The 7th Framework Programme
Theme SPACE
Project POPDAT

D1.4 DATA MINING TOOLS

REPORT

Version 1

Project ref. no.	263240
Project title	POPDAT Problem-oriented Processing and Database Creation for Ionosphere Exploration

Deliverable status	Final
Contractual date of delivery	Project Month 15
Actual date of delivery	Month 15
Deliverable number	D1.4
Deliverable title	Data mining tools
Type	Report
Status & version	Version 1
Number of pages	21
WP contributing to the deliverable	WP1
WP / Task responsible	Eötvös Loránd University
Author(s)	Project Consortium
EC Project Officer	Paula MOTA-ALVES
Keywords	Ionospheric research data file format record structure catalogue waves

Content:

1. Fixing of the tools used in POPDAT, and selected in D1.3

The advanced methods and software's for POPDAT data mining were selected in the D1.3 document titled as **"Report on preliminary output of problem-oriented data processing"**. The Ionosphere Wave Service (IWS) was presented in D3.3 report. Selection of the complete set of wave phenomena appearing in the IWS was given in the D1.3 report. These wave phenomena, which will appear in the present construction of the IWS are the following:

- Electromagnetic wave phenomena with complete automatic searching:
  - Whistlers with a wide range of signal dispersion (WHF).
  - Longer path fractional hop whistlers (LIWHF).
  - Short path fractional hop whistlers (SIWHF).
  - Low-latitude whistler pairs (WHPF).
- Non-electromagnetic wave phenomena with semi-automatic searching:
  - Atmosphere gravity waves (AGW).
  - Traveling ionospheric disturbances, in situ data (TID).
  - Traveling ionosphere disturbances, GPS data (TID-VTEC).

Inside the POPDAT the semi-automatic selection of data performed by researchers to find the connection between the thunderstorm activities, the TLE's, the whistlers and the broadband electromagnetic noises are also tasks.

The tools of these tasks are the content of this report.

Data sets in POPDAT IWS, collected by automatic or semi-automatic detection procedures, may include specific event subclasses of the above mentioned phenomena. These subclasses, listed below, are to be selected by direct application of IWS search engine in data base.

spiky whistlers (SpW), swallow-tailed whistlers (STW), X-shaped whistlers in the WHF, LIWHF and WHPF categories;

ion whistlers in the SIWHF category;

seismic activity and whistlers in WHF, LIWHF and SIWHF categories;

seismic activity and AGW's in AGW category;

seismic activity and TID's in TID and TID-TEC categories.

However, these steps are done by the researchers using the IWS data and these steps are not the parts of the present POPDAT service. The POPDAT opens the way for these investigations.

The solving of the full-automatic selection of the SpW, STW, X-shaped and ion whistlers needs some fundamental research and further development of the present existing methods. In the other tasks which are mainly the investigation of the coupled appearing of seismic, etc. and wave events, as WHF's, AGW's, TID's, the direct activity of the researchers is needed.

A few classes of the wave phenomena were omitted from the scope of the POPDAT, thus not appearing in the further application of the IWS, and the data sets selected in the POPDAT project. These phenomena are electromagnetic phenomena, as

the hiss and chorus signals having diffuse shapes in a wide range of forms,

the triggered emissions having a very wide range of shape character and

the strange signals having unknown, non-defined, non-categorized signal shapes.

These categories have in a very wide variable and/or diffuse shapes without exact quantitative description parameters. Therefore in the POPDAT we have no possibility to add e.g. meta-data description of these signals, we have no well tested and reliable methods for automatic searching of these signal classes in the satellite raw data bases, etc.

In the following we shall present the parameters and application mode of software tools for searching the wave phenomena selected for the IWS in the POPDAT project. The theory and the description of the algorithms were presented in the D1.3 report. Applying these algorithms it is possible to produce the final data base of the IWS. Description of service modes of the IWS is subject of D3 reports already submitted or in preparation.

1.1 Electromagnetic wave phenomena

The selected core method for the POPDAT project is the Automatic Whistler Detection (AWD) method and software, as it was presented in the Chapter 1.1 in the D1.3 report. The original version of this method and software was developed for the automatic investigation of continuously recorded broadband VLF signals in ground based stations. (See more in our earlier reports, in the references cited in these reports and in the reports of the EU FP7 PLASMON project.)

