



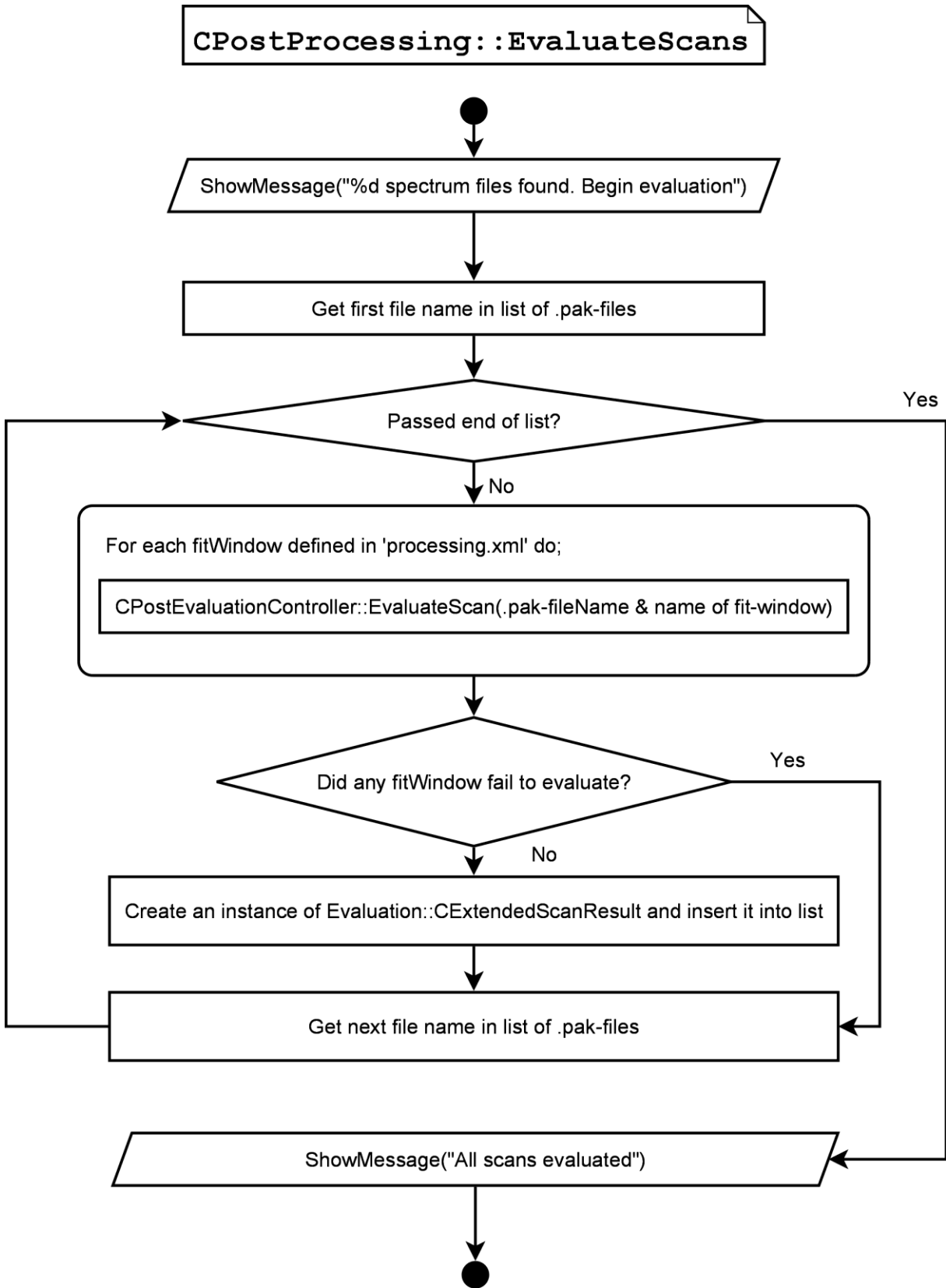
NOVACPPP User Manual

2024-11-15

**Optical Remote Sensing Group,
Department of Radio and Space Science,
Chalmers University of Technology,
Göteborg, Sweden**

Table of Contents

1. OVERVIEW	5
2. SOFTWARE REQUIREMENT	6
3. INITIAL SETUP.....	7
4. CONFIGURATION.....	8
<i>CONFIGURING THE LOCATIONS – SETUP.XML.....</i>	<i>9</i>
<i>CONFIGURING THE EVALUATION – THE .EXML – FILES.....</i>	<i>12</i>
<i>CONFIGURING THE PROCESSING – PROCESSING.XML.....</i>	<i>19</i>
<i>COMMAND LINE ARGUMENTS</i>	<i>27</i>
<i>GIVING THE WIND FIELD – THE .WXML – FILES.....</i>	<i>29</i>
<i>GIVING THE PLUME ALTITUDE</i>	<i>32</i>
5. RUNNING – FLUX MODE.....	33
GENERATED FILES	33
WHAT PARAMETERS ARE USED TO CALCULATE THE FLUXES	34
6. RUNNING – CALIBRATION MODE.....	35
STANDARD CROSS SECTIONS.....	35
GENERATED FILES	36
7. ALGORITHMS – HOW THINGS ARE REALLY DONE.....	37
DARK CURRENT CORRECTION	37
DOAS FIT.....	37
WAVELENGTH CALIBRATION USING A SOLAR SPECTRUM	37
SCAN PROPERTIES.....	38
GEOMETRY CALCULATIONS.....	40
DUAL-BEAM WIND SPEED	43
RETRIEVING THE WIND FIELD AT A GIVEN TIME AND LOCATION	44
8. SOFTWARE DETAILS – A PROGRAMMER’S REFERENCE	45
POST PROCESSING OF FLUXES.....	45



..... 47

POST PROCESSING OF COMPOSITION MEASUREMENTS..... 48

1. OVERVIEW

This program is designed for batch processing of spectra collected in the NOVAC project.

In the NovacPPP, the following conventions are used.

- The word **column** is used for the integrated concentration (of a certain species) along the path that the light has passed before being collected by the instrument. This can be a vertical column density or a slant column density.
- All fluxes are given in **kg/s**.
- All angles (cone angle, compass direction, etc.) are given in **degrees**.
- Wind speeds are always given in **m/s**.
- Wind directions are always defined as the direction from which the wind comes, in **degrees** from north with positive angles clockwise.
- All molecular absorption cross sections are assumed to be in **cm²/molecule**.
- All molecular columns are given in **molecules/cm²**.
- Dates are given in the format **yyyy.mm.dd**.
- Latitudes and longitudes are given in the format **dd.dddddd**.
- Altitudes are in **meters above sea level**.
- An **instrument** is identified by the serial number of its spectrometer. If the spectrometer is replaced, then this should in the configuration be treated as the instrument was removed and a new instrument was installed on the same location. If any other component of an instrument is replaced then this is treated in the configuration as a new location of the same instrument.
- A **channel** (or channel number) is an identifier for spectrometers with the same serial-number. This is only relevant for the SD2000 – series spectrometers, where two spectrometers can share one serial number. In these instruments are the two spectrometers identified as the Master channel (number 0) and the Slave channel (number 1). For all other spectrometer models is the channel number always zero.
- A **fit window** is a set of settings for performing a DOAS fit. I.e. this contains a range of pixels to use, a number of reference files, parameters for order of polynomial to include in the fit etc. Each fit-window is identified by its name and the spectrometer it operates on.

2. SOFTWARE REQUIREMENT

The program does not need any other driver programs. The platform requirements are as follows:

- CPU - Pentium IV, 1 GHz or higher
- Memory - 512 MB or more
- Operating system - Windows XP or Windows Vista
- Screen Resolution - at least 800 by 600 pixels.

These are minimum requirements; a higher performing computer will result in shorter processing time. A multi-core processor is recommended for optimum performance, NovacPPP is optimized for multi-threading.

3. INITIAL SETUP

The software is setup by copying the executable (NovacPPP_YYYY_MM_DD.exe, where YYYY_MM_DD corresponds to a date) to a folder on the local computer. A sub-folder with the name 'configuration' must be created and should contain the files necessary for the configuration (see section 4). No further installation procedure should be necessary.

Make sure that there is enough free space on the source disk to store the results of the evaluation. If data is to be downloaded from the Novac data server, a fast Internet connection is required.

4. CONFIGURATION

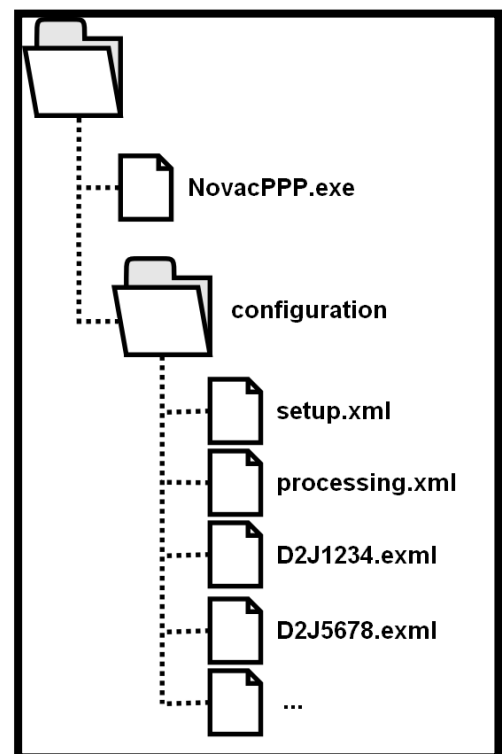
The configuration used for the NovacPPP is found in the directory **configuration** which must be located in the same directory as the executable (NovacPPP.exe).

The **configuration** directory must contain one file named **setup.xml**, in which the location of each instrument is found at each specific instance. Each instrument can be located at arbitrarily many locations, however they must be disjoint in time. The file can contain as many instruments and as many positions as you like. **N.B.** An instrument is identified by the serial number of its spectrometer. If the spectrometer is replaced, then this should in the configuration be treated as the instrument was removed and a new instrument was installed on the same location. If any other component of an instrument is replaced then this is treated in the configuration as a new location of the same instrument.

For each instrument in the **setup.xml** file, there must be one file with the name **SERIAL.xml** in the **configuration** directory which defines the settings for the DOAS evaluation of spectra from the spectrometer identified by the serial-number SERIAL. I.e. if you have defined an instrument with the serial-number *D2J1234* in **setup.xml** then there must be a file with the name **D2J1234.xml** in the configuration-directory. Each .xml file can contain an arbitrary number of fit windows, however these must differ in one of the following; 1) the time period they are valid 2) the channel number they operate on or 3) their name. The .xml file also contains the settings for how to correct the measured spectra for dark counts. This is specified as a special section, separate from the fit windows.

The **configuration** directory must contain one file named **processing.xml**, in which the settings for the processing are given. This file specifies which fit-windows to use, and the values of the parameters that can be selected by the user.

When the processing in NovacPPP is begun all the parsed files from the configuration-directory are copied to a sub-folder (copiedConfiguration) the output-directory. The purpose with this is to make it possible to later inspect the parameters used with each run of the program. The file **processing.xml** that is written to the copiedConfiguration directory contains the values of all parameters that are possible to specify by the user, not only the ones that were specified in the original **processing.xml**.



