Warning and Mitigation Technologies for Travelling Ionospheric Disturbances Effects

TechTIDE

D3.3

Models for the specification of ionospheric background

Version 1.1

Grant agreement no: 776011

The TechTIDE has received funding from the European Commission Horizon 2020 research and innovation programme (2017 – 2020)

Table of Contents

Document Information

Deliverable number:	D3.3
Deliverable title:	Models for the specification of ionospheric background
Date of Delivery:	31 October 2018
Author(s):	Tsagouri I., Belehaki A., Koutroumbas K.
Work Package no.:	3
Work Package title:	TID activity metrics
Work Package leader:	IAP
Dissemination level:	Public
Nature:	Report

Abstract

The precise identification of TIDs requires accurate determination of the background ionospheric conditions. This information will be used by the user of the TechTIDE system to evaluate the criticality of the prevailing conditions in conjunction to the results provided by the system on the identification of TIDs. The deliverable 3.3 describes the algorithms for the calculation of the background conditions in terms of the key ionospheric characteristics (e.g. foF2 and hmF2), as well as the algorithms developed to address additional challenges in the monitoring of the ionospheric activity. These challenges include data quality issues (e.g. data gaps and outliers) and the characterization of the ionospheric disturbance level (weak, moderate, intense).

The deliverable collects relevant information, data, and methodologies fulfilling the scope of the Task 3.2.2 of WP3.

Document history

Version	Date	Edited by	Reason for modification / Remarks
1.0	24/10/2018	Tsagouri I., Belehaki A., Koutroumbas K.	Original version
1.1	30/10/2018	Tsagouri I.	Minor modifications based on the comments received from the TechTIDE Executive Board

Disclaimer

Executive Summary

The reliable identification of TIDs requires accurate determination of the background conditions. The challenge lies in the fact that the TID effects on the ionospheric plasma properties should be cautiously distinguished from large scale space weather effects (i.e., ionospheric storm effects) that may be simultaneously evolved or may act as the triggering mechanism.

The aim of Task 3.2.2 is to develop the algorithms for the calculation of the ambient electron density and relevant key ionospheric characteristics (e.g. foF2 and hmF2). The ionospheric background will be used by the user of the TechTIDE system to evaluate the criticality of the prevailing conditions in conjunction to the results provided by the system on the identification of TIDs.

The tools/methodologies/algorithms for the representation of the background conditions are described in Section 1, including algorithms for the monitoring of large and small scale effects, separately (Sections 1.1 and 1.2). In addition, our analysis demonstrated the need to address additional challenges for the effective implementation of the proposed algorithms. These challenges include data quality issues (i.e. occurrence of data outliers and data gaps) and the requirement to characterize properly the level of the ionospheric activity. In response to these needs, a set of auxiliary tools/algorithms are described in Section 2.

1. Algorithms for the determination of the background conditions

The reliable identification of TIDs requires accurate determination of the background conditions. The challenge lies in the fact that the TID effects on the ionospheric plasma properties should be cautiously distinguished from large scale space weather effects (i.e., ionospheric storm effects) that may be simultaneously evolved or may act as the triggering mechanism.

The aim of Task 3.2.2 is to develop the algorithms for the calculation of the ambient electron density and relevant key ionospheric characteristics (e.g. foF2 and hmF2). The ionospheric background will be used by the user of the TechTIDE system to evaluate the criticality of the prevailing conditions in conjunction to the results provided by the system on the identification of TIDs.

The work plan was based on the assessment of state of the art methodologies during specific time intervals that were selected through the TechTIDE event catalogue available in the project's wiki pages.

Taking into account the TechTIDE specifications, the proposed methodologies should be able to meet the following requirements:

- *Prediction efficiency:* the algorithms should be able to reproduce the normal ionospheric changes (e.g., diurnal, monthly, seasonal, solar cycle, latitudinal, longitudinal dependence), but also large scale disturbances (e.g., large scale storm time effects).
- *Implementation:* the algorithms should be able to provide output in real-time.

Based on the requirements, the running medians/averages are considered as the preferable approaches for the representation of the background conditions against other options such as spectral analysis methods and nowcasting ionospheric models, which mainly meet difficulties in real time response.

1.1 Running medians

For the representation of the normal ionosphere, as well as the monitoring of large scale disturbances it is suggested to use the running medians calculated using the observations obtained over the last 30 days (current day/value and backwards). The time resolution of the running medians follows the time resolution of the observations, so that in each epoch the ionospheric background is estimated as the median of the values recorded in the same epoch during the last 30 days. This option is the close to the concept of the monthly median that is used widely for the determination of the normal ionospheric variation in ionospheric studies, being available in real time at the same time. Indeed, comparison tests between 30-days running medians and monthly medians indicate that there is a satisfactory agreement between them within the uncertainty limits (see Fig. 1).