Since AWD has been optimized for ground recorded event types, different in several characteristics from electromagnetic waves collected in space plasmas, further development was necessary for the application of the original AWD method and algorithm in satellite raw records. The result of this development is the advanced AWD software package, the AWDSat, detailed in Section 2. After the successful finishing the development of the advanced, satellite oriented Automatic Whistler Detector software, the current version is operational. According to target wave phenomena it includes four detection sub-routines in one unit:

a) The **AWDSat-WHF** for the automatic detection of the normal (typical single hop) whistlers in the satellite recordings. This algorithm can detect and select automatically the whistlers in the data sets recorded on board of satellites, if these whistlers have a higher dispersion, i.e.

$$\mathrm{D} > 35\sqrt{s}.$$

The typical dispersion range of the normal whistlers in satellite records is: $\mathrm{D} = 45 \sim 200 / 300\sqrt{s}$.

b) The **AWDSat-LIWHF** for the automatic detection of the longer path fractional hop whistlers in the satellite records. This algorithm can detect, can select automatically the whistlers in the data sets recorded on board of satellites, if these whistlers have smaller dispersion, i.e.

$$\mathrm{D} = 8 \sim 35\sqrt{s}.$$

c) The **AWDSat-SIWHF** for the automatic detection of the short path fractional hop whistlers in the satellite records. This algorithm can detect, can select automatically the whistlers in the data sets recorded on board of satellites, if these whistlers have small dispersion, i.e.

$$\mathrm{D} \leq 8\sqrt{s}.$$

The typical dispersion values of these whistlers in satellite records are: $\mathrm{D} = 0.8 \sim 8\sqrt{s}$.

d) The **AWDSat-WHPF** for the automatic detection of SIWHF and LIWHF signals if they appear simultaneously, the SIWHF signal appears before the LIWHF signal and the time gap between the two signals is in the order of 100 msecond.

The application modes of these software's are given in the paragraph 2, Note that the output formats were given in the D1.3 report, IWS input data structures have been detailed in D2.2 and D3.4 reports. Data processing routines in executable formats are attached electronically to the Appendix of this report or available by download (GPS VTEC).

1.2 Atmosphere gravity waves

Technique for atmosphere gravity waves (AGWs) selection in the satellites measurement data is based on extensive practical experience. There are convincing examples of its efficacy, but as yet there is no theoretical and methodological substantiation of the method necessary to provide automatic wave identification and extraction from the datasets. Currently AGW extraction from the measurement data is performed semi-automatically, under operator control. See this in more detail in D1.3.

Source experimental data are stored at server of the National Space Science Data Center at http://nssdc.gsfc.nasa.gov/atmoweb in free access. The elementary unit of AGWs catalogue is NetCDF Wave Data File related to a single orbit of DE 2 satellite (e.g. Wave Data File at orbit No. 7280). Wave data consist of

$\qquad$ (i) the vector of values of relative wave variation $a = \delta\rho / \overline{\rho}$ and

$\qquad$ (ii) the vector of values of trend $\overline{\rho}$.

The absolute AGW density variation is $\delta\rho = a \cdot \overline{\rho}$. NetCDF file also includes several groups of additional parameters. There are headers which identify type of the wave process (in our case, AGW), data source (in our case, DE 2 mission, the NACS experiment), orbit number, etc. Second, there are parameters of the space-time window (vectors of UT, geographical and magnetic coordinates, altitude etc). The third set of parameters characterizes the AGW itself: the maximal and minimal wave amplitude, the presence and localization of definite the certain kinds of wave events (PAGW, BAGW, LAGW), etc.

The data processing procedure consists in the band pass filtering of source data in the range of medium-scales. First, a broadband filter used. In a case of quasi harmonic variations identification the filter band is contracted to the carrier frequency to clear the waveform of high-frequency noises and any low-frequency trend.

To verify that the resulting variations are AGWs, we check whether the measured characteristics of the signal fit the theory of AGW.

1.3 Traveling ionosphere disturbances (in suit data)

TID recognition by computer code POPTID.exe from satellite in situ data is based on search for specific ionospheric TID signatures at different latitudes and altitudes. Thus, first we separate ionosphere into two major zones,

$\qquad$ - polar cap zone, excluding auroral oval zones
$\qquad$ - middle and low latitude zone, excluding ionospheric trough region

1.3.1. *Polar Cap and Auroral zone.*

        At high latitudes, charged particles represent the local ion/neutral behavior dominated by ionosphere-magnetosphere interaction processes. Ionospheric plasma is highly irregular in a wide range of scales, especially in auroral oval zone where additional rate of ionization is associated with preciritating particles, large shear flows, providing plasma structuring and instabilities. At high geomagnetic activity levels, large scale electric fields drive intense plsma convection which drives neutrals to follow it up with similar dynamical behavior. These ion/neutral interaction processes dominate ionosphere structures at an altitudes up to 350-450km.

        At high magnetic inclination angles large DC Electric field have significant component of apparent vertical ion motion which makes TID recognition questionable. This has maximum at the auroral zone, where horizontal ion velocities often reach few km/s. To avoid misstreating of these variations as TID signatures we exclude highly structured zones in horizontal ion velocity. Thus, working with polar cap data at moderate and low geomagnetic activity leves (smooth Vhor profile) we accepted folloing basic principles concerning development of POPTID high latitude data processing:

                - Within Polar Cap, vertical ion motion (along the magnetic field lines) is accepted to be the most important parameter for us in TID recognition at F-region heights.. It represents the wave-like structures and very well correlate with neutral vertical motion.

                - Depending on satellite altitude, dayside/nightside and/or horizontal plasma convection, Ni and Ti could follow more or less Vver behavior. Relative effectiveness of chemical and dynamical ion/neutral coupling processes decreases with altitude, loosing individual TID signatures for some of the parameteres.