CONFIGURING THE LOCATIONS – SETUP.XML

The file setup.xml contains the configuration of the installed instruments. Only data from instruments that are described in setup.xml can be evaluated.

The file contains the following items;

- **instrument** – this is a section describing the setup of one instrument. The instrument is identified by the serial number of the spectrometer in it. The instrument section contains several sub-sections further describing the instrument.

<INSTRUMENT> SECTION

- **serial** – the serial number of the instrument, used to identify the instrument. If the spectrometer of an instrument has been replaced at any time, then two <instrument> sections must be written for this instrument.
- **location** – a description of a site where the instrument has been located for some time. If any component of the instrument (other than the spectrometer) has been changed then two separate <location> sections must be defined for this instrument. An arbitrary number of locations can be described for each instrument, however for correct analysis of the data it is important that there is no overlap between the times for which they are valid.

<LOCATION> SECTION

- **name** – The name of the site where the instrument was located. This is defined by the user and supplied to make the interpretation of data simpler for the user.
- **time_from** – The date from which this description of a location is valid. This is typically the day when the installation was made, or the instrument was changed. This must be a date in the format yyyy.mm.dd or a functional expression, such as TODAY(), corresponding to today's date, or TODAY(-7) corresponding to seven days ago.
- **time_to** – The date until which this description of a location is valid. This must be a date in the format yyyy.mm.dd. By default, this is 9999.12.31 – which corresponds to that the instrument is still there. If you wish to make sure that no spectra with dates in the future (could happen if the GPS connection is lost) are processed then this can be set to the date today.
- **latitude** – the latitude of the location, in degrees and decimal degrees.
- **longitude** – the longitude of the location, in degrees and decimal degrees.
- **altitude** – the altitude of the location, in meters above sea level.
- **volcano** – the name of the volcano that the instrument monitored during the installation at this location. Should be written without accents (á, à, é, ...) and by replacing spaces with underscores (_).

- ***compass*** – the compass direction of the instrument during this installation period. Only relevant for version I instruments. This is in degrees from north, counting positive angles clockwise.
- ***coneangle*** – the cone angle of the instrument during this installation period. Only relevant for version I instruments.
- ***type*** – the type of instrument. This is an integer describing the type of instrument, 0 corresponds to Version I instrument, 1 corresponds to Version II instrument.
- ***spec*** – string describing the model of the spectrometer. This is by default “S2000”. This is used to identify the maximum possible intensity from the spectrometer and hence also determining the saturation ratio of any measured spectrum. Currently supported spectrometer models are *S2000*, *USB2000*, *USB2000+*, *HR2000*, *HR4000*, *QE65000*, *MayaPro*, *AvaSpec*, and *Flame*.

EXAMPLE OF SETUP.XML

The following is an example of a simple setup.xml file, with only one instrument but where the scanner has at one point been changed from flat to a conical;

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- This is the global configuration file for the evaluation of spectra
      in the NOVAC Post Processing Program -->

<NovacPPPConfiguration>
  <outputdirectory>C:\Novac\Output\</outputdirectory>
  <tempdirectory>C:\Novac\Temp\</tempdirectory>
  <instrument>
    <serial>D2J2142</serial>
    <location>
      <name>Chipiquixtle</name>
      <time_from>2005.01.25</time_from>
      <time_to>2007.12.31</time_to>
      <latitude>19.02</latitude>
      <longitude>-98</longitude>
      <altitude>4000</altitude>
      <volcano>Popocatepetl</volcano>
      <compass>180.0</compass>
      <coneangle>90.0</coneangle>
      <type>0</type>
      <tilt>0.0</tilt>
      <spec>S2000</spec>
    </location>
    <location>
      <name>Chipiquixtle</name>
      <time_from>2008.01.01</time_from>
      <time_to>9999.12.31</time_to>
      <latitude>19.02</latitude>
      <longitude>-98</longitude>
      <altitude>4000</altitude>
      <volcano>Popocatepetl</volcano>
      <compass>180.0</compass>
      <coneangle>60.0</coneangle>
      <type>0</type>
      <tilt>0.0</tilt>
      <spec>S2000</spec>
    </location>
  </instrument>
</NovacPPPConfiguration>
```

CONFIGURING THE EVALUATION – THE .EXML – FILES

For each instrument that is described in `setup.xml`, there **must** be one file describing how to evaluate the spectra from that spectrometer. The name of this file must be the serial-number of the spectrometer + the file ending “.xml”. This file specifies a number of so-called **fit windows** (see the introduction) each one of these include a set of references (i.e. trace gas cross sections), a set of pixels to use for the fitting and some other options for how the fitting should be done. Each .xml file can include arbitrarily many fit windows; however, they must differ in either name, channel they operate on or time they are valid for!

E.g. if an installed instrument has been re-calibrated at a specific time, then one may wish to use a set reference files for all spectra collected before the time of the re-calibration and one set of reference files for all spectra after the time of the re-calibration. The .xml file can then specify two fit-windows, one which is valid from the time of the installation to the day the re-calibration was made and one fit-window which is valid from the day of the re-calibration until the instrument was removed. These two fit windows should in this case have the same name to specify that they do evaluate for the same thing.

Collected spectra can also be evaluated in several wavelength regions to extract further information on e.g. the radiative transfer at the collection time. If several fit windows are valid at a given time, their names are used to differentiate between them. To evaluate each spectrum in two different wavelength regions, the .xml file must contain two fit-windows, both valid at the time when the spectrum was collected, but with different names. An unlimited number of fit-windows can be specified in the .xml file, which ones will be used are specified by the file `processing.xml` (see below).

NOTE: It is very important that there is only one fit-window with a specific name valid at each given time.

The first line in the .xml file is the serial-number of the instrument for which this configuration is valid. Then follows a series of fit windows each describing how the spectra from this spectrometer should be evaluated.

The .xml file also contains the settings for how to correct the measured spectra for dark counts. This is given as a separate section named `<darkCurrent>`.

For performing instrument calibrations (as in section 6) there is also a separate section in the .xml file named `<Calibration>` where an initial wavelength calibration and, optionally, an instrument line shape file are defined. Notice that the `<Calibration>` section is different from the `<wavelengthCalibration>` section; the `<Calibration>` section defines the initial properties of the instrument and is defined once in the file, the `<wavelengthCalibration>` section defines one Fraunhofer reference *per fit window* and is used to determine a shift between the references and the measured spectrum.

<FITWINDOW> - SECTION

The fit window section contains the following items;

- **channel** – The channel in the spectrometer that this fit-window is valid for. This can be omitted and is then default zero. Normally this is equal to 0, however the SD2000 – spectrometers contain two spectrometers with the same serial-number, the two are only separated by their channel number.
- **name** – The name of the fit-window, this is supplied to make it easier for the user to interpret the output.
- **validFrom** – The date from which these parameters are valid from. Must be a date in the format yyyy.mm.dd. This must be a date in the format yyyy.mm.dd or a functional expression, such as TODAY(), corresponding to today's date, or TODAY(-7) corresponding to seven days ago.
- **validTo** – The date until which these parameters are valid. Must be a date in the format yyyy.mm.dd or a functional expression. By default, this is 9999.12.31 – which corresponds to that the instrument is still there. If you wish to make sure that no spectra with dates in the future (could happen if the GPS connection is lost) are processed then this can be set to the date today.
- **specLength** – The length of the spectra, in pixels.
- **interlaceSteps** – Normally equal to 0. This can be omitted and is then default to zero. However, in some cases can spectra be read out in an interlaced fashion, e.g. during wind-measurements with the SD2000 spectrometers, and this parameter would then be equal to 1.
- **polyOrder** – The order of the polynomial to include in the DOAS fit.
- **fitType** – the type DOAS fit to perform. Available options are
 - 0 – spectra will be divided by the sky-spectrum and then high-pass filtered (using a binomial high-pass filtering with 500 iterations). This is the default value and corresponds to the type of fit done in the NovacProgram.
 - 1 – spectra will be high-pass filtered (using a binomial high-pass filtering with 500 iterations), the sky-spectrum is also high-pass filtered and included in the DOAS fit. This makes it possible to compensate for shifts between measured and sky spectrum.
 - 2 – only the polynomial is used to remove the slowly varying component of the spectra.
- **fitLow** – the lower edge of the range of pixels that are used to make the DOAS fit.
- **fitHigh** – the upper edge of the range of pixels that are used to make the DOAS fit.
- **offsetFrom** – The first pixel to use for removing the electronic offset from the spectra. This defaults to 50, which is a suitable option for an USB2000/S2000 device starting at ~290nm.

- ***offsetTo*** – The last pixel to use for removing the electronic offset from the spectra. This defaults to 200, which is a suitable option for an USB2000/S2000 device starting at ~290nm.
- ***reference*** – The reference section contains a description of each of the cross-sections included in the DOAS fit. See below...

<REFERENCE> - SECTION

The reference-section describes the properties of the cross-sections included in the DOAS fit.

Notice that the cross sections should not be filtered, any necessary filtering of the reference-files is done inside the NovacPPP!

This section contains the following items;

- **name** – the name of the reference file. This is used to identify the reference in the program – thus it is necessary to format the name according to the list of identifiable species.
- **path** – The full path to the file containing the cross-section data. Must be on the local computer and must be readable. The file must contain a cross section data with the same length as the size of the spectra.
- **shiftOption** – The option for the shift. Available values are;
 - 0 – Free shift.
 - 1 – Fixed to the value specified by 'shiftValue'
 - 2 – Linked to the reference number identified by 'shiftValue' (e.g. if shiftValue is equal to 0, then this reference will be shifted the same amount as the first reference in the list).
 - 3 – Limited; NOT IMPLEMENTED YET!!!
- **shiftValue** – the value of the shift, see 'shiftOption' for how to interpret this.
- **squeezeOption** – The option for the squeeze. Value interpreted in the same way as for 'shiftOption'.
- **squeezeValue** – The value of the squeeze, see shiftOption for how to interpret this.
- **columnOption** – The option for the column, interpreted with the same values as 'shiftOption'. **THIS SHOULD ALWAYS BE 0 UNLESS YOU REALLY KNOW WHAT YOU ARE DOING!**
- **columnValue** – The value of the column – this should normally not be defined...

<WAVELENGTHCALIBRATION> - SECTION

The wavelengthCalibration – section contain an option for how the best shift of the references against the measured spectra can be calibrated before the DOAS fit is performed (see section 5). Notice that this is not the same as performing an instrument calibration. There is only one element which can be configured in this section;

- **fraunhoferSpec** – The path to a solar spectrum which should already been convolved with the instrument line shape and sampled on the same wavelength grid as all the references. If defined, then this will be used to make a pre-fit of one spectrum in the scan where the shift between the measured spectrum and the references is determined. This shift is then applied to the references for all spectra in the same scan.

<DARKCORRECTION> - SECTION

The darkCorrection settings are separate from the settings in the fit-window (for a description of dark current correction see section 0). This section contains the following items;

- **validFrom** – The date from which these parameters are valid from. Must be a date in the format yyyy.mm.dd. This must be a date in the format yyyy.mm.dd or a functional expression, such as TODAY(), corresponding to today's date, or TODAY(-7) corresponding to seven days ago.
- **validTo** – The date until which these parameters are valid. Must be a date in the format yyyy.mm.dd or a functional expression. By default, this is 9999.12.31 – which corresponds to that the instrument is still there. If you wish to make sure that no spectra with dates in the future (could happen if the GPS connection is lost) are processed then this can be set to the date today.
- **dark** – the option for how to correct for the dark, must be one of the following strings;
 - **SCAN** – this is the default value. This means that the measured spectra will be corrected for dark by subtracting one (dark) spectrum with the same exposure time as the measured spectrum and that this spectrum can be found in the same scan.
 - **MODEL** – this means that the measured spectra will be corrected for dark by modeling a dark spectrum from a dark-current and an offset spectrum. These two can then either be given by the user or found in the same scan, depending on the settings on 'darkCurrent' and 'offset'
 - **USER** – this means that the measured spectra will be corrected for dark by modeling a dark spectrum from a user supplied dark current spectrum and a user supplied offset spectrum.
- **darkCurrent** – this is the option for how to retrieve the dark current spectrum, this is only used if the parameter 'dark' is MODEL. Can be one of the following
 - **SCAN** – means that the dark current spectrum is found in the scan.
 - **USER** – means that the dark current spectrum is specified by the user. The parameter 'darkCurrentSpec' must then contain the path to the dark current spectrum.
- **darkCurrentSpec** – The path to the dark current spectrum that should be used. Only used if the parameter 'darkCurrent' is USER or if 'dark' is USER.
- **offset** – this is the option for how to retrieve the offset spectrum, this is only used if the parameter 'dark' is MODEL. Can be one of the following
 - **SCAN** – means that the offset spectrum is found in the scan.
 - **USER** – means that the offset spectrum is specified by the user. The parameter 'offsetSpec' must then contain the path to the offset spectrum.
- **offsetSpec** – The path to the offset spectrum that should be used. Only used if the parameter 'offset' is USER or if 'dark' is USER.

<CALIBRATION> - SECTION

The <Calibration> section defines two files, one initial wavelength calibration and one initial instrument line shape file. This section is only used in calibration mode (see section 6) and are ignored in flux mode processing.

The calibration section contains the following items;

- ***initialCalibrationFile*** – the full filename and path to a file which contains the initial wavelength calibration (pixel-to-wavelength mapping). This file is mandatory in calibration mode. There are two options for the format in this file:
 - A text file containing one or two columns where the first (or the only) column defines the wavelength (in nanometers air) for each pixel on the spectrometer detector. An old reference file used for evaluation can be used if this contains two columns of data.
 - An Extended STD file saved from performing a instrument calibration in MobileDoas or NovacProgam such that the file contains both a wavelength calibration and an instrument line shape. If *initialCalibrationFile* is an extended STD file and an instrument line shape can be found in the file, then the *initialInstrumentLineshapeFile* will be ignored if it is defined.
- ***initialInstrumentLineshapeFile*** – This must be a text file containing an instrument line shape description in the .SLF format used by QDOAS, i.e. containing two columns of data where the first column is a differential wavelength and the second is the instrument line shape This file is optional in calibration mode.

EXAMPLE OF .EXML FILE

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- This is the configuration file for the evaluation of spectra in the
      NOVAC Post Processing Program -->
<EvaluationConfiguration>
  <serial>D2J2142</serial>
  <fitWindow>
    <channel>0</channel>
    <name>SO2</name>
    <validFrom>1970.01.01</validFrom>
    <validTo>9999.03.13</validTo>
    <specLength>2048</specLength>
    <interlaceStep>1</interlaceStep>
    <polyOrder>5</polyOrder>
    <fitType>0</fitType>
    <fitLow>320</fitLow>
    <fitHigh>460</fitHigh>
    <Reference>
      <name>SO2</name>
      <path>C:\References\D2J2142_SO2_294K_Master.xls</path>
      <shiftOption>1</shiftOption>
      <shiftValue>0.000000</shiftValue>
      <squeezeOption>1</squeezeOption>
      <squeezeValue>1.000000</squeezeValue>
      <columnOption>0</columnOption>
    </Reference>
    <Reference>
      <name>O3</name>
      <path>C:\References\D2J2142_O3_223K_Master.xls</path>
      <shiftOption>1</shiftOption>
      <shiftValue>0.000000</shiftValue>
      <squeezeOption>1</squeezeOption>
      <squeezeValue>1.000000</squeezeValue>
      <columnOption>0</columnOption>
    </Reference>
    <Reference>
      <name>RING</name>
      <path>C:\References\D2J2142_Ring_Master.xls</path>
      <shiftOption>1</shiftOption>
      <shiftValue>0.000000</shiftValue>
      <squeezeOption>1</squeezeOption>
      <squeezeValue>1.000000</squeezeValue>
      <columnOption>0</columnOption>
    </Reference>
  </fitWindow>
  <DarkCorrection>
    <dark>SCAN</dark>
  </DarkCorrection>
</EvaluationConfiguration>

