Technical Note on:

Description of Statistical Tool

Draft
Revision -

29 October, 2003

Delivery of WP 4100 of the CCN#5 of the study:

"Development of an Optimised Algorithm for Routine p, T and VMR Retrieval from MIPAS Limb Emission Spectra"

Contract No: **11717/95/NL/CN**

Prepared by:

Name	Institute
G. Bazzini	Ph.I.
M. Gai	Ph.I.
P. Raspollini	IROE-CNR

Approved by:

Name	Institute
B. Carli	IROE-CNR

CONTENTS

1. Reference documents

[RD1] TN-IROE-GS0101, Issue: 1 - Revision: A
Title: Level 2 Algorithm Characterization & Validation Plan

[RD2] TN-IROE-GS0103, Draft
Title: ORM for Commissioning Phase

2. Introduction

In the frame of ESA contract 11717/95/NL/CN an Optimized forward /retrieval Model (ORM) was developed, suitable for implementation in MIPAS near real-time Level 2 Processor. In particular, version 1.22.1 of the ORM_ABC code (described in [RD2], [RD3] and subsequent e-mails sent to ESA and Astrium]) is the scientific reference for the Retrieval Component Library of MIPAS Level 2 NRT processor.
When the first MIPAS measurements will be available, the impact of the most critical approximations implemented in the ORM must be adequately characterized.
All the procedures planned for validating ORM during the Commissioning Phase, as well as the requirements of the software tools needed for performing this cal./val. activity, were described in [RD1]. The required software tools consist in both a modificated ORM with added functionalities [RD2] and a dedicated software tool for the analysis of ORM products (Statistical Tool).
Objective of this TN is to describe the operations performed by the Statistical Tool, its installation/run procedures, and its interfaces with ORM.

3 Description

Statistical Tool (ST) is a "Stand Alone" code dedicated to display data retrieved by ORM inversion code, and analyse them from a statistical point of view.

In order to collect all the data needed for Statistical Tool a specific output routine has been developed and installed in ORM. This routine is dedicated to the extraction of data and their writing in a specific format. A file using such format is readable by Statistical Tool. No other format is recognized by Statistical Tool [RD2].

The ST performs two main group of functionalities: the first dedicated to internal analysis (HouseKeeping Section) and an other one dedicated to the final results of ORM code (Product Section) [See also paragraph ***8 Functionalities of Statistical Tool*** for further details.]

The routines are coded in C/C++ language, and require an external graphic environment. Actual graphical socket is ROOT, a package developed and freely distribuited by CERN.

Runtime module has been tested over REDHAT and DEBIAN Operative Systems; since code is supplied in source version, the user is able to work on all systems consistent with ROOT package. [See also paragraph ***6 Graphic Environment*** for further details.]

4 Appeareance

Statistical Tool is supplied as a compressed *tar* archive:

StatisticalTool_0.8.tar.gz for Unix / Linux operative systems;

5 Requirement

The user needs a work-station machine provided with:

- Windows / Unix / Linux operative system;
- Root Package;
- C++ - compiler;

The code is able to run properly [graphic and calculation] on a low resource PC [for instance Intel Pentium 133 MHz, 128 MByte RAM.]

Memory has to be increased according to the amount of data treated. As a reference, an entire orbit takes about 350 MBytes.

6 Graphical Environment: about Root package.

Root is a free distribution software, available on CERN internet site.

One can find last (and previous) realese at the following internet address: http:\root.cern.ch. On the internet site Root package is available in binary version (already compiled for some platforms) or in source files.

We hardly recommend reading the Root Internet Site, as well as the user guide,

7 How to install Statistical Tool

1 First install ROOT package.
A detailed guide to install ROOT is contained in README file, inside of ROOT package.

If you have downloaded binary version, you need only to set environment variables and paths. An useful extract of the guide is tracked below:

> *Setting the Environment Variables*
> ---
> > *Setting the ROOTSYS and PATH Variables*
> > *Define the variable ROOTSYS pointing to the directory where you unpacked the ROOT distribution:*

export ROOTSYS=/root (in bash or ksh)
setenv ROOTSYS /root (in csh or tcsh)

Add ROOTSYS/bin to your PATH:
export PATH=$PATH:$ROOTSYS/bin (in bash or ksh)
set path=($path $ROOTSYS/bin) (in csh or tcsh)

Setting the Library Path

On HP-UX, before executing the interactive module, you must set the library path:
export SHLIB_PATH=$SHLIB_PATH:$ROOTSYS/lib (in bash or ksh)
setenv SHLIB_PATH ${SHLIB_PATH}:${ROOTSYS}/lib (in csh or tcsh)
On AIX, before executing the interactive module, you must set the library path:
[-z "$LIBPATH"] && export LIBPATH=/lib:/usr/lib
export LIBPATH=$LIBPATH:$ROOTSYS/lib
or the equivalent setenv.
On Linux, Solaris, Alpha OSF and SGI, before executing the interactive module, you must set the library path:
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$ROOTSYS/lib
or the equivalent setenv.
On Solaris, in case your LD_LIBRARY_PATH is empty, you should set it like this:
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$ROOTSYS/lib:/usr/dt/lib
where $ROOTSYS is an environment variable pointing to the ROOT directory. For example, if you use the HPUX-10 AFS version (see below), you should set:
export ROOTSYS=/afs/cern.ch/na49/library.4/ROOT/v2.23/hp700_ux102/root (in bash or ksh)
setenv ROOTSYS /afs/cern.ch/na49/library.4/ROOT/v2.23/hp700_ux102/root (in csh or tcsh)

If you have downloaded source version, you need first to compile the code and then set environment variables and paths.