                - At high altitudes Ni, Ti variations should be partially and with caution addressed to AGW at low altitudes, as they compose the influence of great number of additional sources from below and far above magnetosphere.

1.3.2. *Middle and Low Latitude (MLL) zone:*

        At MLLs, it is important to exclude Main Inospheric Trough (MIT) region from data processing, because of it presence in data generates broad wave spectrum just because of MIT steap density gradient at it poleward side. At this edge it is often observed large Sub-Auroral Ion Drifts (few km/s called SAID or subauroral plarization jets- SAPS or streams-SAS), which strongly affects chemistry and dynamics within whole atmosphere in vertical direction. Equatorward from trough region, due to rapid decrease of the magnetic inclination, vertical ion drift motion and wave-like structures are controled by magnetic field line configuration. Vertical ion motion is not simply AGW produced by local ion/neutral interaction processes, but a subject of global DC electric field distribution, especially it zonal component. Toward equator, Ni and Ti variations become more important and representative AGW generated TIDs at 240-450k altitude. Thus, at MLL zone POPTID data processing represents:

                - Ion density variations are taken to be the indicator for presence of wave-like structures (TID) in data

                - MLL distribution of wave-like structures in Ti and Vver are accepted to be subject of Ni variations.

                - The effect of MIT is eliminated by equatorward shift of MLL boundary toward smaller density gradients

                - Equatorial Bubble/Blob structures could be eliminated by the existing enhancement in their spectral power toward high frequencies

        These basic principles were accepted as input for POPTID code development. It concerns mainly the Dynamics Explorer-2 data and partially DEMETER and AE-C and AE-E data processing because of

orbital specifics of AE and DEMETER satellites. Here attached POPTID code provides semiautomatic data processing of IDM/RPA data from DE-2 satellite. Automated version of the POPTID software is under development and should be accomplished after semiautomatic DE-2 data processing.

1.4 GPS VTEC analysis of traveling ionosphere disturbances

        The detailed method used to detect TID in TEC time series is described in *Chapter 3.4*. However the parameterization of the algorithm is presented here.
        Two parts of the processing require setting parameters: the pre-processing of SPECTRE database (TEC time series from the service SPECTRE) and the detection of TIDs.

1.4.1 *Pre-processing of SPECTRE database*
        SPECTRE service relies on well-proof algorithms to compute VTEC maps on daily basis. In addition to the final "VTEC maps" products the SPECTRE service produces intermediate products which contain biased TEC time series, estimated IFB and TGD, ionosphere pierce point positions, obliquity factor and many other useful quantities.
        In the frame of the SPECTRE pre-processing of the GPS data the detection of cycle slips (*Blewitt*, 1990) and the resolution of phase ambiguities impact the continuity of TEC time series. First, the cycle slips are detected but not corrected which means that a cycle slip without data gap induces a discontinuity in the data. Second, the resolution of the phase ambiguity is based on the computation of the average of a continuous segment of data. However, the SPECTRE service process daily RINEX files so that an additional discontinuity is introduced in a continuous data time series that cover two days. This additional discontinuity looks like a cycle slip.
        Consequently we decided to apply a cycle slip detection and correction algorithm as a preprocessing of the SPECTRE data before applying the TIDs detection method.
        After several tests, we decided to apply a cycle slip correction when a variation of the TEC rate occurs. The variation has to be greater than 0.2 TEC units per minute in comparison to the local trend of the TEC time series. This threshold is quite small in comparison to usual threshold for cycle slips detection. As example the threshold is set to 6 TECu per minute in SPECTRE cycle slip detection algorithm. The difference is because we noticed that the additional cycle slips, induced by merging two continuous data segment with different phase ambiguity resolution, are smaller than common cycle slips.

1.4.2 *Detection of TIDs*
        In the frame of the POPDAT project we decided to detect two types of TIDs: the medium scale TIDs (MSTIDs) and the large scale TIDs (LSTIDs).
        The method of detection has been already described in the report D3.1. The method is the same for the detection of MSTIDs and LSTIDs, we apply only different parameters.
        One may note that the frequency bands used for the TIDs detection are quite large. Moreover we decided to set continuous band from 3 minutes to 2 hours and we split the band around 30 minutes in order to separate the MSTIDs and the LSTIDs. Consequently, the method detects all ionosphere perturbations from highest to lowest frequencies so that the method can be useful for TIDs induced by seismic activity as well as induced by geomagnetic storms or inner atmosphere dynamics.