```

CONFIGURING THE PROCESSING – PROCESSING.XML

The file `processing.xml` is used to give the options for how the processing should be done.

E.g. is this file used to determine which fit-windows that should be used to evaluate the scans. The fit-window specified as the 'main' fit-window is the one which will be used to calculate the fluxes, to calculate the plume speed or the geometries.

Each parameter that is possible to specify has a default value which will be used if nothing is specified in `processing.xml`. When the processing is done will a file called `processing.xml` be created in the 'copiedConfiguration' directory in the output directory. This file contains the value of all the parameters, both the ones specified in the original `processing.xml` file and the ones that were taken as default values.

The file contains the following items (here sorted by category);

GENERAL SETTINGS		
	MaxThreadNum	The maximum number of threads that will be started by the software during the processing. This should be \leq the number of processors (or cores) on the computer for maximum performance (but at least 1). Default: 2
	StartNow	Deprecated: If non-zero then the processing will start immediately without waiting for the user to press the button. Useful for scripting. This option was removed when the program was converted to a console only application, the processing will always start immediately when the application starts.

OUTPUT FILES		
	tempdirectory	The directory that can be used for temporary files (on the local computer). Default: {empty}
	outputdirectory	The directory where the processed files and evaluation results should be stored (on the local computer). Default: {empty}

PROCESSING MODE		
	mode	The processing mode, typically flux. Must be one of the defined processing modes: FLUX, CALIBRATION, COMPOSITION (not fully implemented) or STRATOSPHERE (not fully implemented) Default: FLUX
	molecule	The molecule of main interest, typically this is SO ₂ . Must be one of the pre-defined molecules. Default: SO ₂

THE VOLCANO		
	Volcano	The volcano that the processing is to handle. This is either the full name, the simplified name or the Si-code of the volcano. Default: {empty}

DATA TIME RANGE		
	FromDate	The first day that we should look for data (inclusive). Default: 2005.10.01
	ToDate	The last day that we should look for data (inclusive). Default: Today's date

LOCATION OF THE SPECTRUM FILES		
	LocalDirectory	The directory on the local computer where we should look for spectrum files to process. Notice that it is possible to search for data both on the local computer and on the FTP server. Default: {empty}
	IncludeSubDirs_Local	This is 'true' if we should include sub-directories to 'LocalDirectory' in our search for data. Default: 1 Possible values: 0, 1, false, true
	FilenamePatternMatching_Local	This is non-zero if we should only include .pak files with filenames containing a serial number of a configured device and a date and time in the processing time range. If this is zero (false) then all .pak files in the scanned directories are included. Default: 1 Possible values: 0, 1, false, true
	FTPDiretory	The full path to a directory on a FTP - server where we should scan for data files. Notice that it is possible to search for data both on the local computer and on the FTP server. Default: {empty}
	IncludeSubDirs_FTP	This is non-zero if we should include sub-directories to 'FTPDiretory' in our search for data Default: 1
	FTPUsername	The username to log in to the FTP-server. Default: Log in to the Novac FTP-server
	FTPPassword	The password to log in to the FTP-server Default: Log in to the Novac FTP-server

WHAT TO DO WITH THE PROCESSED RESULTS		
	UploadResults	<p>This is true if we should upload the results (FluxLogs etc) to the Novac FTP server.</p> <p>Default: 0</p> <p>Possible values: 0, 1, false, true</p>

THE WIND FIELD		
	WindFieldFile	<p>The full path and filename of the wind field file OR</p> <p>The directory where to search for wind field files.</p> <p>This can be a local file/directory or found on the FTP-server</p> <p>Default:{empty}</p>
	WindFileOption	<p>How to interpret the WindFieldFile</p> <p>0 <=> WindFieldFile is an ordinary .wxml file</p> <p>1 <=> WindFieldFile is a directory containing .wxml files with the name "VOLCANO_analysis_YYYYMMDD.wxml"</p> <p>Default: 0</p>

THE PLUME ALTITUDE		
	PlumeAltitudeFile	<p>The full path and filename of the .txt file which contains prior information on the plume altitudes.</p> <p>This must be a local file. Default:{empty}</p>

THE <GEOMETRYCALC> SECTION		
	completenessLimit	<p>Only scans with calculated completeness higher than this given value will be used to calculate the geometries.</p> <p>Default: 0.7</p>
	validTime	<p>The time a geometry measurement is valid. In seconds. Half of this time is before the measurement is made (which in turn is the average of the start times of the two scans that are involved) and half is after.</p> <p>Default:600</p>
	maxStartTimeDifference	<p>The maximum time difference (in seconds) between the start-time of two scans that can be combined to make a</p>

		<p>plume altitude calculation.</p> <p>This does not affect geometry calculations which are done using only one instrument.</p> <p>Default:900</p>
	minInstrumentDistance	<p>The minimum distance between two instruments that can be used to make a geometry calculation. In meters.</p> <p>This does not affect geometry calculations which are done using only one instrument.</p> <p>Default:200</p>
	maxInstrumentDistance	<p>The maximum distance between two instruments that can be used to make a geometry calculation. In meters.</p> <p>This does not affect geometry calculations which are done using only one instrument.</p> <p>Default:10000</p>
	maxPlumeAltitudeError	<p>The maximum error in the plume altitude calculation that we can tolerate.</p> <p>This does not affect geometry calculations which are done using only one instrument.</p> <p>Default:500</p>
	maxWindDirectionError	<p>The maximum error in the wind direction calculation that we can tolerate</p> <p>Default:10</p>

THE <DUALBEAM> SECTION

	useMaximumTestLength	<p>Non-zero if we should use the maximum test length possible</p> <p>Default:1</p>
	maxWindSpeedError	<p>The maximum acceptable error in the wind-speed as determined from the dual-beam measurements</p> <p>Default:10</p>
	validTime	<p>The time a geometry measurement is valid. In seconds. Half of this time is before the measurement is made and half is after</p> <p>Default:900</p>

THE <FITWINDOWS> SECTION		
	item	Each 'item' is the name of a fit-window that we should evaluate the spectra for. Notice that there must be a fit-window with this name defined for all spectrometers that we are to evaluate spectra from. By default there is only one <item> element with the string SO2.
	main	The name of the fit window that is the most important one. This must also be defined as an <item>. By default this is SO2.

