

FULL PAPER

Open Access



Ionospheric electron density modelling using B-splines and constraint optimization

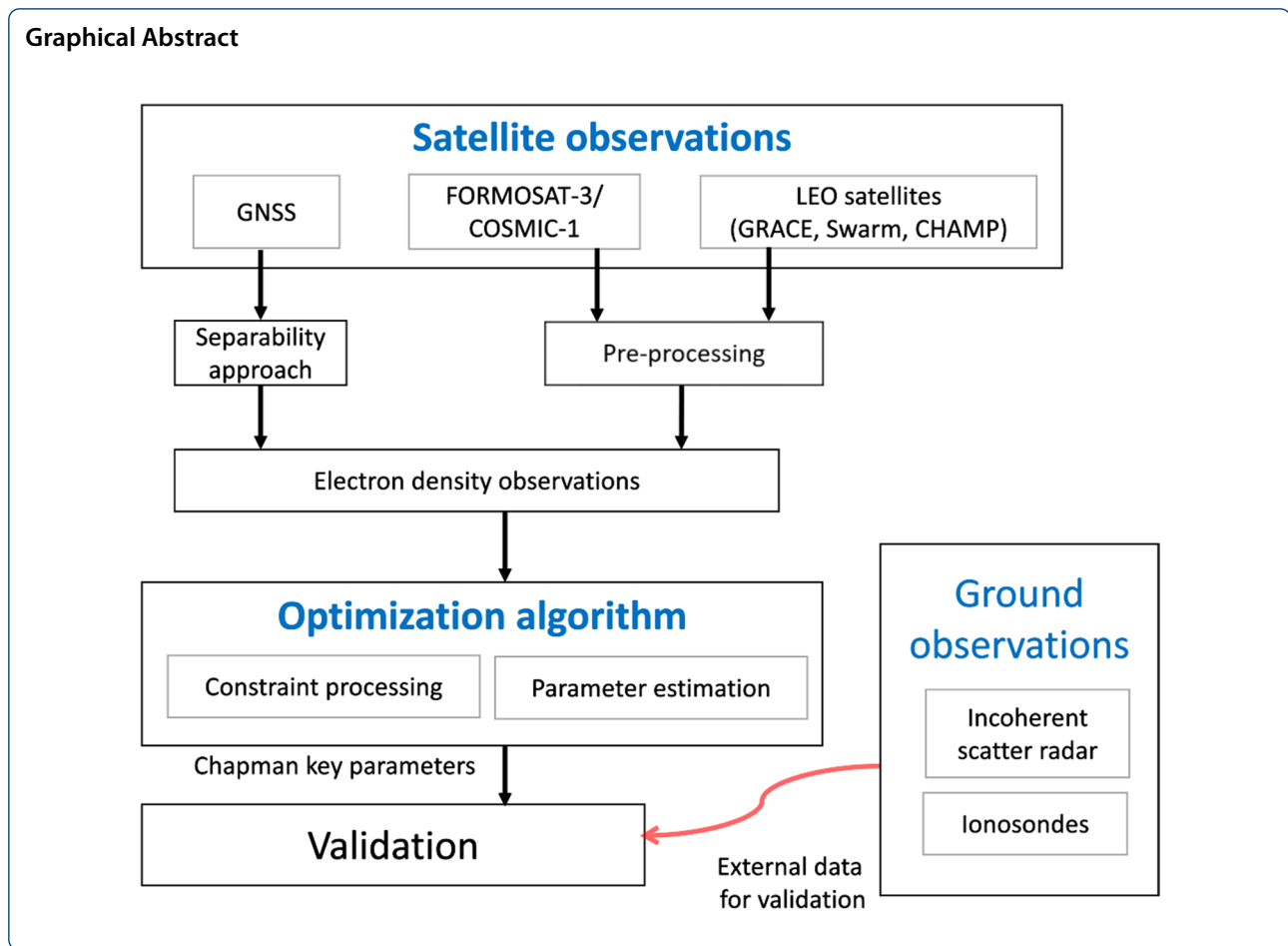
Ganesh Lalgudi Gopalakrishnan*  and Michael Schmidt

Abstract: Many modern applications, such as precise point positioning, autonomous driving or precision agriculture would benefit significantly if a high-precision and high-resolution model of electron density in the ionosphere and the plasmasphere would be globally available. Since the development of such a model still relies on data with insufficient and uneven global coverage, the consideration of background information and the introduction of equality and inequality constraints on Chapman key parameters are essential. In this work, we develop a multi-layer Chapman model based on B-spline expansions of selected key parameters of the electron density. The unknown series coefficients of the key parameters are subject to equality and inequality constraints. Finally, the developed model is applied to a combination of real and semi-simulated input data; the results are validated through ionosonde measurements.

Keywords: Ionosphere layers, Multi-layer Chapman model, Inequality constraints, B-spline expansions, Optimization approach

*Correspondence: ganesh.lalgudi-gopalakrishnan@tum.de

DGFI-TUM, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany



Introduction

The Sun's radiation has important impact on different components of the Earth system, for instance on the sub-components plasmasphere and ionosphere. The ionization of neutral gas molecules results in ionized plasma and free electrons. The four-dimensional (4D) space and time-dependent electron density need to be accurately known for satellite navigation and telecommunication, since electromagnetic waves are strongly affected on their path through the charged part of the atmosphere¹. The 4D electron density is a highly variable function: amongst others it reflects plasma fluctuations with periods less than a few minutes, diurnal and seasonal variations, long-period changes following the solar cycle of 11 years as well as the impact of space weather events, such as solar flares and coronal mass ejections (CME).

If the 4D electron density would be known everywhere within the plasmasphere and the ionosphere at any moment in time, we would be able to correct every measurement of space-geodetic observation techniques, such as the Global Navigation Satellite Systems (GNSS), Satellite Altimetry, the French satellite tracking system Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS) or Very Long Baseline Interferometry (VLBI) for the plasmaspheric and ionospheric impact independently if we use single or multi-frequency measurements. Vice versa, all these observation techniques including, e.g. the space-based GNSS technique of measuring Ionospheric Radio Occultations (IRO) as well as the Langmuir probe measurements on-board CHAMP and Swarm provide valuable information on the state of the plasmasphere and the ionosphere, and thus for modelling the electron density. To be more specific, in this paper we aim on modelling the electron density function

$$N_e = N_e(\varphi, \lambda, h, t) \quad (1)$$

¹ Availability of ionospheric corrections also improve convergence time in precise point positioning (PPP) solutions.

depending on latitude φ and longitude λ —both defined with respect to a geocentric Earth-bounded coordinate system Σ_E —on the height $h = r - R_E$ above the surface of a spherical Earth with a constant radius R_E and r being the radial distance and on the time t with high accuracy and high resolution. Since modelling a 4D function is rather challenging, the electron density is usually transferred to the 3D Vertical Total Electron Content (VTEC) depending on latitude and longitude as well as the time. Integrating the electron density (1) along the ray-path, e.g. between a transmitting GNSS satellite P^S and a ground-based receiver P_R the Slant Total Electron Content (STEC)

$$STEC(P^S, P_R, t) = \int_{P^S}^{P_R} N_e(\varphi, \lambda, h, t) ds \quad (2)$$

is obtained. Due to the dispersive property of the plasma-sphere and the ionosphere the geometry of the ray-path with the differential line element ds depends on the carrier frequency f of the electromagnetic wave, and thus deviates from the straight line between P^S and P_R . To simplify the notation and bring vector notation into play, we introduce the 3D position vector

$$\mathbf{x} = r \cdot [\cos \varphi \cos \lambda, \cos \varphi \sin \lambda, \sin \varphi]^T \quad (3)$$

defined in Σ_E and rewrite Eq. (2) as

$$STEC(\mathbf{x}^S, \mathbf{x}_R, t) = \int_{P^S}^{P_R} N_e(\mathbf{x}, t) ds. \quad (4)$$

Multiplying the right-hand side of Eq. (4) with the factor $\pm 40.3/f^2$, we obtain the ionospheric delay $d_{\text{ion}} = (\pm 40.3/f^2) \cdot STEC$. This relation is an approximation, valid for standard conditions of the geomagnetic field and signal frequencies f larger than 1 GHz. It neglects the higher-order effects which may reach up to 0.2 cm in zenith direction for GPS signals. Note, the sign of d_{ion} changes with reference to carrier phase observations (“−”) or for pseudorange measurements (“+”). The conversion of STEC into VTEC is achieved by introducing an isotropic mapping function $M(z)$ depending on the zenith angle z . It performs the projection

$$VTEC(\mathbf{x}_{IPP}, t) = M(z) \cdot STEC(\mathbf{x}^S, \mathbf{x}_R, t) \quad (5)$$

under the assumption that all electrons are concentrated in a spherical shell Ω_H of an infinitesimal thickness with fixed height $H = r - R_E$ above the spherical Earth. This procedure is denoted as the single layer model (SLM) approach; the spatial reference for the VTEC observation is given by the ionospheric piercing point P_{IPP} with position vector \mathbf{x}_{IPP} . It is defined as the intersection point of the straight line between P^S and P_R and the sphere Ω_H ;

see e.g., Schaer (1999). From Eq. (4), we obtain from the integration of the electron density along the height h the result

$$VTEC(\mathbf{x}_{IPP}, t) = \int_{\text{height}} N_e(\varphi_{IPP}, \lambda_{IPP}, h, t) dh. \quad (6)$$

Note, that if the electron density is precisely known the result (6) usually outperforms the corresponding VTEC value from Eq. (5) because applying the mapping function $M(z)$ can introduce a significant error.

Several approaches exist for representing VTEC globally by using input data from GNSS. The Ionospheric Associated Analysis Centers (IAAC) of the International GNSS Service (IGS) routinely provide Global Ionosphere Maps (GIM) representing VTEC based on different mathematical approaches, e.g. the series expansion or the discretization technique. Whereas in the latter case, usually the voxel-based tomography approach is chosen (see e.g. Hernández-Pajares et al. (1999)), in the case of a 2D series expansion spherical harmonics (SH) are often used (see e.g. Schaer (1999)). However, since SHs are oscillating over the sphere and, thus, are characterized by a *global* behaviour, they cannot represent data of heterogeneous density and quality in a proper way; see e.g. Schmidt et al. (2011). Consequently, at DGFI-TUM, we rely for more than two decades on 2D B-spline basis functions for modelling VTEC, see e.g. Zeilhofer (2008), Schmidt (2007), Schmidt et al. (2015), Limberger (2015) and Erdogan et al. (2017a) for two main reasons, namely (1) they are characterized by their localizing feature and (2) they can be used to generate a multi-scale representation (MSR) based on wavelet theory; see e.g. Schmidt et al. (2015) and Goss et al. (2019). In the case of global modelling, the 2D B-spline basis functions are defined as tensor products of 1D endpoint-interpolating B-spline functions depending on the latitude φ and 1D trigonometric B-spline functions depending on the longitude λ ; see Schmidt et al. (2011), Schmidt et al. (2015), and Goss et al. (2019); general introductions to both kinds of B-spline basis functions can be found in Schumaker and Traas (1991) as well as Lyche and Schumaker (2000).

Multiplying a 2D series expansion, e.g. in terms of SHs or 2D tensor product B-spline functions dependent on latitude φ and longitude λ with a vertical profile function dependent on the height h , a mathematical model is obtained for describing a 3D function, such as the spatial part of the electron density N_e , i.e. $N_e(\varphi, \lambda, h)$. Howe et al. (1998) and Gao and Liu (2002) realized this concept—also known as the *separability approach* Hernández-Pajares et al. (2000)—by combining a SH expansion with empirical orthogonal functions (EOF); Schmidt et al. (2008) transferred this global approach to a regional

application by multiplying the 2D tensor product endpoint-interpolating B-splines with the first five EOFs obtained from a principal component analysis (PCA) of a reference model, namely the International Reference Ionosphere (IRI). As an alternative to the aforementioned empirical approach, in ionosphere research often so-called “physics-motivated” functions, e.g. the *Chapman function* and the Epstein function are used; see Feltens (2007) and Rawer (1988). As will be shown by Eq. (8), we rely in this work on the Chapman function. A further approach is defining the vertical profile function as another 1D series expansion in terms of endpoint-interpolating B-spline functions along the height. This way the electron density $N_e(\varphi, \lambda, h)$ is spatially modelled by a series expansion in terms of 3D tensor product B-spline functions; This approach increases the number of B-spline coefficients to be estimated significantly, but remains free from any assumptions about the physical distribution of the electrons along the vertical. A synergy between this 3D B-spline representation and the voxel-based approach mentioned before was shown by Olivares-Pulido et al. (2019). An extensive discussion of the aforementioned and further approaches to model the electron density (1) only spatially or even spatio-temporally, i.e. in three or four dimensions, can be found in the PhD thesis of Limberger (2015).

Each Chapman function (8) is defined by three key parameters, two of them are included in a strongly non-linear manner and are also highly correlated with each other. Since such strong correlations also occur between key parameters of different ionospheric layers, the application of the classical least squares estimation often leads—if at all—to physically unrealistic estimation results for the unknown key parameters, e.g. a negative value for the maximum of the electron density within a specified layer; see Liang (2017). Such problems are often handled by introducing “inequality constraints” for the unknown parameters and solved by using optimization algorithms, such as the ones described in Koch (1985), Roese-Koerner (2015), Necedal and Wright (2006), Boyd and Vandenberghe (2004). Inequality-constrained optimization problems are relatively difficult to solve compared to either unconstrained or equality constrained optimization problems; see Chapter 16 of Necedal and Wright (2006).

Although there are optimization procedures available (see e.g. Lötstedt (1984), Gill et al. (1984), Coleman and Li (1996) and Mead and Renaut (2010)), the detail of their applied algorithms, the applied assumptions and the implementation along with the possibility to obtain accuracy information, i.e. to estimate standard deviations are either not described or not always explicitly transparent

to the user or not available at all. In many cases these algorithms are made for smaller dimensional problems—within 100 to 500 unknown parameters—and are not easily “scalable” for relatively higher-dimensional problems, e.g. a global 4D electron density modelling.

In this paper, due to the consideration of inequality constraints, the estimators of the selected Chapman key parameters are lying within physically realistic limits, for example, non-negative values for the maximum value of the electron density within a specified ionosphere layer. We will develop in the following a general concept where both equality and inequality constraints can be included.

This paper is organized as follows: in the second Section titled “*Electron density modelling*”, we derive the multi-layer Chapman model of the global electron density. In the following third Section, we set up the series expansions in 2D B-spline basis functions for the selected key parameters and define the inequality and equality constraints related to given bound functions in matrix and vector form. In the next section titled “*optimization approach*”, we describe the optimization algorithm in detail and present a recipe for applying our developed ICOA. In the next section, we then apply the ICOA to a combination of real and semi-simulated input data and describe the validation of the estimated key parameters using the IRI model as well as ionosonde observations. Finally, we provide a summary and outlook for follow-up work.

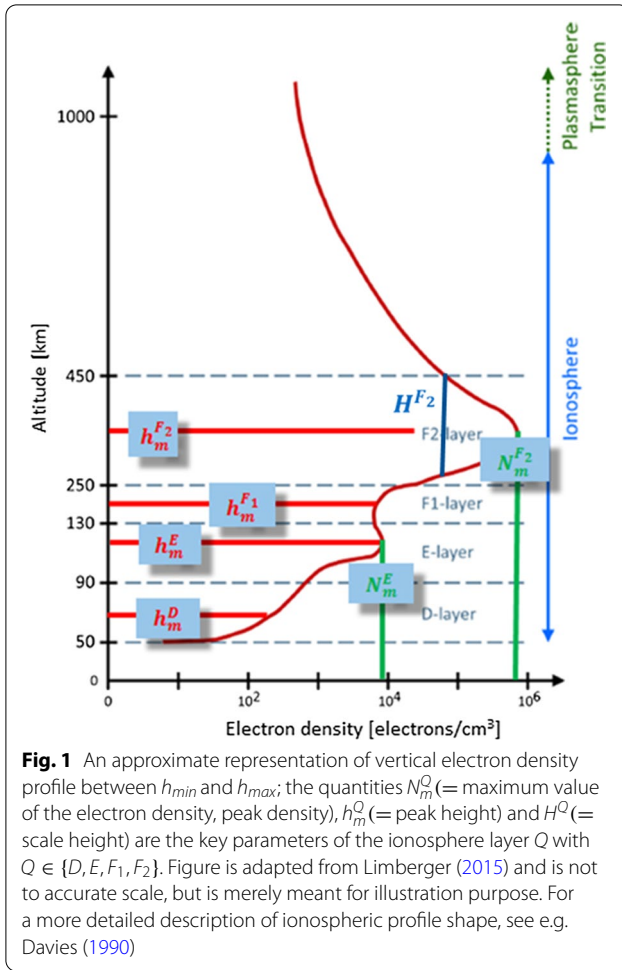
Electron density modelling

As can be seen in Fig. 1 the ionosphere defined between the height values h_{min} and h_{max} is stratified into the D -, the E -, the F_1 - and the F_2 -layer with a smooth transition into the plasmasphere. Following Limberger (2015), we may set $h_{min} = 50$ km and $h_{max} = 1000$ km. Each layer has its physical significance, due to the dominating atmospheric processes, such as ionization, recombination and the overall energy balance of the ionosphere; see e.g. Chapman (1931).