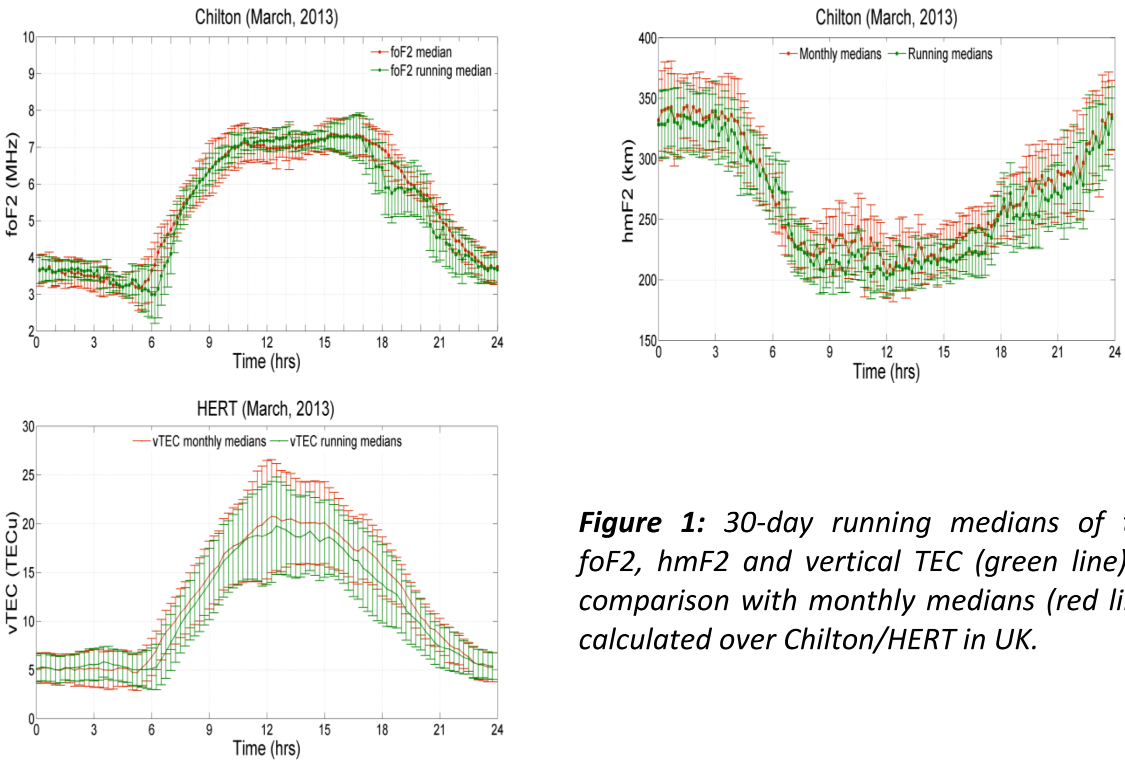

Figure 1: *30-day running medians of the foF2, hmF2 and vertical TEC (green line) in comparison with monthly medians (red line) calculated over Chilton/HERT in UK.*

An example of the running medians representations is provided in Fig. 2 for foF2 and hmF2 over Dourbes for the disturbed period 6 - 13 November 2017.

The ionospheric activity level can then be evaluated through the residuals/differences (DIFFs) of the observations from the running medians (see also Fig 3).

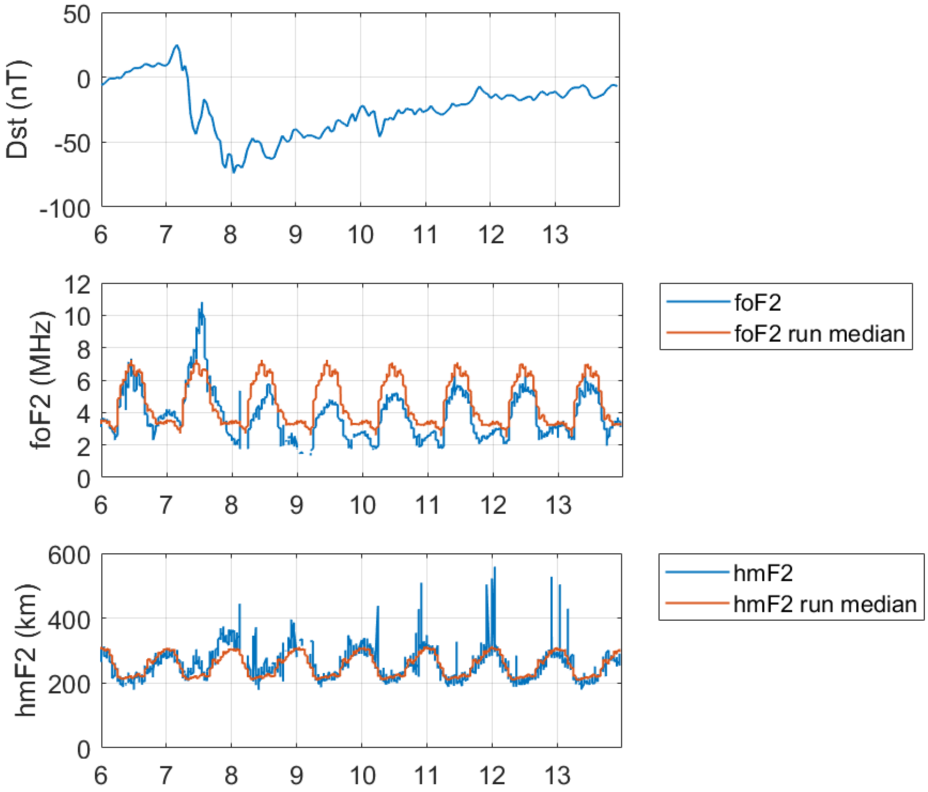

Figure 2: *Actual observations and running medians representations for foF2 and hmF2 over Dourbes for the disturbed period 6 - 13 November 2017. Dst records are also provided in the top panel as an indicator of the geomagnetic activity level.*

1.2 Running averages

Running averages could respond to the needs for the monitoring of small scale effects, i.e. TIDs. Following the results obtained previously (e.g. [1]), but also on the analysis performed here **we suggest the background representation for the monitoring of small scale ionospheric effects with 60 min running averages** (current value and backwards). **The ionospheric activity level can then be evaluated again through the residuals/differences (DIFFs) of the observations from the running averages** (see Fig. 3).