*Table 1: List of parameters for the detection of MSTIDs and LSTIDs. Note that the
data sampling rate (dt) is 30 seconds.*

Parameter description	Parameter	value
Minimum elevation angle of data used to reseach TIDs (in degree) for multipath mitigation	el_mask	20 degrees
Minimum period of researched MSTIDs (in minutes)	mstid.t1	3 minutes
Maximum period of researched MSTIDs (in minutes)	mstid.t2	30 minutes
Minimum duration of a TEC time series to search MSTIDs in it (in epoch)	mstid.seg_min_nb_ep	(mstid.t2 * 60 / dt) * 3, e.g. 180 epochs or 90 minutes
Threshold for the detection of MSTIDs, relative zero-to-peak amplitude (in pourcents)	mstid.thd_tid_amp	2%
Minimum duration of a valid MSTID (in epoch)	mstid.thd_tid_nep	10 epochs or 5 minutes
Minimum period of researched LSTIDs (in minutes)	lstid.t1	30 minutes
Maximum period of researched LSTIDs (in minutes)	lstid.t2	120 minutes
Minimum duration of a TEC time series to search LSTIDs in it (in epoch)	lstid.seg_min_nb_ep	(lstid.t2 * 60 / dt) * 2, e.g. 480 epochs or 240 minutes
Threshold for the detection of LSTIDs, relative zero-to-peak amplitude (in pourcents)	lstid.thd_tid_amp	6%
Minimum duration of a valid LSTID (in epoch) [use -1 for default value: mean]	lstid.thd_tid_nep	mean([lstid.t1,lstid.t2])*60/dt, e.g. 150 epochs or 75 minutes

1.5 Noise and EMC disturbance elimination

Satellite records are exhibiting noise of quasi monochromatic and broadband character. Their elimination prior to detailed analysis is essential.

Extracting *broadband noise* from the dynamic spectra was one of the tasks in POPDAT project. Detailed description of the method can be obtained from D1.3 "Report on preliminary output of problem-oriented processing" document, Section 1.3.1. and references therein. An automatic code was developed in MATLAB for POPDAT project purposes. The code application was presented in document D1.3, Section 3.5.1. It works on the ASCII data sample from DEMETER satellite ICE experiment. The code is attached as a p-file: **b_noise.p** that can be run on Linux platform with Matlab software (version: MATLAB R2012a). In addition the input file with data sample is provided.

Considering the method of *monochromatic interference* elimination from Dynamic Fourier Spectrum, MATLAB software package was developed and implemented.

The method is based on the application of Butterworth bandstop digital filter. Before run the bandstop filter it is necessary to input to the program the following parameters: noise frequency, filter order and frequency interval for bandstop digital filter. For determination of the monochromatic noise frequency the Discrete Fourier Transform is used. Its gives the possibility for the proper choice of the filter order and frequency range for noise suppression.

The example of the method application and the program code are presented in points 2.5 and 3.5 of this document.

1.6 Electromagnetic effect manifestation of thunderstorms

Method used for presenting electromagnetic effects caused by thunderstorms was presented in the D1.3 in Section 1.3.2. and references therein. In order to identify the thunderstorm manifestation there is a need of ground-based measurements. In pilot-project example presented in D1.3 report, Section 3.5.2 both the ground-based and satellite recordings had been analyzed. In frame of the POPDAT project and as requirements of deliverable presented in this report a p-file code was developed: wf2sp.p. As in previous example code was written in Matlab and can be run on Linux platform with Matlab software (version: MATLAB R2012a). The input data file is provided.

2. Description of the application mode of the computer codes used in POPDAT

2.1 Electromagnetic wave phenomena

Since AWD concept is based on adaptive filtering, the procedure is highly sensible for 'noise' patterns, and handled in the first phase. This procedure section had to be changed according to the difference of on-board and ground based recording characteristics. Most remarkably – as it was demonstrated e.g. in the Fig. 1.3 in the D1.3 report – in the beginning phase the original AWD omits the 'vertical' spectral lines in dynamic spectra which are spherics with great amplitude and not whistlers, therefore these lines in ground based measurements are dominant noises to get rid of. However, in the satellite records the SIWHF phenomena show columnar pattern in VLF spectra above 5-10 kHz. These signals are the principal electromagnetic phenomena measured in situ in the terrestrial plasma environment, related to thunderstorm activity research. Note that local, manmade EMC-noises may also appear in satellite recording, elimination of these signal parts, overlapping short whistler impulses require careful analysis.

Satellite recordings exhibit typically much larger number of whistler occurrences than received on ground. This affects the minimum temporal separation between independently detected events, thus the detection efficiency. New detector settings had to be found and set so that frequent sequences of short path whistlers are possible to detect as separate events.

Further, whistlers detected in space-born datasets have dispersion values falling in a rather large range, due to diverse possible propagation paths. These waves may appear overlapped according to their variety of possible signal dispersions. This condition forced to modify the ground application, where recorded whistlers exhibit similar dispersions in a given period/station. Consecutive detection of different dispersion classes (listed in *Paragraph 1.1*) could be thus performed. In this case all other whistler types represent noise.

As a completely new concept, applied and tested first in satellite recordings in frame of POPDAT database preparation, after one above mentioned detection turn on a given type, values representing the detected events are nullified from the spectra, lowering the noise level and allowing a better detection in its next loop turns. This is best performed if the most dominant type, determined also automatically is processed first. This kind of data processing is running in the present developing, testing phase.