In this work, we approximate according to Feltens (2007) the vertical electron density profile by the so-called multi-layer approach

$$\begin{aligned} N_e(h) &= N_e^D(h) + N_e^E(h) + N_e^{F1}(h) + N_e^{F2}(h) + N_e^P(h) \\ &= \sum_{Q=1}^4 N_m^Q p^Q(h) + N_0^P p^P(h), \end{aligned} \quad (7)$$

where the summation index $Q \in \{1 = D, 2 = E, 3 = F_1, 4 = F_2\}$ refers to the four layers of the ionosphere and P means the plasmasphere. We use the Chapman function



$$N_e^Q(h) = N_m^Q \exp\left(\frac{1}{2}\left(1 - \frac{h - h_m^Q}{H^Q} - \exp\left(-\frac{h - h_m^Q}{H^Q}\right)\right)\right) \quad (8)$$

to describe the vertical electron density distribution of the Q -layer. Herein, the three quantities N_m^Q (= peak density), h_m^Q (= peak height) and H^Q (= scale height) are the key parameters of the Q -layer. The peak density N_m^Q corresponds to the maximum value of the electron density within the Q -layer and the peak height h_m^Q means the altitude corresponding to the peak density. The scale height H^Q refers to the thickness of the profile around the peak height. Whereas N_m^Q means a scale factor in Eqs. (7) and (8), the key parameters h_m^Q and H^Q define the shape of the profile function $p^Q(h)$ introduced in Eq. (7). Note, each profile function $p^Q(h)$ is defined within the full height range of the ionosphere, i.e. between h_{min} and h_{max} , and not restricted to the sub-range of the Q -layer. This way no jumps or discontinuities occur at the transition points of two adjacent layers along the vertical.

The exponential decay of the electron density within the plasmasphere P above the height value $h_m^{F_2}$ will be described by the function

$$N_e^P(h) = N_0^P \exp\left(-\frac{|h - h_m^{F_2}|}{H^P}\right), \quad (9)$$

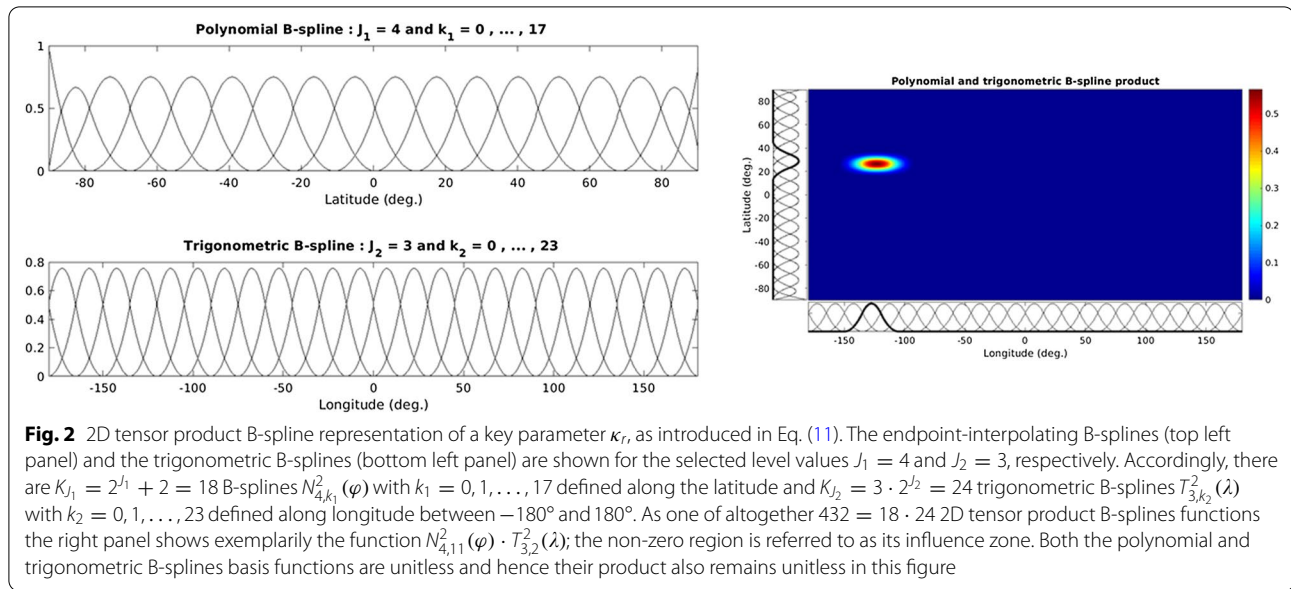
where the quantities N_0^P (= basis density) and H^P (= scale height) are the two key parameters of the plasmasphere. Whereas N_0^P means a scale factor in Eqs. (7) and (9) the parameters $h_m^{F_2}$ and H^P define the shape of the profile function $p^P(h)$ introduced in Eq. (7). From Eqs. (7), (8) and (9), we conclude that our *multi-layer Chapman model (MLCM)* is based on the set

$$\mathcal{K} = \{N_m^D, h_m^D, H^D, N_m^E, h_m^E, H^E, N_m^{F_1}, h_m^{F_1}, H^{F_1}, N_m^{F_2}, h_m^{F_2}, H^{F_2}, N_0^P, H^P\} \quad (10)$$

of altogether 14 key parameters κ_r with $r = 1, 2, \dots, 14$. This set can be further extended, e.g. by choosing different scale heights for the top and the bottom side of each layer. In the following we will stay with the set as defined in Eq. (10).

Both the D -layer and the E -layer are thin compared to the F_1 -layer and to the F_2 -layer; see Berkner and Wells (1934) and Bremer (1998). Whereas the peak densities N_m^D and $N_m^E > N_m^D$ of the D - and the E -layer, respectively, are at least two orders of magnitude lower compared to the peak density $N_m^{F_2}$ of the F_2 -layer under nominal conditions, i.e. $N_m^D < N_m^E \ll N_m^{F_2}$, for the F_1 -layer the definition $N_m^{F_1} = 0$ holds during nighttime; see Liang (2017). The 14 key parameters $\kappa_r \in \mathcal{K}$ defined in Eq. (10) vary on spatio-temporal scales and our aim is their global modelling using an appropriate mathematical approach. The unknown model parameters of this approach should then be estimated by a sufficient number of observations which are assumed to be functionals of the 4D electron density $N_e(\varphi, \lambda, h, t)$ as defined in Eq. (1), e.g. STEC measurements as introduced in Eq. (4).

Earlier studies, e.g., by Limberger (2015) and Liang (2017) highlighted the two main challenges for the selection of an appropriate estimation procedure, namely (1) an inhomogeneous and insufficient global coverage of observations and (2) significant correlations between most of the key parameters. The first item refers to large spatial gaps of observations, from ground-based GNSS receivers, in particular over oceans and deserts as well as dense clusters of observations, for example, over Europe, the United States and Japan; see Erdogan et al. (2017b) and Goss et al. (2020). The different measurement techniques we use in this study and the spatial distribution of their observations will be discussed briefly in Section “Application to real-data”.



The second item states that most of the key parameters $\kappa_r \in \mathcal{K}$ are physically correlated to each other, e.g. due to the electron–ion recombination and different coupling processes in the thermosphere–ionosphere system; see e.g., Stankov et al. (2003), Stankov and Jakowski (2006) as well as Belehaki et al. (2006). Statistical correlations between the key parameters arise due to their observability and sensitivity to electron density. This was shown in Limberger et al. (2014a) with regard to regional ionosphere modelling considering only the F_2 -layer in the South American region by using a combination of real and simulated observations. The incorporation of inequality constraints partially decorrelates the unknown parameters by localizing them to be within specific magnitude bounds. Accordingly, Limberger (2015), Liang (2017) highlighted the need for stochastic modelling and the consideration of inequality constraints on the key parameters. We address this issue in detail in Section “Optimization approach”.

Keeping in view the aforementioned challenges, in this paper we use (1) electron density modelling considering the D -, the E -, the F_1 -, and the F_2 -layer as well as the plasmasphere; (2) the multi-layer Chapman approach for modelling the altitude dependency of the electron density; (3) 2D B-spline tensor product basis functions for the representation of the selected key parameters along latitude and longitude, and (4) the combination of Formosat-3/COSMIC and GRACE radio occultation measurements as well as so-called semi-simulated observations based on the separability approach developed by Hernández-Pajares et al. (2000). Furthermore, the two most important novelties

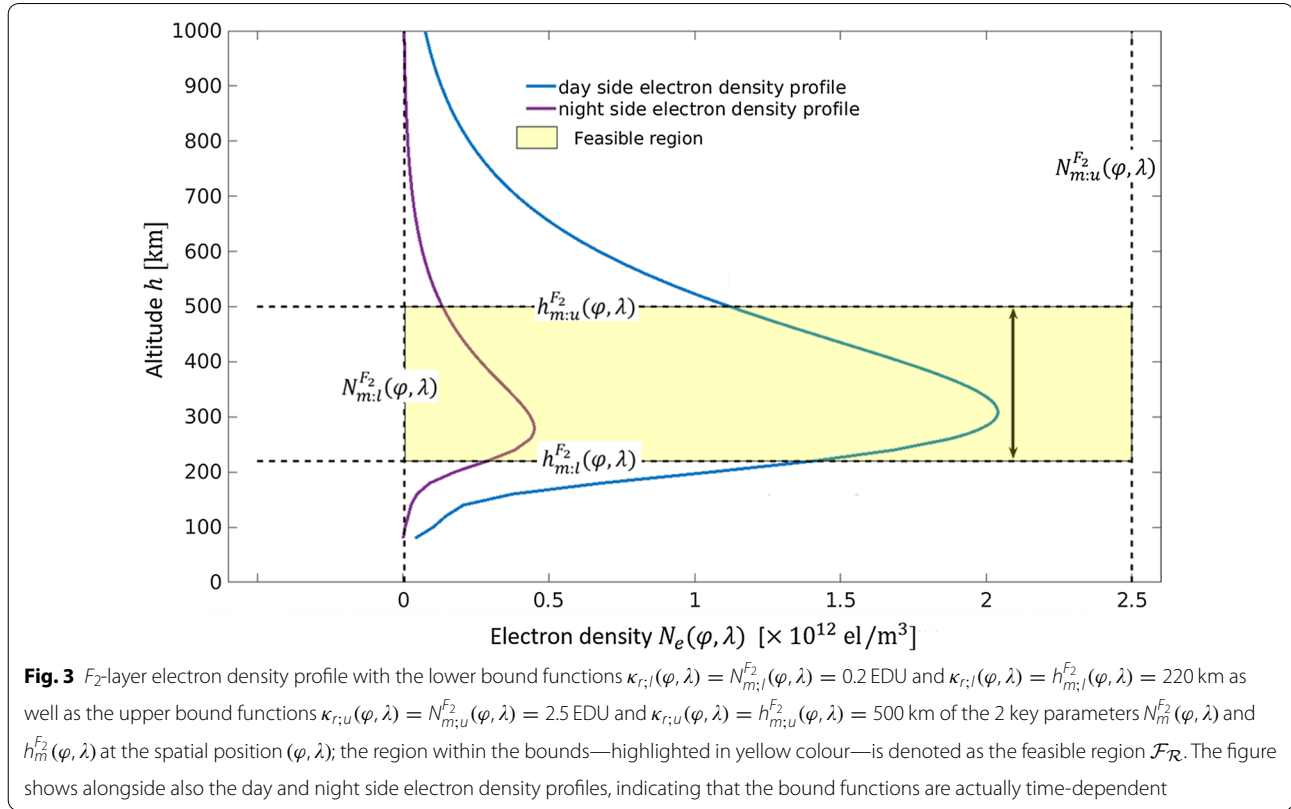
of this paper are (1) the consideration of inequality constraints on selected key parameters of the multi-layer Chapman model (MLCM) and (2) the application of a constraint optimization algorithm for parameter estimation. To the knowledge of the authors, this approach has not yet been applied to ionosphere modelling.

Parameterization of the multi-layer Chapman model (MLCM)

Each key parameter $\kappa_r \in \mathcal{K}$ of the set \mathcal{K} introduced in Eq. (10) of the MLCM as defined by Eqs. (7), (8) and (9) can be represented globally by the 2D B-spline expansion:

$$\kappa_r(\varphi, \lambda, t) = \sum_{k_1=0}^{K_{J_1}-1} \sum_{k_2=0}^{K_{J_2}-1} d_{k_1, k_2; \kappa_r}^{J_1, J_2}(\varphi) T_{J_2, k_2}^{d_T}(\lambda) + v_{\kappa_r}(\varphi, \lambda, t) \quad (11)$$

in terms of the $K_{J_1} \cdot K_{J_2} = u$ 2D tensor products $N_{J_1, k_1}^{d_N}(\varphi) T_{J_2, k_2}^{d_T}(\lambda)$ of the endpoint-interpolating polynomial B-spline functions $N_{J_1, k_1}^{d_N}(\varphi)$ of degree d_N and the trigonometric B-spline functions $T_{J_2, k_2}^{d_T}(\lambda)$ of degree d_T . The shift parameters $k_1 \in \{0, 1, \dots, K_{J_1} - 1\}$ and $k_2 \in \{0, 1, \dots, K_{J_2} - 1\}$ for the endpoint-interpolating and the trigonometric B-spline functions define the position of the 2D tensor product B-spline on the sphere Ω_H . The total numbers $K_{J_1} = 2^{J_1} + 2$ and $K_{J_2} = 3 \cdot 2^{J_2}$ of the B-spline functions depend on the B-spline levels J_1 and J_2 with respect to latitude φ and longitude λ . Whereas the function $v_{\kappa_r}(\varphi, \lambda, t)$ denotes the truncation error, the



u time-dependent series coefficients $d_{k_1, k_2; \kappa_r}^{J_1, J_2}(t)$ are the unknown parameters of the expansion (11) and have to be determined. In the sequel we select $d_N = d_T = 2$ for the degree values d_N and d_T . Consequently, Fig. 2 shows in the right panel exemplarily the 2D quadratic B-spline basis function $N_{4,11}^2(\varphi) \cdot T_{3,2}^2(\lambda)$. Due to its strong localization feature only observations available within the so-called influence zone, i.e. the non-zero area of this 2D B-spline function, contribute to the estimation of the related series coefficient $d_{11,2;\kappa_r}^{4,3}(t_i)$ at the time moment $t = t_i$ for the selected key parameter $\kappa_r \in \mathcal{K}$.

Although different B-spline level values can be considered for each key parameter $\kappa_r \in \mathcal{K}$, we choose for all of them the same level values, resulting in the same number of coefficients for each key parameter. This is due to the fact that according to Goss et al. (2019) the numerical values for the levels J_1 and J_2 are depending on the sampling intervals $\Delta\varphi$ and $\Delta\lambda$ of the input data and on the spectral content of the individual key parameters.

Since the key parameters of the set \mathcal{K} describe the electron density of the ionosphere and the plasmasphere according to the MLM, significant correlations $\rho_{r,s} = \rho_{s,r}$ exist between the estimated sets of B-spline coefficients $d_{k_1, k_2; \kappa_r}^{J_1, J_2}(t_i)$ and $d_{k_1, k_2; \kappa_s}^{J_1, J_2}(t_i)$ of various pairs (κ_r, κ_s) of two key parameters $\kappa_r \in \mathcal{K}$ and $\kappa_s \in \mathcal{K}$ with

$r, s = 1, 2, \dots, 14$ at the time moment $t = t_i$, e.g. for $\kappa_r = N_m^{F_2}$ and $\kappa_s = H^{F_2}$; see Limberger et al. (2014a).

For demonstrating the proposed optimization approach, a subset $\mathcal{K}_1 \subset \mathcal{K}$ of altogether r_1 key parameters $\kappa_r \in \mathcal{K}_1$ will be chosen in our electron density modelling as also done earlier by Limberger et al. (2014b), Liang et al. (2015) and Tsai et al. (2016). For that purpose we will neglect the dependence on time t in the sequel. As already mentioned in the context of the definition of the set \mathcal{K} in Eq. (10) we will assume now that all r_1 key parameters $\kappa_r \in \mathcal{K}_1$ are subject to inequality constraints², such as

$$-\kappa_{r;u}(\varphi, \lambda) + \kappa_r(\varphi, \lambda) \leq 0, \quad (12a)$$

$$\kappa_{r;l}(\varphi, \lambda) - \kappa_r(\varphi, \lambda) \leq 0, \quad (12b)$$

where the lower bound function $\kappa_{r;l}(\varphi, \lambda)$ and the upper bound function $\kappa_{r;u}(\varphi, \lambda)$ are given as physically realistic limits. Examples for the key parameters $\kappa_r \in \mathcal{K}_1 = \{\dots, N_m^{F_2}, h_m^{F_2}, \dots\}$ according to Eqs. (12a) and (12b) with $220 \text{ km} \leq h_m^{F_2}(\varphi, \lambda) \leq 500 \text{ km}$ and

² As the key parameters in Eq. (10) are modelled globally on a sphere, the constraints (12a), (12b), (14a) and (14b) apply and must be satisfied at all latitude, longitude pairs on the global grid.