THE <SKYSPECTRUM> SECTION		
	option	<p>The settings for the sky spectrum to use, can be either one of:</p> <ul style="list-style-type: none"> • SCAN – The sky spectrum will be taken as the designated sky spectrum in each scan. • AverageOfGood – The average of all good spectra in the scan will be used as sky spectrum. • Index – Spectrum number 'value' in each scan will be used as sky spectrum. Requires that the parameter 'value' is specified with an appropriate spectrum index. • User – A spectrum file (with an already dark corrected spectrum) will be used as sky spectrum for all spectra. Requires that the 'value' parameter is specified. <p>Default: SCAN</p>
	value	<p>If option is INDEX then this is the index of the spectrum that will be used as sky-spectrum (e.g. spectrum number 5 in each scan).</p> <p>If option is USER then this is the full file-name of the spectrum file that we should use as a sky-spectrum.</p>

THE <DISCARDING> SECTION		
	completenessLimit	Only flux measurements with a calculated completeness higher than this given value will be used to calculate a flux. Default: 0.9
	minimumSaturationInFitRegion	<p>All spectra with so little light that the pixel with the highest intensity in the fit-region has a saturation level less than this limit will be ignored in the evaluation.</p> <p>This judgement is done after the dark-current & offset has been removed.</p> <p>Range is 0.0 (reject none) to 1.0 (reject all spectra)</p>

		Default: 0.05
	MaxExpTime_Got	The maximum exposure-time for a spectrum from a version 1 instrument for us to consider it good and to evaluate it (Note in milliseconds) Default: 900
	MaxExpTime_Hei	The maximum exposure-time for a spectrum from a version 2 instrument for us to consider it good and to evaluate it (Note in milliseconds) Default: 4000

The Calibration section is only applicable when the Mode is *Calibration*.

THE <CALIBRATION> SECTION	
solarSpectrumFile	Full path and filename of a high-resolution solar spectrum file. This must be a two-column spectrum file with the first column being the wavelength in nanometer air. Default: {empty}
instrumentLineShapeFitOption	Option for what type of instrument line shape should be fitted to the measured spectrum during the calibration. Available options are: <i>0: Do not fit an instrument line shape during the instrument calibration.</i> This requires knowledge of an initial instrument line shape for good results. If no initial instrument line shape is provided then a Gaussian instrument line shape will be estimated from the shape of the spectrum and used. <i>1: Fit a super-gaussian instrument line shape.</i> This will fit a super-gaussian to the measured spectrum. Make sure to setup <i>instrumentLineShapeFitRegionLow</i> and <i>instrumentLineShapeFitRegionHigh</i> . Default: 1
instrumentLineShapeFitRegionLow	The lower edge of the wavelength range over which an instrument line shape is fitted when <i>instrumentLineShapeFitOption</i> is 1. Default: 330 [nm]
instrumentLineShapeFitRegionHigh	The upper edge of the wavelength range over which an instrument line shape is fitted when

	<i>instrumentLineShapeFitOption</i> is 1. Default: 350 [nm]
<code>generateEvaluationSettings</code>	<p>If set to non-zero value then new .exml files will be generated based on the previously setup .exml files and the references generated by the performed instrument calibrations.</p> <p>Default: 1</p>
<code>intervalHours</code>	<p>How often instrument calibrations should be performed on each spectrometer. This is a floating-point value, hence 0.5 represents every 30 minutes.</p> <p>Default: 0</p>
<code>intervalTimeOfDayLow</code>	<p>The time of day when the calibrations should be started, in order to avoid performing calibrations right after sunrise. This is given in number of seconds since midnight (range 0 to 86400) and given in UTC and compared to the gps-timestamp in the measured spectra.</p> <p>Default: 32400</p>
<code>intervalTimeOfDayHigh</code>	<p>The time of day when the calibrations should be stopped, in order to avoid performing calibrations too close to sunset. This is given in number of seconds since midnight (range 0 to 86400) and given in UTC and compared to the gps-timestamp in the measured spectra.</p> <p>Default: 54000</p>

EXAMPLE OF PROCESSING.XML FILE

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- This is the configuration file for the processing of scans in the NOVAC Post
Processing Program -->
<NovacPostProcessing>
  <MaxThreadNum>4</MaxThreadNum>
  <outputdirectory>D:\Tungurahua\NPPP\Output\</outputdirectory>
  <tempdirectory>D:\Tungurahua\NPPP\Temp\</tempdirectory>
  <mode>Flux</mode>
  <molecule>SO2</molecule>
  <Volcano>1502-08=</Volcano>
  <FromDate>2005.10.01</FromDate>
  <ToDate>2009.07.31</ToDate>
  <LocalDirectory>D:\Tungurahua\NPPP\Temp\</LocalDirectory>
  <IncludeSubDirs_Local>1</IncludeSubDirs_Local>
  <FTPDirectory></FTPDirectory>
  <IncludeSubDirs_FTP>1</IncludeSubDirs_FTP>
  <FTPUsername>novacUser</FTPUsername>
  <FTPPassword>iht-links.</FTPPassword>
  <WindFieldFile>D:\Tungurahua\NPPP\Tungurahua_analysis.wxml</WindFieldFile>
  <GeometryCalc>
    <completenessLimit>0.60</completenessLimit>
    <validTime>600</validTime>
    <maxStartTimeDifference>900</maxStartTimeDifference>
    <minInstrumentDistance>200</minInstrumentDistance>
    <maxInstrumentDistance>120000</maxInstrumentDistance>
  </GeometryCalc>
  <FitWindows>
    <item>SO2</item>
    <item>SO2_high</item>
    <main>SO2</main>
  </FitWindows>
  <SkySpectrum>
    <option>SCAN</option>
  </SkySpectrum>
  <Discarding>
    <completenessLimit>0.800</completenessLimit>
    <minimumSaturationInFitRegion>0.020</minimumSaturationInFitRegion>
    <MaxExpTime_Got>1000</MaxExpTime_Got>
    <MaxExpTime_Hei>10000</MaxExpTime_Hei>
  </Discarding>
</NovacPostProcessing>
```

COMMAND LINE ARGUMENTS

Several of the parameters set up in *processing.xml* can be overridden using command line arguments. The arguments are on the form: --NAME=VALUE.

E.g. the following command line argument will override the FromDate:

--FromDate=2024.05.01

Parameter Name	Example of value	Comment
--FromDate	2024.05.01	The first day that we should look for data (inclusive).
--ToDate	2024.05.01	The last day that we should look for data (inclusive).
--Volcano	0101-04= or Masaya	The volcano that the processing is to handle. This is either the full name, the simplified name or the Si-code of the volcano.
--WorkDir		Simulates the directory where the executable file is located. The program will attempt to read the configuration files in the sub directory 'configuration' in this directory.
--MaxThreadNum	15	The maximum number of threads that will be started by the software during the processing. This should be <= the number of processors (or cores) on the computer for maximum performance (but at least 1).
--IncludeSubDirs_Local	1	This is non-zero if we should include sub-directories to LocalDirectory in our search for data
--FTPDiretory		The full path to a directory on a FTP - server where we should scan for data files. Notice that it is possible to search for data both on the local computer and on the FTP server.
--IncludeSubDirs_FTP	1	This is non-zero if we should include sub-directories to FTPDirectory in our search for data.
--FTPUsername		The username to log in to the FTP-server.
--FTPPassword		The password to log in to the FTP-server
--UploadResults	1	This is true if we should upload the results (FluxLogs etc) to the Novac FTP server.
--outputdirectory		The directory where the processed files and evaluation results should be stored (on the local

		computer)
--tempdirectory		The directory that can be used for temporary files (on the local computer).
--WindFieldFile		The full path and filename of the wind field file OR the directory where to search for wind field files.
--mode	2	The processing mode, typically flux. Must be one of the defined processing modes: FLUX, CALIBRATION, COMPOSITION (not fully implemented) or STRATOSPHERE (not fully implemented)
--molecule	SO ₂	The molecule of main interest, typically this is SO2. Must be one of the predefined molecules.

GIVING THE WIND FIELD – THE .WXML – FILES

The wind field to be used in the batch processing is given in a new file format, known as .wxml. This is an XML-based file format, stored in normal ASCII format.

The main wrapper of the file is the <Wind> item, this contains one argument, the name of the volcano that the wind-field is valid for. This must agree with the name of the volcano chosen. When starting the post-processing, the NovacPPP will search for a file with the name VOLCANONAME.wxml, where VOLCANONAME is the name of the chosen volcano, in the configuration directory and use the wind field defined in this file as the wind field to use.

<WINDFIELD> - SECTION

The .wxml file can contain an unlimited number of <windfield> sections. Each specifies the wind field at a given altitude and within a given time interval. This section also contains the source of the wind field. The source of wind-information must be one of the following;

- user
- default
- ecmwf_forecast
- ecmwf_analysis
- dual_beam_measurement
- model_wrf
- geometry_calc
- geometry_calc_single_instr
- noaa_gdas
- noaa_fnl
- unknown

The <windfield> section can then contain an unlimited number of <item> lines.

<ITEM> - SECTION

The <item> section specifies the wind speed and wind direction at a given position (at the same altitude and time range that is given in the surrounding <windfield> section). This field must specify the following arguments;

- **lat** – the latitude (in degrees and decimal degrees) where the wind-information is valid
- **lon** – the longitude (in degrees and decimal degrees) where the wind-information is valid
- **ws** – the wind speed, in m/s.
- **wd** – the direction of the wind. Counted as the direction from which the wind comes in degrees from north.

The following arguments are optional;

- **wse** – the estimated error in wind speed, in m/s.
- **wde** – the estimated error in wind direction, in degrees.

EXAMPLE OF .WXML FILE

```
<?xml version="1.0" encoding="ISO-8859-1"?>

<!-- This file defines the wind field for a given volcano. To be used
for the calculation of fluxes in the NOVAC Post Processing Program -->


<Wind volcano="san_cristobal" >

  <windfield>

    <source>ecmwf_analysis</source>

    <valid_from>2009.01.14T03:00:00</ valid_from >

    <valid_to>2009.01.14T09:00:00</valid_to>

    <altitude>1800.0</altitude>

    <item lat="16.24" lon="-79.865" ws="11.60" wse="5.00" wd="206.00" wde="12.5"/>

  </windfield>

  <windfield>

    <source>ecmwf_analysis</source>

    <valid_from>2009.01.14T09:00:00</ valid_from >

    <valid_to>2009.01.14T12:00:00</valid_to>

    <altitude>1800.0</altitude>

    <item lat="16.24" lon="-79.865" ws="10.20" wse="5.00" wd="213.00" wde="21.5"/>

  </windfield>

</Wind >
```


GIVING THE PLUME ALTITUDE

If there is prior knowledge about the altitude of the plume at any given time, then this can be given to the NovacPPP through a plume altitude file. This is a simple ASCII text file containing the following (tab or space separated) header:

```
#ValidFrom      ValidTo          PlumeAltitude    PlumeAltitudeError  Source
```

The four columns in the file contain the following values (in this exact order)

ValidFrom	The date and time from which this piece of information is valid. Must be in UTC in the format YYYY.MM.DDTHH:MM (e.g. 2009.01.01T06:00)
ValidTo	The date and time until which this piece of information is valid. Must be in UTC in the format YYYY.MM.DDTHH:MM (e.g. 2009.01.01T06:00)
PlumeAltitude	The altitude of the plume, in meters above sea level, at this time interval.
PlumeAltitudeError	The estimated error in the given altitude of the plume, in meters, at this time interval.
Source	The source of this information, must be either of; user, default, geometry_calc, or geometry_calc_single_instr

Example:

```
#ValidFrom      ValidTo          PlumeAltitude    PlumeAltitudeError  Source
2009.01.01T00:00 2009.01.01T23:59 5000             200                  USER
2009.01.02T00:00 2009.01.02T23:59 5100             300                  USER
2009.01.03T00:00 2009.01.03T23:59 5200             250                  USER
2009.01.04T00:00 2009.01.04T23:59 5300             453                  USER
2009.01.05T00:00 2009.01.05T23:59 5400             928                  USER
2009.01.06T00:00 2009.01.06T23:59 5500             182                  USER
```

5. RUNNING – FLUX MODE

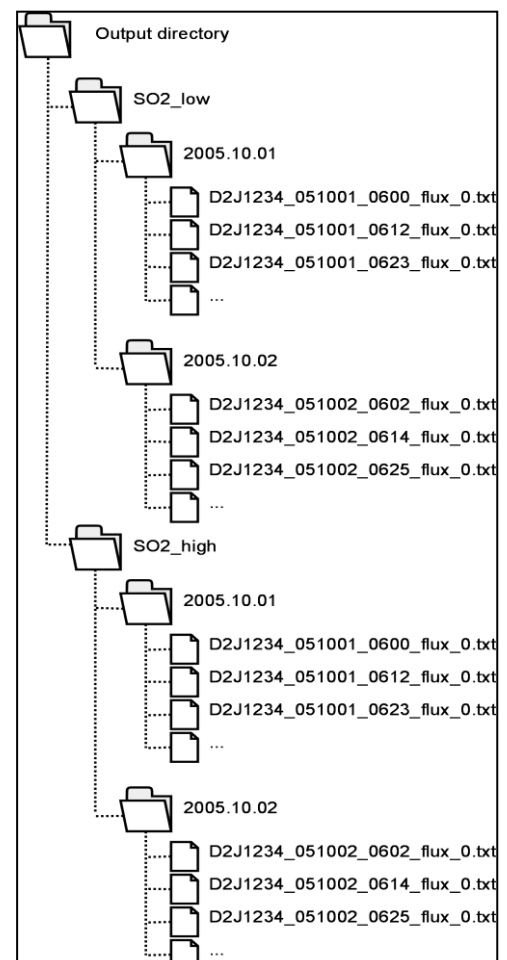
To run the post processing of spectral data, start by making sure that you have prepared the configuration as you want it and set mode to FLUX in *processing.xml*. To start the processing, run the NovacPPP executable from a command line, PowerShell or shell terminal. There is no user interface of the program and the processing will start automatically. The program will start with a sanity check of the settings and if any errors are detected, the program will stop with an error message describing what setting is incorrect.

GENERATED FILES

As the post processing of the files progresses, several files will be generated in the specified output directory of the software (as specified in *processing.xml*). First of all, the settings used in *processing.xml* will be copied to the output directory to make it possible to check the settings used afterwards. The files generated during the processing are;

- **StatusLog.txt** – This file contains the messages that are printed to the console while the processing is running, for later inspection.
- **GeometryLog.txt** – This is generated after all geometrical calculations have been made. This contains the calculated plume heights which are judged as acceptable and which are used in the calculation of the fluxes. The file contains information on which two measurements were combined to calculate the plume height.
- **FluxLog.txt** – This file is generated after all fluxes have been calculated and contains the calculated fluxes and the parameters used to calculate them.
- **ProcessingStatistics.txt** – This file is generated when the post processing finishes and contains some statistics for how many flux measurements were accepted for flux calculation and how many were rejected for different reasons.

Together with these files are a number of directories created (see the example the figure to the right), one for each fit-window that is specified in *processing.xml*. Each of these directories contains a large number of folders, one for each day there were measurements in the data set of spectrum files. In each of these folders are the evaluation log files of all measurements generated that specific day located. The name of the evaluation log files contains the spectrometer that produced the measurement, the date the measurement was done, the time of day (UTC) that the measurement started and the mode of the measurement (indicated by a four-letter combination, e.g., 'flux', 'wind' or 'comp'). For flux-mode measurements will an evaluation log file be generated only if the measurement is judged to be so good that a flux will be calculated.



WHAT PARAMETERS ARE USED TO CALCULATE THE FLUXES

- **Setup:** the setup of the instrument (position, compass, cone angle etc.) is taken from setup.xml. Any additional information that is present in the spectrum files is ignored.
- **Wind speed:** The best available wind speed is used (see the section “Retrieving the wind field at a given time and location”).
- **Wind direction:** The best available wind direction is used (see the section “Retrieving the wind field at a given time and location”).
- **Plume height:** The best available estimate available of the plume height at the time of the flux measurement is used. Here plume height measurements made by combining two scans are given the highest priority.

6. RUNNING – CALIBRATION MODE

The calibration mode can be used as a preparatory step, before running any other mode of processing, to prepare good references for the spectrometers. The input of this mode is the same set of files as for flux mode, including setup.xml, processing.xml and a set of .xml files defining the evaluation settings. The output of the mode is a set of reference files and set of .xml files defining how these reference files can be used. The .xml files can then be used as input into running the program in e.g., flux mode, thereby evaluating the spectra and calculating fluxes using the generated references.

To run the post processing of spectral data, start by making sure that you have prepared the configuration as you want it, set mode to CALIBRATION, made sure that there is a *Calibration* – section in processing.xml (see section 4) and made sure that there is a *Calibration* section in each .xml file defining an initial wavelength calibration, and/or an initial instrument line shape file.

To start the processing, run the NovacPPP executable from a command line, PowerShell or shell terminal. There is no user interface of the program and the processing will start automatically. The program will start with a sanity check of the settings and if any errors are detected, the program will stop with an error message describing what setting is incorrect.

STANDARDCROSSSECTIONS

The NovacPPP does make use of the *StandardCrossSections* data set in order to create the references. This is a sub-folder named *StandardCrossSections* in the same directory as the NovacPPP executable file. This must contain a number of high resolved cross sections absorbers and optional pseudo absorbers and a *StandardCrossSections.xml* file which lists them and their properties. An example of an *StandardCrossSections.xml* can be seen below:

```
<StandardCrossSections>
  <CrossSection>
    <File>S02_Bogumil(2003)_293K_239-395nm.xs</File>
    <Name>S02_Bogumil_293K</Name>
    <Medium>vacuum</Medium>
  </CrossSection>
  <CrossSection>
    <File>O3_Voigt(2001)_223K_230-851nm(100mbar).xs</File>
    <Name>O3_Voigt_223K</Name>
    <Medium>vacuum</Medium>
  </CrossSection>
  <AdditionalAbsorber>
    <File>Ring_short_SOLARFL_DOASIS.xs</File>
    <Name>Ring</Name>
    <Medium>air</Medium>
  </AdditionalAbsorber>
  <FraunhoferSpectrum>
    <File>SOLARFL_296-440nm.xs</File>
    <Medium>air</Medium>
  </FraunhoferSpectrum>
</StandardCrossSections>
```

Each cross-section tag contains a reference to an absorber, with a file name in the local directory, a friendly name of the absorber (will be used in the name of the references) and the medium for the wavelength of the high-resolution cross section (all references need to be in nanometer air and the medium is necessary to know if the wavelength should be converted or not).

The *AdditionalAbsorber* defines pseudo-absorbers, typically a Ring spectrum.

The *FraunhoferSpectrum* defines a high-resolution solar spectrum which will be used to create a Fraunhofer reference spectrum, i.e., it will be convolved with the derived instrument line shape and sampled on the instrument wavelength calibration grid and saved to the output together with the remaining references.

Each file listed in the *StandardCrossSections.xml* will be used to create a reference file and will be included in the generated .xml file.

GENERATED FILES

Reference files. The program will generate one reference for each high-resolution cross section found in *StandardCrossSections.xml*. These references are located in the folder {OUTPUT_DIRECTORY}/YYYY.MM.DD/{SERIAL}, where the date is the date of the measurement of each calibrated measurement.

Exml-files: the program will generate one .exml file for each spectrometer which has been calibrated. The calibrations are performed for each spectrometer for which there are spectrum files in the input and for which there are user-defined .exml files as input. These files will be saved in a folder named *calibration* which will be created in the in the specified output directory of the software (as specified in *processing.xml*). The contents of these .exml file is based on the user-defined .exml files but with some updates:

- Each calibration performed will result in fit-window if the output file. Each such fit-window will be valid from time half-way between the current calibration and the previously performed calibration and valid to a time half-way between the current calibration and the next performed calibration. Exceptions are the first and the last calibrations performed which will be valid indefinitely before and after the calibration, respectively.
- The first fit-window defined (for the evaluated channel) in the user-supplied .exml files will be duplicated for each calibration performed. The remaining fit-windows in the user-supplied .exml files will be copied as-is.
- Each created fit-window will contain each reference found in the *StandardCrossSections.xml*.
- Each created fit-window will also contain a *<wavelengthCalibration>* section with a convolved Fraunhofer reference. If you do not want this section in the output files, then remove the Fraunhofer reference from *StandardCrossSections.xml*.

7. ALGORITHMS – HOW THINGS ARE REALLY DONE

DARK CURRENT CORRECTION

The dark current in the detector can be compensated for in two ways;

- **Modeling:** This requires the collection of two special spectra with no light entering the spectrometer. One offset spectrum which is collected using the shortest possible exposure time and one dark current spectrum which is collected using a very long exposure time. The dark current of the detector is then modeled as a combination of the dark current and the offset spectra.
- **Direct measurement:** This requires the collection of only one special spectrum with no light entering the spectrometer. This spectrum (usually referred to as 'dark') can only be used to correct spectra with exactly the same exposure time as was used to collect the dark spectrum.

In the Novac – project are both these ways of compensating for the dark current applied. The most common (and fastest) way is by direct measurement, by collecting a dark spectrum in the beginning of the scan and then using the same exposure time for all subsequently collected spectra in the scan.

The option of modeling has the advantage that it is possible to have different exposure times throughout the scan, however at the expense of longer time to prepare the measurement.

DOAS FIT

WAVELENGTH CALIBRATION USING A SOLAR SPECTRUM

NovacPPP is capable of determining if there is a shift (or squeeze) between the wavelength calibration of the measured spectrum and the wavelength calibration of the reference spectrum by using a solar spectrum.

When a spectrometer is calibrated then the pixel to wavelength mapping is determined at that specific time (let's call this Γ_0) as well as the instrumental slit function. The reference files are then created using this slit function and pixel to wavelength mapping. When the instrument at a later time make measurements can the pixel to wavelength mapping have changed, e.g. due to changes in temperature. Let's call this new pixel to wavelength mapping Γ_M . The difference between Γ_0 and Γ_M can be approximated by a shift and a squeeze between the two, this difference will induce a miss-fit between the references and the measured spectrum in the DOAS-fit. The difference between Γ_0 and Γ_M can be determined and compensated for if a solar spectrum is provided to the NovacPPP. This solar spectrum (let's call this S_0) must be created using the same slit function and pixel to wavelength mapping as the reference files (Γ_0).

The shift between Γ_0 and Γ_M in the wavelength region of interest is determined by performing a DOAS fit where the solar spectrum S_0 is fitted to the measured spectrum with the shift of S_0 set to

free. The value of the shift that provides the best fit of the solar spectrum to the measured spectrum is determined.

NovacPPP determines the shift from the spectrum in the scan which has intensity closest to 75%. The first fit is performed in the same wavelength range as the second DOAS fit, thus this method requires that the region is wide enough to provide enough Fraunhofer lines.

After determining the optimum shift does the NovacPPP apply the same shift to the reference spectra when evaluating all the spectra in the scan.

SCAN PROPERTIES

From each collected scan are the following properties calculated;

- **Scan offset**
- **Plume centre**
- **Plume centre error**
- **Completeness**
- **The lower plume edge**
- **The upper plume edge**

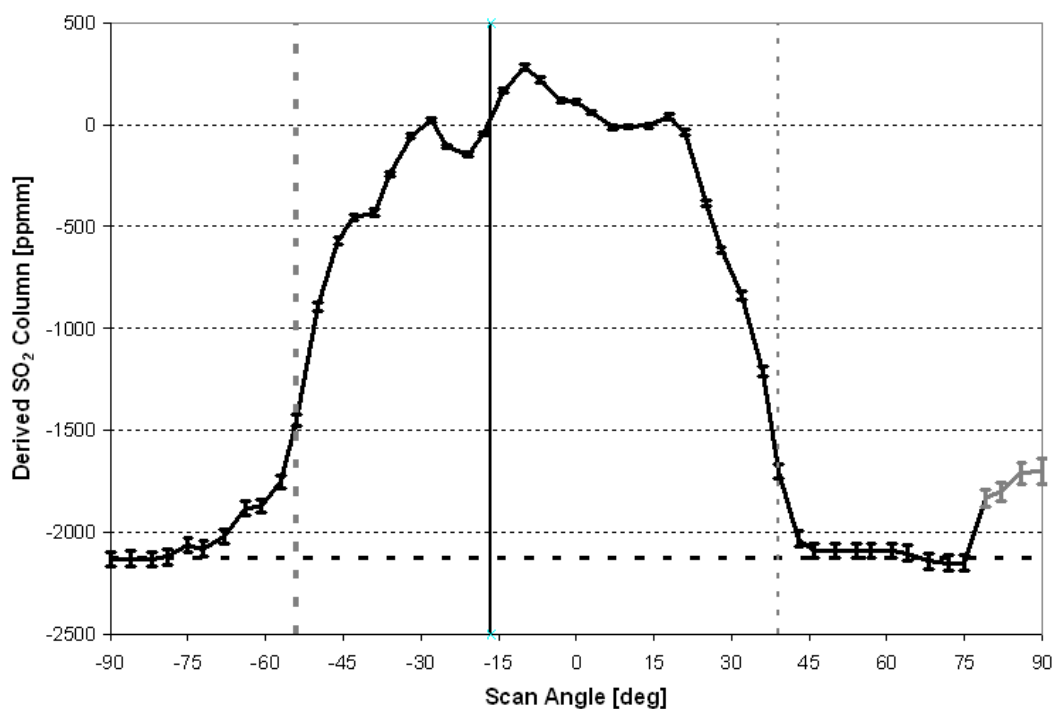


FIGURE 1. EXAMPLE OF ONE SCAN COLLECTED AT POPOCATÉPETL. BLACK SOLID LINE REPRESENTS MEASURED SO_2 COLUMNS LABELLED AS "GOOD" BY THE NOVACPROGRAM, GREY SOLID LINE REPRESENTS MEASURED SO_2 COLUMNS LABELLED AS "BAD" BY THE NOVACPROGRAM. HORIZONTAL DASHED LINE IS CALCULATED OFFSET LEVEL, VERTICAL SOLID LINE IS CALCULATED PLUME CENTRE AND VERTICAL DASHED LINES ARE CALCULATED PLUME EDGES.

SCAN OFFSET