2.2 Atmosphere gravity waves

AGWs selection performs by using a special filter. Its advantages are:
- the small duration in time (or space)
- the narrow transition region
- wide area in spectrum whith an amplification coefficient is equal to one (signal is not distorted in the given frequency range)

Here is a simple example of such a filter in the language IDL. The function takes only two arguments:
- Ls – determines the cutoff frequency cutoff filter
- Lg – determines the width of the transition region of the filter (Partially dependent on the length of the filtered data series and the cutoff frequency, therefore it determined at the

time of its application to specific data series).

During the band-pass filtering the signal is passed through this filter twice to highlight the appropriate band.

The procedure of AGWs catalogues creation is automated only on the reading of source data and on the creation of XML and NetCDF files. AGWs recognizing is performed manually. During this process human has to control a set of parameters which also shows that the wave process is precisely the AGW.

An example of AGW-filter realization in the language IDL:

```
      function mSincCompress, arr, Ls, Lg
  N = N_ELEMENTS(Arr)
  ans = fltarr(N)
  for i = 0, N-1 do begin
   msi = (  sinc(( findgen(N) - i)*2*!PI/Ls )*2/Ls  )*(  (1/(sqrt(!PI)*Lg))*exp( -( findgen(N) - i
)^2 /Lg^2 )  )
   ans[i] = total(arr*msi/total(msi))
   to = total(msi)
  end
  return, ans
end


      function sinc, arr
 ans = fltarr(N_ELEMENTS(arr))
 for i = 0, N_ELEMENTS(arr) - 1 do begin
   ans[i] = sincn(arr[i])
 endfor
 return, ans
end


      function sincn, x
 if (x eq 0) then return, 1
 if (x ne 0) then return, sin(x)/x
end
```

2.3 Traveling ionosphere disturbances (in suit data)

DE2_POPTID_SA.exe FORTRAN routine comprises selection and processing of DE-2 thermal plasma data from IDM/RPA instrument onboard DE-2 satellite. It consists following subroutines:

DATE_AND_TIME – calculates forward/backward UT epoch since given date
RDATA_DE – input data read from intermediate DE-2 data files with already confirmed regions
RUNMED – performs running median with variable length over selected data parts
DRAW_RPA – creates input data summary plot for operator's input control
DRAW_HHTr – creates intermediate graphics of Hilbert-Huang Transformation of input data
HHTR – provides Hilbert-Huang Transformation of selected input data
DCTAB – provides discrete cosine transformation and corresponding spectrum of input data
FSPLH – Spline interpolation/filtering over input data

XML_OUTPUT – creates event XML output files for IWS operational service of end users

POPDAT_TID_NetCDF_transfer.pro is a GDL routine providing transfer of POPTID output data files into NetCDF format for further implementation to POPDAT IWS system

2.4 GPS VTEC analysis of traveling ionosphere disturbances

In this case it was detailed in paragraph 1.4.

2.5 Noise and EMC disturbance elimination

Wide-band electromagnetic noise detection is performed with inspection of the pseudonorm value defined as:

$$\| f \| = \sum_{j,\,kw} w(j)\, d_{jk}^2$$

[2.5.1]

computed for each time instant, where d_{jk} are the wavelet coefficients for the spectrum at a certain time instant and $w(j)$ is a weight factor that depends on scale (the larger scale the larger weight factor). The high value of the pseudonorm relative to the whole time domain (the time interval the dynamic spectrum is computed on) indicates possible time position of the broadband spectrum.   Application developed used as an input ASCII file with VLF power spectrum. Converting raw data and preparing sample of an ASCII input file is performed by external procedures developed and used beyond the POPDAT project. Time resolution for output data is 6.763s and frequency resolution is 19.531 Hz. Calculated and extracted broadband noise spectrum is written to the file with name specified by the user. In addition quicklook image is produced and written to the EPS file (name is defined by the user).

