) \cdot \hat{\boldsymbol{z}} \in [a - b/2, a + b/2]\right\}$. This defines a layer of thickness b at height a above some reference point $\mathbf{x}_0$ (see **Figure** 1). Within this layer the FED is realised from,

$$n_e \sim \mathscr{N}[\mu, K], \tag{8}$$

where $\mu : \mathscr{X} \to \mathbb{R}_{>0}$ is the mean function, and $K : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ is the covariance kernel function. That is, the ionospheric FED is regarded to be a uncountable infinite set of random variables (RVs) indexed by spatial locations in $\mathscr{X}$, 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. 6 to FED, we impose that for all $\mathbf{x} \in \mathbb{R}^3 \setminus \mathscr{X} : n_e(\mathbf{x}) = 0$. This simply means that we take electron density to be zero outside the layer, and makes $G_i^{\hat{\mathbf{k}}}$ well-defined. To further simplify the model, we will suppose that the mean FED in the layer is constant; i.e. for all $\mathbf{x} \in \mathscr{X} : \mu(\mathbf{x}) = \bar{n}_e$.

One immediate question that should be asked is, why use 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 displays a multitude of behaviours. Jordan et al. (2017) have 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 – including being highly expressive, being easy to interpret, and (in some cases) allowing closed-form analytic integration over hypotheses (Rasmussen & Williams 2006).

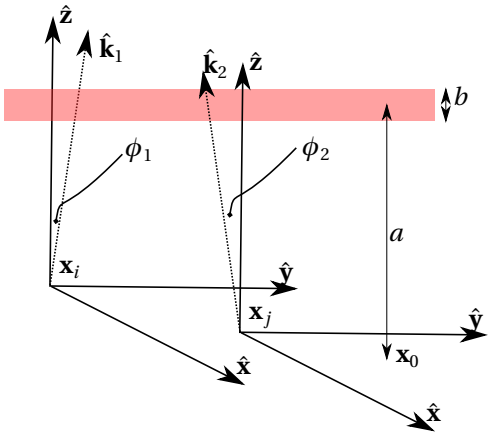

Fig. 1: The geometry of the toy model. The ionosphere is a layer of thickness b at height a above a reference location $\mathbf{x}_0$. In general, ΔTEC is the TEC along one geodesic minus the TEC along another parallel geodesic. Usually, these geodesics are originating at antennae i and j (locations $\mathbf{x}_i$ 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.

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 $\rho(\mathbf{x}) \to 0$, 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 4, we determine estimates of σ_{n_e} and $\bar{n}_e$ by fitting our models to actual observed calibrator data, IRI and observations taken from Kivelson & Russell (1995). This places the ratio at $\sigma_{n_e}/\bar{n}_e \lesssim 0.06$. This suggests 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 thus become $\mathscr{R}_{\mathbf{x}}^{\hat{\mathbf{k}}}[n](s) = \mathbf{x} + s\hat{\mathbf{k}}$. 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 & Koopmans 2015). For frequencies $\lesssim$ 50 MHz however, this simplification becomes problematic. Under the straight-ray assumption, Equation 7 becomes

$$\tau_{ij}^{\hat{\mathbf{k}}} = \int_{s_i^{\hat{\mathbf{k}}-}}^{s_i^{\hat{\mathbf{k}}+}} n_e(\mathbf{x}_i + s\hat{\mathbf{k}})\, ds - \int_{s_j^{\hat{\mathbf{k}}-}}^{s_j^{\hat{\mathbf{k}}+}} n_e(\mathbf{x}_j + s'\hat{\mathbf{k}})\, ds'. \tag{9}$$

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

$$s_i^{\hat{\mathbf{k}}\pm} = \left(a \pm \frac{b}{2} - (\mathbf{x}_i - \mathbf{x}_0) \cdot \hat{\mathbf{z}}\right) \sec\phi. \tag{10}$$

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

Equation 7 shows directly that if n_e is a GP, then ΔTEC is too. This can be understood by viewing the RI as the limit of a Riemann sum. Recall 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, $\mathscr{S} = \left\{(\mathbf{x}, \hat{\boldsymbol{k}}) \,|\, \mathbf{x} \in \mathbb{R}^3, \hat{\boldsymbol{k}} \in \mathbb{S}^2\right\}$. This is analogous to saying that the coordinates of the ΔTEC GP are a tuple of antenna location and calibration direction. Thus, given any $\mathbf{y} = (\mathbf{x}, \hat{\boldsymbol{k}}) \in \mathscr{S}$ the ΔTEC is denoted by $\tau^{\hat{\boldsymbol{k}}}_{\mathbf{x}0}$. Because ΔTEC is a GP, its distribution is completely specified by its first two moments.

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

$$m_{\Delta\text{TEC}}(\mathbf{y}) = 0, \tag{11}$$

where $\mathbf{y} = (\mathbf{x}_i, \hat{\boldsymbol{k}}) \in \mathscr{S}$. 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 lengths of two parallel rays' intersection with the ionosphere layer are not the same, 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** 1.

$$\begin{aligned} K_{\Delta\text{TEC}}(\mathbf{y}, \mathbf{y}') &= \mathbb{E}\left[\tau^{\hat{\mathbf{k}}}_{i0} \tau^{\hat{\mathbf{k}}'}_{j0}\right] && (12)\\ &= \mathbb{E}\left[(G^{\hat{\mathbf{k}}}_i n_e - G^{\hat{\mathbf{k}}}_0 n_e)(G^{\hat{\mathbf{k}}'}_j n_e - G^{\hat{\mathbf{k}}'}_0 n_e)\right] && (13)\\ &= I^{\hat{\mathbf{k}}\hat{\mathbf{k}}'}_{ij} + I^{\hat{\mathbf{k}}\hat{\mathbf{k}}'}_{00} - I^{\hat{\mathbf{k}}\hat{\mathbf{k}}'}_{i0} - I^{\hat{\mathbf{k}}\hat{\mathbf{k}}'}_{0j}, && (14) \end{aligned}$$

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

$$I^{\hat{\mathbf{k}}\hat{\mathbf{k}}'}_{ij} = \int_{s^{\hat{\mathbf{k}}-}_i}^{s^{\hat{\mathbf{k}}+}_i} \int_{s^{\hat{\mathbf{k}}'-}_j}^{s^{\hat{\mathbf{k}}'+}_j} K\left(\mathbf{x}_i + s\hat{\boldsymbol{k}}, \mathbf{x}_j + s'\hat{\boldsymbol{k}}'\right) \mathrm{d}s \mathrm{d}s'. \tag{15}$$

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. 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 which is well known since antennas near the reference antenna typically have small ionospheric phase corrections. Going forward we shall refer to Eq. 14 as the physical kernel, or our kernel.

3.1. 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 based on the observation of scintillation of radio waves from satellites (Yeh & Swenson 1959). One of their results is a simplified variance function, which can be related to the phase structure functions in Section 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) have 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.

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. A high signal-to-noise calibration corresponds to better determination of ΔTEC from gains in a real calibration program. Secondly, we characterise two ionosphere varieties as introduced in Section 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. 7 thereby producing ΔTEC. We then add an amount of white noise to ΔTEC which mimics the uncertainty in a real calibration program with given calibration signal-to-noise. Finally, we then 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 shall call the other kernels, which we compare our kernel to, the competitor kernels, and the models that they induce, the competitor models.