$0.02 \text{ EDU} \leq N_m^{F_2}(\varphi, \lambda) \leq 2.50 \text{ EDU}$ are exemplar inequality constraints

$$\begin{aligned} -500 \text{ km} + h_m^{F_2}(\varphi, \lambda) &\leq 0 && \text{km} \\ 220 \text{ km} - h_m^{F_2}(\varphi, \lambda) &\leq 0 && \text{km} \\ -2.50 \text{ EDU} + N_m^{F_2}(\varphi, \lambda) &\leq 0 && \text{EDU} \\ 0.02 \text{ EDU} - N_m^{F_2}(\varphi, \lambda) &\leq 0 && \text{EDU} . \end{aligned}$$

For convenience, we define electron density units (EDU) as $1 \text{ EDU} = 10^{12} \text{ el/m}^3$. The example realized by the four equations above for the upper and lower bounds on the F_2 -layer peak density $N_m^{F_2}(\varphi, \lambda)$ and the corresponding peak height $h_m^{F_2}(\varphi, \lambda)$ at a specific spatial location (φ, λ) is illustrated in Fig. 3. The region enclosed by the constraint bounds $N_{m;l}^{F_2}$, $N_{m;u}^{F_2}$, $h_{m;l}^{F_2}$ and $h_{m;u}^{F_2}$ is called the feasible region $\mathcal{F}_{\mathcal{R}}$ for the key parameters $N_m^{F_2}$ and $h_m^{F_2}$. The final solution of the estimation process must be a pair $(N_m^{F_2}, h_m^{F_2}) \in \mathcal{F}_{\mathcal{R}}$. Figure 3 also shows that the lower and upper bounds for the two selected key parameters can be chosen differently for day and night, and thus are time-dependent.

Note, that we did not consider any inequality constraint among the same type of key parameter, but related to different layers, e.g. the condition $h_m^{F_1} < h_m^{F_2}$ for the two peak heights of the F_1 - and the F_2 -layer. This condition must be fulfilled and could be realized by two non-overlapping constraints, such as $h_m^{F_1} < h_{m;u}^{F_2} = h_{m;l}^{F_2}$ and $h_{m;l}^{F_2} \leq h_m^{F_2}$.

Besides the inequality constraints, such as introduced in Eqs. (12a) and (12b), the equality constraints

$$\kappa_r(\varphi, \lambda) = \kappa_{r;e}(\varphi, \lambda) \quad (13)$$

must hold for the remaining $r_2 = 14 - r_1$ key parameters $\kappa_r \in \mathcal{K}_2$ with given physically realistic values $\kappa_{r;e}(\varphi, \lambda)$. Note, that the subset $\mathcal{K}_2 = \mathcal{K} \setminus \mathcal{K}_1$ means the complementary set of \mathcal{K}_1 .

Inserting the series expansion (11) into the inequality constraints (12a), and (12b) and neglecting the truncation error yields:

$$-\kappa_{r;u}(\varphi, \lambda) + \sum_{k_1=0}^{K_{J_1}-1} \sum_{k_2=0}^{K_{J_2}-1} d_{k_1, k_2; \kappa_r}^{J_1, J_2} N_{J_1, k_1}^2(\varphi) T_{J_2, k_2}^2(\lambda) \leq 0, \quad (14a)$$

$$\kappa_{r;l}(\varphi, \lambda) - \sum_{k_1=0}^{K_{J_1}-1} \sum_{k_2=0}^{K_{J_2}-1} d_{k_1, k_2; \kappa_r}^{J_1, J_2} N_{J_1, k_1}^2(\varphi) T_{J_2, k_2}^2(\lambda) \leq 0. \quad (14b)$$

Defining the $u \times 1$ vector

$$\mathbf{d}_{\kappa_r} = [d_{0,0;\kappa_r}^{J_1, J_2}, d_{1,0;\kappa_r}^{J_1, J_2}, \dots, d_{K_{J_1}-1, K_{J_2}-1; \kappa_r}^{J_1, J_2}]^T \quad (15)$$

of the B-spline series coefficients $d_{k_1, k_2; \kappa_r}^{J_1, J_2}$ as well as the $u \times 1$ vector

$$\begin{aligned} \boldsymbol{\phi}_{J_1, J_2}(\varphi, \lambda) &= [N_{J_1, 0}^2(\varphi) T_{J_2, 0}^2(\lambda), N_{J_1, 1}^2(\varphi) \\ &T_{J_2, 0}^2(\lambda), \dots, N_{J_1, K_{J_1}-1}^2(\varphi) \\ &T_{J_2, K_{J_2}-1}^2(\lambda)]^T \end{aligned} \quad (16)$$

of the 2D tensor product B-spline basis functions $N_{J_1, k_1}^2(\varphi) T_{J_2, k_2}^2(\lambda)$ Eqs. (14a) and (14b) can be rewritten as:

$$-\kappa_{r;u}(\varphi, \lambda) + (\boldsymbol{\phi}_{J_1, J_2}(\varphi, \lambda))^T \cdot \mathbf{d}_{\kappa_r} \leq 0, \quad (17a)$$

$$\kappa_{r;l}(\varphi, \lambda) - (\boldsymbol{\phi}_{J_1, J_2}(\varphi, \lambda))^T \cdot \mathbf{d}_{\kappa_r} \leq 0. \quad (17b)$$

In the next step, we evaluate the lower bound function $\kappa_{r;l}(\varphi, \lambda)$ and the upper bound function $\kappa_{r;u}(\varphi, \lambda)$ at altogether m_1 points $P(\varphi_i, \lambda_i) = P_i$ with $i = 1, \dots, m_1$. Applying Eq. (11) to $\kappa_{r;l}(\varphi_i, \lambda_i)$ and assuming that with $m_1 > u$ the points P_i are defined, for instance, on a regular grid or a Reuter grid on the sphere Ω_H , the linear equation system

$$\mathbf{A}_{\kappa_r} \mathbf{d}_{\kappa_r; l} = \kappa_{r;l} + \mathbf{e}_{r;l} \quad (18)$$

results, where the $m_1 \times u$ matrix $\mathbf{A}_{\kappa_r} = [(\boldsymbol{\phi}_{J_1, J_2}(\varphi_i, \lambda_i))^T]_{i=1, \dots, m_1}$ is of full column rank, the $m_1 \times 1$ vector $\kappa_{r;l} = [\kappa_{r;l}(\varphi_i, \lambda_i)]_{i=1, \dots, m_1}$ collects the given lower bound function values and the $m_1 \times 1$ vector $\mathbf{e}_{r;l} = [-v_{\kappa_r; l}(\varphi_i, \lambda_i)]_{i=1, \dots, m_1}$ the corresponding negative truncation error values. The $u \times 1$ vector

$$\mathbf{d}_{\kappa_r; l} = [d_{k_1, k_2; \kappa_r; l}^{J_1, J_2}, d_{1, 0; \kappa_r; l}^{J_1, J_2}, \dots, d_{K_{J_1}-1, K_{J_2}-1; \kappa_r; l}^{J_1, J_2}]^T \quad (19)$$

of the series coefficients $d_{k_1, k_2; \kappa_r; l}^{J_1, J_2}$ is defined analogously to the vector \mathbf{d}_{κ_r} in Eq. (15). Under the assumption that all lower bound function values $\kappa_{r;l}(\varphi_i, \lambda_i)$ are weighted equally the application of the least-squares estimation (LSE) to Eq. (18) yields the solution

$$\bar{\mathbf{d}}_{\kappa_r; l} = (\mathbf{A}_{\kappa_r}^T \mathbf{A}_{\kappa_r})^{-1} \mathbf{A}_{\kappa_r}^T \kappa_{r;l}, \quad (20)$$

and thus the vector

$$\bar{\kappa}_{r;l} = \mathbf{A}_{\kappa_r} \bar{\mathbf{d}}_{\kappa_r; l} \quad (21)$$

of the approximate values $\bar{\kappa}_{r;l}(\varphi_i, \lambda_i)$ for the lower bound function values $\kappa_{r;l}(\varphi_i, \lambda_i)$ at the m_1 grid points P_i . If we perform now in the same manner as in the Eqs. (18) to (21) for the upper bound function $\kappa_{r;u}(\varphi, \lambda)$, we obtain the vector

$$\bar{\kappa}_{r;u} = \mathbf{A}_{\kappa_r} \bar{\mathbf{d}}_{\kappa_r;u} \quad (22)$$

of the approximate values $\bar{\kappa}_{r;u}(\varphi_i, \lambda_i)$ for the lower bound function values $\kappa_{r;u}(\varphi_i, \lambda_i)$ at the m_1 grid points P_i from the estimation $\bar{\mathbf{d}}_{\kappa_r;u}$ of the $u \times 1$ vector

$$\mathbf{d}_{\kappa_r;u} = [d_{k_1, k_2; \kappa_r; u}^{J_1, J_2}, d_{1, 0; \kappa_r; u}^{J_1, J_2}, \dots, d_{K_1-1, K_2-1; \kappa_r; u}^{J_1, J_2}]^T \quad (23)$$

of the series coefficients $d_{k_1, k_2; \kappa_r; u}^{J_1, J_2}$. In addition to the optimization algorithm (presented in the next section), the proposed approach also requires least-squares adjustment for determining the constraint bounds on B-splines coefficients; see Eq. (20). The equality and inequality constraints are originally applied to Chapman key parameters and are then transformed to the B-spline coefficients using least squares estimate of the constraint bound. Combining least square and optimization together is the most direct way to represent the unknown B-spline coefficients of Chapman key parameters as well as their constraint bounds in the Gauss–Markov Model (GMM)³ described in the following section. Of course other options (e.g. spherical harmonics) to represent the transformed constraint bounds (20) are also possible. But for the sake of utilizing the advantages of B-spline basis functions, we wish to use it here. Defining the $(2 \cdot m_1) \times u$ constraint coefficient matrix $\mathbf{B}_{\kappa_r} = [-\mathbf{A}_{\kappa_r}^T, \mathbf{A}_{\kappa_r}]^T$ as well as the $(2 \cdot m_1) \times 1$ constraint bound vector $\mathbf{b}_{\kappa_r} = [-\bar{\kappa}_{r;u}^T, \bar{\kappa}_{r;u}^T]^T$ the inequality constraints (12a) and (12b) can be expressed as the linear constraint matrix inequality (LMI)

$$\mathbf{B}_{\kappa_r} \mathbf{d}_{\kappa_r} \leq \mathbf{b}_{\kappa_r} \quad (24)$$

for the $u \times 1$ vector \mathbf{d}_{κ_r} from Eq. (15) of the series coefficients $d_{k_1, k_2; \kappa_r}^{J_1, J_2}$ for an unknown key parameter $\kappa_r \in \mathcal{K}_1$. Since the Chapman function (8) and the exponential decay function (9) for the plasmasphere depend nonlinearly on a large number of key parameters $\kappa_r \in \mathcal{K}$, the coefficient vector

$$\mathbf{d}_{\kappa_r} = \mathbf{d}_{\kappa_r|0} + \Delta \mathbf{d}_{\kappa_r} \quad (25)$$

needs to be decomposed into the vector $\mathbf{d}_{\kappa_r|0}$ of the approximate, i.e., initial values $d_{k_1, k_2; \kappa_r|0}^{J_1, J_2}$ of the series coefficients $d_{k_1, k_2; \kappa_r}^{J_1, J_2}$ and the vector $\Delta \mathbf{d}_{\kappa_r}$ of the corresponding correction values $\Delta d_{k_1, k_2; \kappa_r}^{J_1, J_2}$. Inserting Eq. (25) into Eq. (24) yields the LMI

$$\mathbf{B}_{\kappa_r} \Delta \mathbf{d}_{\kappa_r} \leq \Delta \mathbf{b}_{\kappa_r} \quad (26)$$

³ Only then the constraint and unconstrained terms of the Lagrangian function (38) can be represented using the same unknown parameter vector $\boldsymbol{\beta}$ to be later used in (34).

for the $u \times 1$ correction vector $\Delta \mathbf{d}_{\kappa_r}$ with $\Delta \mathbf{b}_{\kappa_r} = \mathbf{b}_{\kappa_r} - \mathbf{B}_{\kappa_r} \mathbf{d}_{\kappa_r|0}$. In the same manner as we derived the LMIs (24) and (26), we obtain from Eq. (13) the linear constraint matrix equality (LME)

$$\mathbf{E}_{\kappa_r} \mathbf{d}_{\kappa_r} = \mathbf{e}_{\kappa_r}, \quad (27)$$

where \mathbf{E}_{κ_r} is a $m_2 \times u$ constraint coefficient matrix and $\mathbf{e}_{\kappa_r} = \mathbf{E}_{\kappa_r} \bar{\mathbf{d}}_{\kappa_r;e}$ is the $m_2 \times 1$ corresponding constraint vector for the key parameter $\kappa_r \in \mathcal{K}_2$ with $\bar{\mathbf{d}}_{\kappa_r;e} = (\mathbf{E}_{\kappa_r}^T \mathbf{E}_{\kappa_r})^{-1} \mathbf{E}_{\kappa_r}^T \kappa_{r;e}$ calculated⁴ analogously to Eq. (20) and $m_2 \geq u$. Inserting Eq. (25) into Eq. (27) yields the LME

$$\mathbf{E}_{\kappa_r} \Delta \mathbf{d}_{\kappa_r} = \Delta \mathbf{e}_{\kappa_r} \quad (28)$$

for the $u \times 1$ correction vector $\Delta \mathbf{d}_{\kappa_r}$ with $\Delta \mathbf{e}_{\kappa_r} = \mathbf{e}_{\kappa_r} - \mathbf{E}_{\kappa_r} \mathbf{d}_{\kappa_r|0}$. Note, that in case of $\mathbf{d}_{\kappa_r|0} = \bar{\mathbf{d}}_{\kappa_r;e}$ the right-hand side of Eq. (28) reduces to $\Delta \mathbf{e}_{\kappa_r} = \mathbf{0}_{(m_2 \times 1)}$, where $\mathbf{0}_{(m_2 \times 1)}$ means the $m_2 \times 1$ vector of zero values. Furthermore, in case $m_2 = u$, we may set $\mathbf{E}_{\kappa_r} = \mathbf{I}_u$, where \mathbf{I}_u means the $u \times u$ identity matrix. However, in the following we will stay at the more general LME formulation (28).

To demonstrate the developed procedure, we introduce the subset

$$\mathcal{K}_1 = \{\kappa_1 = N_m^{F_2}, \kappa_2 = h_m^{F_2}, \kappa_3 = H^{F_2}, \kappa_4 = N_0^P, \kappa_5 = H^P\} \quad (29)$$

of the $r_1 = 5$ key parameters of the F_2 -layer and the plasmasphere. Consequently, the set $\mathcal{K}_2 = \mathcal{K} \setminus \mathcal{K}_1$ contains the remaining $r_2 = 9$ key parameters of the D -, E - and F_1 -layer. For setting up the parameter estimation model, we define the $(14 \cdot u) \times 1$ vector

$$\mathbf{d}_{\mathcal{K}} = [\mathbf{d}_{\mathcal{K}_1}^T, \mathbf{d}_{\mathcal{K}_2}^T]^T, \quad (30)$$

with the $(5 \cdot u) \times 1$ vector $\mathbf{d}_{\mathcal{K}_1} = [d_{\kappa_1}^T, d_{\kappa_2}^T, \dots, d_{\kappa_5}^T]^T$ and the $(9 \cdot u) \times 1$ vector $\mathbf{d}_{\mathcal{K}_2} = [d_{\kappa_6}^T, d_{\kappa_7}^T, \dots, d_{\kappa_{14}}^T]^T$ for the coefficient vectors \mathbf{d}_{κ_r} as defined in Eq. (15), with $r = 1, 2, \dots, 14$. In the same manner we define the two $(14 \cdot u) \times 1$ vectors

$$\mathbf{d}_{\mathcal{K}|0} = [\mathbf{d}_{\mathcal{K}_1|0}^T, \mathbf{d}_{\mathcal{K}_2|0}^T]^T, \quad (31a)$$

$$\Delta \mathbf{d}_{\mathcal{K}} = [\Delta \mathbf{d}_{\mathcal{K}_1}^T, \Delta \mathbf{d}_{\mathcal{K}_2}^T]^T \quad (31b)$$

of the vectors $\mathbf{d}_{\kappa_r|0}$ and $\Delta \mathbf{d}_{\kappa_r}$ of the initial series coefficient values and the corresponding correction values, respectively, as defined in Eq. (25).

⁴ The equality constraint bounds transformation will lead to an approximation due to the truncation error in B-spline series expansion. This is perhaps one of the few disadvantages of transforming the constraint bounds from the space of key parameters to B-spline coefficients.