Regarding to monochromatic noises, let us consider as an example the noise suppression method for the wave probe signal in VARIANT experiment (Sich-1M satellite, 22.03.2005, orbit 1363, Bx channel). The fragment of Dynamic Fourier Spectrum of this signal is shown in Fig. 1. Let us select the frequency range from 8500 to 9500 Hz (Fig. 2). As it is seen from Fig. 2, the monochromatic interference harmonics are near 9000 Hz. (Apparently they related to ground-based ELF radio stations). For interference detailing, the Discrete Fourier Spectrum is also built for chosen frequency range (Fig. 3, 4).
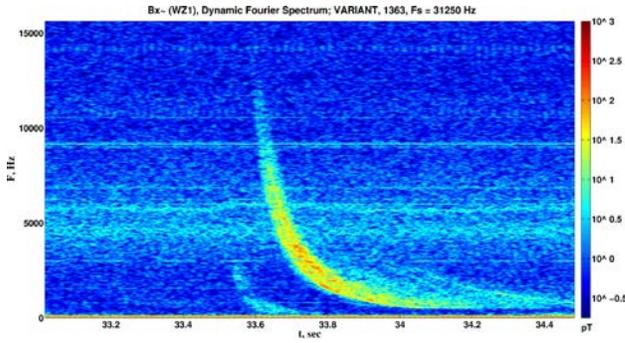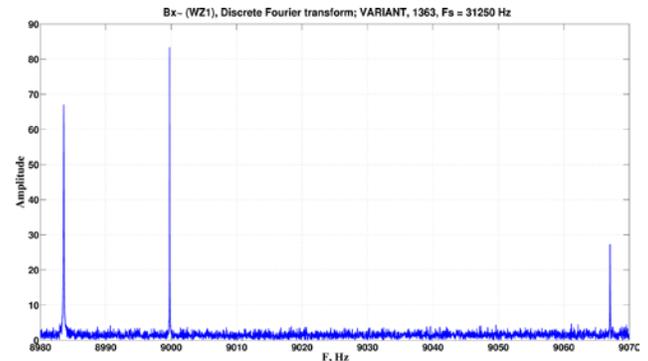
Fig. 1. *Dynamic Fourier Spectrum of the magnetic field sensor signal in full frequency range.*
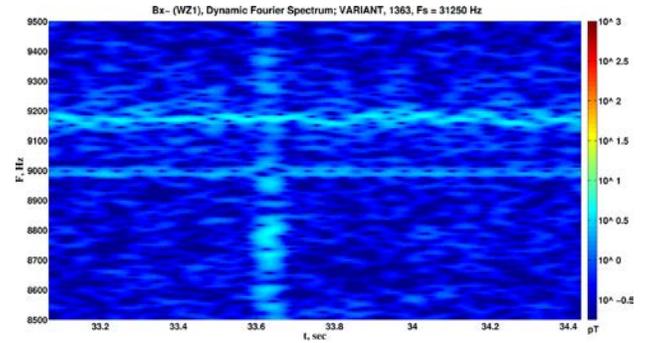
Fig. 2. *Dynamic Fourier Spectrum of the magnetic field sensor signal in the frequency range from 8500 to 9500 Hz.*
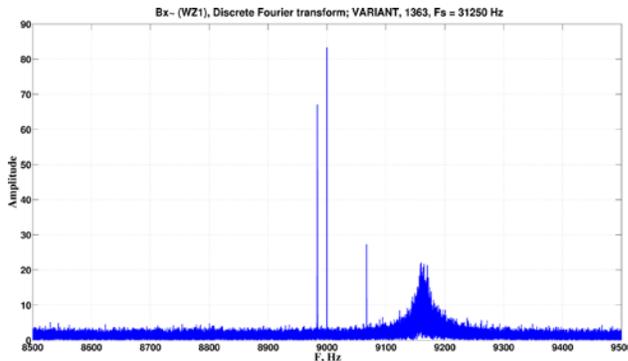
Fig. 3. *Discrete Fourier Spectrum of the magnetic field sensor signal in the frequency range from 8500 to 9500 Hz.*

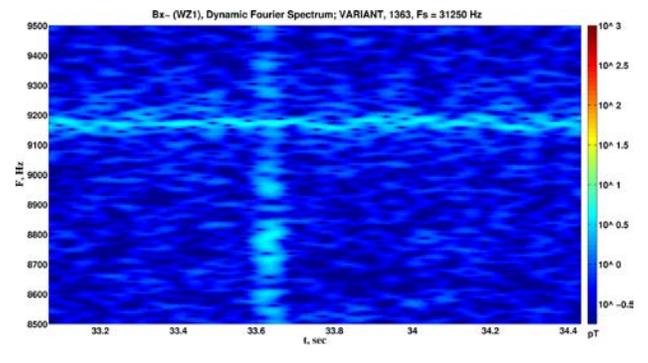**Fig. 4.** *Discrete Fourier Spectrum of the magnetic field sensor signal in the frequency range from 8980 to 9070 Hz.*

Results of the application of the filtration procedure to the chosen three monochromatic interference harmonics are shown in Figs. 5 and 6.
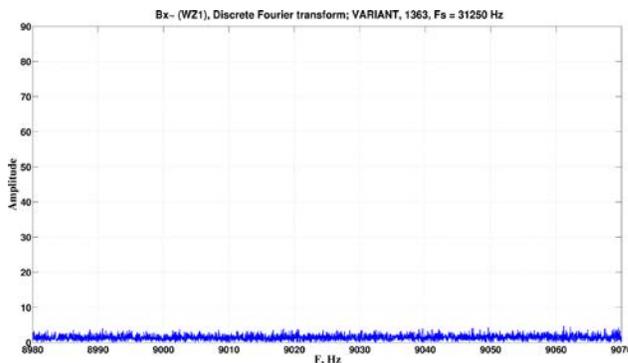
Fig. 5. Discrete Fourier Spectrum after application of the filtration procedure.

Fig. 6. Dynamic Fourier Spectrum after proposed method application.