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 antennae are within 150 m of others. 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 center of the field of view, which mimics selecting bright calibrators from a primary-beam uncorrected radio source model. We con-

sider a range of calibration signal-to-noise levels, which corresponding 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 mTECU 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, so 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, e.g. length-scale and variance. As mentioned in Section 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 will designate the *dawn* and *dusk* ionosphere varieties. These ionosphere varieties are summarised in **Table** 1. In Section 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 & 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. For the dawn ionosphere we choose the Matérn-3/2 (M32) kernel,

$$K_{\text{M32}}(\mathbf{x},\mathbf{x}') = \sigma^2_{n_e}\left(1+\frac{\sqrt{3}}{l_{\text{M32}}}|\mathbf{x}-\mathbf{x}'|\right)\exp\left[\frac{-\sqrt{3}}{l_{\text{M32}}}|\mathbf{x}-\mathbf{x}'|\right], \tag{16}$$

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

$$K_{\text{EQ}}(\mathbf{x},\mathbf{x}') = \sigma^2_{n_e}\exp\left[\frac{-|\mathbf{x}-\mathbf{x}'|^2}{2l^2_{\text{EQ}}}\right], \tag{17}$$

which produces realisations that are infinitely differentiable and smooth.

Both kernels have two hyperparameters, variance $\sigma^2_{n_e}$ and length-scale l. In order to estimate the FED variation, σ_{n_e}, we used observations from Kivelson & 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 $h \approx 1.177 l_{\text{EQ}}$ for the EQ and $h \approx 0.969 l_{\text{M32}}$ 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 visually similar to typical Dutch LOFAR-HBA calibration data.

Table 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	$6\cdot10^9$	15
dusk	350	200	EQ	$3\cdot10^9$	15

For a given ionosphere variety, the FED are realised from the corresponding GP at sampling points along the geodesics. As per our definition in Section 3, the FED is zero for points outside the layer. The FED along each geodesic is then numerically integrated using either trapezoid quadrature or Simpson's Rule, which produces TEC, cf. Eq. 3. Differential TEC is computed by spatially referencing the TEC, cf. Eq. 4. White noise corresponding to calibration signal-to-noise is added to the ΔTEC in the observed dataset. The spacing of samples along the geodesic is chosen in order to guarantee $< 1\%$ error in the resulting ΔTEC. Note, that this requires a much higher relative precision in the absolute TEC calculations.

Due to computational limits, we only realise one simulation per experimental condition – i.e. 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 analysis.

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 & Williams 2006). 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 result in models that spatially smooth – and therefore interpolate – the observed data, but with a different assumption on the underlying roughness of the function. In order to use these kernels to model ΔTEC, we give each subspace of $\mathscr{S}$ its own kernel and take the product. For example, if K_C is the competitor kernel type, and $(\mathbf{x},\hat{\boldsymbol{k}}),(\mathbf{x}',\hat{\boldsymbol{k}}')\in\mathscr{S}$, then we form the kernel $K_C((\mathbf{x},\hat{\boldsymbol{k}}),(\mathbf{x}',\hat{\boldsymbol{k}}')) = K^1_C(\mathbf{x},\mathbf{x}')K^2_C(\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}')$ thereby giving each subspace of the index-set, $\mathscr{S}$, its own kernel with associated hyperparameters.

Figure 3 shows each kernel profile with the same HPD and **Figure** 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.

Note, the evaluation of the physical kernel requires performing a double integral, 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 4 equal intervals without loss of effectiveness, however we choose to use 7 partitions. We discuss this choice in Section 6.5.

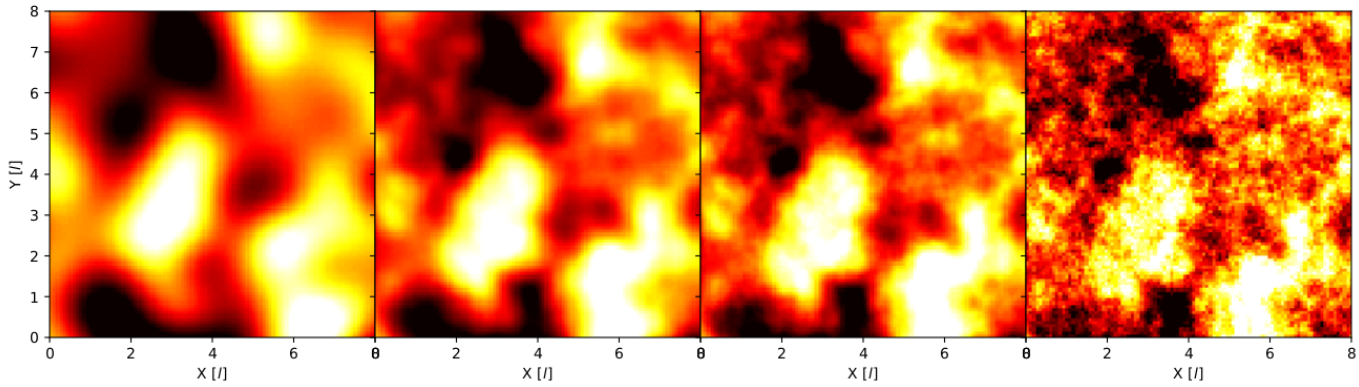

Fig. 2: Example realisations from exponential quadratic, Matérn-5/2, Matérn-3/2, and Matérn-1/2 kernels. The same HPD was used in all kernels, however the smoothness of the resulting process realisation is different for each.

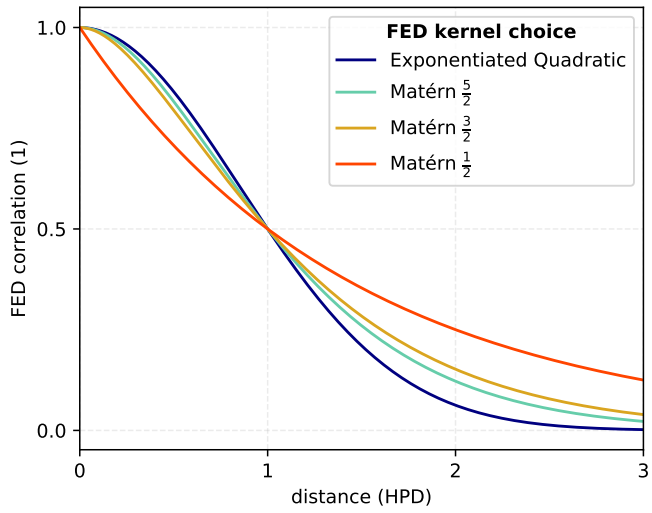

Fig. 3: This shows the shape of several kernels as a function of separation in units of the kernel's HPD.

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 this is often called respectively 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. A positive outcome would be all of these aspects being found to be true.