Consequently, we obtain from Eq. (26) the LMI $\mathbf{B}_{\mathcal{K}_1} \Delta \mathbf{d}_{\mathcal{K}_1} \leq \Delta \mathbf{b}_{\mathcal{K}_1}$ with $\mathbf{B}_{\mathcal{K}_1} = [\mathbf{B}_{\kappa_1}, \mathbf{B}_{\kappa_2}, \dots, \mathbf{B}_{\kappa_5}]$ and $\Delta \mathbf{b}_{\mathcal{K}_1} = [\Delta \mathbf{b}_{\kappa_1}^T, \Delta \mathbf{b}_{\kappa_2}^T, \dots, \Delta \mathbf{b}_{\kappa_5}^T]^T$ or

$$\mathbf{B}_{\mathcal{K}} \Delta \mathbf{d}_{\mathcal{K}} \leq \Delta \mathbf{b}_{\mathcal{K}_1}, \quad (32)$$

with $\mathbf{B}_{\mathcal{K}} = [\mathbf{B}_{\mathcal{K}_1}, \mathbf{0}_{(10 \cdot m_1) \times (9 \cdot u)}]$. In the same manner it follows from Eq. (28) the LME $\mathbf{E}_{\mathcal{K}_2} \Delta \mathbf{d}_{\mathcal{K}_2} = \Delta \mathbf{e}_{\mathcal{K}_2}$ with $\mathbf{E}_{\mathcal{K}_2} = [\mathbf{E}_{\kappa_6}, \mathbf{E}_{\kappa_7}, \dots, \mathbf{E}_{\kappa_{14}}]$ and $\Delta \mathbf{e}_{\mathcal{K}_2} = [\Delta \mathbf{e}_{\kappa_6}^T, \Delta \mathbf{e}_{\kappa_7}^T, \dots, \Delta \mathbf{e}_{\kappa_{14}}^T]^T$ or

$$\mathbf{E}_{\mathcal{K}} \Delta \mathbf{d}_{\mathcal{K}} = \Delta \mathbf{e}_{\mathcal{K}_2}, \quad (33)$$

with $\mathbf{E}_{\mathcal{K}} = [\mathbf{0}_{(9 \cdot m_2) \times (5 \cdot u)}, \mathbf{E}_{\mathcal{K}_2}]$.

To the knowledge of the authors, there are only a few relevant works, where an inequality constrained optimization algorithm (ICOA) is applied in geodesy, e.g. the geodetic levelling and datum definition by Koch (1985) and Koch (1988), the non-negative variance component estimation (VCE) as well as the zenith troposphere delay (ZTD) estimation by Roesse-Koerner (2015). In the former two papers, the Lagrange multiplier and slack variable concepts are already used (described in detail in the following section) and the latter uses the so-called ‘‘Active-set methods’’; see Boyd and Vandenberghe (2004), Nocedal and Wright (2006). Inequality-constrained least squares was used by Zhang et al. (2013) to avoid negative VTEC in global ionosphere maps.

Optimization approach

To estimate the vector $\mathbf{d}_{\mathcal{K}}$ as defined in Eq. (30) we need a sufficient number of observations, well distributed over the region of investigation, i.e. in our study given globally within the ionosphere and plasmasphere. To simplify the procedure we assume that electron density measurements $N_e(\mathbf{x}_j)$ at observations points $P(\mathbf{x}_j) = P_j$ are given. Thus, we introduce the $n \times 1$ vector $\mathbf{y} = \mathbf{n}_e - \mathbf{n}_{e|0}$ as the difference of the $n \times 1$ vector $\mathbf{n}_e = [N_e(\mathbf{x}_j)]_{j=1, \dots, n}$ of the electron density observations $N_e(\mathbf{x}_j)$ and the $n \times 1$ vector $\mathbf{n}_{e|0} = [N_{e|0}(\mathbf{x}_j)]_{j=1, \dots, n}$ of the corresponding approximate values $N_{e|0}(\mathbf{x}_j)$ computed from the vector $\mathbf{d}_{\mathcal{K}|0}$ of the initial B-spline coefficients as defined in Eq. (31a). Note, that in this study we deal exclusively with electron density observations $y_j = N_e(\mathbf{x}_j)$ and not with measurements $y_j = f_j(N_e(\mathbf{x}))$ which are functionals of the electron density, such as GNSS STEC observations as introduced in Eq. (4). In this latter case further linearization steps have to be performed as well as a numerical integration along the ray-path between the transmitting satellite P^S and the receiver P_R ; for more details concerning these steps we refer to Limberger (2015), Liang (2017) or Zeilhofer (2008). Since in case of electron density measurements

$y_j = N_e(\mathbf{x}_j)$ the observation equation is non-linearly depending on most of the key parameters, we have to set up the $n \times (14 \cdot u)$ matrix $\mathbf{X} = [\mathbf{X}_{\kappa_1}, \mathbf{X}_{\kappa_2}, \dots, \mathbf{X}_{\kappa_{14}}]$ of the partial derivatives of the electron density with respect to the B-spline coefficients of all 14 key parameters κ_1 to κ_{14} .

Following Koch (1999) and using the aforementioned matrices and vectors we introduce the Gauss–Markov model (GMM)

$$\mathbf{y} + \mathbf{e} = \mathbf{X}\boldsymbol{\beta} \quad \text{with} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}, \quad (34)$$

where $\boldsymbol{\beta} = \Delta \mathbf{d}_{\mathcal{K}}$ is the $(14 \cdot u) \times 1$ vector of the unknown corrections for the B-spline coefficients as defined in Eq. (31b); \mathbf{e} is the $n \times 1$ vector of the observation errors, D means the dispersion operator, \mathbf{P} is the given $n \times n$ positive definite weight matrix of the observations and σ^2 is the unknown variance factor. Note, that besides the B-spline coefficients of all 14 key parameters we do not consider any further unknown parameter in the vector $\boldsymbol{\beta}$, e.g. no bias parameter. We apply the LSE such that the objective function

$$J(\boldsymbol{\beta}) = \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{P} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (35)$$

needs to be minimized subject to the LMI (32) and the LME (33). With the $(10 \cdot m_1) \times 1$ vector

$$\mathbf{s}_{\mathcal{K}_1} = [s_1^2, s_2^2, \dots, s_{10 \cdot m_1}^2]^T \quad (36)$$

of the non-negative slack variables s_i^2 we rewrite the LMI (32) as the LME

$$\mathbf{B}_{\mathcal{K}} \Delta \mathbf{d}_{\mathcal{K}} + \mathbf{s}_{\mathcal{K}_1} = \Delta \mathbf{b}_{\mathcal{K}_1} \quad \text{with} \quad \mathbf{s}_{\mathcal{K}_1} \geq \mathbf{0}_{(10 \cdot m_1) \times 1}. \quad (37)$$

To minimize the objective function (35) subject to the LMEs (33) and (37), we define the Lagrange function

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{s}) = \frac{1}{\sigma^2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{P} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + 2 \cdot \boldsymbol{\lambda}^T (\mathbf{B}\boldsymbol{\beta} - \mathbf{b} + \mathbf{s}) \right]; \quad (38)$$

see e.g. Koch (1985) and Koch (1999). In Eq. (38) we introduced the $(10 \cdot m_1 + 9 \cdot m_2) \times (14 \cdot u)$ constraint coefficient matrix $\mathbf{B} = [\mathbf{B}_{\mathcal{K}}^T, \mathbf{E}_{\mathcal{K}}^T]^T$, the $(10 \cdot m_1 + 9 \cdot m_2) \times 1$ constraint bound vector $\mathbf{b} = [\Delta \mathbf{b}_{\mathcal{K}_1}^T, \Delta \mathbf{e}_{\mathcal{K}_2}^T]^T$, the $(10 \cdot m_1 + 9 \cdot m_2) \times 1$ vector

$$\mathbf{s} = [s_{\mathcal{K}_1}^T, s_{\mathcal{K}_2}^T]^T \quad (39)$$

of the slack variables with $\mathbf{s}_{\mathcal{K}_2} = \mathbf{0}_{(9 \cdot m_2) \times 1}$, and thus $\mathbf{s} \geq \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times 1}$, as well as the $(10 \cdot m_1 + 9 \cdot m_2) \times 1$ vector

$$\lambda = [\lambda_{\mathcal{K}_1}^T, \lambda_{\mathcal{K}_2}^T]^T \tag{40}$$

of the Lagrange multipliers λ_i . Setting $\mathbf{Q} = 2 \cdot \mathbf{X}^T \mathbf{P} \mathbf{X}$ and $\mathbf{q} = 2 \cdot \mathbf{X}^T \mathbf{P} \mathbf{y}$, Eq. (38) can be rewritten as

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{s}) = \frac{1}{\sigma^2} \left[\frac{1}{2} \cdot \boldsymbol{\beta}^T \mathbf{Q} \boldsymbol{\beta} - (\mathbf{q}^T - \boldsymbol{\lambda}^T \mathbf{B}) \boldsymbol{\beta} - \boldsymbol{\lambda}^T (\mathbf{b} - \mathbf{s}) + \mathbf{y}^T \mathbf{P} \mathbf{y} \right]. \tag{41}$$

It shall be noted that in the special case that for the vector \mathbf{s} of the slack variables $\mathbf{s} = \mathbf{0}$ holds, the problem (41) becomes an optimization problem with equality constraints only. In case that for the vector $\boldsymbol{\lambda}$ of the Lagrange multipliers $\boldsymbol{\lambda} = \mathbf{0}$ holds, the problem means an unconstrained optimization problem.

To estimate the optimal solution $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{s}})$, the derivatives of the Lagrange function (41) with respect to the three unknown parameter vectors $\boldsymbol{\beta}, \boldsymbol{\lambda}$, and \mathbf{s} have to be set to the zero vector, i.e.

$$\frac{\partial L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{s})}{\partial \boldsymbol{\beta}} = \mathbf{0}_{(14 \cdot u) \times 1}, \tag{42a}$$

$$\frac{\partial L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{s})}{\partial \boldsymbol{\lambda}} = \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times 1}, \tag{42b}$$

$$\frac{\partial L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{s})}{\partial \mathbf{s}} = \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times 1}. \tag{42c}$$

Evaluating these three derivatives, we obtain the so-called Karush–Kuhn–Tucker (KKT) conditions:

$$\begin{aligned} \mathbf{S} &= \begin{bmatrix} \mathbf{S}_{\mathcal{K}_1} & \mathbf{0}_{(10 \cdot m_1) \times (9 \cdot m_2)} \\ \mathbf{0}_{(9 \cdot m_2) \times (10 \cdot m_1)} & \mathbf{S}_{\mathcal{K}_2} \end{bmatrix}; \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathcal{K}_1} & \mathbf{0}_{(10 \cdot m_1) \times (9 \cdot m_2)} \\ \mathbf{0}_{(9 \cdot m_2) \times (10 \cdot m_1)} & \boldsymbol{\Lambda}_{\mathcal{K}_2} \end{bmatrix}, \\ \mathbf{S}_{\mathcal{K}_1} &= \text{diag}(\tilde{s}_1^2, \dots, \tilde{s}_{10 \cdot m_1}^2); \quad \mathbf{S}_{\mathcal{K}_2} = \text{diag}(\tilde{s}_{10 \cdot m_1 + 1}^2 = 0, \dots, \tilde{s}_{10 \cdot m_1 + 9 \cdot m_2}^2 = 0), \\ \boldsymbol{\Lambda}_{\mathcal{K}_1} &= \text{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{10 \cdot m_1}); \quad \boldsymbol{\Lambda}_{\mathcal{K}_2} = \text{diag}(\tilde{\lambda}_{10 \cdot m_1 + 1}, \dots, \tilde{\lambda}_{10 \cdot m_1 + 9 \cdot m_2}). \end{aligned}$$

of each pair of a Lagrange multiplier $\tilde{\lambda}_i$ and the associated slack variable \tilde{s}_i^2 has to be zero. In other words, if the slack variable is different from zero, the corresponding Lagrange multiplier must be zero and vice versa. Furthermore, for the equality constraints (33) the slack variables are according to Eq. (39) zero by definition. Since the matrix \mathbf{Q} is at least positive semi-definite, the KKT conditions (43a) to (43c) are not only necessary, but also sufficient. Consequently, the optimization problem (41) can be solved by evaluating the system (43a) to (43c); see e.g. Koch (1985) or Nocedal and Wright (2006).

There are different techniques to solve the KKT conditions (43a) to (43c); see e.g. Borwein and Lewis (2000), Gass and Harris (2001), Roese-Koerner (2015). We use the IPM (Nocedal and Wright (2006), pg.481), where a solution pair $\{\tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{s}}\}$ is estimated iteratively from the assumption $\tilde{\lambda}_i \tilde{s}_i^2 = \sigma_{\text{cen}} \cdot \mu_{\text{com}}$ for all $i = 1, 2, \dots, (10 \cdot m_1)$, using the so-called “complimentarity measure” μ_{com} and the “centering parameter” σ_{cen} ; see e.g. Mehrotra (1992), Tütüncü et al. (2003) and Nocedal and Wright (2006). μ_{com} and σ_{cen} are tuning parameters and responsible for the intermediate solution path taken by the optimization parameters $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\mathbf{s}}$. This way, according to Nocedal and Wright (2006), we obtain the perturbed KKT condition

$$\mathbf{S} \boldsymbol{\Lambda} \mathbf{1} - \sigma_{\text{cen}} \cdot \mu_{\text{com}} \mathbf{1} = \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times 1}, \tag{44}$$

also known as complimentary slack condition, where $\mathbf{1} = [1, 1, \dots, 1]^T$ means an $(10 \cdot m_1 + 9 \cdot m_2) \times 1$ vector of ones. The other matrices and vectors in Eq. (44) are defined as

$$\mathbf{Q} \tilde{\boldsymbol{\beta}} - \mathbf{q} + \mathbf{B}^T \tilde{\boldsymbol{\lambda}} = \mathbf{0}_{(14 \cdot u) \times 1}, \tag{43a}$$

$$\mathbf{B} \tilde{\boldsymbol{\beta}} - \mathbf{b} + \tilde{\mathbf{s}} = \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times 1}, \tag{43b}$$

$$\begin{aligned} \tilde{\boldsymbol{\lambda}}^T \tilde{\mathbf{s}} &= 0 \quad \text{with} \quad \tilde{\mathbf{s}} \geq \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times 1} \\ \text{and} \quad \tilde{\boldsymbol{\lambda}} &\geq \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times 1}; \end{aligned} \tag{43c}$$

see e.g. Karush (1939) and Kuhn and Tucker (1951). The consequence of Eq. (43c) is $\tilde{\lambda}_i \tilde{s}_i^2 = 0$ subject to $\tilde{s}_i^2 \geq 0$ and $\tilde{\lambda}_i \geq 0$ for $i = 1, 2, \dots, (10 \cdot m_1 + 9 \cdot m_2)$, i.e. the product

For simplification we introduce the three vectors

$$\bar{\boldsymbol{\beta}} = [\tilde{\boldsymbol{\beta}}^T, \tilde{\boldsymbol{\lambda}}^T, \tilde{\mathbf{s}}^T]^T, \tag{45a}$$

$$\bar{\boldsymbol{\beta}}_0 = [\tilde{\boldsymbol{\beta}}_0^T, \tilde{\boldsymbol{\lambda}}_0^T, \tilde{\mathbf{s}}_0^T]^T, \tag{45b}$$

$$\Delta \bar{\boldsymbol{\beta}} = [\Delta \tilde{\boldsymbol{\beta}}^T, \Delta \tilde{\boldsymbol{\lambda}}^T, \Delta \tilde{\mathbf{s}}^T]^T = \bar{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}_0. \tag{45c}$$

Whereas the vector $\bar{\boldsymbol{\beta}}_0$ contains the given initial values, the vector $\Delta \bar{\boldsymbol{\beta}}$ of the corresponding corrections can be calculated from the linear equation system

$$\begin{bmatrix} \mathbf{Q} & & \mathbf{B}^T & & \mathbf{0}_{(14 \cdot u) \times (10 \cdot m_1 + 9 \cdot m_2)} \\ \mathbf{B} & & \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times (10 \cdot m_1 + 9 \cdot m_2)} & & \mathbf{I}_{(10 \cdot m_1 + 9 \cdot m_2)} \\ \mathbf{0}_{(10 \cdot m_1 + 9 \cdot m_2) \times (14 \cdot u)} & & \mathbf{S}_0 & & \mathbf{\Lambda}_0 \end{bmatrix} \cdot \begin{bmatrix} \Delta \tilde{\boldsymbol{\beta}} \\ \Delta \tilde{\boldsymbol{\lambda}} \\ \Delta \tilde{\boldsymbol{s}} \end{bmatrix} = \begin{bmatrix} -\mathbf{Q} \boldsymbol{\beta}_0 + \mathbf{q} - \mathbf{B}^T \boldsymbol{\lambda}_0 \\ -\mathbf{B} \boldsymbol{\beta}_0 + \mathbf{b} - \mathbf{s}_0 \\ -\mathbf{S}_0 \boldsymbol{\Lambda}_0 \mathbf{1} + \sigma_{\text{cen}} \cdot \boldsymbol{\mu}_{\text{com}} \mathbf{1} \end{bmatrix}, \quad (46)$$

which follows from the KKT equations (43a), (43b) and (44) applying Newton’s method. In Eq. (46) the matrices \mathbf{Q} and \mathbf{B} as well as the vectors \mathbf{q} and \mathbf{b} are known, the $(10 \cdot m_1 + 9 \cdot m_2) \times (10 \cdot m_1 + 9 \cdot m_2)$ matrices \mathbf{S}_0 and $\mathbf{\Lambda}_0$ are evaluated at the initial positions $\tilde{\mathbf{s}}_0$ and $\tilde{\boldsymbol{\lambda}}_0$; see e.g. Nocedal and Wright (2006). Note, in the context of the definition of the GMM (34) we set $\boldsymbol{\beta} = \Delta \mathbf{d}_{\mathcal{K}}$, and thus we may identify the vector $\tilde{\boldsymbol{\beta}}_0$, Eq. (45b), of the initial values for the B-spline coefficients as the zero vector.