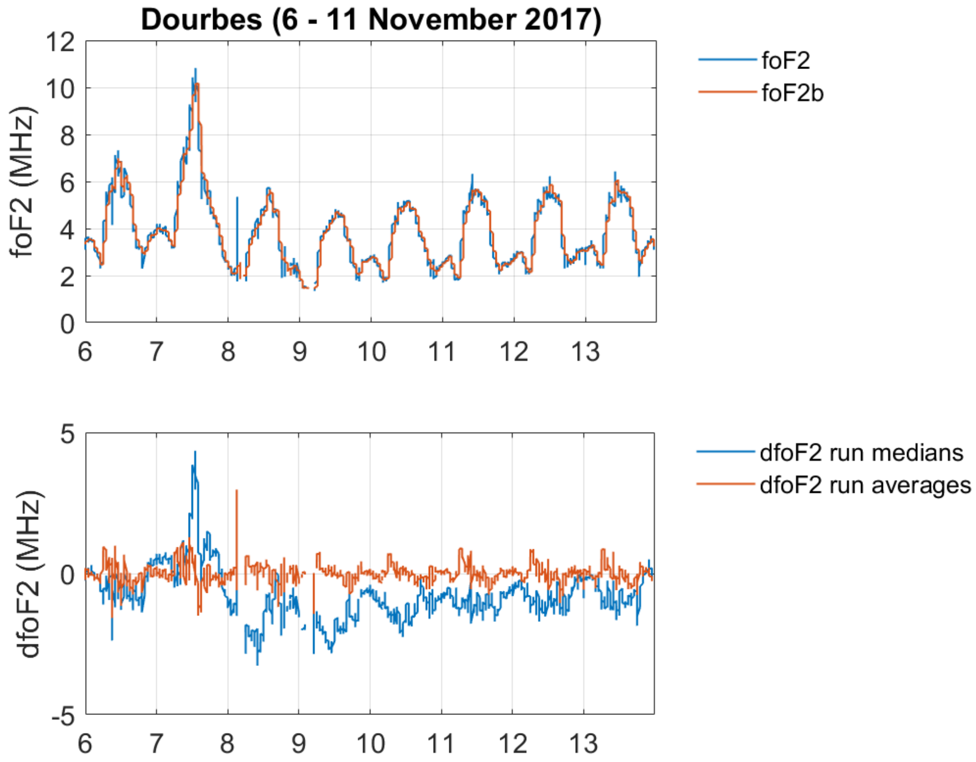

Figure 3: *Top: The background foF2 values, foF2b (red line) estimated as 60 min running average and the absolute foF2 (blue line) for the time interval 6 - 13 November 2017, estimated with data from Dourbes. Bottom: The residuals of foF2 from their running median values (blue line) plotted together with the de-trended foF2 using the 60 min running averages as background (red line) for the interval and location. It is obvious that the two quantities differ largely since the de-trended foF2 does not include large scale ionospheric variations.*

2. Additional challenges

The analysis demonstrated also the need to address additional challenges for the effective implementation of the proposed algorithms. These challenges include data quality issues (i.e. occurrence of data outliers and data gaps) and the requirement to characterize properly the level of the ionospheric activity. In response to these needs, a set of auxiliary tools/algorithms are described below.

2.1 Data quality

The reliable monitoring of the ionospheric conditions requires the continuous availability of high quality data. Some of the challenges that should be addressed include the occurrence of data outliers (probably attributed to autoscaling errors) and data gaps (see also Fig. 2 and Fig. 3). To address the needs a set of Data Filtering Algorithms (DFAs) was developed as below.

2.1.1 Removing outliers

We suggest removing the outliers in foF2 and hmF2 in two steps as below:

1st step: a) Remove outliers in foF2: a value is removed if it differs more than 20% than its previous one (continuous counting).

dfoF2 (%) = [(foF2_current - foF2_previous)/foF2_previous] * 100

b) Remove spikes in hmF2: Remove the values that correspond to foF2 values removed in a.

2nd step: Remove **additional** outliers in hmF2 by removing a value if it differs more than 20% than the previous one (continuous counting).

The threshold of 20% was determined through the analysis of the distributions of the relative deviations of any "current" value with respect to the previous one. Fig. 4 and Fig. 5 present the corresponding distributions in Dourbes and Ebre for foF2 and hmF2. The analysis includes two years of different solar activity level - 2017 for low solar activity and 2012 for high solar activity - and has been applied to data with 5 min time resolution.