To measure how well a model, given a particular choice of kernel K and hyperparameters, represents the observed data 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 $\mathbf{X} \in \mathscr{S}$ according to $\boldsymbol{\tau}_{\rm obs} = \boldsymbol{\tau}(\mathbf{X}) + \epsilon$ where $\epsilon \sim \mathcal{N}[0, \sigma^2]$ and $\boldsymbol{\tau}(\mathbf{X}) \sim \mathcal{N}[0, K(\mathbf{X},\mathbf{X})]$ then the LPO is,

$$\log P_K(\boldsymbol{\tau}_{\rm obs}) = \log \mathcal{N}[0, \boldsymbol{B}] \tag{18}$$

where $\boldsymbol{B} = K(\mathbf{X},\mathbf{X}) + \sigma^2 \boldsymbol{I}$. 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 $\mathbf{X}^* \in \mathscr{S}$ according to $\boldsymbol{\tau}^*_{\rm obs} = \boldsymbol{\tau}(\mathbf{X}^*) + \epsilon^*$ with $\epsilon^* \sim \mathcal{N}[0, \sigma^2]$ then the LPH conditional on observed $\boldsymbol{\tau}_{\rm obs}$ is,

$$\log P_K\left(\boldsymbol{\tau}^*_{\rm obs} \mid \boldsymbol{\tau}_{\rm obs}\right) = \log \mathcal{N}[K(\mathbf{X}^*,\mathbf{X})\boldsymbol{B}^{-1}\boldsymbol{\tau}_{\rm obs}, \boldsymbol{B}^* - K(\mathbf{X}^*,\mathbf{X})\boldsymbol{B}^{-1}K(\mathbf{X},\mathbf{X}^*)] \tag{19}$$

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

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

$$\Delta\mathrm{LPO}_{\rm C}(\eta) \triangleq \frac{P_{\Delta\rm TEC}\left(\boldsymbol{\tau}_{\rm obs} \mid \eta\right)}{P_{\rm C}\left(\boldsymbol{\tau}_{\rm obs} \mid \eta\right)} \tag{20}$$

$$\Delta\mathrm{LPH}_{\rm C}(\eta) \triangleq \frac{P_{\Delta\rm TEC}\left(\boldsymbol{\tau}^*_{\rm obs} \mid \boldsymbol{\tau}_{\rm obs}, \eta\right)}{P_{\rm C}\left(\boldsymbol{\tau}^*_{\rm obs} \mid \boldsymbol{\tau}_{\rm obs}, \eta\right)} \tag{21}$$

where $P_{\Delta\rm TEC}$ is the probability distribution using the physical kernel and $P_{\rm C}$ is the distribution using a competitor kernel. The variable η represents a particular choice of experimental conditions, e.g. 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 $\Delta\mathrm{LPO}_{\rm C}(\eta)$ value of 1 implies that, for the given experimental conditions η, both models represent the observed data equally probable, and a value of 1.5 would imply that the physical kernel representation is 50% more probable than the competitor kernel. Note that considering the ratio of marginal probabilities is the canonical way of model selection (Rasmussen & Williams 2006). 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 & Williams 2006), which is equivalent to maximising the LPO of that model on the available observed dataset. We

Table 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

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 $\sigma^2_{n_e}$, and for the competitor kernels this corresponds to learning a variance and the length-scales for each subspace.

5. Results

In **Table** 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. In all experiments the optimisation converged. We observe that for both ionosphere varieties the discrepancy of a is on the order of a $\sim$ 10 km, or a few percent, implying that a can be learned from data. The discrepancy of HPD, is on the order of a 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. 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 a enough variation in the secant of the zenith angle. In a sparse and noisy observation of ΔTEC, the secant variation is poor and it's 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, on the order of 35%.

In **Table** 3 we summarise the performance of the physical kernel against each competitor kernel. We display the mean of $\Delta\text{LPO}_\text{C}(\eta)$, and $\Delta\text{LPH}_\text{C}(\eta)$ 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 have used bold font in **Table** 3 to indicate the best competitor model, which we term the best competitor model.

Consider first the ability of each model to represent the observed data. For the dawn ionosphere, the M52 competitor kernel has the best (lowest) $\langle\Delta\text{LPO}_\text{C}\rangle_\eta = 1.55$ and $\Delta\text{LPO}_\text{C}^{\eta_\text{nom}} = 1.46$, 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. 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.

Table 3: Shows the probability ratio FOMs (see text) averaged over experiemental conditions and at nominal conditions. Larger values indicate the physical model is more probable. Bold face indicates the best performing competitor model (lower number).

	$\langle\Delta\text{LPO}_\text{C}\rangle_\eta$	$\Delta\text{LPO}_\text{C}^{\eta_\text{nom}}$	$\langle\Delta\text{LPH}_\text{C}\rangle_\eta$	$\Delta\text{LPH}_\text{C}^{\eta_\text{nom}}$
		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**

Next, consider the ability of each model to infer the held-out data. For the dawn ionosphere, the M52 competitor kernel has the best (lowest) $\langle\Delta\text{LPH}_\text{C}\rangle_\eta = 1.49$ and $\Delta\text{LPO}_\text{C}^{\eta_\text{nom}} = 1.31$, 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. 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 that 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 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 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 best competitor and physical models' predictive means 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 the physical models'.