2.6 Electromagnetic effect manifestation of thunderstorms

Whistlers are electromagnetic waves that can be generated by lightning discharges and observed at VLF range. In order to study thunderstorm manifestation in electromagnetic environment in Earth ionosphere ground-based and satellite measurements must be combined. Correlation of these two methods is being developed in SRC. The pilot project described in D1.3 report, Section 1.3.2 briefly presents this approach. Application provided by SRC takes as an input wave form recordings and returns spectrogram with signals related to particular lightning event with use of FFT algorithm. In POPDAT purposes application uses only one electric field component (E12 from DEMETER VLF data APID 1131 in particular example) and as an output produces spectrogram with time resolution 0.315 s and frequency resolution 115 Hz. Application produces a quicklook image and write the image to the EPS file (name is defined by the user).

Converting raw data and preparing sample of an ASCII input file is performed by external procedures developed and used beyond the POPDAT project.

Note: The raw DEMETER data are a property of LPCE-CNRS Orleans. Therefore, in order to obtain access to the DEMETER repository, users are encouraged to visit the website: http://demeter.cnrs-orleans.fr/

3. Appendices: Computer codes in executable form

Computer codes, added to this report in electronic form are stand-alone routines. They are executable programs, performing the same data processing on a limited set of input satellite data as during POPDAT database building procedures is done. Executing these routines demonstrate the capabilities of applied procedures, the yielded outputs conform to file descriptions given e.g. in D1.3 report.

3.1 Electromagnetic wave phenomena

Electromagnetic waves, detected in satellite recordings are restricted to four, well distinguishable classes, listed in *Section 1*. Each wave type is detected using the same core algorithm with specific settings. Detection af all listed phenomena is included in the provided software. Sample satellite recordings are raw, time domain VLF waveform data sections. EM wave detection has been implemented in matlab environment. Code execution requiers installed matlab with Signal Processing included. No specific requirement for OS/platform.

Executable is:
    AWDSat_POPDAT.p

Raw satellite datasets are:
    sampleVLFwave1.DAT, sampleVLFwave2.DAT

Usage, in matlab command prompt e.g.:
    AWDSat_POPDAT sampleVLFwave1.DAT

Sample data and provided matlab p-code should be copied to the same folder. Ascii output file is written during program execution, containing identifiers of the detected events. (Make sure that there is write permisson on the folder.) This output is converted to netCDF as IWS database input.

3.2 Atmosphere gravity waves

        The AGW extraction method has been implemented in MATLAB. The following files are needed to extraction AGW in your work directory:
    1. make_AGW_NetCDF.m (main executable file)
    2. FindRange.m (internal function for main file)
    3. source DE-2 file (for example: 1982332T163140_0_DE2_NACS_1S_V01.ASC)

Execution: type *"make_AGW_NetCDF"* in MATLAB command window.

3.3 Traveling ionosphere disturbances (in situ data)

DE2_POPTID_SA.exe
POPDAT_TID_NetCDF_transfer.pro

3.4 GPS VTEC analysis of traveling ionosphere disturbances

GPS VTEC input data, TI2D analyzer software package and the related data set is attached to this report. Due to the large file size users may download and uncompress before the application. The complete package can be obtained from:
www.noveltis.com/spectre/popdat/TI2D-v1.1_pcode.zip

3.4.1 *TI2D function description*

The TID Detection method has been implemented as a Matlab function: ti2d().This function read intermediate SPECTRE product in format MAT and create the POPDAT data (NetCDF data and XML metadata) dedicated to large scale and medium scale TIDs (catalogue TID-TEC).

The input of the function is a text file describing all requested inputs parameters. An example is provided in the TI2D package and in the following table.

Table 2: Example of text file containing input parameters of the ti2d function.

C:\POPDAT\TID-TEC\data\input	# Input data path
C:\POPDAT\TID-TEC\data\ouput	# Output data path
Default	# netcdf toolbox path (use "Default" to use default Matlab netcdf functions)
Default	# mexcdf toolbox path (use "Default" to use default Matlab netcdf functions)
C:\POPDAT\TID-TEC\src\convdate	# convdate toolbox path (use convdate package included in code source)
C:\POPDAT\TID-TEC\src\utc	# utc toolbox path (use utc package included in code source)
30	# Data sampling rate period (in seconds)
10	# Number of epoch used to compute smooth derived data
100	# Number of epoch without cycle slip at the beginning and the end of a segment (induced by SPECTRE algorithm)
0.1	# Threshold for cycle slip detection between two epochs (in TECu)
9.969209968386869e+036	# Default value for empty data in NetCDF/Mat input file TL (this value appears also in .mat file created from NetCDF files TL)
20	# Minimum elevation angle of data used to reseach TIDs (in degree) for multipath mitigation
03	# Minimum period of researched MSTIDs (in minutes)
30	# Maximum period of researched MSTIDs (in minutes)
-1	# Minimum duration of a TEC time series to search MSTIDs in it (in epoch) [use -1 for default