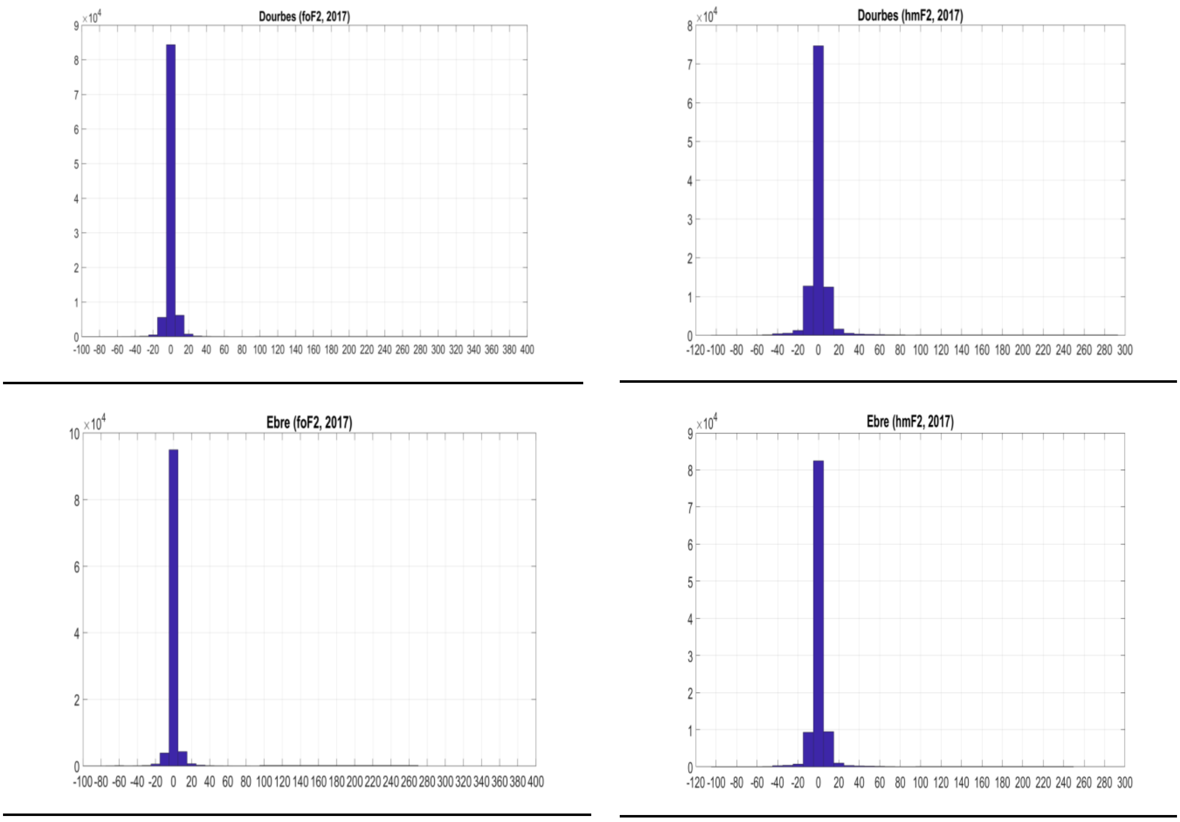

Figure 4: *The distributions of the relative deviations calculated over Dourbes (top) and Ebre (bottom) locations for foF2 (left) and hmF2 (right) in 2017 (low solar activity conditions).*