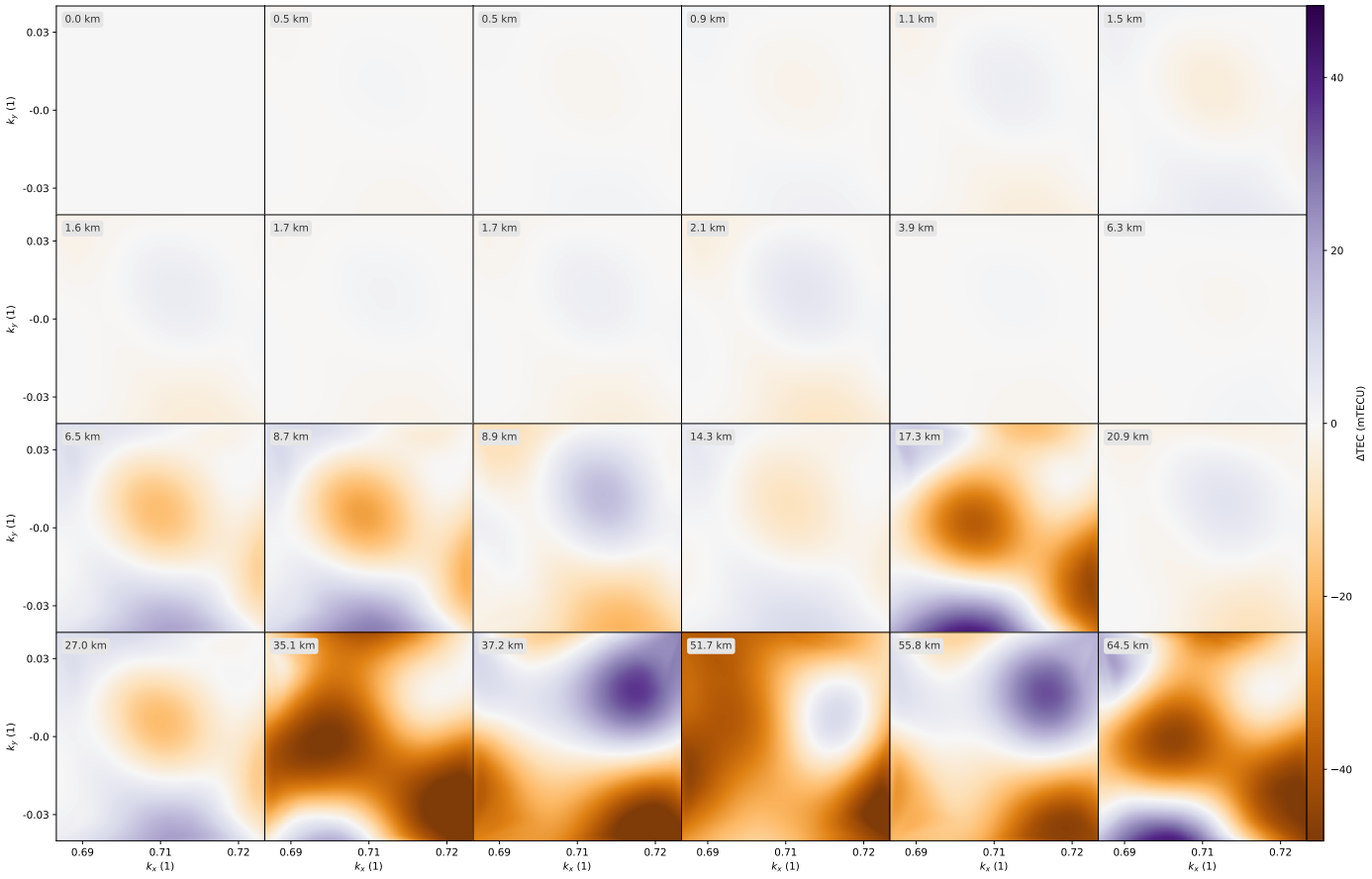

Fig. 4: Example of antenna-based ΔTEC screens from the dusk ionosphere simulation. Each plot show the simulated ground truth (noise-free) ΔTEC for each geodesic originating from that station with axes given in direction components k_x and k_y. The inset label gives how far the antenna is from the reference antenna. Antennae further from the reference antenna tend to have larger magnitude ΔTEC as expected. Each plot box bounds a circular 12.6 deg^2 field of view.

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

$$\delta l \approx \frac{q^2}{\epsilon_0 m_e \nu^2 r}\delta\tau \tag{22}$$

$$\approx 1.16'' \left(\frac{r}{10\text{km}}\right)^{-1}\left(\frac{\nu}{150\text{MHz}}\right)^{-2}\left(\frac{\delta\tau}{\text{mTECU}}\right) \tag{23}$$

In **Figure** 5 we have plotted the mean linear regression of the absolute equivalent source shift of the 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's about $4''$. Because there are few nearest calibrator distance 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 have a smaller slope and 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 separations comes from the fact that it's impossible on average to do better than the observation noise. At 1 degrees of separation, the physical model mean equivalent source shift is approximately half of the best competitor model. At 0 degrees of separation, the mean source shift is the same for both models as expected.

6. Discussion

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 be obvious that it would 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. In fact, it should be seen as a reason for prefering 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 & 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.

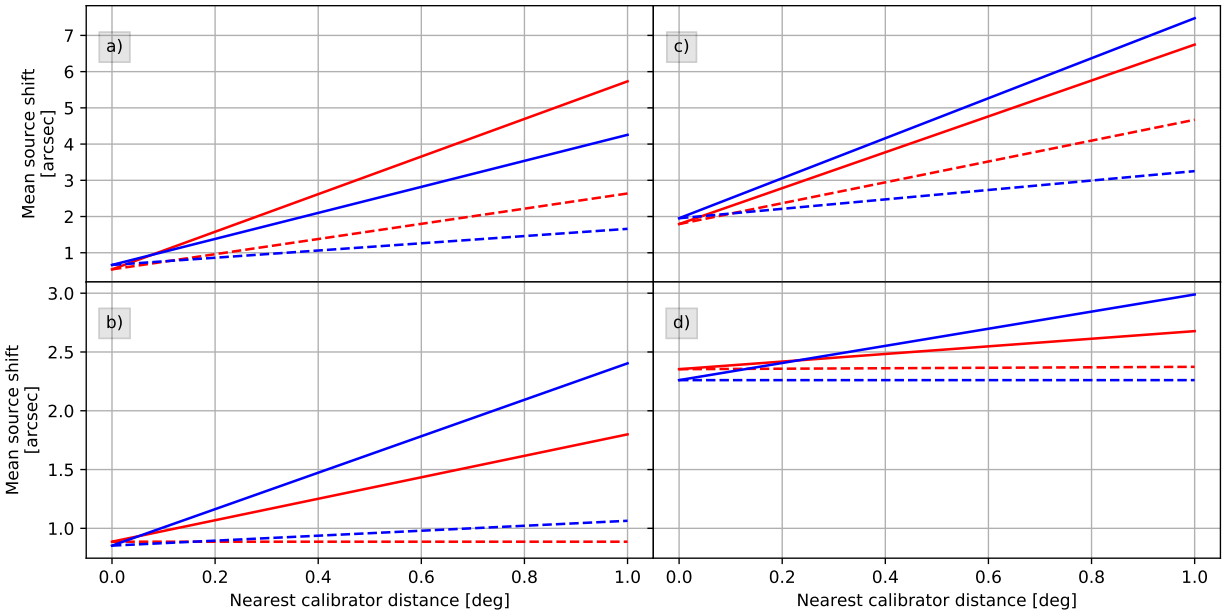

Fig. 5: This shows the mean equivalent source shift as a function of angular distance from the nearest calibrator caused by inference errors from the ground truth for **a)** remote stations (RS; > 3 km from the reference antenna) at nominal conditions (30 calibrators for 12.6deg^2 and 1 mTECU noise), **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 *solid line styles* are the best competitor models (see text), the *dashed line styles* are the physical model. The red lines are dawn ionospheres, and the blue lines are dusk ionospheres.

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.

6.2. Implicit tomography

The results of Section 5 indicate that the physical model provides a better explanation of ΔTEC data than any of the competitor models. One might ask, why does it perform 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 FED,

$$P(\boldsymbol{\tau} \mid \boldsymbol{\tau}_{\rm obs}) = \int_{\boldsymbol{n_e}} P(\boldsymbol{\tau} \mid \boldsymbol{n_e}) P(\boldsymbol{n_e} \mid \boldsymbol{\tau}_{\rm obs}) \, \mathrm{d}\boldsymbol{n_e}, \tag{24}$$

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

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

$$P(\boldsymbol{\tau} \mid \boldsymbol{\tau}_{\rm obs}) = \int_{\boldsymbol{n_e}} P(\boldsymbol{n_e}, \boldsymbol{\tau} \mid \boldsymbol{\tau}_{\rm obs}) \, \mathrm{d}\boldsymbol{n_e} \tag{25}$$

$$= \int_{\boldsymbol{n_e}} P(\boldsymbol{n_e} \mid \boldsymbol{\tau}_{\rm obs}) P(\boldsymbol{\tau} \mid \boldsymbol{n_e}, \boldsymbol{\tau}_{\rm obs}) \, \mathrm{d}\boldsymbol{n_e} \tag{26}$$

where in the second line we used the product rule of probability distributions (Kolmogorov 1956). Equating Eqs. 24 and 26, we discover that if, $P(\boldsymbol{\tau} \mid \boldsymbol{n_e}) = P(\boldsymbol{\tau} \mid \boldsymbol{n_e}, \boldsymbol{\tau}_{\rm obs})$ 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. 24 and 26 are equal due to the linear relation between FED and ΔTEC and 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 term this 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 seismic physics community for performing tomography of the Earth's interior.