As already mentioned the above presented solution of the KKT equations is based on the IPM. The reader is referred to Mehrotra (1992), Anderson et al. (1996), Billups and Ferris (1996), Potra and Wright (2000), Boyd and Vandenberghe (2004), Nocedal and Wright (2006), Mitchell et al. (2006), Forst and Hoffmann (2010) for alternative KKT solution techniques.

The choice of the step size values between two adjacent iteration steps it and $it + 1$ is one of the most important specifications, because they are needed to calculate the updated values of the elements of the parameter vector $\bar{\boldsymbol{\beta}}$ defined in Eq. (45a). In an unconstrained optimization problem, a “full step” can be taken, i.e. $\bar{\boldsymbol{\beta}}_{it+1} = \bar{\boldsymbol{\beta}}_{it} + \Delta \bar{\boldsymbol{\beta}}_{it}$ or in other words the step size matrix $\mathbf{T}_{it} = \mathbf{I}$ equals the identity matrix along the descent direction. However, the choice of the step size values in presence of inequality constraints is not trivial, e.g. due to the risk of overshooting the feasible region. In general, the solution of the linear equation system (46) has to be updated from the iteration step it to the next step $it + 1$ by

$$\bar{\boldsymbol{\beta}}_{it+1} = \bar{\boldsymbol{\beta}}_{it} + \mathbf{T}_{it} \Delta \bar{\boldsymbol{\beta}}_{it} \quad (47)$$

until convergence is reached. In Eq. (47) the update matrix \mathbf{T}_{it} is defined as

$$\mathbf{T}_{it} = \text{diag}(\mathbf{T}_{\boldsymbol{\beta}}, \mathbf{T}_{\boldsymbol{\lambda}}, \mathbf{T}_{\boldsymbol{s}}), \quad (48)$$

with the diagonal block matrices

$$\mathbf{T}_{it} = \text{maximize} \left\{ t_{\beta_j}, t_{\lambda_k}, t_{s_k^2} \mid (\tilde{\mathbf{s}} + \mathbf{T}_s \Delta \tilde{\mathbf{s}} \geq \mathbf{0}) \text{ and } (\tilde{\boldsymbol{\lambda}} + \mathbf{T}_{\lambda} \Delta \tilde{\boldsymbol{\lambda}} \geq \mathbf{0}) \right\}, \quad (50)$$

such that $0 < \{t_{\beta_j}, t_{\lambda_k}, t_{s_k^2}\} \leq 1$ for all $j = \{1, \dots, 14 \cdot u\}$, $k = \{1, \dots, 10 \cdot m_1 + 9 \cdot m_2\}$,

$$\mathbf{T}_{\boldsymbol{\beta}} = \text{diag}(t_{\beta_1}, \dots, t_{\beta_{14 \cdot u}}), \quad (49a)$$

$$\mathbf{T}_{\boldsymbol{\lambda}} = \text{diag}(t_{\lambda_1}, \dots, t_{\lambda_{10 \cdot m_1}}, t_{\lambda_{10 \cdot m_1 + 1}}, \dots, t_{\lambda_{10 \cdot m_1 + 9 \cdot m_2}}), \quad (49b)$$

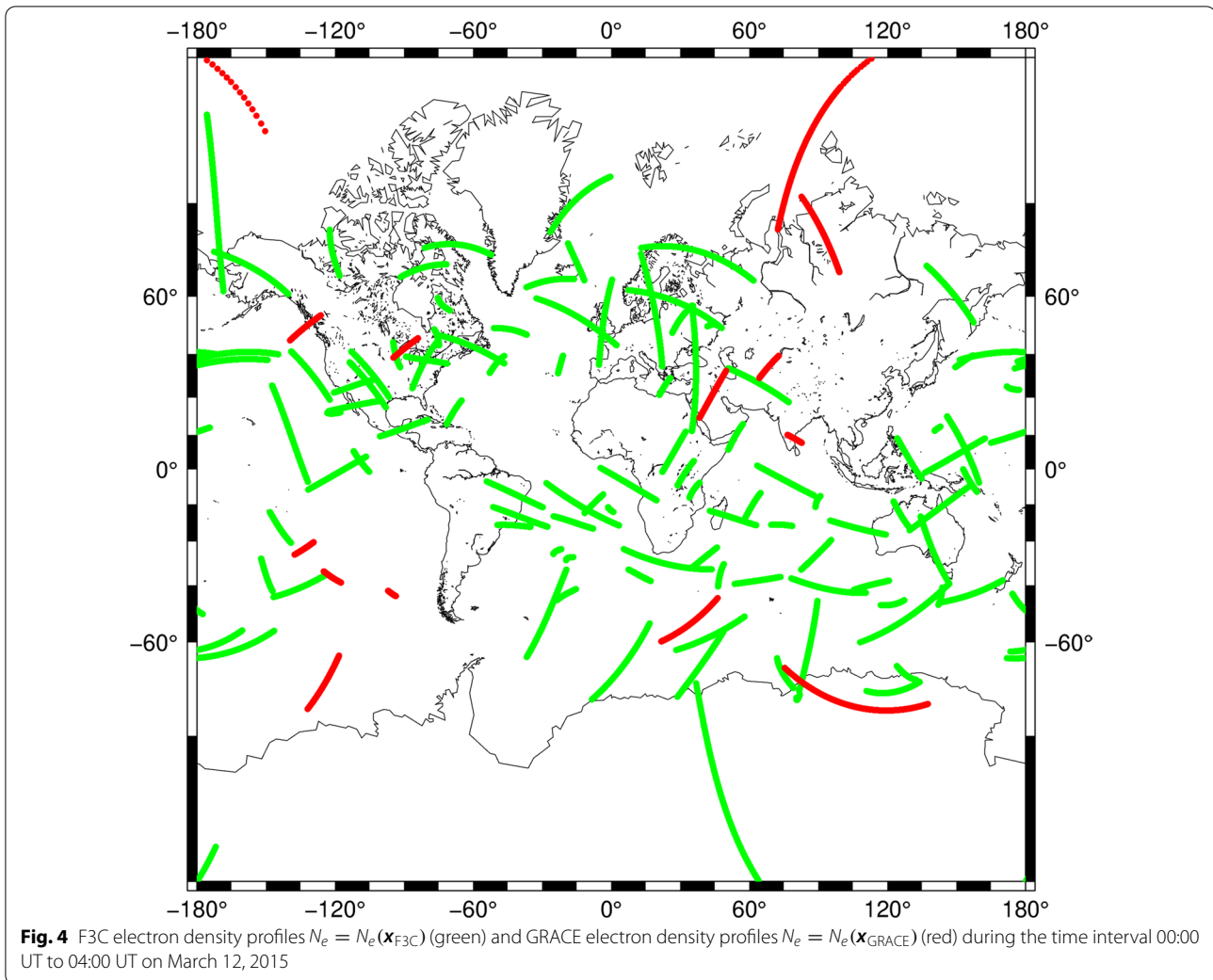
$$\mathbf{T}_{\boldsymbol{s}} = \text{diag}(t_{s_1^2}, \dots, t_{s_{10 \cdot m_1}^2}, t_{s_{10 \cdot m_1 + 1}^2}, \dots, t_{s_{10 \cdot m_1 + 9 \cdot m_2}^2}), \quad (49c)$$

such that the diagonal elements are the respective step sizes used along $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and \boldsymbol{s} satisfying the conditions

$$\begin{aligned} & [t_{\beta_1}, \dots, t_{\beta_{14 \cdot u}}] \in (0, 1]; \\ & [t_{\lambda_1}, \dots, t_{\lambda_{10 \cdot m_1}}, t_{\lambda_{10 \cdot m_1 + 1}}, \dots, t_{\lambda_{10 \cdot m_1 + 9 \cdot m_2}}] \in (0, 1]; \\ & [t_{s_1^2}, \dots, t_{s_{10 \cdot m_1}^2}, t_{s_{10 \cdot m_1 + 1}^2}, \dots, t_{s_{10 \cdot m_1 + 9 \cdot m_2}^2}] \in (0, 1]. \end{aligned}$$

Although according to the definition (36) of the vector \boldsymbol{s} the slack variables for the equality constraints are included in \mathbf{T}_{it} , they can be omitted from Eq. (48). Since the non-negativity of the Lagrange multipliers λ_i and the slack variables s_i^2 has to be checked in each iteration, the chosen step size values along the descent direction are an important consideration. This procedure is one of the main modifications in our algorithm compared to the classical Newton method; see e.g. Nocedal and Wright (2006).

The descent direction (45c) is estimated by solving the linear equation system (46) with the initial update matrix $\mathbf{T}_{it} = \mathbf{I}$. Further on, the step size estimation becomes a sub-problem where the goal is the estimation of the largest diagonal elements t_{β_j} for all $j \in \{1, \dots, 14 \cdot u\}$, $t_{\lambda_k}, t_{s_k^2}$ for all $k \in \{1, \dots, 10 \cdot m_1 + 9 \cdot m_2\}$, of \mathbf{T}_{it} which when used in Eq. (47), satisfy the KKT conditions (43c), and thus



where $\Delta\tilde{\mathbf{s}}$ and $\Delta\tilde{\boldsymbol{\lambda}}$ are the descent directions along the slack variables and the Lagrange multiplier vectors. Note, in the most general case, different step size matrices corresponding to each of the three vectors $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and \mathbf{s} could be chosen. However, in this work, it was found that using only two step size matrices, i.e. one for the unknown model parameter vector $\boldsymbol{\beta}$ and the other for the vector pair $\boldsymbol{\lambda}$ and \mathbf{s} , is sufficient, and thus \mathbf{T}_λ and \mathbf{T}_s are assumed to be equal. These choices are made as part of the “design” of the developed ICOA. Furthermore, the step sizes are estimated in each iteration. This aspect is one of the most computationally demanding properties of the developed procedure, i.e. it becomes one of the disadvantages of adding more inequality constraints to our modelling problem; see Nesterov and Nemirovskii (1994), Todd et al. (1998), Potra and Wright (2000), Halldórsson and Tütüncü (2003), Mitchell et al. (2006)). During this work, it was found that an approximate step size is also generally

sufficient in the developed ICOA. Having estimated the step size matrix (48), the complementarity measure

$$\mu_{\text{com}} = \frac{(\tilde{\boldsymbol{\lambda}} + \mathbf{T}_\lambda \Delta\tilde{\boldsymbol{\lambda}})^T (\tilde{\mathbf{s}} + \mathbf{T}_s \Delta\tilde{\mathbf{s}})}{10 \cdot m_1 + 9 \cdot m_2} \quad (51)$$

is computed; for details see Tütüncü et al. (2003).

In the following, we present the altogether 8 steps of the developed iterative algorithm for the combined step size and descent direction estimation within the ICOA:

- (1) Set the initial values $\tilde{\boldsymbol{\beta}}_0$, $\tilde{\mathbf{s}}_0$, $\tilde{\boldsymbol{\lambda}}_0$ from Eq. (45b) of the unknown optimization parameters and solve for the descent directions $\Delta\tilde{\mathbf{s}}$ and $\Delta\tilde{\boldsymbol{\lambda}}$ considering $\tilde{\mathbf{s}}_0 \geq \mathbf{0}$ and $\tilde{\boldsymbol{\lambda}}_0 \geq \mathbf{0}$.
- (2) Initialize the centering parameter σ_{cen} and use the initial step size matrix $\mathbf{T}_{\text{it}} = \mathbf{I}$. Compute μ_{com} using Eq. (51); this value of complementarity measure

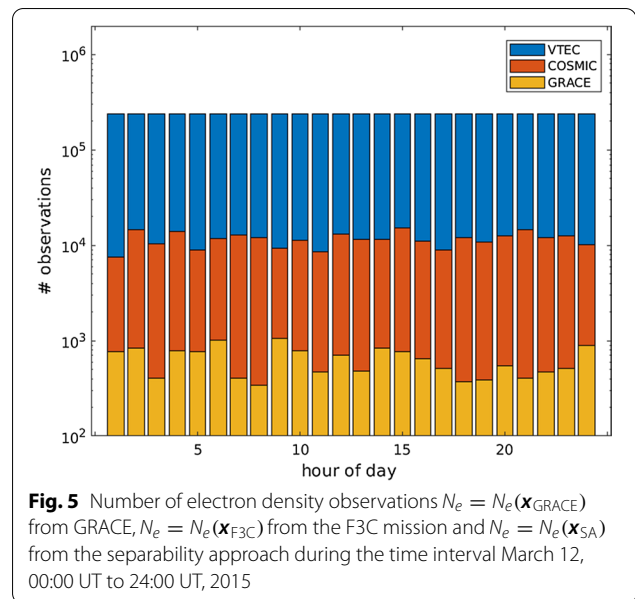
is called $\mu_{\text{com,full}}$ because it means the "full step" assumption $\mathbf{T}_{\text{it}} = \mathbf{I}$.

- (3) Estimate $\Delta\tilde{\boldsymbol{\beta}}$, $\Delta\tilde{\boldsymbol{\lambda}}$ and $\Delta\tilde{\boldsymbol{s}}$ from the linear equation system (46) with the initial value vectors and matrices $\tilde{\boldsymbol{\beta}}_0$, $\tilde{\boldsymbol{\lambda}}_0$, $\tilde{\boldsymbol{s}}_0$, \mathbf{T}_{it} and the tuning parameters $\mu_{\text{com,full}}$ and σ_{cen} .
- (4) Estimate the step size \mathbf{T}_{it} from Eq. (50).
- (5) Compute the complementarity measure μ_{com} using Eq. (51) with the update matrix \mathbf{T}_{it} from Step 4.
- (6) Set the centering parameter $\sigma_{\text{cen}} = (\mu_{\text{com,full}}/\mu_{\text{com}})^3$ as proposed by Tütüncü et al. (2003).
- (7) Compute the updated optimization parameters with Eq. (47) using \mathbf{T}_{it} computed in Step 4.
- (8) Repeat the Steps 4 to 7 until convergence is reached.

We have derived in this work the linear equation system (46) to solve the inequality constrained optimization problem, in which, \mathbf{Q} , \mathbf{B} and \mathbf{P} are known matrices; \mathbf{y} and \mathbf{b} are given vectors or calculable from given values. We used Newton's method with the initial value $\tilde{\boldsymbol{\beta}}_0$ and the step size matrix \mathbf{T}_{it} . For our investigations, i.e. for the global electron density modelling problem, the choice of initial values defined within the feasible region is necessary. In the next section, we will apply our proposed constrained optimization procedure to real data.

Application to real-data, validation and discussion

In this section we apply the developed ICOA, as described in previous two sections, to model the electron density N_e from direct measurements $y_j = N_e(\mathbf{x}_j)$ given at observation sites $P(\mathbf{x}_j) = P_j$. To be more specific, we choose electron density profiles from GRACE and Formosat-3/COSMIC (F3C) ionospheric radio occultation (IRO) measurements profiles. As reported in Yue et al. (2014), F3C produced on average 1000 to 1500 occultations per day during the years 2014 and 2015. We refer to Tsai et al. (2016) for the role of the F3C mission and the follow-up mission Formosat-7/COSMIC-2 (F7C) in global electron density modelling and recommend the papers of Tsai et al. (2009), Hernández-Pajares et al. (1998), Limberger (2015) for detailed description of IRO retrieval methods. For the scope of this work, the focus was on the development of ICOA and therefore the IRO observations have been directly taken from the COSMIC Data Analysis and Archive Center (CDAAC); see Schreiner et al. (2003). Figure 4 shows the projections of the electron density profiles on the surface of Earth for a 4-h time interval at March 12, 2015. Note, in the context of IRO processing an electron density profile cannot be interpreted as a set of electron density values given along the vertical, i.e. $N_e = N_e(h_j)$, but given along the tangent points $P(\mathbf{x}_{T_j})$



which are defined as a series of the closest points to Earth along the line-of-sight between the occulting GPS and the LEO satellite at specific time moments; see e.g. Limberger et al. (2013). Since both the F3C mission and the GRACE mission provide only a relatively small number of electron density observations, we introduce a third type the electron density observations derived from the so-called separability approach Hernández-Pajares et al. (2000). In this approach electron density values at globally distributed 3D grid points $P(\varphi_k, \lambda_k, h_j)$ at time moment t_i are defined as

$$N_e(\varphi_k, \lambda_k, h_j, t_i) = VTEC(\varphi_k, \lambda_k, t_i) \cdot \bar{p}_k(h_j, t_i), \quad (52)$$

where the VTEC values $VTEC(\varphi_k, \lambda_k, t_i)$ are evaluated from VTEC models at the 2D points $P(\varphi_k, \lambda_k)$ at time t_i and $\bar{p}_k(h_j, t_i)$ means a value of a vertical profile function at the horizontal grid point $P(\varphi_k, \lambda_k)$ evaluated at height h_j at time moment t_i fulfilling the normalization condition

$$\bar{p}_k(h_j, t_i) = \frac{p_k(h_j, t_i)}{\int_{h_{\min}}^{h_{\max}} p_k(h, t_i) dh}. \quad (53)$$