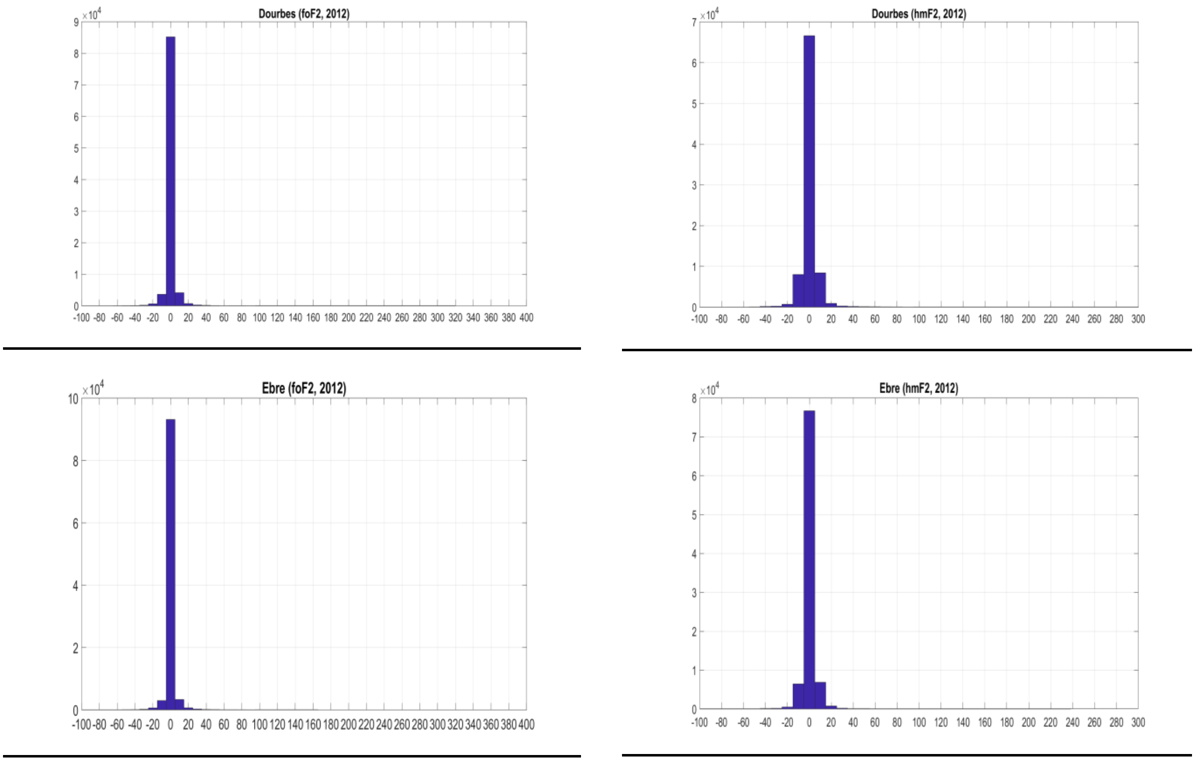

Figure 5: *The distributions of the relative deviations calculated over Dourbes (top) and Ebre (bottom) locations for foF2 (left) and hmF2 (right) in 2012 (high solar activity conditions).*

An example of how the proposed algorithm works is provided in Fig. 6.

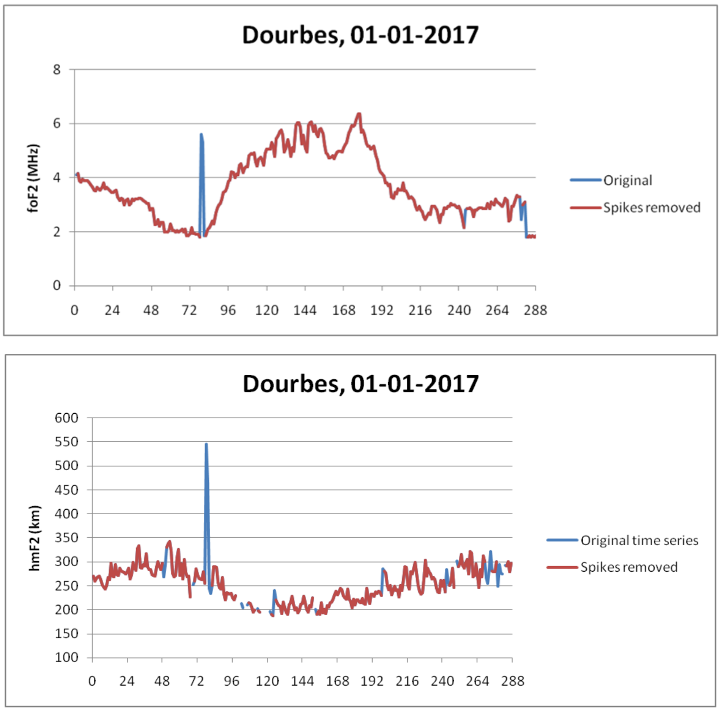

Figure 6: *The original time series of foF2 (top) and hmF2 (bottom) values (blue lines) over Dourbes on 1 January 2017 together with the resulted ones after the removal of the outliers (red lines).*

2.1.2 Data gaps

The proposed methodology is a general purpose methodology that can be used for any time series. Its usefulness is twofold: it can be used (a) either for filling gaps or (b) for performing one-step ahead predictions. Actually, it is based on a non-parametric local linear function approximation (LLFA) method ([2], [3]). In the next section we describe the LLFA method and then, show how it can be adapted in the time series estimation framework.

2.1.2.1 Description of the proposed method

The LLFA method

Assume that it is available a set of data $X = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)\}$ with $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \dots, x_{il}]^T \in R^l$ and $y_i \in R, i = 1, \dots, n,$ [1] so that $\boldsymbol{x}_i$, and y_i are related via a function f, whose general formula is unknown. That is, $y_i = f(\boldsymbol{x}_i)$. The problem now is: “given a certain vector $\boldsymbol{x}^*$ estimate its associate y^*via f, based on the information included in X.”

A solution for this problem is the following:

Determine the k nearest neighbors of $\boldsymbol{x}$ in the array X. $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,...,$\boldsymbol{x}_k$).

[1] Note that, in general, $\boldsymbol{y}_i$ could be a vector in a k-dimensional space, that is, $\boldsymbol{y}_i \in R^k$.

1. Among $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$, determine the k nearest neighbors of $\boldsymbol{x}^*$. That is, compute the (Euclidean) distances of $\boldsymbol{x}^*$ with all $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ vectors and keep those vectors that correspond to the k smallest distances (without loss of generality and in order to keep the notation as simple as possible, let us assume that the nearest to $\boldsymbol{x}^*$ vectors are the$\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_k$ $(k \ll n)$).
2. Construct the table $X = \begin{bmatrix} \boldsymbol{x}_1^T & 1 \\ \boldsymbol{x}_2^T & 1 \\ \vdots & \\ \boldsymbol{x}_k^T & 1 \end{bmatrix}$ and the vector $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}$.
3. If $X^T X$ is invertible, then
   3.1. Compute $\boldsymbol{\theta} = (X^T X)^{-1}(X^T \boldsymbol{y})$(least square error solution)
   3.2. Estimate y^* as $\widehat{y^*} = [\boldsymbol{x}^* \ 1]^T \boldsymbol{\theta}$(first order approximation)

   Else

   3.3. Estimate y^* as $\widehat{y^*} = mean(y_1, \dots, y_k)$ (zero order approximation)

In step 3.1, $\boldsymbol{\theta}=[\theta_1,\dots,\theta_d, \theta_0]^T$is the column vector modeling the (regressor) hyperplane *H*: $\theta_1 x_1 + \dots + \theta_l x_l + \theta_0 = 0$defined by the points $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_k$, through the minimization of the sum of error squares function.