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 do so is 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 $\boldsymbol{v}$, so that $n_e(\mathbf{x}, t) = n_e^0(\mathbf{x} - \boldsymbol{v}\, t)$. 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.

6.4. Structure function turnover and anisotropic diffractive scale

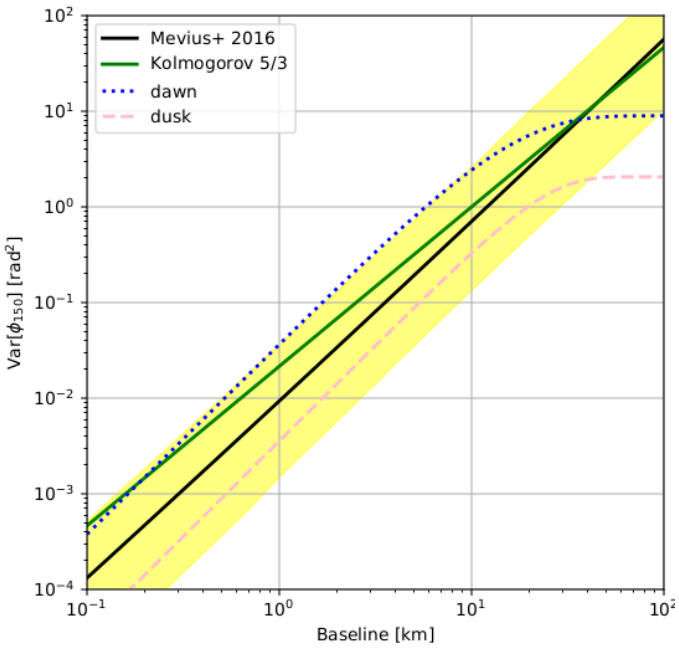

Fig. 6: The dotted and dashed lines show the phase structure function corresponding to the physical kernel, with the dawn and dusk configurations respectively (see **Figure** 1). Along side is the predicted structure function of Kolmogorov turbulence with a diffraction scale of 10 km, and the structure function constrained from observations in Mevius et al. (2016) with 1σ confidence region in yellow. Note, that Mevius et al. (2016) observes a turnover, but does not characterise it, thus we do not attempt to plot it here.

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 number 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 exponent $-5/3$. A convenient related measurable function for the ionosphere is the phase structure function (van der Tol 2009),

$$D(r) = \langle (\phi_\nu(R) - \phi_\nu(r+R))^2 \rangle_R \tag{27}$$

$$\triangleq \left(\frac{r}{r_{\text{diff}}}\right)^\beta \tag{28}$$

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 called the diffractive 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 the β to be 1.89 ± 0.1, slightly higher than the one predicted by Kolmogorov's theory, and the diffractive scale to range from 5 km to 30 km (Mevius et al. 2016).

In **Figure** 6 the structure functions of the physical kernel are shown for the dawn and dusk varieties, alongside Kolmogorov's $\beta = 5/3$ and the Mevius et al. (2016) observations. Though not plotted, Mevius 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 they 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 decaying 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.

We provide here an alternative explanation for the turnover, and slope deviating from Kolmogorov's 5/3, viz. that FED correlations are stationary, isotropic, and monotonically decreasing (SIMD). Moreover, as shown in Appendix B, our model in conjunction with 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 Earth's magnetic field lines. 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.

6.5. Low-accuracy numerical integration

The numerical integration required to compute Eq. 14 is performed using the 2D Trapezoid rule. This requires choosing 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 2 partitions, 20% with 3 partitions, 10% with 4 partitions, and 6% with 7 partitions. After experimentation it was surprisingly found that 2 partitions was sufficient to beat all competitor models, and that marginal improvement occurs after 5 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 4 partitions for our trials.

7. Conclusion

In this work, we have 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 evinced from ionosphere models. We have shown that under these assumptions the differential total electron content (ΔTEC), which is directly observable, 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 our model's efficacy, 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 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 to generalise 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 the best competitor models'. We prove that our model is cheaply performing implicit tomographic inference, which follows because ΔTEC is a linear projection of FED and the FED is a GP. We have suggested 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 deviating from Kolmogorov's 5/3-turbulence, 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.

Acknowledgements. J. G. A. and H. T. I. acknowledge funding by NWO under 'Nationale Roadmap Grootschalige 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.

References

Arora, B. S., Morgan, J., Ord, S. M., et al. 2016, PASA, 33, e031
Bilitza, D. & Reinisch, B. W. 2008, Advances in Space Research, 42, 599
Bouman, K. L., Johnson, M. D., Zoran, D., et al. 2016, in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
Byrd, R., Lu, P., Nocedal, J., & Zhu, C. 1995, SIAM Journal on Scientific Computing, 16, 1190
Cargill, P. J. 2007, Plasma Physics and Controlled Fusion, 49, 197
Cohen, M. H. 1973, IEEE Proceedings, 61, 1192
de Gasperin, F., Mevius, M., Rafferty, D., Intema, H., & Fallows, R. 2018, A&A, 615
Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&AS, 117, 137
Harrison, I., Camera, S., Zuntz, J., & Brown, M. L. 2016, MNRAS, 463, 3674
Hendriks, J. N., Jidling, C., Wills, A., & Schön, T. B. 2018, arXiv e-prints, arXiv:1812.07319
Hurley-Walker, N., Callingham, J. R., Hancock, P. J., et al. 2017, MNRAS, 464, 1146
Intema, H. T., van der Tol, S., Cotton, W. D., et al. 2009, A&A, 501, 1185
Jeffreys, H. 1925, Proceedings of the London Mathematical Society, s2-23, 428
Jidling, C., Hendriks, J., Wahlström, N., et al. 2018, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 436, 141
Jones, R. C. 1941, Journal of the Optical Society of America (1917-1983), 31, 488
Jordan, C. H., Murray, S., Trott, C. M., et al. 2017, MNRAS, 471, 3974
Kazemi, S., Yatawatta, S., Zaroubi, S., et al. 2011, MNRAS, 414, 1656
Keller, J., Bellman, R., & Society, A. M. 1964, Stochastic Equations and Wave Propagation in Random Media, Proceedings of symposia in applied mathematics (American Mathematical Society)
Kivelson, M. G. & Russell, C. T. 1995, Introduction to Space Physics, 586
Kolmogorov, A. N. 1956, Foundations of the theory of probability, 2nd edn. (Chelsea Pub Co)
Kolmogorov, A. N. 1991, Proceedings of the Royal Society of London Series A, 434, 9
Koopmans, L. V. E. 2010, ApJ, 718, 963
Loi, S. T., Murphy, T., Cairns, I. H., et al. 2015, Geophys. Res. Lett., 42, 3707
Matthews, A. G. d. G., van der Wilk, M., Nickson, T., et al. 2017, Journal of Machine Learning Research, 18, 1
Mevius, M., van der Tol, S., Pandey, V. N., et al. 2016, Radio Science, 51, 927
Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, ApJ, 838, 65
Rasmussen, C. E. & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning) (The MIT Press)
Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, A&A, 622, A1
Spoelstra, T. A. T. 1983, A&A, 120, 313
Tasse, C., Hugo, B., Mirmont, M., et al. 2018, A&A, 611, A87
van der Tol, S. 2009, PhD thesis, TU Delft
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
van Weeren, R. J., de Gasperin, F., Akamatsu, H., et al. 2019, Space Sci. Rev., 215, 16
van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016, ApJS, 223, 2
Vedantham, H. K. & Koopmans, L. V. E. 2015, MNRAS, 453, 925
Vernstrom, T., Gaensler, B. M., Brown, S., Lenc, E., & Norris, R. P. 2017, MNRAS, 467, 4914
Weiss, Y. & Freeman, W. T. 2001, Neural Computation, 13, 2173
Wilcox, C. H. 1962, SIAM Review, 4, 55
Wolf, E. 1969, Optics Communications, 1, 153
Wu, L., Wang, D., & A. Evans, J. 2019, Nature, 566, 1
Yeh, K. C. 1962, Journal of Research of National Bureau of Standards, 5, 621
Yeh, K. C. & Swenson, Jr., G. W. 1959, J. Geophys. Res., 64, 2281