As VTEC model we choose DGFI-TUM's high-resolution global ionosphere map (GIM) 'othg' which comprises a much higher spectral content (up to spherical harmonic degree $n = 33$) as the GIMs from the Ionospheric Associated Analysis Centers (IAAC) of the International GNSS Service (IGS). As stated in the IGS Technical Report 2018 Villiger A (2019) 'othg' is one of the worldwide best GIMs; it is based on a 2D B-spline representation according to Eq. (11) and characterized by a spatial resolution of $2.5^\circ \times 5^\circ$ with respect to latitude φ and longitude λ as well as a temporal resolution of 10 minutes; for more

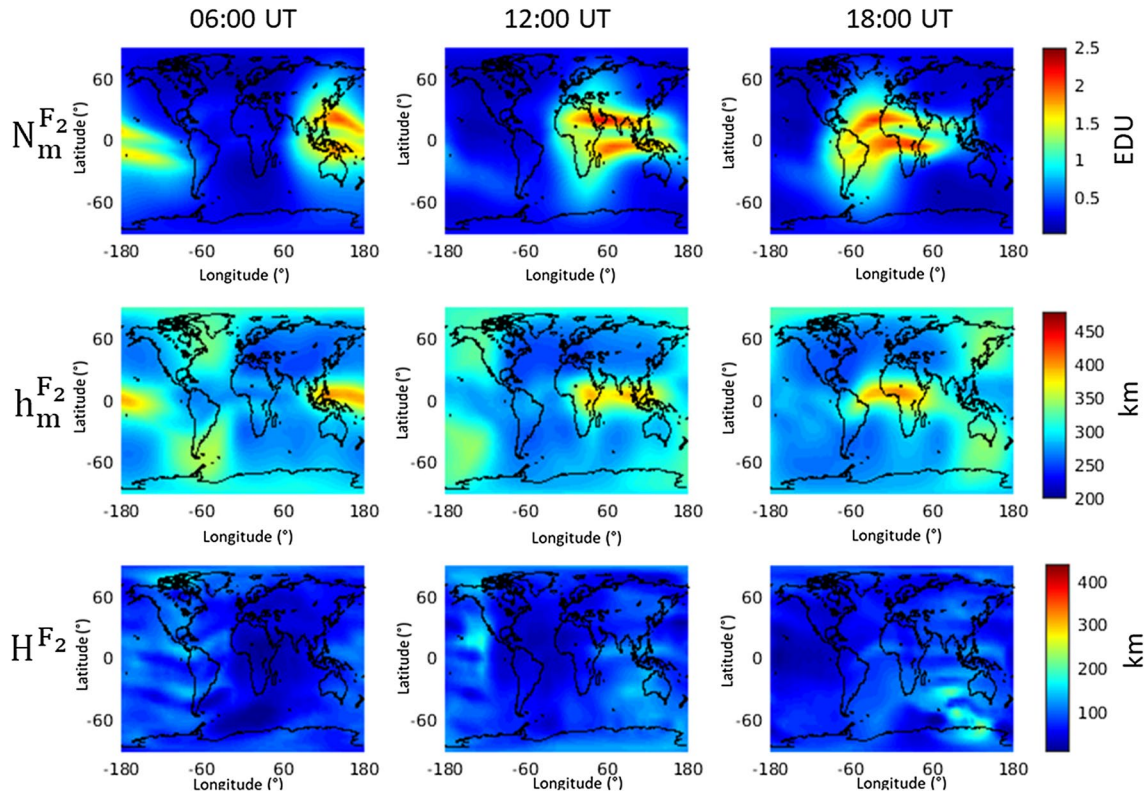


Fig. 6 Maps of the lower bound of the three key parameters $\kappa_{1;l} = N_{m;l}^{F_2}(\varphi, \lambda)$ for the peak density (top row), $\kappa_{2;l} = h_{m;l}^{F_2}(\varphi, \lambda)$ for the corresponding peak height (middle) and the scale height $\kappa_{3;l} = H_{l;l}^{F_2}(\varphi, \lambda)$ (bottom) at three different time epochs on March 12, 2015

details concerning 'othg' we refer to Goss et al. (2019). The profile functions $p_k(h, t_i)$ in Eq. (53) are computed from the International Reference Ionosphere (IRI) 2012 Bilitza, Dieter et al. (2014) at around 90 discrete heights h_j between 100 km and 1000 km at the regular grid points $P(\varphi_k, \lambda_k)$ at time moments t_i between March 12, 2015, 0:00 UT and March 14, 2015, 24:00 UT. This is the reason for the large number of electron density observations from the separability approach (SA) compared to the GRACE and F3C input data, as shown in Fig. 5. Note, that the electron density observations (52) can be seen as semi-simulated data, since 'othg' is computed from real GNSS data and could be supplemented by DORIS and satellite altimetry observations; the profile functions are in opposite generated by the empirical model IRI2012.

It needs to be mentioned that we chose this way of creating a sufficient large observation vector \mathbf{y} for the GMM (34), because the number of real electron density measurements from IRO, the Langmuir probe measurement system on CHAMP and Swarm, ionosonde data, etc., is much too less for estimating the 4D electron density function (1) by applying our developed ICOA. It is possible that in the near future, when using large amounts

of data from GNSS (constellations), IRO missions and CubeSats, it will no longer be necessary to consider simulated or semi-simulated input data. As shown in Fig. 8, the input block top left contains electron density observations $N_e = N_e(\mathbf{x}_{SA})$ generated accordingly to Eq. (52) from GNSS observations. The next block top mid comprises the electron density observations $N_e = N_e(\mathbf{x}_{F3C})$ from the IRO measurements of the F3C mission. Finally, the input data block top right contain electron density observations $N_e = N_e(\mathbf{x}_{GRACE})$ derived from GRACE and/or GRACE-FO IRO measurements. Additionally, electron density information from GRACE K-band measurements, from CHAMP and Swarm Langmuir probe data can also be supported here. After running some preprocessing steps, such as outlier detection the data are stored in the $n \times 1$ vector \mathbf{y} of the GMM (34). In a first run of the developed ICOA, we identify the weight matrix \mathbf{P} with the identity matrix \mathbf{I} . In our modelling approach, we set the B-spline levels J_1 and J_2 in the series expansion (11) for the key parameter κ_r as well as for the corresponding upper and lower bound functions $\kappa_{r;u}$ and $\kappa_{r;l}$ to the values $J_1 = 4$ and $J_2 = 3$; cf. Fig 2. Consequently, with $u = 432$ our model consists of $14 \cdot 432 = 6048$ series

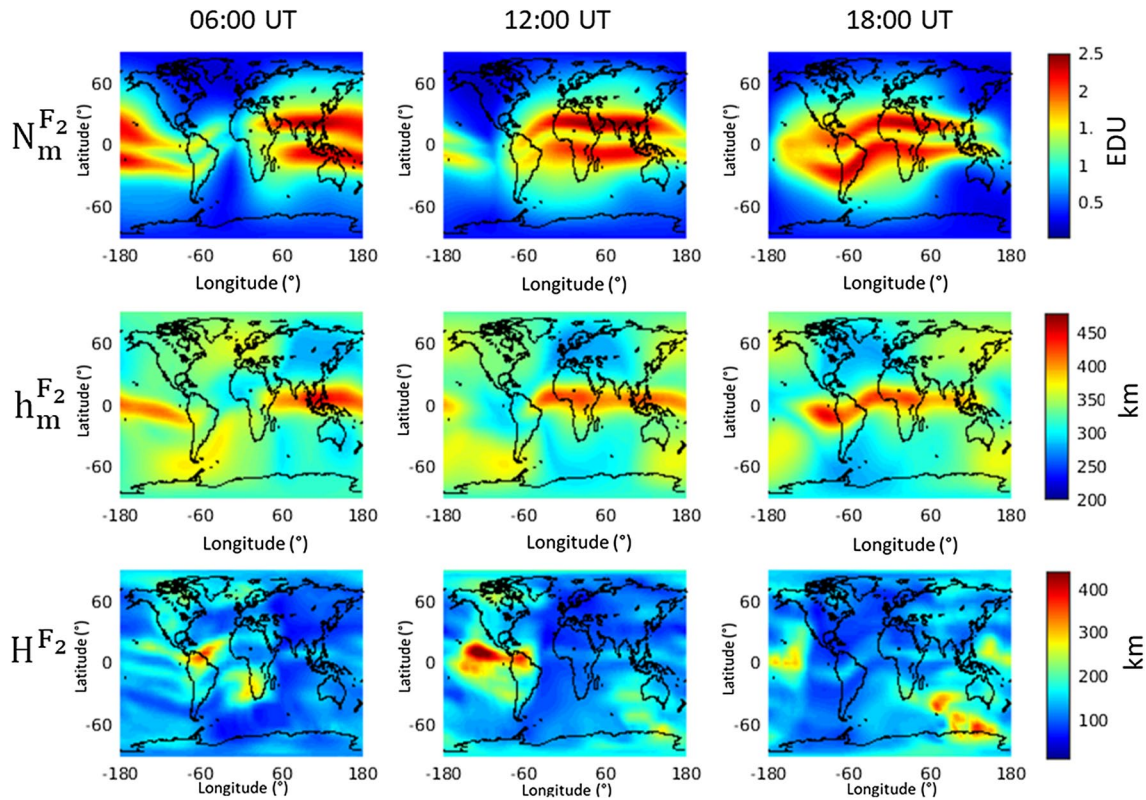


Fig. 7 Maps of the upper bound of the three key parameters $\kappa_{1;u} = N_{m;u}^{F_2}(\varphi, \lambda)$ for the peak density (top row), $\kappa_{2;u} = h_{m;u}^{F_2}(\varphi, \lambda)$ for the corresponding peak height (middle) and the scale height $\kappa_{3;u} = H_u^{F_2}(\varphi, \lambda)$ (bottom) at three different time epochs on March 12, 2015

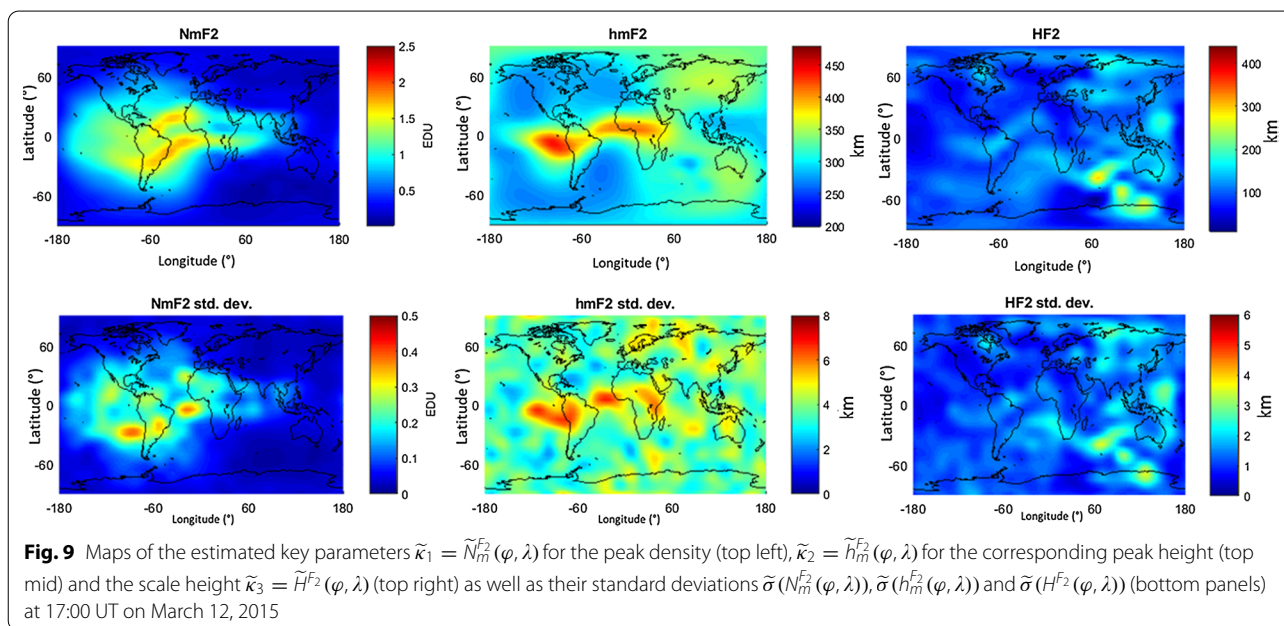
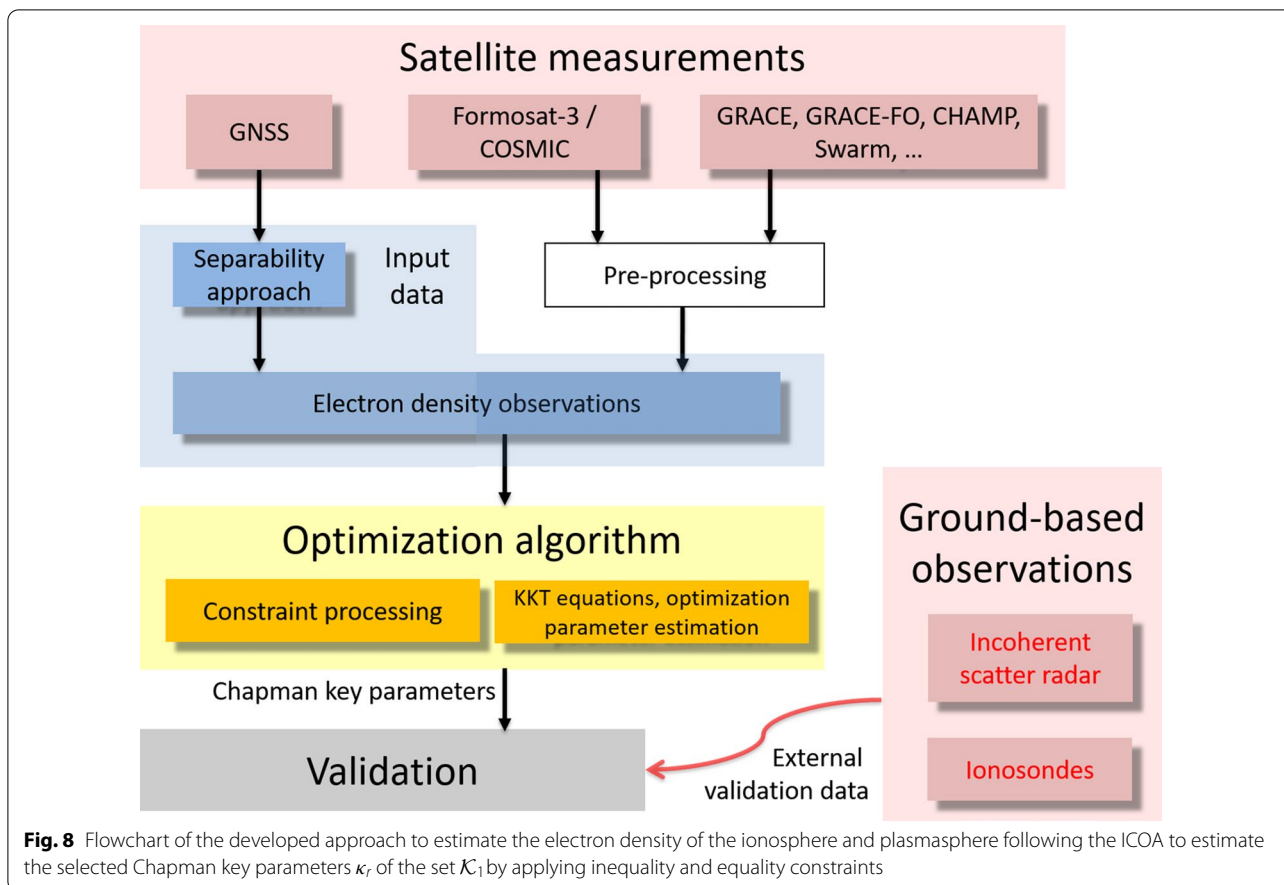
coefficients comprised in the vector $\mathbf{d}_{\mathcal{K}}$ introduced in Eq. (30). Due to the decomposition of the set \mathcal{K} into the subsets \mathcal{K}_1 and \mathcal{K}_2 according to Eq. (29), we have to estimate altogether $5 \cdot 432 = 2,160$ unknown series coefficients collected in the coefficient vector $\mathbf{d}_{\mathcal{K}_1}$. The remaining key parameters collected in the subset \mathcal{K}_2 , i.e. the peak densities, the corresponding peak heights and the scale heights of the D -, the E - and the F_1 -layer are assumed to be given. More specifically, the functions $N_m^D(\varphi, \lambda)$, $h_m^D(\varphi, \lambda)$, $N_m^E(\varphi, \lambda)$, $h_m^E(\varphi, \lambda)$, $N_m^{F_1}(\varphi, \lambda)$ and $h_m^{F_1}(\varphi, \lambda)$ have been taken from IRI2012 and the scale heights $H^D(\varphi, \lambda)$, $H^E(\varphi, \lambda)$ and $H^{F_1}(\varphi, \lambda)$ are derived from the empirical ionospheric slab-thickness relation from Limberger (2015), Liang (2017). The lower bound functions $N_{m;l}^{F_2}(\varphi, \lambda)$, $h_{m;l}^{F_2}(\varphi, \lambda)$, $H_l^{F_2}(\varphi, \lambda)$, $N_{0;l}^P(\varphi, \lambda)$ and $H_l^P(\varphi, \lambda)$ as well as the upper bound functions $N_{m;u}^{F_2}(\varphi, \lambda)$, $h_{m;u}^{F_2}(\varphi, \lambda)$, $H_u^{F_2}(\varphi, \lambda)$, $N_{0;u}^P(\varphi, \lambda)$ and $H_u^P(\varphi, \lambda)$ are also obtained from IRI 2012. Examples of the lower and upper constraint maps are shown in Figs. 6 and Fig. 7.