The intuition behind the above method is that the unknown function f is linearly approximated close to $\boldsymbol{x}^*$ (step 3.1 computes the parameters of the associated hyperplane *H*) and the estimate of y^* associated to $\boldsymbol{x}^*$, is performed based on this approximation (step 3.2). In the case where linear approximation is not possible, zero order approximation is adopted and y^* is estimated as the mean of $y_1, y_2, \dots, y_k$that correspond to$\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_k$, respectively.

LLFA for estimating time series

Consider a time-series $\{x(n)\}, n = 1, \dots, N$, and assume that the memory of the system is d (that is $x(n)$ is dependent on the vector $[x(n-d), x(n-d+1), \dots, x(n-1)]^T$).

1st stage (creating the database): Defining $\boldsymbol{x}_i{}^T = [x(i), \dots, x(i+d-1)], i = 1, \dots, N-d,$ we construct the following data set

$$X = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_{N-d}, y_{N-d})\}$$

where $y_i \equiv x(d+i), i = 1, \dots, N-d$. That is, each sequence of d consecutive measurements in the time series is associated with its next value. It is noted that no missing values are encountered in X(if a certain measure $x(i)$ in $\{x(n)\}, n = 1, \dots, N$ is missing, all pairs in Xthat contain $x(i)$ are discarded).

2nd stage (prediction): Assume now that the (successive) measurements $x(i-d), \dots, x(i-1)$ are available and we want to estimate the value $x(i)$ based on the previous d measures.

Defining $\boldsymbol{x} = [x(i-d), \dots, x(i-1)]^T$, we simply employ the LLFA method described before.

2.1.2.2 Evaluation of the proposed method

To evaluate the efficiency of the proposed method we followed the methodology described below (for d=5):

- We selected 15 periods of 4-hours duration over Dourbes for the time interval 6 - 13 November 2017 (1 meas./5mins), from which we created artificial gaps of various durations. The 15 periods have been selected from three LT sectors: 01:00 - 05:00 LT (Morning), 11:00 - 15:00 (Noon) and 17:00 - 21:00 (Afternoon) (5/LT sectors).
- For each of the 4-hour data gap we performed 16 runs:
  At the 1st run, we considered only the first 3 meas. (15 mins gap) as missing.
  At the 2nd run, we considered only the first 6 meas. (30mins gap) as missing. . . .
  At the 16th run, we considered all the 48 meas. (240 mins gap) as missing.
- We compared the method's estimates with the actual measurements for each run using the MSE (mean square error) metric, calculated in each data gap/step.
- The same tests were repeated for median values.

Indicative results of the proposed method's evaluation are presented in Fig. 7.

Based on the evaluation results, **it is suggested to use the proposed method in filling data gaps of duration up to 45 min in order to efficiently keep reliability of the estimates.**