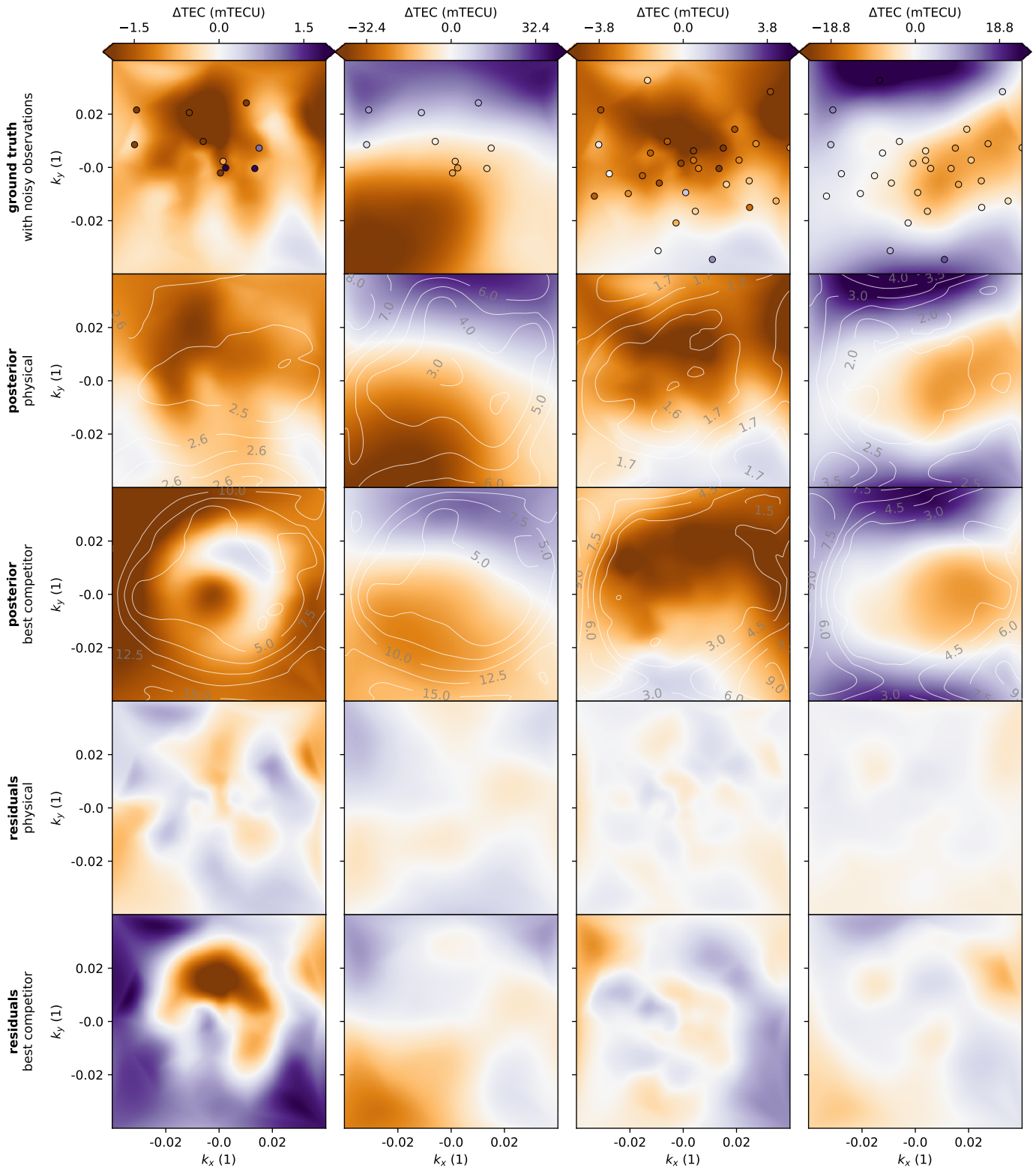

Fig. 7: Example visual comparison of the predictive performance of our physical model versus that of the best competitor model for the dawn ionosphere. *First row*, is the ground truth ΔTEC overlaid noisy draws from the ground truth which are the observations; *Second/third rows*, the posterior mean with uncertainty contours for the physical model and best competitor model respectively. *Fourth/fifth rows*, the 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.

Appendix A: Derivation of tomographic equivalence

We now explicitly prove the assertion that Eq. 24 is equal to Eq. 26, that is,

$$\int P(\boldsymbol{\tau} \mid \boldsymbol{n_e}) P(\boldsymbol{n_e} \mid \boldsymbol{\tau}_{\text{obs}}) \, \mathrm{d}\boldsymbol{n_e} = \int P(\boldsymbol{n_e}, \boldsymbol{\tau} \mid \boldsymbol{\tau}_{\text{obs}}) \, \mathrm{d}\boldsymbol{n_e}. \tag{A.1}$$

Note that we'll sometimes use the notation $\mathcal{N}[a \mid m_a, C_a]$ which is equivalent to $a \sim \mathcal{N}[m_a, C_a]$.