Due to the division of each measured spectrum by a Fraunhofer reference, the derived slant columns are relative to the absorption in the Fraunhofer reference. Thus, if the Fraunhofer reference is collected in the zenith and a volcanic gas plume passes straight over the instrument, there will be a significant absorption from SO₂ in the Fraunhofer reference. All spectra with less absorption will then be evaluated as having negative columns. To correct for this, an artificial offset is introduced as the difference between the apparent column zero level and the actual column zero level.

In the development of the NovacProgram was an algorithm for automatically finding the offset developed. This is done by first removing all bad spectra in the scan, the remaining N spectra are then sorted in order of decreasing SO₂ column. The offset is then taken as the average column in the $N/5$ spectra with lowest SO₂ column. The same algorithm is used in the NovacPPP.

PLUME CENTRE, IT'S ERROR AND THE EDGES OF THE PLUME

The software can also approximately calculate the viewing direction which points to the centre of the plume, given that the whole plume is visible from the instrument. After first removing the offset, the scan is regarded to be able to see the plume if there is a region, R , with at least five consecutive scan angles, with good spectra, where the average column value is at least five times the average column error in the same region.

The scan angle where the centre of mass of the plume is found is then calculated as a weighted average of the scan angles within the region R where the weights are the derived column at each scan angle. To estimate the error in plume centre are the two measurement directions closest to the centre of the plume found where the column value has dropped below 90% of the maximum column value in the scan. The error in the plume centre position is then calculated as half of this distance (in degrees).

The edges of the plume are calculated as the first measurement angles, counted from the measurement angle where centre of the plume is calculated to be, where the column values has dropped to $1/e$ of the maximum column value. This algorithm works well for most single plumes and does not make any further assumptions on the shape of the plume.

COMPLETENESS

When the software has determined that a specific scan sees the plume, it is of interest to determine if the scan crossed the entire gas plume or if the plume is seen at too low scan angles to be complete. For this purpose a quality parameter known as the completeness is calculated as:

$$Completeness = 1.0 - \frac{1}{2} \times \frac{\max\left(\frac{1}{5} \sum_{i=1}^5 SCD_i; \frac{1}{5} \sum_{i=n-5}^n SCD_i\right)}{\max(SCD)}$$

The completeness varies between 0.5 when the plume is seen at the scan angles closest to the horizon and 1.0 when the plume is seen inside the scan. Thus, a scan which has completeness less than 1.0 does not capture the entire plume and that the resulting flux might not correctly estimate the actual flux. Note that this is a quality parameter, and not necessarily a correct counter of the

percentage of the plume that is seen. The formula for the completeness uses the average value of the five lowest, good, measurement directions which corresponds to the lowest 18° in the normal case of a measurement every 3.6° . This is a quite large angle to use and thus rejects a large number of scans. The choice of using five measurements has been determined empirically as the number of angles needed to guarantee the presence of clean air below the plume.

GEOMETRY CALCULATIONS

With geometry calculations is meant calculating the direction, altitude or width of the plume using one or more scans collected by one or several instruments.

Currently there are two types of geometry calculations implemented in the NPPP, the direction of the plume can be calculated from a single scan from a single instrument (this gives good results only if the plume passes nearly over the instrument) or the direction and altitude of the plume can be calculated by combining two scans which see the plume from two different directions.

When the evaluation of all spectra is done, the list of scan results is sorted in order of increasing start-time (this is very important since this speeds up the geometry calculations enormously). The scans are then combined two and two to try to calculate the plume direction and altitude. Only scans which fulfill the criteria specified in processing.xml (see page 19) and are flux-measurements (i.e. they contain a wide range of measurement angles from one horizon to the other) are used. Only if a specific scan cannot be combined with any other scan to generate a geometry calculation then the program will attempt to make a geometry calculation using only the given scan.

COMBINING TWO SCANS FROM DIFFERENT INSTRUMENTS

Before the calculation is made are the locations of the two instruments checked against the criteria given in processing.xml, if the instruments are too far apart or too close then no further calculations are made. Also the start times of the two scans are checked, if too long time has passed between them then no further calculations are made.

The algorithm used to calculate the plume direction/altitude is taken from the NovacProgram, this is done as follows: for each of the two scans the measurement direction corresponding to the centre of mass of the plume is determined, and lines originating in the instrument (Γ_1 and Γ_2) are constructed in these directions. By using an initial assumption of the plume height (h_0), these two lines are used to calculate two plume propagation directions of the plume by first calculating the interception points (p_1 and p_2) between each of the lines Γ_1 and Γ_2 and an imagined plane at altitude h_0 then calculating the direction of the line from the emission source to each of the two points p_1 and p_2 . If the assumed plume height differs from the true plume height then these two directions will differ. An iterative Gauss-Newton search is then performed in which the assumed plume height h is updated in each step. The search is considered to have converged when the assumed plume height yields two plume directions which differ by less than one degree, giving an estimate of the height and direction of the plume. This algorithm was selected as it is fairly robust and does allow for different scanning geometries of the two instruments used.

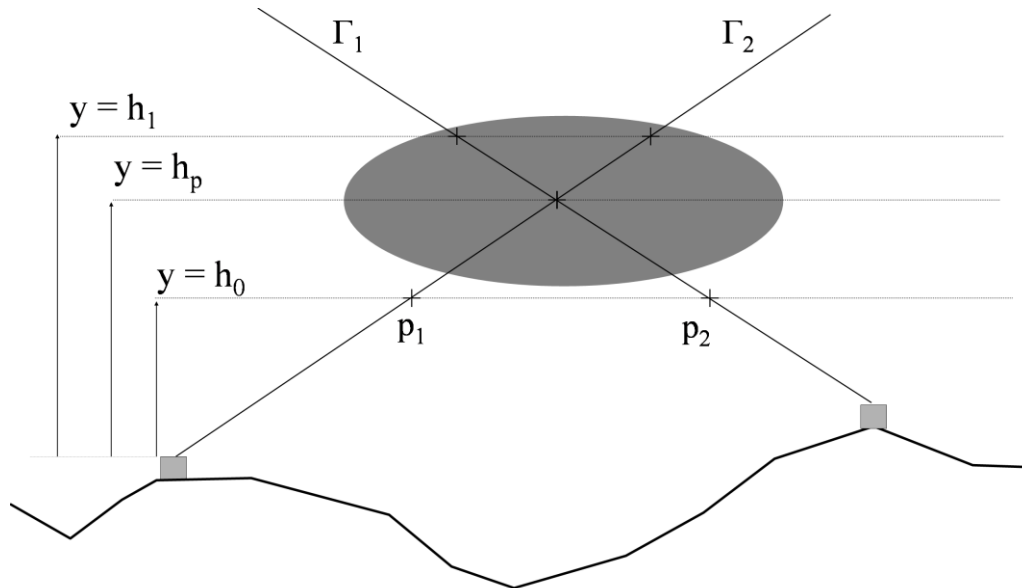


FIGURE 2. THE ALTITUDE OF THE GAS PLUME CAN BE DETERMINED IF TWO SCANNING INSTRUMENTS SEE THE PLUME SIMULTANEOUSLY. FIRST A LINE IS DRAWN FROM EACH OF THE INSTRUMENTS IN THE DIRECTION OF THE CENTRE OF THE PLUME. THEN, A PLUME HEIGHT IS ASSUMED AND, FOR EACH OF THE INSTRUMENTS, THE WIND DIRECTION IS CALCULATED UNDER THE ASSUMED PLUME HEIGHT. WHEN THE ASSUMED PLUME HEIGHT EQUALS THE TRUE PLUME HEIGHT, THE TWO CALCULATIONS GIVE THE SAME WIND DIRECTION.

This calculation results in a direction and an altitude of the plume. However, for these to be useful must the error in the calculation be estimated. This is done by, for each of the two scans, perturbing the angle at which the centre of the plume is found with the estimated error in the plume centre position. Doing this four times (adding or subtracting the error to/from the plume centre position for each of the two scans) results in four different estimates of the plume direction/altitude. The error in the calculated plume altitude is then taken as the average of the four calculated plume altitudes. The error in the calculated plume direction is also taken as the average of the four calculated plume altitudes – **THIS SHOULD BE UPDATED!**