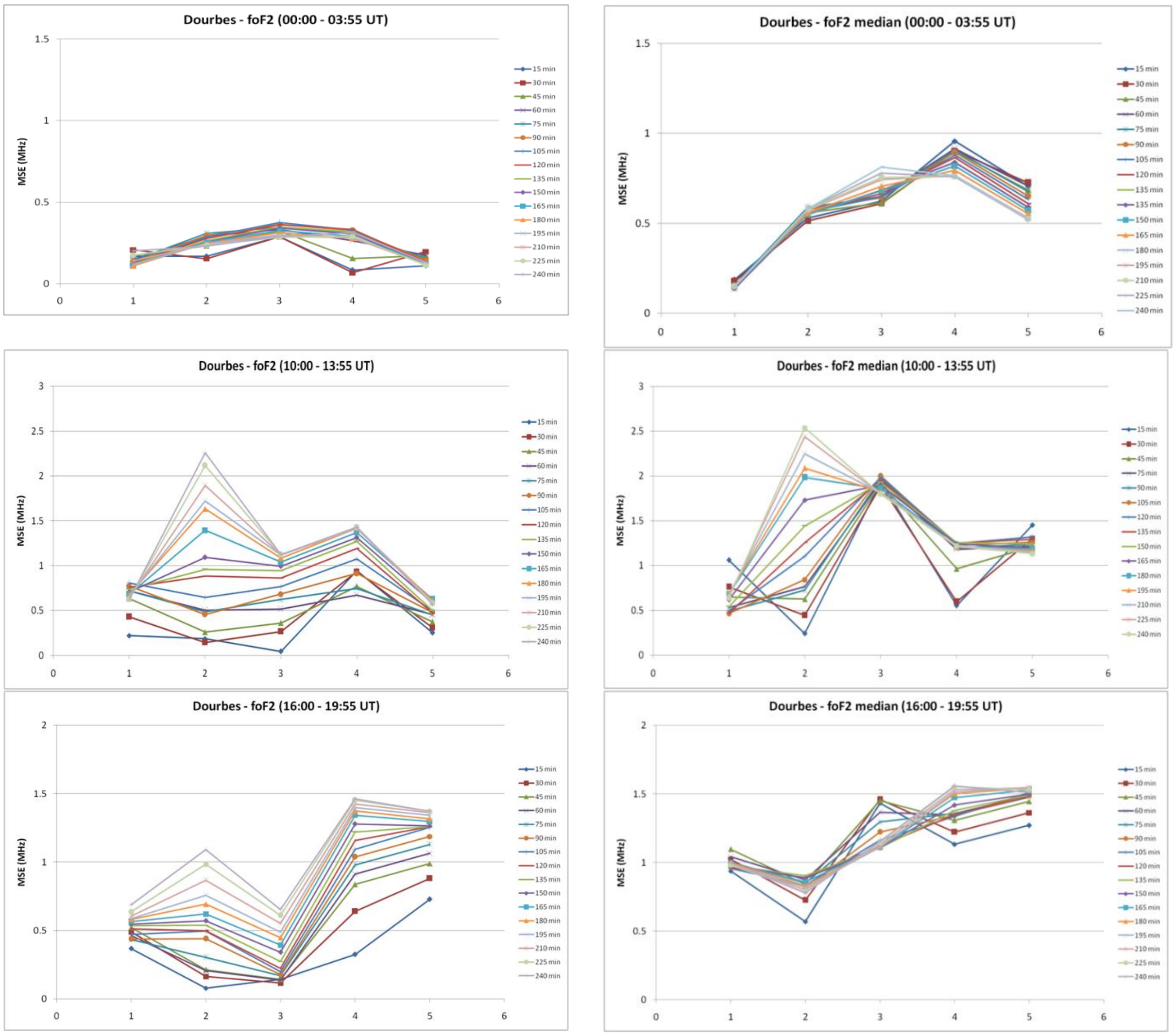

Figure 7: *The MSE estimates in resulted from the comparison between the method's estimates with the actual measurements for each run in each data gap/step for Morning (top), Noon (middle) and Afternoon (bottom) local time sectors in Dourbes.*

2.2 Characterization of the ionospheric activity level

The ionospheric activity level should be characterized e.g. as weak, moderate or intense especially for large scale effects for users' awareness. This is suggested to be realized in two steps:

<u>1st step:</u> Defining a threshold for significant disturbances. In case of running medians, one may use the relative standard deviation, STD (%) of the values taken into account in the calculation of the medians.

The relative STD (%) for foF2 is calculated as follows:

STD (%)=(STD_foF2 run median/foF2 run median)*100

where *STD_foF2 run median* is the STD of the values used for the calculation of the foF2 run median. The STD (%) is calculated in each epoch and it is depended on the characteristic, location, season and local time (see Fig. 8).

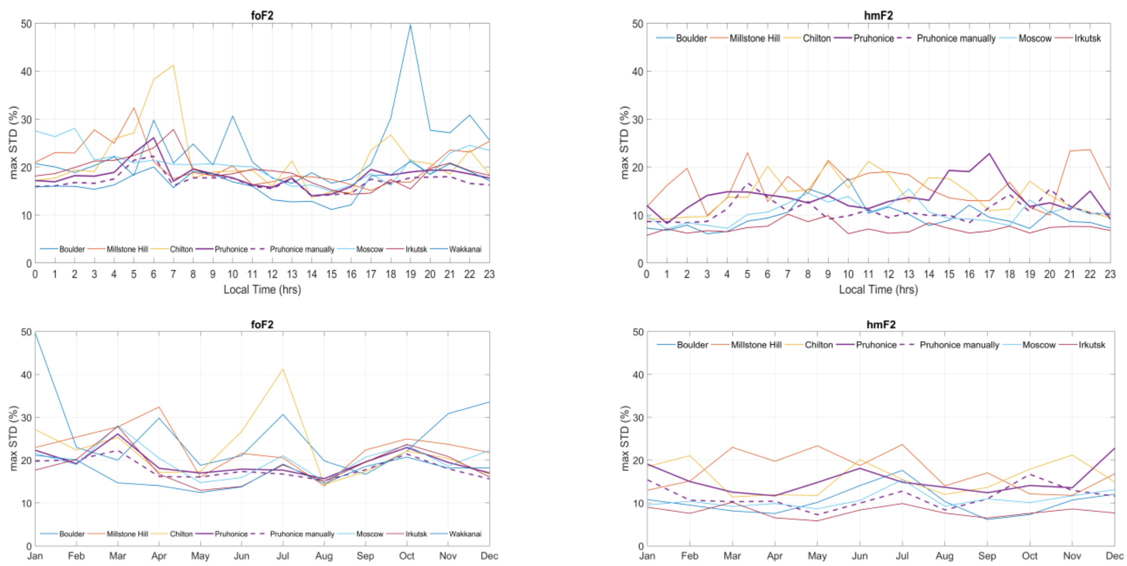

***Figure 8:** The maximum relative STD (%) for foF2 (left) and hmF2 (right) versus local time (top panels) and month (bottom panels) obtained for 2012 (high solar activity).*

<u>2nd step:</u> Quantification of the disturbances intensity (e.g. weak, moderate, intense). It is suggested to be realized in terms of users' requirements. For this purpose, we suggest to compare the residuals of each ionospheric characteristics from the background conditions (taking into account the STD) with other TID activity indicators related to users requirements, as for example the AATR (see Fig. 9).