We define the matrix representation of the DRI operator in Eq. 6, $\boldsymbol{\Delta}_* \boldsymbol{n_e} = \{\Delta^{\hat{\boldsymbol{k}}}_{\mathbf{x}} n_e \mid (\mathbf{x}, \hat{\boldsymbol{k}}) \in \mathscr{S}_*\}$, and likewise let $\boldsymbol{\Delta}$ be the matrix representation over the index-set $\mathscr{S}_{\text{obs}}$. Similarly, the matrix representation of the FED kernel – the Gram matrix – is $\boldsymbol{K} = \{K(\mathbf{x}, \mathbf{x}') \mid \mathbf{x}, \mathbf{x}' \in \mathscr{X}\}$. Using these matrix representation we have the following joint distribution,

$$P(\boldsymbol{n_e}, \boldsymbol{\tau}, \boldsymbol{\tau}_{\text{obs}}) = \mathcal{N}\begin{bmatrix} \bar{n}_e & \boldsymbol{K} & \boldsymbol{K}\boldsymbol{\Delta}_*^T & \boldsymbol{K}\boldsymbol{\Delta}^T \\ 0\,, & \boldsymbol{\Delta}_*\boldsymbol{K} & \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}_*^T & \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}^T \\ 0 & \boldsymbol{\Delta}\boldsymbol{K} & \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}_*^T & \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I} \end{bmatrix}. \tag{A.2}$$

Let us first work out the left-hand side (LHS) of Eq. A.1. Because $\boldsymbol{\tau} = \boldsymbol{\Delta}_* \boldsymbol{n_e}$, and using standard Gaussian identities we have,

$$P(\boldsymbol{\tau} \mid \boldsymbol{n_e}) = \mathcal{N}[\underbrace{\boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{K}^{-1}(\boldsymbol{n_e} - \bar{n}_e)}_{\boldsymbol{\Delta}_*(\boldsymbol{n_e} - \bar{n}_e)}, \underbrace{\boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}_* - \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{K}^{-1}\boldsymbol{K}\boldsymbol{\Delta}_*}_{\mathbf{0}}] \tag{A.3}$$

Similarly, the second distribution on the LHS is,

$$P(\boldsymbol{n_e} \mid \boldsymbol{\tau}_{\text{obs}}) = \mathcal{N}[\bar{n}_e + \boldsymbol{K}\boldsymbol{\Delta}^T(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\tau}_{\text{obs}}, \boldsymbol{K} - \boldsymbol{K}\boldsymbol{\Delta}^T(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\Delta}\boldsymbol{K}]. \tag{A.4}$$

We now apply belief propagation of Gaussians (Weiss & Freeman 2001) to evaluate the integral on the LHS,

$$\int P(\boldsymbol{\tau} \mid \boldsymbol{n_e}) P(\boldsymbol{n_e} \mid \boldsymbol{\tau}_{\text{obs}}) \, \mathrm{d}\boldsymbol{n_e} \tag{A.5}$$

$$= \int \mathcal{N}[\boldsymbol{\tau} \mid \boldsymbol{\Delta}_*(\boldsymbol{n_e} - \bar{n}_e), \mathbf{0}] \mathcal{N}[\boldsymbol{n_e} \mid \bar{n}_e + \boldsymbol{K}\boldsymbol{\Delta}^T(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\tau}_{\text{obs}}, \boldsymbol{K} - \boldsymbol{K}\boldsymbol{\Delta}^T(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\Delta}\boldsymbol{K}] \, \mathrm{d}\boldsymbol{n_e} \tag{A.6}$$

$$= \mathcal{N}[\underbrace{-\boldsymbol{\Delta}_*\bar{n}_e + \boldsymbol{\Delta}_*(\bar{n}_e + \boldsymbol{K}\boldsymbol{\Delta}^T(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\tau}_{\text{obs}})}_{\boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}^T(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\tau}_{\text{obs}}}, \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}_*^T - \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}^T(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}_*^T] \tag{A.7}$$

In order to work out the right-hand side (RHS), we simply condition Eq. A.2 on $\boldsymbol{\tau}_{\text{obs}}$ and then marginalise $\boldsymbol{n_e}$ by selecting the corresponding sub-block of the Gaussian,

$$P(\boldsymbol{n_e}, \boldsymbol{\tau} \mid \boldsymbol{\tau}_{\text{obs}}) \tag{A.8}$$

$$= \mathcal{N}\left[\begin{pmatrix} \bar{n}_e \\ 0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{K}\boldsymbol{\Delta}^T \\ \boldsymbol{\Delta}_*^T\boldsymbol{K}\boldsymbol{\Delta}^T \end{pmatrix}\left(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{\tau}_{\text{obs}}, \begin{pmatrix} \bar{K} & \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}_*^T \\ \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}^T & \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}_*^T \end{pmatrix} - \begin{pmatrix} \boldsymbol{K}\boldsymbol{\Delta}^T \\ \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}^T \end{pmatrix}\left(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I}\right)^{-1}\begin{pmatrix} \boldsymbol{\Delta}\boldsymbol{K} & \boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}_*^T \end{pmatrix}\right] \tag{A.9}$$

Marginalising over $\boldsymbol{n_e}$ is equivalent to neglecting the sub-block corresponding to $\boldsymbol{n_e}$. Therefore, the RHS is,

$$\int P(\boldsymbol{n_e}, \boldsymbol{\tau} \mid \boldsymbol{\tau}_{\text{obs}}) \, \mathrm{d}\boldsymbol{n_e} = \mathcal{N}\left[\boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}^T\left(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{\tau}_{\text{obs}}, \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}_*^T - \boldsymbol{\Delta}_*\boldsymbol{K}\boldsymbol{\Delta}^T\left(\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}^T + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{\Delta}\boldsymbol{K}\boldsymbol{\Delta}_*^T\right]. \tag{A.10}$$

■

Appendix B: Derivation of the differential TEC variance function and its limits

We derive the ΔTEC variance function $\sigma^2_{\Delta\text{TEC}}(d)$ for zenith observations ($\boldsymbol{k} = \boldsymbol{k}' = \hat{\boldsymbol{z}}$) by considering a baseline between an antenna-of-interest at $\mathbf{x}_i = \mathbf{x}_j$ and a reference antenna at $\mathbf{x}_0 = \mathbf{0}$. To use the Pythagorean theorem later, we assume that this baseline lies in the plane of the local horizon, i.e. perpendicular to the zenith. Without loss of generality, we can orient the coordinate axes such that this baseline lies along the $\hat{\mathbf{x}}$ direction, so that $\mathbf{x}_i - \mathbf{x}_0 = d\hat{\mathbf{x}}$. Here $d \triangleq ||\mathbf{x}_i||$ is the distance between the two antennae. We then take the general covariance function $K_{\Delta\text{TEC}}\left(\left[\mathbf{x}_i, \mathbf{x}_0, \hat{\mathbf{k}}\right], \left[\mathbf{x}_j, \mathbf{x}_0, \hat{\mathbf{k}}'\right]\right)$, and find that in this particular case

$$\sigma^2_{\Delta\text{TEC}}(d) \triangleq K_{\Delta\text{TEC}}([\mathbf{x}_i, \mathbf{x}_0, \hat{\mathbf{z}}], [\mathbf{x}_i, \mathbf{x}_0, \hat{\mathbf{z}}]) \tag{B.1}$$

$$= \sum_{p_1=0}^{1}\sum_{p_2=0}^{1}(-1)^{p_1+p_2}\int_0^b\int_0^b K_{n_e}\left(||\mathbf{x}_{(1-p_1)i} - \mathbf{x}_{(1-p_2)i} + \hat{\mathbf{z}}(s_1 - s_2)||\right) \mathrm{d}s_1 \mathrm{d}s_2, \tag{B.2}$$

where K_{n_e} is an arbitrary stationary and isotropic kernel (such as the Exponentiated Quadratic and Matérn $\frac{3}{2}$ kernels considered earlier) for the FED. The two terms where p_1 and p_2 are equal give the same contribution, as do the two terms for which p_1 and p_2 are unequal. By subsequently applying the Pythagorean theorem in this last case (i.e. $p_1 = 0$ and $p_2 = 1$, and vice versa), we find

$$\sigma^2_{\Delta\mathrm{TEC}}(d) = 2\int_0^b \int_0^b K_{n_e}(|s_1 - s_2|) - K_{n_e}\left(\sqrt{d^2 + (s_1 - s_2)^2}\right) \mathrm{d}s_1 \mathrm{d}s_2. \tag{B.3}$$