We have estimated the B-spline coefficients corresponding to the 5 key parameters κ_r of the set \mathcal{K}_1 between March

12 to 14, 2015 with a temporal resolution of 1 hour. As examples, maps of the estimated F_2 -layer key parameters $\tilde{\kappa}_1 = \tilde{N}_m^{F_2}(\varphi, \lambda)$ for the peak density, $\tilde{\kappa}_2 = \tilde{h}_m^{F_2}(\varphi, \lambda)$ for the corresponding peak height and the scale height $\tilde{\kappa}_3 = \tilde{H}^{F_2}(\varphi, \lambda)$ at 17:00 UT on March 12, 2015, computed from the series expansion (11) at regular grid points, are shown in the top row of Fig. 9 along with maps of their estimated standard deviations $\tilde{\sigma}(N_m^{F_2}(\varphi, \lambda))$, $\tilde{\sigma}(h_m^{F_2}(\varphi, \lambda))$ and $\tilde{\sigma}(H^{F_2}(\varphi, \lambda))$ computed via a Monte Carlo approach; for details see Roese-Koerner et al. (2014)).

For the validation of these results, we evaluate the series expansion (11) for the key parameters $\hat{\kappa}_1 = \hat{N}_m^{F_2}$ and $\hat{\kappa}_2 = \hat{h}_m^{F_2}$ at the exact locations of the selected ionosonde stations using the estimated B-spline coefficients; see Fig. 10. For the same time period, we also obtained the $N_m^{F_2}$ and $h_m^{F_2}$ directly measured in those stations⁵ for validation. The prevailing space weather situation for the month of March 2015 is also shown using the daily averaged time series of Kp and F10.7 indices in Fig. 11. We define the deviations $\Delta N_m^{F_2}$ and $\Delta h_m^{F_2}$ as the differences

⁵ The ionosonde observations are obtained from the NGDC open source repository ftp.ngdc.noaa.gov/ionosonde/



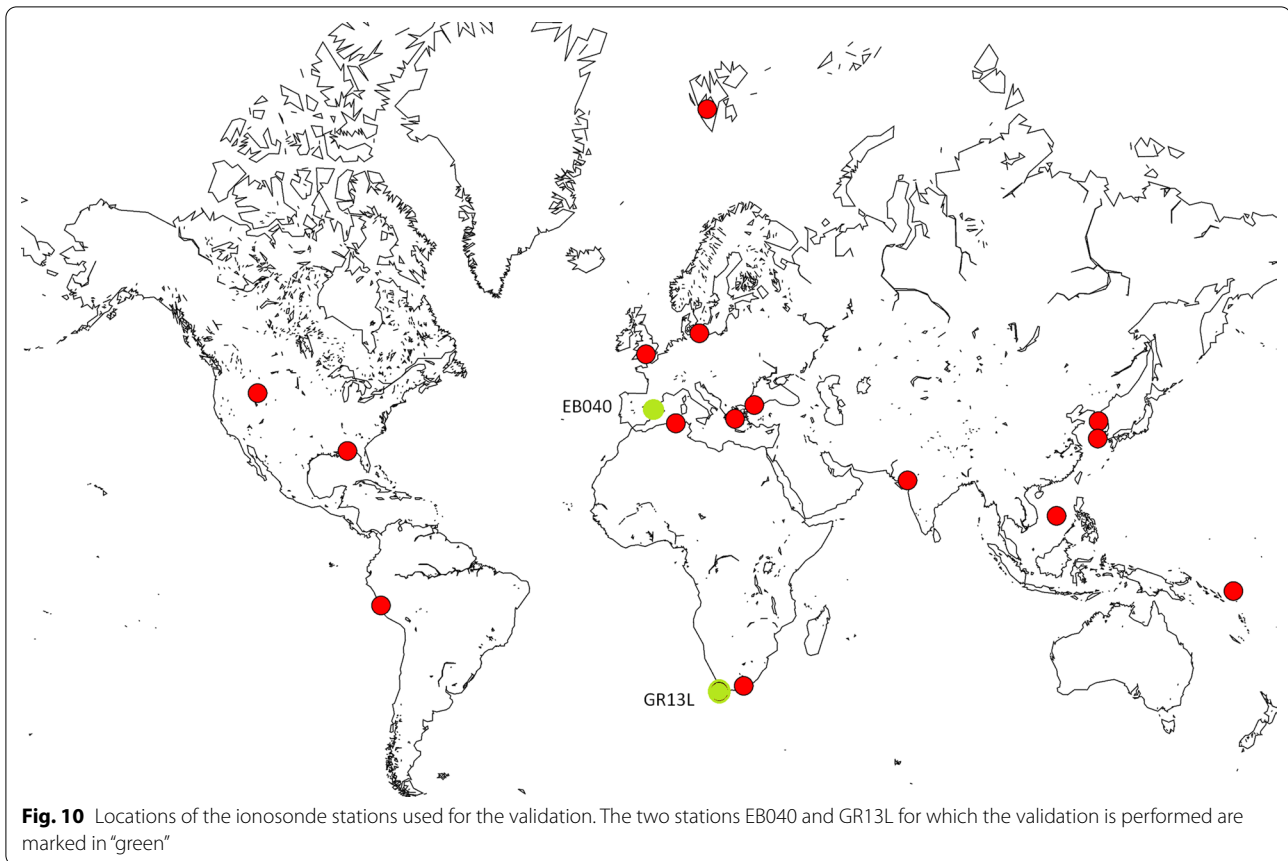


Table 1 Mean deviations with respect to $N_m^{F_2}$ (Columns 2–4) and to $h_m^{F_2}$ (Columns 5–7) at the station GR13L between March 12, 00:00 UT to March 14, 24:00 UT, 2015

Time window (UT)	12 Mar (EDU)	13 Mar (EDU)	14 Mar (EDU)	12 Mar (km)	13 Mar (km)	14 Mar (km)
0–6	0.09 ± 0.06	0.30 ± 0.07	0.16 ± 0.06	-29.87 ± 7.66	-19.57 ± 6.87	-21.51 ± 9.56
6–12	0.06 ± 0.20	0.07 ± 0.31	0.04 ± 0.18	-10.07 ± 6.13	-8.27 ± 7.33	-9.65 ± 10.40
12–18	0.16 ± 0.19	0.07 ± 0.13	0.08 ± 0.16	-7.18 ± 6.98	-6.43 ± 7.10	-3.43 ± 8.50
18–24	-0.01 ± 0.22	-0.01 ± 0.24	-0.01 ± 0.18	-27.33 ± 7.31	-24.65 ± 8.19	-22.56 ± 7.08

between the estimated and ionosonde-derived key parameters at the station locations. As selected results, we present (1) the time series of the estimated key parameters $N_m^{F_2}$ and $h_m^{F_2}$ in comparison with them from the ionosonde station EB040 between March 12 and 14, 2015, in Fig. 12, (2) the time series of the key parameter deviations $\Delta N_m^{F_2}$ and $\Delta h_m^{F_2}$ at the ionosonde station EB040 during the same time interval, in Fig. 12 as well as (3) the average values of the deviations $\Delta N_m^{F_2}$ and $\Delta h_m^{F_2}$ between March 12 and 14, 2015 at the station GR13L for temporal interval of 6 hours in Table 1.

As shown in Fig. 4, there is an insufficient global observation coverage of F3C and GRACE IRO observations for

modelling the key parameters in \mathcal{K}_1 . This leads to a non-uniformity between different regions, thus leading to a unusually larger key parameter deviation. But the largest $N_m^{F_2}$ deviation is observed during 0–6 UT on two of the three (13, 14 March 2015) days due to the possible inaccuracies of the approximate given values in \mathcal{K}_2 , especially for the layers that are presumed inactive during the night. The daily averaged $h_m^{F_2}$ deviation across all stations was determined⁶ to be ± 30 km without any systematic offset. Similar deviation values are also reported by Lei and Chuo (2014), Limberger et al. (2014b), Liang et al. (2015), McNamara et al. (2008).

⁶ Note, we show only results for the ionosonde station EB040 and not for the others.

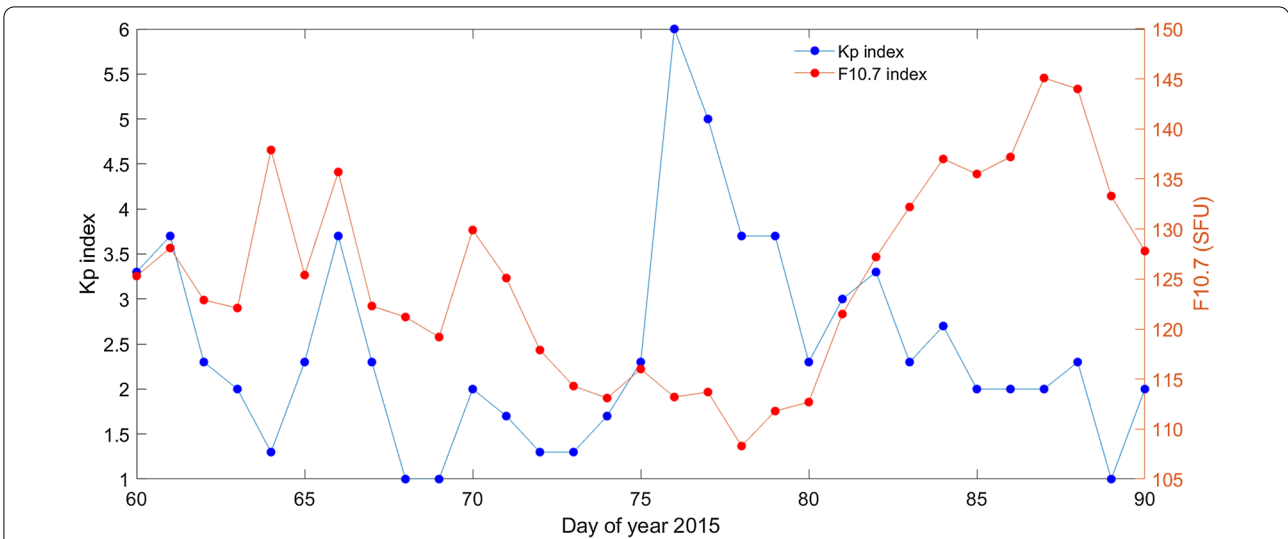


Fig. 11 Space weather situation in March 2015 is shown using the daily averaged time series of Kp and F10.7 indices. While Kp index is unitless, F10.7 has the unit of solar flux units (SFU) with $1 \text{ SFU} = 10^{-22} \text{ Wm}^{-2} \text{ Hz}^{-1}$. The values corresponding to the time window in Table 1 correspond to day of year (DOY) 71, 72 and 73 of year 2015. Data for Kp and F10.7 are downloaded from NASA omniweb explorer portal <https://omniweb.gsfc.nasa.gov/form/dx1.html> accessed on 26 June 2022

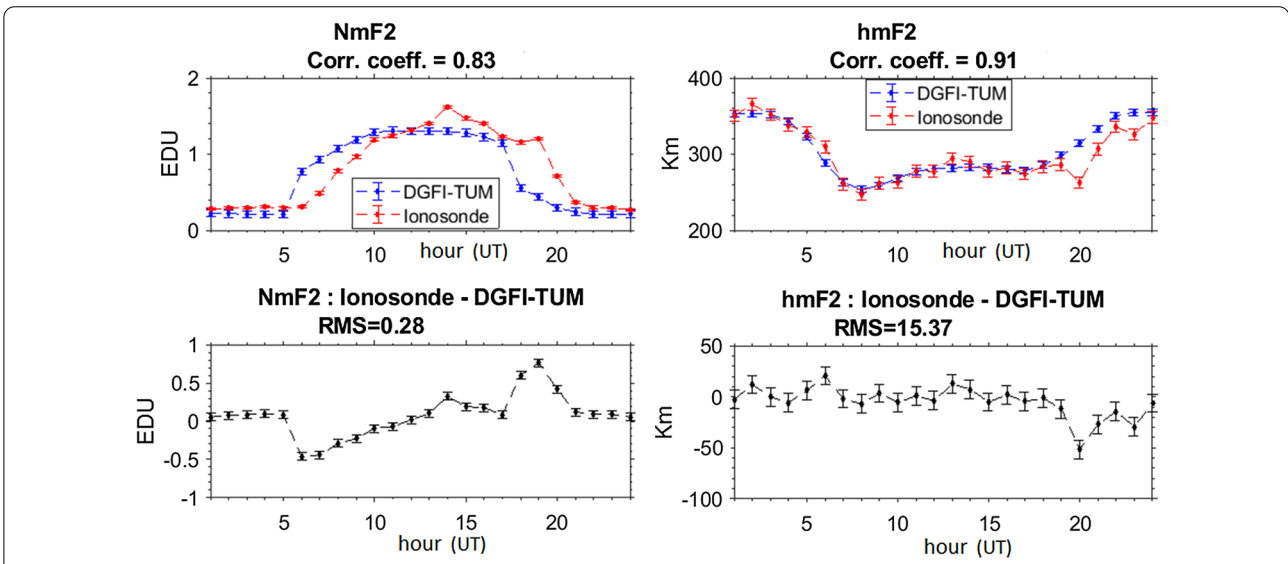


Fig. 12 Comparison of estimated key parameters \hat{N}_m^{F2} (top left panel) and \hat{h}_m^{F2} (top right panel) with the ionosonde data for the station EB040 at March 14, 2015. The lower panels show the corresponding deviations ΔN_m^{F2} (left) and Δh_m^{F2} (right)

Furthermore, Hunsucker (1991), McNamara et al. (2008) mention that the effect of the ionosonde instrument and the technology types, e.g. digisonde, dynasonde, incoherent scatter radar is also significant for the analysis of deviation, which depends on several factors, such as the time of the day, season, location of ionosonde, the phase of the solar cycle, sunspot number and other modelling error(s). To analyse if there is a systematic offset in the deviation due to these reasons, ideally a longer validation campaign is required. Although

we have assumed given values of the E-layer in our modelling approach but did not consider the sporadic E-layer, see Arras et al. (2009) and Tsai et al. (2018), which also needs to be considered in further work.

The difference between the estimated and the ionosonde-derived peak density as well as the peak height for the station EB040 is shown in Fig. 12, where the ionosonde captures the local variations of the electron density and the estimated key parameters follow a smooth

transition with time. The ionosonde data has the higher temporal resolution of 10 minutes compared with our estimates at 1 hour. While it is necessary to use VTEC data for electron density modelling in our approach, its relative number of observations compared to that of F3C and GRACE is higher due to the larger number of discrete altitudes used for defining the profile function. This situation is expected to change with more observations from the F7C mission and the availability of other IRO observations⁷.

For the GR13L station, the maximum deviation with respect to the peak density $N_m^{F_2}$ is 0.49 ± 0.23 EDU which is approximately 22% of the peak value and for $h_m^{F_2}$ we obtain -42.91 ± 7.97 km, which is approximately 9% of the peak value at this location. The corresponding mean values of the two key parameter deviations are 0.17 EDU and 14.3 km, respectively. Both the maxima and the mean deviations are computed during the validation period 12 to 14 March 2015. At the EB040 station, the corresponding maximum deviations are 0.77 ± 0.09 EDU, 48.4 ± 12.5 km, respectively, and the trend is in the same order of magnitude at the other stations as well. Furthermore, we also noticed a difference between the deviations during the day and during the night. As already discussed, they are likely caused due to a non-uniform observation coverage for ED modelling. Similar trends between the IRI model and GPS derived TEC values were also reported in Endeshaw (2020) and Bjoeland et al. (2016). Due to the relatively large number of observations from the separability approach as well as a higher IRI model accuracy near the F_2 peak height, IRI has a significant impact on all the modelled key parameters. Similar conclusions have been made by Limberger et al. (2014b) and Liang et al. (2016). It shall be noted that the larger deviations of the estimated peak density with respect to the ionosonde measurements are mainly due to the sensitivity, i.e. the linear relation between the electron density N_e and the peak density $N_m^{F_2}$ according to the Chapman function (8), compared to the double exponential relation of $h_m^{F_2}$ within the Chapman function. Also considering the truncation error involved in using B-spline expansions, the differences in the temporal resolution of the ionosonde measurements and the estimated key parameters, the choice of the profile function $p(h)$ obtained from the climatology in IRI, the approximate given values for the elements of the set \mathcal{K}_2 as well as the ionosonde observation quality, we conclude the deviations are within acceptable limits.