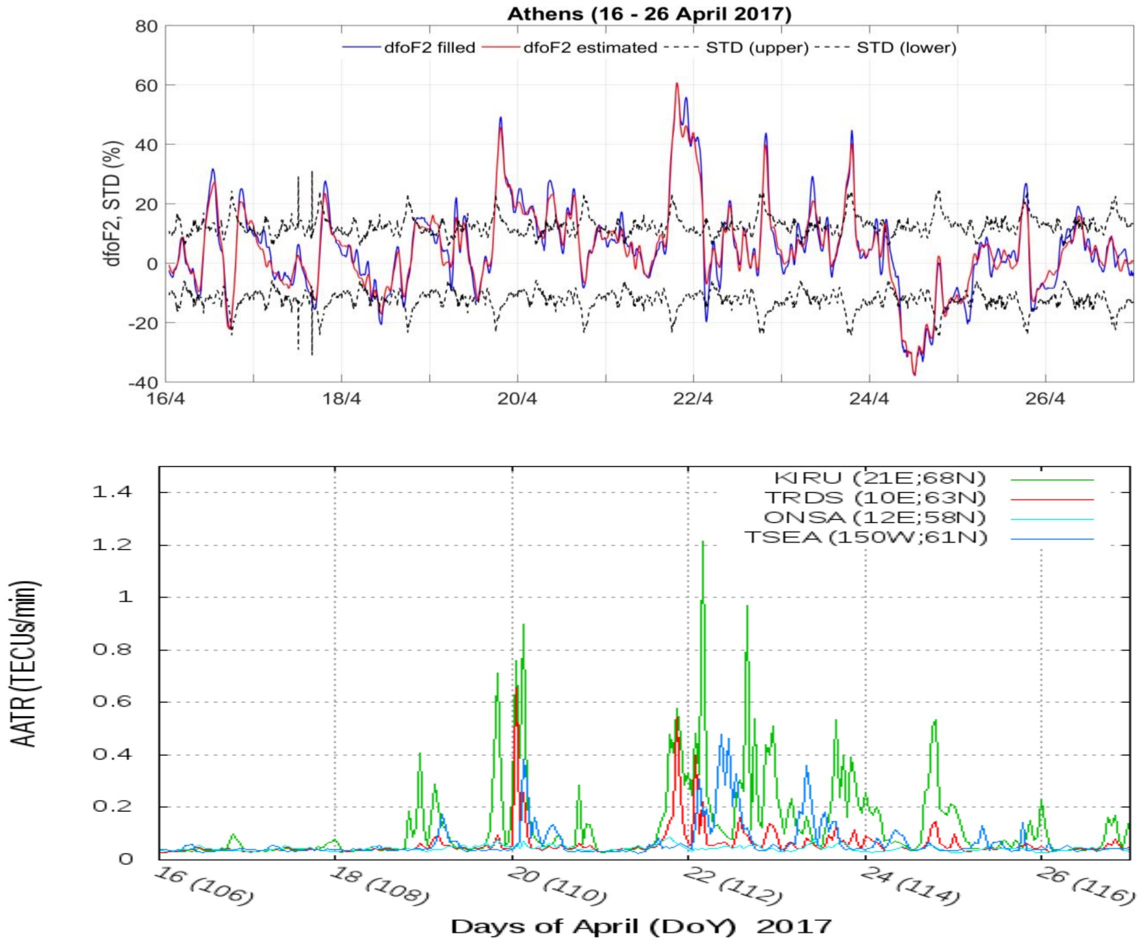

Figure 9: *The foF2 deviation (dfoF2) together with to STD (%) (top) in comparison with the AATR index results for the time interval 16 April - 26 April 2017.*

3. Recommended products for the specification of the ionospheric conditions

A number of products suggested to be provided to the TechTIDE system for the specification of the ionospheric conditions are listed in Table 1.

Table 1: *List of products suggested to be provided for the specification of the ionospheric conditions.*

Products	Type of file	Time resolution	Provider
foF2 and hmF2 from 14 stations (raw and filtered data)	TXT	5min – 15 min	NOA
Running median & DIFFs for foF2 and hmF2 parameters from 14 stations	TXT	5min – 15 min	NOA
Running averages & DIFFs for foF2 and hmF2 parameters from 14 stations	TXT	5min – 15 min	NOA
Geographic maps with indication of DIFFs from running median and de-trended values for foF2 and hmF2 over stations (Europe and Africa)	JPG	5min	NOA

A more quantitative specification of the ionospheric activity level will be given for the 2nd TechTIDE prototype in October 2019.

References

[1] Belehaki A., I. Kutiev, P. Marinov, I. Tsagouri, K. Koutroumbas, P. Elias, Ionospheric electron density perturbations during the 7-10 March 2012 geomagnetic storm period, Advances in Space Research, Volume 59, Issue 4, Pages 1041-1056, 2017, https://doi.org/10.1016/j.asr.2016.11.031.

[2] Farmer J.D., Sidorowich J.J., "Predicting chaotic time series", *Physics Review Letters*, 59, 845-848, 1987.

[3] Farmer J.D., Sidorowich J.J., "Exploiting chaos to predict the future and reduce noise", in *Evolution, Learning and Cognition*, ed. Y.C. Lee, World Scientific Press, New York, 1988.