We manipulate this result to obtain a more insightful expression. First, we note the (implicit) presence of three parameters with dimension length: ionospheric thickness b, reference antenna distance d, and FED kernel half-peak distance h. We perform transformations to dimensionless coordinates $u_1 = \frac{s_1}{h}$ and $u_2 = \frac{s_2}{h}$ to reveal that the *shape* - though not the absolute *scale* - of the function $\sigma^2_{\Delta\mathrm{TEC}}(d)$ is governed only by the length-scale ratios $\frac{b}{h}$ and $\frac{d}{h}$, and the particular functional form of K_{n_e}. Furthermore, for stationary covariance functions, we have $K_{n_e} = \sigma^2_{n_e} C_{n_e}$, where C_{n_e} is the corresponding dimensionless correlation function.
These considerations enable us to express the ΔTEC structure function as a dimensionless, shape-determining double integral appended by dimensionful prefactors; i.e.

$$\sigma^2_{\Delta\mathrm{TEC}}(d) = 2\sigma^2_{n_e} h^2 \int_0^{\frac{b}{h}} \int_0^{\frac{b}{h}} C_{n_e}(h|u_1 - u_2|) - C_{n_e}\left(h\sqrt{\left(\frac{d}{h}\right)^2 + (u_1 - u_2)^2}\right) \mathrm{d}u_1 \mathrm{d}u_2. \tag{B.4}$$

First note that the variance of ΔTEC is simply proportional to the variance of n_e. Note also that $h|u_1 - u_2| < h\sqrt{\left(\frac{d}{h}\right)^2 + (u_1 - u_2)^2}$ for any non-zero d, so that $C_{n_e}(h|u_1 - u_2|) > C_{n_e}\left(h\sqrt{\left(\frac{d}{h}\right)^2 + (u_1 - u_2)^2}\right)$ for all monotonically decreasing correlation functions C_{n_e} (or, equivalently, covariance functions K_{n_e}). With the integrand always positive, we see that the *integral* must be a strictly increasing function of $\frac{b}{h}$ (which occurs in the integration limits). Thus we conclude that for stationary, isotropic and monotonically decreasing (SIMD) FED kernels with HPD h, the ΔTEC variance increases monotonically with the thickness of the ionosphere b. Simply put: thicker SIMD ionospheres cause larger ΔTEC variations.

Let us now consider three limits of the ΔTEC zenith variance function, that all do not require K_{FED} to decrease monotonically.

- In the short-baseline limit, i.e. $\frac{d}{h} \to 0$, we have $C_{n_e}\left(h\sqrt{\left(\frac{d}{h}\right)^2 + (u_1 - u_2)^2}\right) \to C_{n_e}(h|u_1 - u_2|)$. We therefore find that $\sigma^2_{\Delta\mathrm{TEC}} \to 0$ irrespective of other parameters, recovering that the variance of ΔTEC vanishes near the reference antenna.
- In the long-baseline limit, i.e. $\frac{d}{h} \gg \frac{b}{h} > 1$, we see that $\sqrt{\left(\frac{d}{h}\right)^2 + (u_1 - u_2)^2} \approx \frac{d}{h}$, since $(u_1 - u_2)^2 < \left(\frac{b}{h}\right)^2 \ll \left(\frac{d}{h}\right)^2$. Assuming $C_{n_e}(d) \approx 0$ when $\frac{d}{h} \gg 1$, the integrand reduces to $C_{n_e}(h|u_1 - u_2|) - C_{n_e}\left(h \cdot \frac{d}{h}\right) \approx C_{n_e}(h|u_1 - u_2|)$. We find that in this case,

  $$\sigma^2_{\Delta\mathrm{TEC}} \approx 2\sigma^2_{n_e} h^2 \int_0^{\frac{b}{h}} \int_0^{\frac{b}{h}} C_{n_e}(h|u_1 - u_2|) \mathrm{d}u_1 \mathrm{d}u_2. \tag{B.5}$$

  This is the plateau value of the ΔTEC variance that our model predicts for the long-baseline limit.
  Another way to arrive at the plateau value expression of Equation B.5 is by considering the statistical properties of TEC first. In a computation analogous to the one for ΔTEC in Section 3, one can derive the general TEC covariance function K_{TEC}. The variance of $\tau_i^{\hat{\boldsymbol{z}}}$ (the TEC of antenna i while observing towards the zenith $\hat{\boldsymbol{z}}$) is straightforwardly shown to be

  $$\mathbb{V}\left(\tau_i^{\hat{\boldsymbol{z}}}\right) = \sigma^2_{n_e} h^2 \int_0^{\frac{b}{h}} \int_0^{\frac{b}{h}} C_{n_e}(h|u_1 - u_2|) \mathrm{d}u_1 \mathrm{d}u_2. \tag{B.6}$$

  Note the absence of a dependence on i at the RHS. As a ΔTEC is simply a TEC differenced with a TEC for a reference antenna observing in the same direction, we have

  $$\sigma^2_{\Delta\mathrm{TEC}} = \mathbb{V}\left(\tau_i^{\hat{\boldsymbol{z}}} - \tau_0^{\hat{\boldsymbol{z}}}\right) = \mathbb{V}\left(\tau_i^{\hat{\boldsymbol{z}}}\right) + \mathbb{V}\left(\tau_0^{\hat{\boldsymbol{z}}}\right) \tag{B.7}$$

  where the second equality only holds when the TECs are independent. This is exactly the scenario considered in the long-baseline limit. Plugging in Equation B.6 recovers the plateau level.
- We can find a general upper bound to the variance of ΔTEC in terms of physical parameters. To this end, note that the integrand in Equation B.4 is maximised when, over the full range of integration, the first term is 1 whilst the second term equals the infimum of the correlation function. Calling $\inf_{\mathbb{R}}\{C_{n_e}(r) : r \in \mathbb{R}_{>0}\} \triangleq I$, we find the inequality

  $$\sigma^2_{\Delta\mathrm{TEC}} \leq 2\sigma^2_{n_e} h^2 \int_0^{\frac{b}{h}} \int_0^{\frac{b}{h}} 1 - I \, \mathrm{d}u_1 \mathrm{d}u_2 = 2(1 - I)\sigma^2_{n_e} b^2. \tag{B.8}$$

  For strictly positive FED kernels that decay to 0 at large distances (such as the EQ and Matérn kernels considered in this work), we thus find $\sigma^2_{\Delta\mathrm{TEC}} \leq 2\sigma^2_{n_e} b^2$. Kernels resulting in anticorrelated FEDs produce the constraint $\sigma^2_{\Delta\mathrm{TEC}} \leq 4\sigma^2_{n_e} b^2$ or tighter. By measuring $\sigma_{\Delta\mathrm{TEC}}(d)$, one can bound the product $\sigma_{n_e} b$ from below. The strongest bound is obtained for large d.