Summary and outlook

Electron density is the most important geodetic parameters in modelling the upper atmosphere, since it has a significant impact on many applications, such as precise point positioning and satellite navigation. Therefore, its precise modelling is both essential for applications as well as challenging for atmosphere sciences including space weather studies. However, a limited number of observation techniques with global coverage, especially along the altitude and the mutual correlation among different Chapman key parameters are the two main limitations. That means robust estimation estimates are required for 4D global electron density modelling. In this paper, we have developed an inequality constrained optimization algorithm, the ICOA. We have used the multi-layer Chapman approach, where the electron density profiles of the different layers are defined by a set of altogether 14 key parameters. Depending on the chosen layers, their key parameters are then modelled using series expansions in terms of 2D tensor product B-spline functions; the corresponding coefficients are then estimated using ICOA. This approach allows for modelling and estimating multiple Chapman key parameters simultaneously, thereby considering their mutual correlations. Furthermore, we have applied inequality constraints to physically localize the selected key parameters in physically realistic bounds. This way we defined the so-called feasibility region in which we have to search for the final solution. A validation with both simulated and ionosonde data was performed to assess the overall performance. Note, that the results from the simulations are not shown in this paper. Some important findings from our investigations can be summarized as follows:

- The estimated optimization parameters are sensitive to the initial values β_0 , the total number as well as the spatial distribution and quality of the observations, the number of inequality constraints, their chosen lower and upper bound functions as well as on the selected B-spline levels for the Chapman key parameters.
- Particularly, within the optimization algorithm, the initial values $\bar{\beta}_0$ as defined in Eq. (45b) and the step size matrix \mathbf{T} from Eq. (47) have significant impact on the estimated optimization parameters. Therefore, a systematic procedure is required and has to be applied for both aspects; cf. the presented recipe below Eq. (51). In this paper, initial values were chosen from within the feasible region \mathcal{F}_R .
- The choice of the lower bound functions $\kappa_{r;l}$ and the upper bound functions $\kappa_{r;u}$ are of particular importance when key parameters from two or more ionospheric layers have to be considered in the set \mathcal{K}_1 . As

⁷ F7C data are already available since October 2019, but the time period we selected for our study is March 2015.

shown in Fig. 3, the bound functions are generally time varying and numerically in different orders of magnitude. As an example, while the peak density of the F_1 -layer drops to zero at night, that of the F_2 -layer and the plasmasphere still have a significant value; see e.g. Bremer (1998) and Berkner and Wells (1934). Therefore, if the F_1 -layer key parameters are considered in the set \mathcal{K}_1 , their corresponding constraint bounds need to be adapted with time. Accordingly, a reasonable balance between extremely restrictive or more relaxed constraints for both the upper and the lower bound functions is needed. Qualitatively, restrictive constraints are those characterized by relatively small differences $\kappa_{r;u}(\varphi, \lambda) - \kappa_{r;l}(\varphi, \lambda)$; relaxed constraints, however, allow for larger differences.

- In our study, the key parameters of the D -, the E - and the F_1 -layer are obtained from the IRI 2012 model. It shall be noted that IRI also makes assumptions, such as the existence of F_1 layer only within mid- and low-latitudes; see Bilitza (2000), Coisson et al. (2005). Although these assumptions may be reasonable from a long-term climatology perspective, for precise global 4D electron density modelling we need improved models for these layers.

In future, the authors will further investigate the performance of the constraint optimization technique for modelling different sets of key parameters in \mathcal{K}_1 , especially including the bottom layers. Furthermore, different B-spline levels for the individual key parameters as well as Formosat-7/COSMIC-2 (F7C2) electron density observations are also planned in near future. Since the scope of this paper is the ICOA, we have not investigated on the exact nature of mutual correlations between the key parameters of set \mathcal{K}_1 , which depends on the source and duration of data, space weather conditions and also the choice of profile function $p(h)$ in Eq. (52). These aspects shall be investigated in future work as well.

Acknowledgements

This work was done in the scope of the project titled “Development of an operational prototype for the determination of the thermospheric density on the basis of a thermosphere–ionosphere coupling model” with the acronym TIK (original German title “Entwicklung eines operationellen Prototyps zur Bestimmung der thermosphärischen Dichte auf Basis eines Thermosphären-Ionosphären Kopplungsmodells”). The authors would like to acknowledge the German Federal Ministry for Economic Affairs and Energy (BMWi) for sponsoring this project via the German Aerospace Center (DLR Space administration) under the research number: FKZ50LZ1605.

Author contributions

GLG developed the inequality constrained optimization algorithm described in this paper and applied it to model the 4D electron density using the estimated B-spline coefficients and wrote the major part of the manuscript. MS reviewed the manuscript and supported the study with possible validation campaigns. Both authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was done in the scope of the project titled “Development of an operational prototype for the determination of the thermospheric density on the basis of a thermosphere–ionosphere coupling model” with the acronym TIK (original German title “Entwicklung eines operationellen Prototyps zur Bestimmung der thermosphärischen Dichte auf Basis eines Thermosphären-Ionosphären Kopplungsmodells”). The authors would like to acknowledge the German Federal Ministry for Economic Affairs and Energy (BMWi) for sponsoring this project via the German Aerospace Center (DLR Space administration) under the research number: FKZ50LZ1605.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 5 April 2022 Accepted: 21 August 2022

Published online: 19 September 2022

References

- Anderson ED, Gondzio J, Mészáros C, Xu X (1996) Implementation of interior-point methods for large scale linear programs. Springer US, Boston, pp 189–252
- Arras C, Jacobi C, Wickert J (2009) Semidiurnal tidal signature in sporadic e occurrence rates derived from GPS radio occultation measurements at higher midlatitudes. *Ann Geophys* 27(6):2555–2563
- Belehaki A, Marinov P, Kutiev I, Jakowski N, Stankov S (2006) Comparison of the topside ionosphere scale height determined by topside sounders model and bottomside digisonde profiles. *Adv Space Res* 37(5):963–966
- Berkner LV, Wells HW (1934) F-region ionosphere-investigations at low latitudes. *Terr Magn Atmos Electr* 39(3):215–230
- Bilitza D (2000) International reference ionosphere 2000. *Radio Sci* 36(2):261–275
- Bilitza Dieter, Altadill David, Zhang Yongliang, Mertens Chris, Truhlik Vladimir, Richards Phil, McKinnell Lee-Anne, Reinisch Bodo (2014) The international reference ionosphere 2012—a model of international collaboration. *J Space Weather Space Clim* 4:A07
- Billups SC, Ferris MC (1996) Convergence of an infeasible interior-point algorithm from arbitrary positive starting points. *SIAM J Optim* 6(2):316–325
- Bjoland LM, Belyey V, Løvhaug UP, La Hoz C (2016) An evaluation of international reference ionosphere electron density in the polar cap and cusp using Eiscat Svalbard radar measurements. *Ann Geophys* 34:751–758
- Borwein JM, Lewis AS (2000) Inequality constraints. Springer, New York, pp 15–32
- Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge
- Bremer J (1998) Trends in the ionospheric E and F regions over Europe. *Ann Geophys* 16(8):986–996
- Chapman S (1931) The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth. *Proc Phys Soc* 43(1):26
- Coisson P, Radicella S, Leitinger R, Nava B (2005) Topside electron density in IRI and Nequick: features and limitations. *J Adv Space Res* 37:937–942
- Coleman T, Li Y (1996) A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM J Optim* 6(4):1040–1058
- Davies K (1990) Ionospheric radio. *Electromagnetic waves*, Number 31. Institution of Engineering and Technology. <https://digital-library.theiet.org/content/books/ew/pbew031e>
- Endeshaw L (2020) Testing and validating IRI-2016 model over Ethiopian ionosphere. *Astrophys Space Sci* 365:49

- Erdogan E, Schmidt M, Seitz F, Durmaz M (2017) Near real-time estimation of ionosphere vertical total electron content from GNSS satellites using B-splines in a kalman filter. *Ann Geophys* 35(2):263–277
- Feltens J (2007) Development of a new three-dimensional mathematical ionosphere model at European space agency/European space operations centre. *Space Weather*. <https://doi.org/10.1029/2006SW000294>
- Forst W, Hoffmann D (2010) Optimization-Theory and practice. eBook packages mathematics and statistics mathematics and statistics. Springer, New York
- Gao Y, Liu Z (2002) Precise ionosphere modeling using regional GPS network data. *J Glob Position Syst* 1:18–24
- Gass SI, Harris CM (2001) Complementary slackness theorem. Springer, New York, pp 118–119
- Gill PE, Murray W, Saunders MA, Wright MH (1984) Procedures for optimization problems with a mixture of bounds and general linear constraints. *ACM Trans Math Softw* 10(3):282–298
- Goss A, Schmidt M, Erdogan E, Görres B, Seitz F (2019) High-resolution vertical total electron content maps based on multi-scale B-spline representations. *Ann Geophys* 37(4):699–717
- Goss A, Schmidt M, Erdogan E, Seitz F (2020) Global and regional high-resolution VTEC modelling using a two-step B-spline approach. *Remote Sens* 12:1198
- Halldórsson BV, Tütüncü RH (2003) An interior-point method for a class of saddle-point problems. *J Optim Theory Appl* 116(3):559–590
- Hernández-Pajares M, Juan J, Sanz J (1999) New approaches in global ionospheric determination using ground GPS data. *J Atmos Sol Terr Phys* 61(16):1237–1247
- Hernández-Pajares M, Juan JM, Sanz J (2000) Improving the ABEL inversion by adding ground GPS data to LEO radio occultations in ionospheric sounding. *Geophys Res Lett* 27(16):2473–2476
- Hernández-Pajares M, Juan JM, Sanz J, Solé JG (1998) Global observation of the ionospheric electronic response to solar events using ground and LEO GPS data. *J Geophys Res Space Phys* 103(A9):20789–20796
- Howe BM, Runciman K, Secan JA (1998) Tomography of the ionosphere: four-dimensional simulations. *Radio Sci* 33(1):109–128
- Hunsucker RD (1991) Vertical Sounders—the Ionosonde. Springer, Berlin Heidelberg, Berlin, pp 67–93
- Karush W (1939) Minima of functions of several variables with inequalities as side conditions. Master's thesis, Department of Mathematics, University of Chicago
- Koch KR (1985) First order design: optimization of the configuration of a network by introducing small position changes. In: Grafarend EW, Sansó F (eds) Optimization and design of geodetic networks. Springer Berlin Heidelberg, Berlin, pp 56–73
- Koch KR (1988) Optimization of the configuration of geodetic networks. Springer Verlag, Berlin
- Koch KR (1999) Parameter estimation and hypothesis testing in linear models. Springer, Berlin Heidelberg, Berlin
- Kuhn HW, Tucker AW (1951) Nonlinear programming. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press, pp. 481–492
- Lei J, Chuo Y-J (2014) Variations of scale height at f-region peak based on ionosonde measurements during solar maximum over the crest of equatorial ionization anomaly region. Hindawi Publishing Corporation, The Scientific World Journal, London
- Liang W (2017) A regional physics-motivated electron density model of the ionosphere. Dissertation, Technische Universität München, München
- Liang W, Limberger M, Schmidt M, Dettmering D, Hugentobler U (2016) Combination of ground- and space-based gps data for the determination of a multi-scale regional 4-d ionosphere model. In: Rizos C., Willis P. (Eds.) IAG 150 Years, IAG Symposia, 143:751–758
- Liang W, Limberger M, Schmidt M, Dettmering D, Hugentobler U, Bilitza D, Jakowski N, Hoque MM, Wilken V, Gerzen T (2015) Regional modeling of ionospheric peak parameters using GNSS data-an update for IRI. *Adv Space Res* 55(8):1981–1993
- Limberger M (2015) Ionosphere modeling from gps radio occultations and complementary data based on b-splines. Deutsche Geodätische Kommission, Reihe C, 755, Muenchen
- Limberger M, Liang W, Schmidt M, Dettmering D, Hernandez-Pajares M, Hugentobler U (2014) Correlation studies for b-spline modeled f2 chapman parameters obtained from formosat-3/cosmic data. *Ann Geophys* 32:1533–1545
- Limberger M, Liang W, Schmidt M, Dettmering D, Hugentobler U (2013) Regional representation of F2 chapman parameters based on electron density profiles. *Ann Geophys* 31(12):2215–2227
- Limberger M, Liang W, Seitz F (2014) Correlation studies for B-spline modeled F2 chapman parameters obtained from formosat-3/cosmic data. *Ann Geophys* 32(12):1533–1545
- Lötstedt P (1984) Solving the minimal least squares problem subject to bounds on the variables. *BIT Num Math* 24(2):205–224
- Lyche T, Schumaker LL (2000) A multiresolution tensor spline method for fitting functions on the sphere. *SIAM J Sci Comput* 22(2):724–746
- McNamara LF, Baker CR, Decker DT (2008) Accuracy of USU-GAIM specifications of FOF2 and m(3000)f2 for a worldwide distribution of ionosonde locations. *Radio Sci*. <https://doi.org/10.1029/2007RS003754>
- Mead JL, Renaut RA (2010) Least squares problems with inequality constraints as quadratic constraints. *Linear Algebra and its Applications*, 432(8):1936 – 1949. [Special issue devoted to the 15th ILAS Conference at Cancun, Mexico, June 16–20, 2008]
- Mehrotra S (1992) On the implementation of a primal-dual interior point method. *SIAM J Optim* 2(4):575–601
- Mitchell JE, Farwell K, Ramsden D (2006) Interior point methods for large-scale linear programming. Springer, US, Boston, pp 3–25
- Nesterov Y, Nemirovskii A (1994) Interior-point polynomial algorithms in convex programming. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611970791>
- Nocedal J, Wright SJ (2006) Numerical optimization, 2nd edn. Springer, Berlin
- Olivares-Pulido G, Terkildsen M, Arsov K (2019) A 4-D tomographic ionospheric model to support PPP-RTK. *J Geod* 93:1673–1683
- Potra FA, Wright SJ (2000) Interior-point methods. *J Comput Appl Math*, 124(1):281 – 302. [Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations]
- Rawer K (1988) Synthesis of ionospheric electron density profiles with Epstein functions. *Adv Space Res* 8(4):191–199
- Roese-Koerner (2015) Convex optimization for inequality constrained adjustment problems. PhD dissertation, Schriftenreihe des Instituts fuer Geodaesie und Geoinformation
- Roese-Koerner L, Devaraju B, Schuh W-D (2014) Convex optimization under inequality constraints in rank-deficient systems. *J Geodesy* 88(5):415–426
- Schaefer S (1999) Mapping and predicting earth's ionosphere using global positioning system. PhD dissertation, Astronomical Institute, University of Bern
- Schmidt M (2007) Wavelet modelling in support of IRI. *Adv Space Res* 39(5):932–940
- Schmidt M, Bilitza D, Shum CK, Zeilhofer C (2008) Regional 4-D modeling of the ionospheric electron density. *Adv Space Res* 42(4):782–790
- Schmidt M, Dettmering D, Mößner M, Wang Y (2011) Comparison of spherical harmonic and B-spline models for the vertical total electron content. *Radio Sci* 46:RS0D11
- Schmidt M, Dettmering D, Seitz F (2015) Using B-spline expansions for ionosphere modeling. *J JG* 89:939–983
- Schreiner B, Hunt D, Rocken C, Sokolovskiy S (2003) Radio occultation data processing at the COSMIC data analysis and archival center (CDAAC). Springer, Berlin Heidelberg, Berlin, pp 536–544
- Schumaker LL, Traas C (1991) Fitting scattered data on spherelike surfaces using tensor products of trigonometric and polynomial splines. *Num Math* 60(1):133–144
- Stankov S, Jakowski N (2006) Topside ionospheric scale height analysis and modelling based on radio occultation measurements. *J Atmos Sol Terr Phys* 68(2):134–162
- Stankov SM, Jakowski N, Heise S, Muhtarov P, Kutiev I, Warnant R (2003) A new method for reconstruction of the vertical electron density distribution in the upper ionosphere and plasmasphere. *J Geophys Res Space Phys* 108(A5):1164
- Todd MJ, Toh KC, Tütüncü RH (1998) On the nesterov-todd direction in semidefinite programming. *SIAM J Optim* 8(3):769–796
- Tsai L, Su S, Liu C (2016) Ionospheric electron density profiling and modeling of cosmic follow-on simulations. *J Geod* 90:129–142
- Tsai L, Su S, Liu C, e. a. (2018) Global morphology of ionospheric sporadic E layer from the formosat-3/cosmic GPS radio occultation experiment. *GPS Solut*. <https://doi.org/10.1007/s10291-018-0782-2>

- Tsai L-C, Liu C, Y Hsiao T (2009) Profiling of ionospheric electron density based on formosat-3/cosmic data: results from the intense observation period experiment. *Terrestrial Atmospheric and Oceanic Sciences—TERR ATMOS OCEAN SCI*, 20
- Tütüncü RH, Toh KC, Todd MJ (2003) Solving semidefinite-quadratic-linear programs using SDPT3. *Math Program* 95(2):189–217
- Villiger A DR (2019) Technical report 2018. Technical report, IGS Central Bureau, Astronomical Institute, University of Bern
- Yue X, Schreiner WS, Pedatella N, Anthes RA, Mannucci AJ, Straus PR, Liu J-Y (2014) Space weather observations by GNSS radio occultation: From formosat-3/cosmic to formosat-7/cosmic-2. *Space Weather* 12(11):616–621
- Zeilhofer C (2008) Multi-dimensional b-spline modeling of spatio-temporal ionospheric signals. *Deutsche Geodatische Kommission, Reihe A*, 123, Muenchen
- Zhang H, Xu P, Han W, Ge M, Shi C (2013) Eliminating negative VTEC in global ionosphere maps using inequality-constrained least squares. *Adv Space Res* 51:988–1000

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