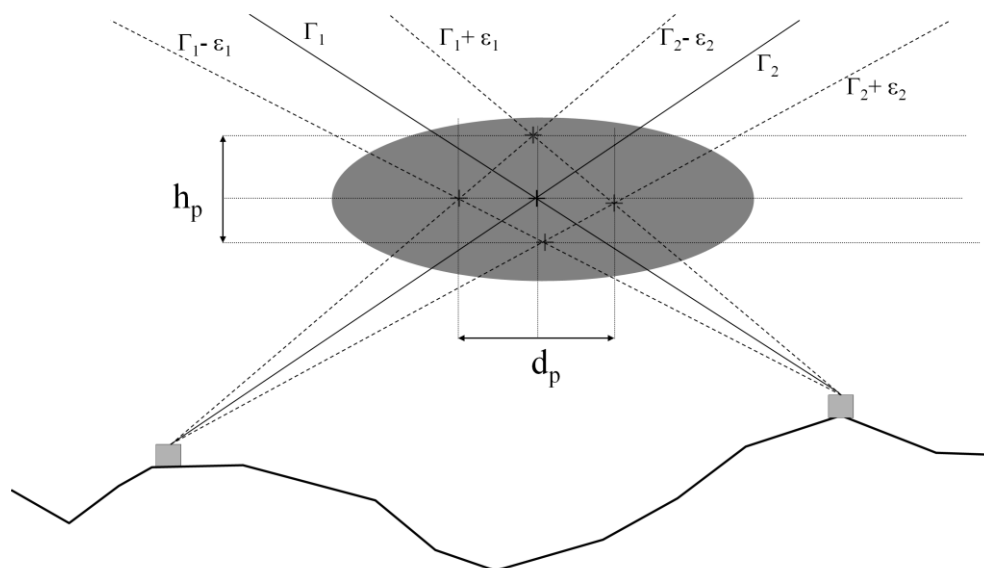


FIGURE 3. THE ERROR IN THE ESTIMATED PLUME ALTITUDE AND PLUME DIRECTION ARE CALCULATED BY VARYING THE PLUME CENTRE POSITIONS OF THE TWO SCANS WITHIN THEIR ERROR RANGE.

USING ONE SCAN FROM A SINGLE INSTRUMENT

Using data from only one instrument does not allow for a full determination of the plume altitude and plume direction, however one of these parameters can be determined if the other is known at the same time. In the NovacPPP, only the direction of the plume is calculated in this way since the conditions for this calculation allows for more accurate calculations than calculating the plume altitude from the plume direction.

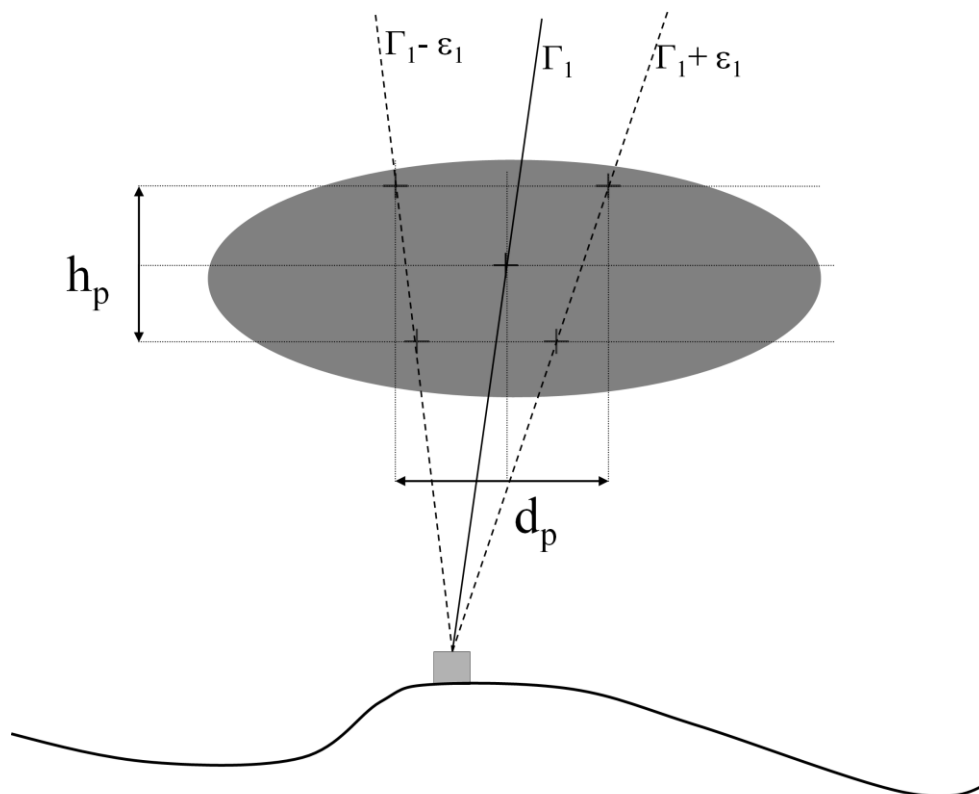


FIGURE 4. WHEN CALCULATING A PLUME DIRECTION USING DATA FROM ONE INSTRUMENT, THE ERROR IN PLUME DIRECTION IS ESTIMATED FROM BOTH THE ERROR IN PLUME CENTRE POSITION AND THE ERROR IN PLUME ALTITUDE.

When performing this type of calculation, it is assumed that the altitude of the plume is known with relatively high accuracy. The calculation is made by drawing a line from the instrument in the direction where the centre of the plume is judged to be up to the estimated altitude of the plume. By then drawing a line from the volcano to the endpoint of this line can the direction of the plume be estimated. The error in the plume direction calculation is estimated by varying the plume centre direction within its estimated error range and by varying the plume altitude within its estimated error range. The total error in plume direction is then calculated as the square root of the sum of 1) the squared difference between the calculated plume directions when varying the plume centre and 2) the squared difference between the calculated plume directions when varying the plume altitude. This calculation typically gives acceptable errors in the plume direction only when the plume passed directly (or almost directly) above the instrument.

DUAL-BEAM WIND SPEED

Some of the Novac instruments are capable of performing measurements of the plume speed using the Dual-Beam technique. These measurements are initialized by the NovacProgram running in the volcano observatory and thus require a direct-link to the instrument to be able to, in near real-time, judge where the plume is at all times.

Dual-beam measurements can be performed by the Version I instruments installed with a dual-channel spectrometer (SD2000) or by the Version II instruments. The measurements are slightly different for the two instruments:

- Version I: The upwind and downwind time series are collected by the two spectrometers built into the SD2000 – package. Thus the spectra for the two time series are stored in two separate .pak-files, one ending with “_0.pak” emerging from the master channel and one ending with “_1.pak” emerging from the slave channel. Both of these files are necessary to calculate a wind speed.
- Version II: The upwind and downwind time series are collected by the same spectrometer, the spectra are thus stored in the same .pak-file. The two time series are collected using different positions for the two motors, thus the up- and down-wind time series are identified by their scan-angles.

The process of calculating wind speeds from the dual-beam measurements performed after all possible geometry calculations have been made. This to make sure that any estimated of plume altitude that can be derived from the scans is calculated before the wind speed measurements are considered, this since the dual-beam technique requires the plume height as input parameter and the lower the error in the plume height the lower the error in the calculated wind speed.

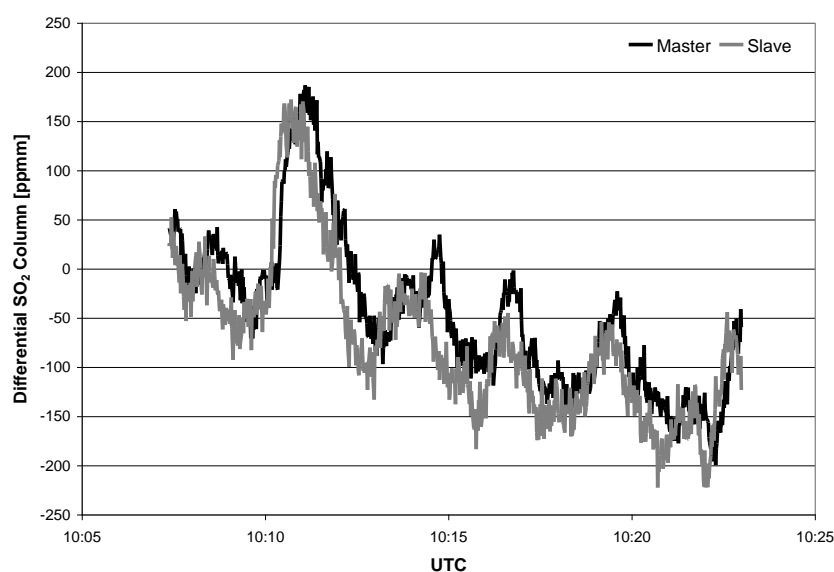


FIGURE 5. DUAL-BEAM WIND MEASUREMENT PERFORMED ON MT. ETNA ON THE 18TH OF JULY 2008. THE MEASUREMENT WAS INITIALIZED AUTOMATICALLY BY THE NOVACPROGRAM.

RETRIEVING THE WIND FIELD AT A GIVEN TIME AND LOCATION

The NovacPPP manages internally a database containing the wind speed and wind direction that was read in from the .wxml file at startup, the wind directions that have been calculated by geometrical calculations, the wind speeds calculated using dual-beam measurements etc. When the processing is finished will the contents of this database be written to the file GeneratedWindField.wxml in the output directory.

Each piece of wind information inserted into the database is labeled with a time range for which it is valid. When a wind field is to be extracted from the database then the database is searched for the 'best' type of wind information that is valid for the requested time. With 'better' type of wind information is meant a source of wind information that is likely to be more accurate than another type of wind info. E.g. is the wind direction calculated from a geometrical calculation considered to be more accurate than a modeled wind direction which in turn is considered to be more accurate than just using the default wind direction. If several pieces of wind information (of the same type, the 'best' available) are valid at the requested time then they will be averaged together.

An example of how this can be done can be seen in Figure 6, where the available pieces of wind information during a part of a day is shown. In the figure are wind information that is judged as 'better' show higher up in the plot. In this example there is a modeled wind field available all the time, however due to limited time resolution in the model does this wind field not change very often. With some intervals has the geometrical calculations managed to get the wind direction from geometrical calculations using a single instrument. From around 10:30 to 12:00 has the plume direction and plume altitude been calculated on six occasions, and on one occasion at around 11:30 has the speed of the plume been measured from a dual-beam measurement. A list of what wind fields would be extracted from this database at some given times can be seen in Table 1.

TABLE 1. WHAT WIND SPEED AND WIND DIRECTION WOULD BE EXTRACTED FROM THIS DATABASE AT GIVEN TIMES.

Time	Wind direction	Wind speed
09:00	Geometry – single instr.	Model
11:00	Geometry	Geometry
11:20	Average of two geometry calculations	Dual-beam

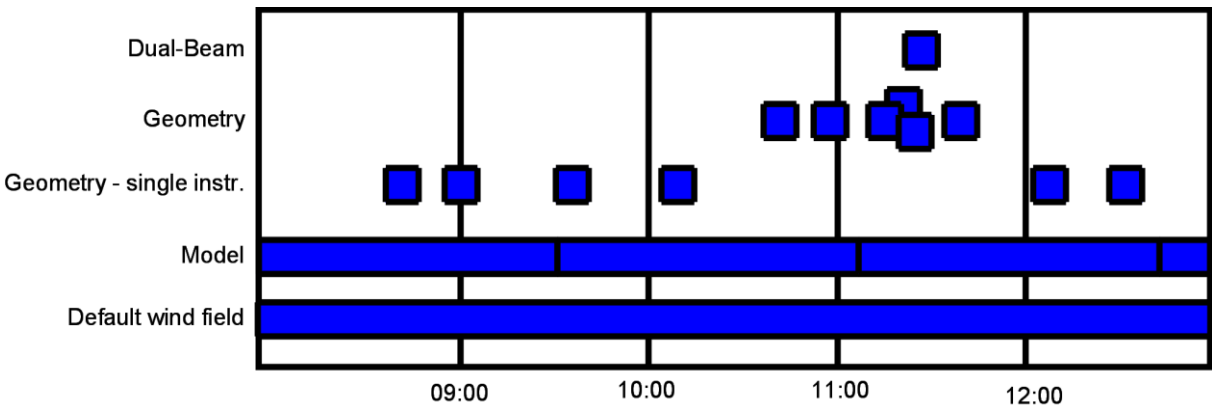


FIGURE 6.

8. SOFTWARE DETAILS – A PROGRAMMER’S REFERENCE

POST PROCESSING OF FLUXES

The controlling code for the post processing is found in the class `CPostProcessing`.

The post processing of fluxes is found in the function

`CPostProcessing::DoPostProcessing_Fluxes`, the general flow of this function is shown in Figure 7

CPostProcessing::DoPostProcessing_Flux

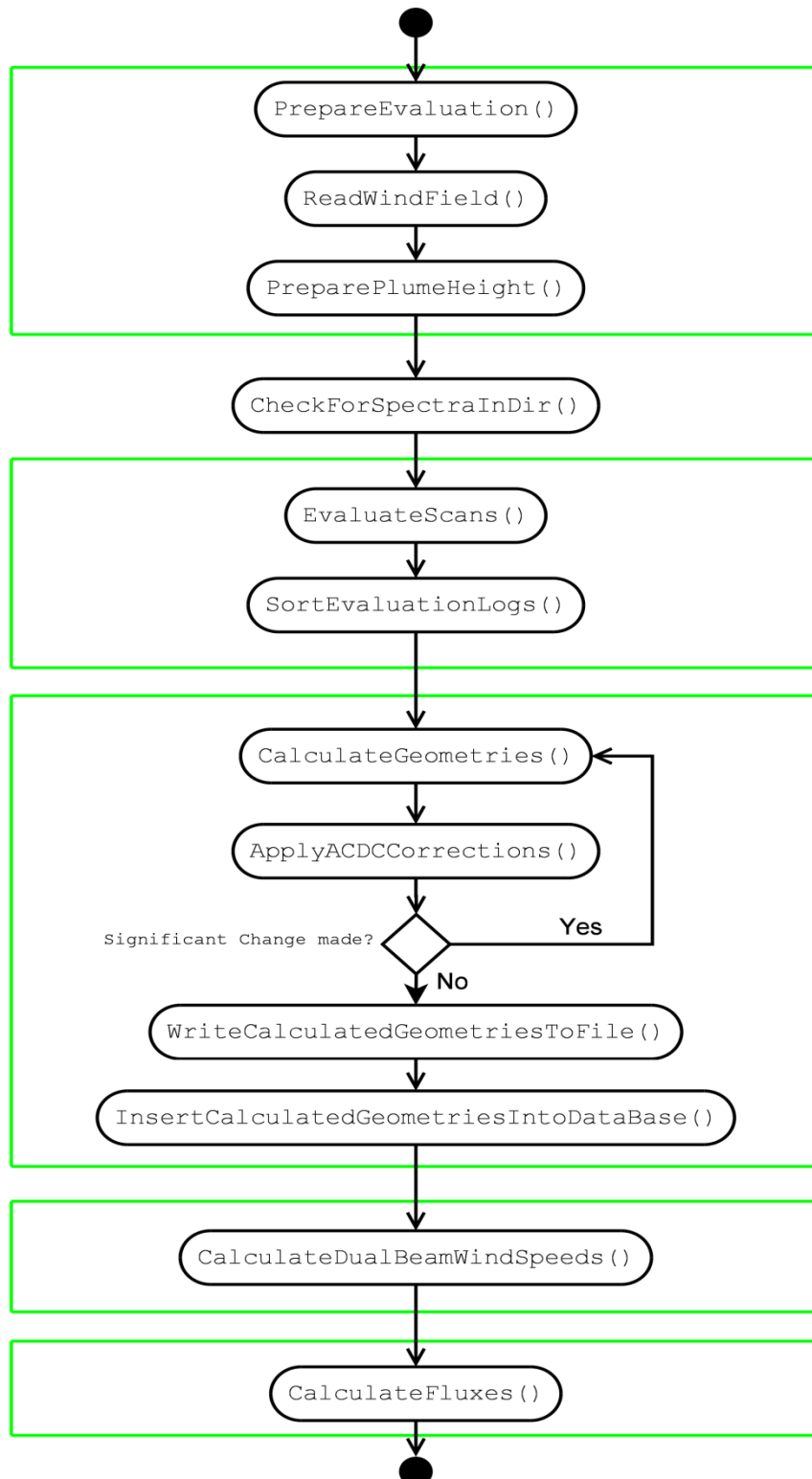


FIGURE 7. GENERAL FLOW OF THE POST PROCESSING OF FLUXES, THE FUNCTIONS SHOWN ARE ALL FOUND IN THE CLASS CPOSTPROCESSING.

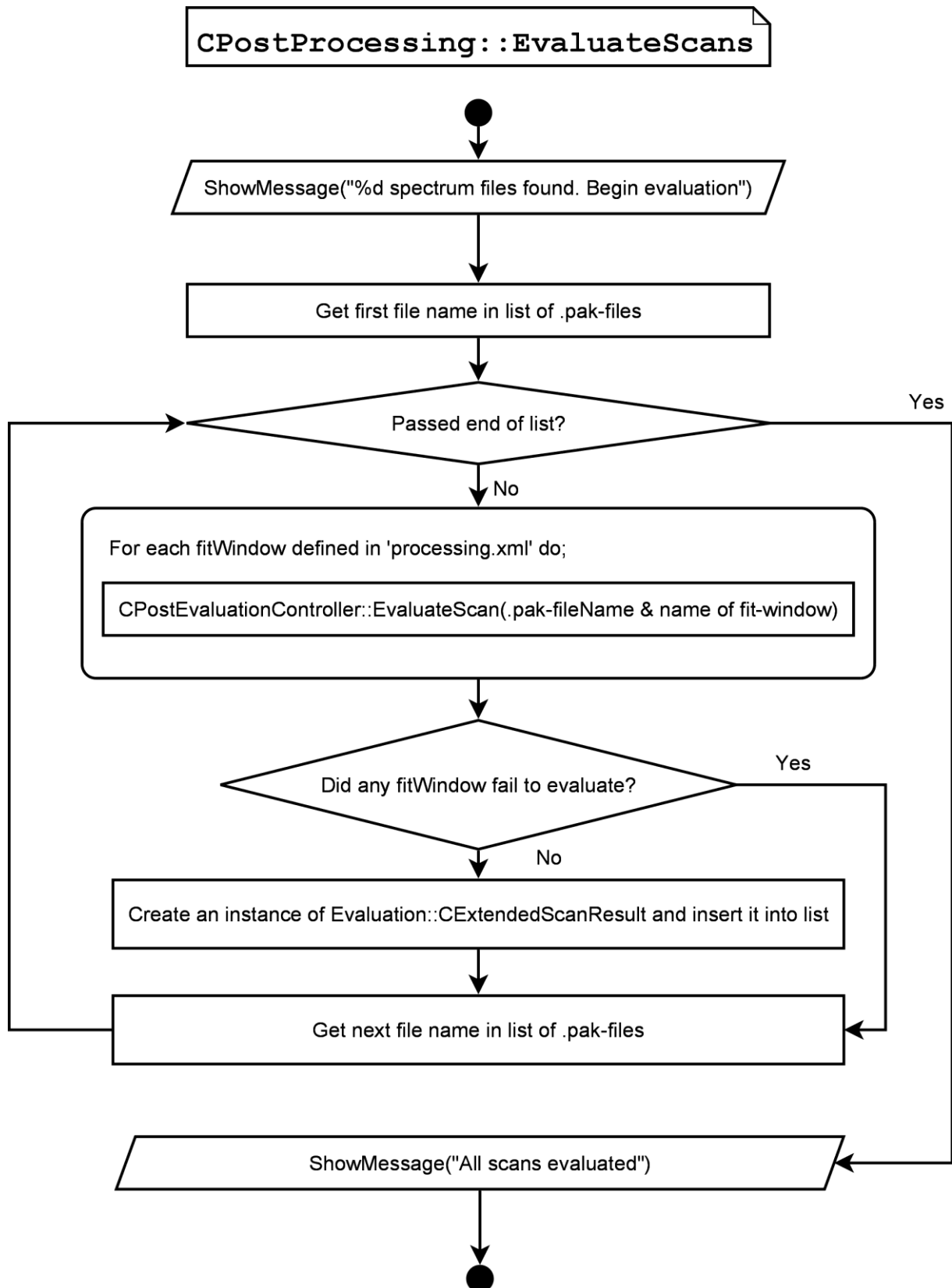


FIGURE 8. GENERAL FLOW OF POST PROCESSING OF FLUXES.

POST PROCESSING OF COMPOSITION MEASUREMENTS

This is not yet implemented... Please be patient...