	parameter: (mstid.t2 * 60 / dt) * 3]
2	# Threshold for the detection of MSTIDs, relative zero-to-peak amplitude (in pourcents)
10	# Minimum duration of a valid MSTID (in epoch) [use -1 for default value: mean([mstid.t1,mstid.t2])*60/dt]
30	# Minimum period of researched LSTIDs (in minutes)
120	# Maximum period of researched LSTIDs (in minutes)
-1	# Minimum duration of a TEC time series to search LSTIDs in it (in epoch) [use -1 for default parameter: (lstid.t2 * 60 / dt) * 2]
6	# Threshold for the detection of LSTIDs, relative zero-to-peak amplitude (in pourcents)
-1	# Minimum duration of a valid LSTID (in epoch) [use -1 for default value: mean([lstid.t1,lstid.t2])*60/dt]
2005	# Year for the strating date
01	# Month for the starting date
18	# Day for the starting date
2005	# Year for the ending date
01	# Month for the ending date
19	# Day for the ending date
1	# Verbose mode on (1) or off (0)
0	# Display mode on (1) or off (0)
1	# Overwrite output mode on (1) or off (0)
ACOR	# Station name used for the data processing (use "ALL" or "All" or "all" to process all station)
15	# Satellite number used for the data processing (use -1 to process all satellites)

An example of data input (SPECTRE intermediate product in .MAT format) is also provided in the TI2D package.

WARNING: it is important to note that the sub-directories structure of the input data directory path have to respect the following format: [Input data directory path]/[yyyy]/[mm] . For example, the input data (SPECTRE TL product in MAT format) of January 2005 have to be in the directory C:/POPDAT/TID-TEC/data/input/2005/01/. The TI2D package provides an example.

The outputs of the function are several files in NetCDF and XML format respecting the data and metadata structures described in the deliveries "D2.2 Accepted files structure for each individual data catalogue" and "D3.4 Metadata Description Report".

3.4.2 *How to use the TI2D function in Matlab?*
   Requirements:
   - Matlab software version 7.12.0.635 (R2011a) or newer with the Signal Processing toolbox;
   - Matlab Java engine on (default). Do not use the option –nojvm to execute Matlab;
   - Operating System: Windows Seven Professional (or compatible) / Debian GNU/Linux 6.0 (or compatible).

Installation:
- Unzip the TI2D package in your work directory.

Execution:
- In Matlab command window: add your work directory (where you installed TI2D package) to the Matlab path (with sub-directory). For example: addapth('MyWorkDirectory');
- In Matlab command window: call the TI2D function with a parameter file as function parameter. For example: ti2d('input_param.txt').

If you don't activate verbose mode and display mode (in the parameter file) the output displayed in the Matlab command window looks like:

```
#######################################################
Travelling Ionosphere Distrubances Detector (TI2D)
     Version: TI2D-v1.1
     Date:August 2012
     Author: NOVELTIS
     Contact: contact@noveltis.fr
#######################################################
 # Read input parameters...
 # Toolbox loading...
 # Parameters checking...
 # First processed date: 2010/01/01
 # Last processed date : 2010/03/31
#######################################################
     Overall progress: 01%
     Processing date: 2010/01/01
#######################################################
 # Loading input data...
 # Processing...
#######################################################
     Overall progress: 02%
     Processing date: 2010/01/02
#######################################################
 # Loading input data...
 # Processing...
#######################################################
     Overall progress: 03%
     Processing date: 2010/01/03
#######################################################
```

The verbose and display mode have been design for software developers and debugging purpose. They provide additional information through the Matlab command window and several Matlab figures. The activation of the display mode should be restricted to the one station / one satellite case otherwise a huge number of figures can pop-up and leads to a computer freeze or crash.

3.5 Noise and EMC disturbance elimination

Routine for *broadband noise* extraction takes as an input prepared dataset (from DEMETER satellite)-sample provided. As an output routine gives an image of an extracted noise and a data file in ASCII format. Input data file is also provide in ASCII format.

Executable file:
      b_noise.p
Input dataset:
      widmo.dat
Matlab command line:
      >> b_noise
User will be asked to provide:
   - input file name (*widmo.dat*)
   - output file name (defined by user)
   - eps image file name (defined by user).

Description of attached MATLAB program code for *monochromatic noise* elimination is given below:
Input Values:

D - signal;
Fs - sampling rate;
fnoise - noise frequency (Hz);
flim - frequency min/max (Hz) for bandstop digital filter;
order - filter order.

Output Values:

D - clear signal.

Program code:

```
w1 = fnoise - flim; w2 = fnoise + flim;
Wn = [w1 w2]/(Fs/2);
[a,b] = butter(order,Wn,'stop');
D = filtfilt(a,b,D).
```

3.6 Electromagnetic effect manifestation of thunderstorms

Routine provided in this subsection takes as an input previously prepared ASCII data file data contains a waveform in VLF range. As an output produces ASCII file with spectrogram and an image that presents whistler caused by lightning.

Executable file:
     wf2sp.p
Input dataset:
     wf.dat
Matlab command line:
     >> wf2sp
User will be asked to provide:
    4. input file name *(wf.dat)*
    5. output file name (defined by user)
    6. eps image file name (defined by user).