To run the ROOT Environment just type: root from the prompt.

2 Create a directory dedicated to Statistical Tool.
Directory name is worthless for Statistical Tool to work properly.

3 Expand supplied archive into dedicated directory.

During installation phase the system will create the following structure under dedicated directory:

Makefile
*.cxx
*.hxx
*.h
/nominalOM
*
/DBMicrowindows
*
/data
*

where

/nominalOM contains nominal Occupation Matrix for each retrieval

/DBMicrowindows contains details referred to specific MWs, such as expected Chi square, and so on.

/data contains some example files generated by ORM, and related to a hypotetical sequence.

Makefile is a batch file useful for compiling the code.

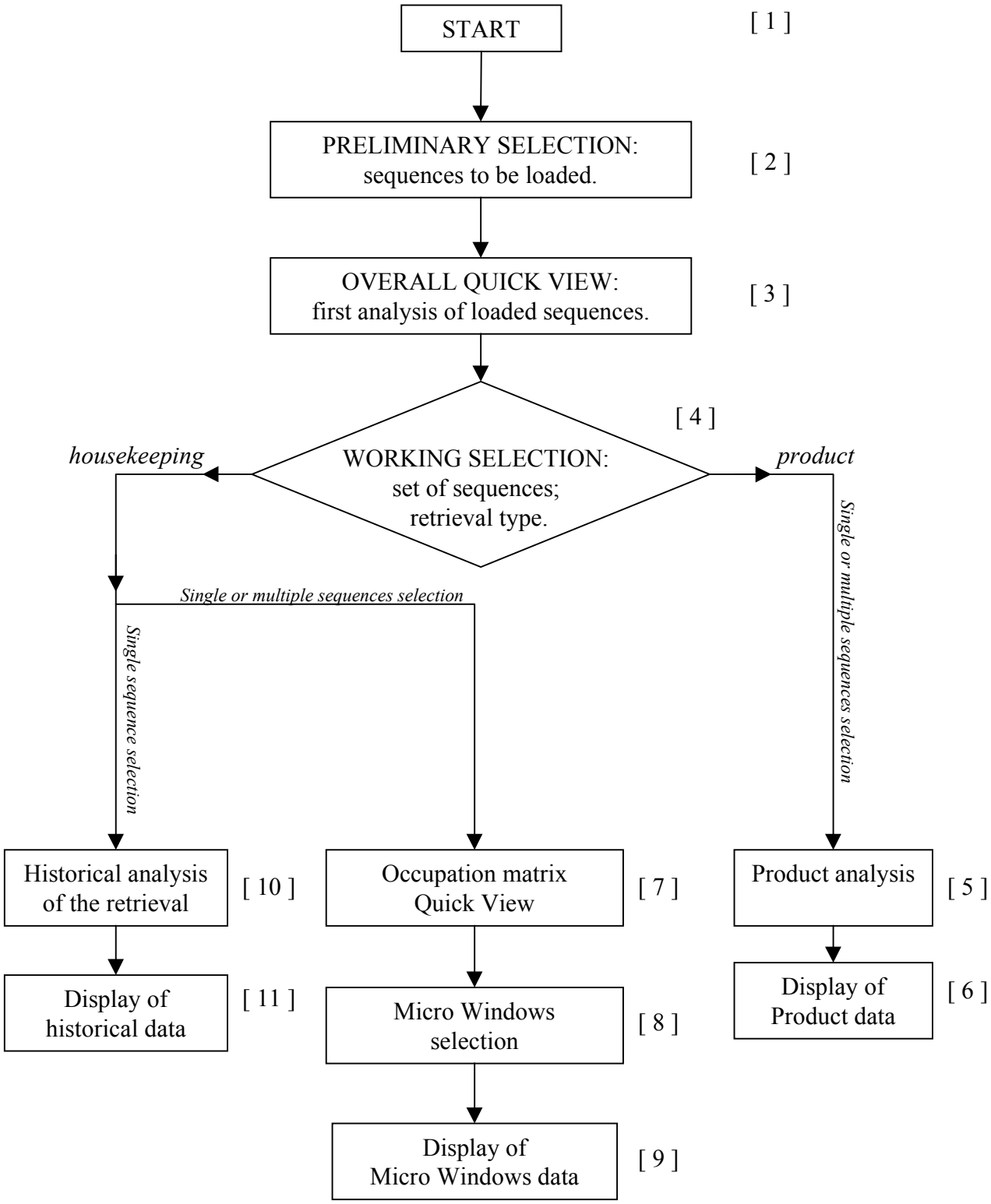
4 Compile Statistical Tool sources. a "Make file" is provided.

8 Functionalities of Statistical Tool:

Fig. 1 shows the logical architecture of Statistical Tool.

To enter Statistical Tool press ST at prompt.

Fig. 1 The Block Logical view



[1] Start

Logic level to enter Statistical Tool procedure.

The user can enter Statistical Tool when ST package is correctly installed on an adequate platform.

[2] Preliminary Selection

Logic level to select the sequences that will be loaded.

The user is allowed to choose a whole orbit or scattered sequences also referred to different orbit. The maximum number of files is TBD.
It will be possible to display all data written in the header of each selected file.

[3] Overall Quick View

Logic level to quickly analyse result of retrieval over a large number of sequences.

The latitude of any selected sequence will be displayed as a function of an index related to the sequence; moreover for each retrieval will be shown:

- χ^2 as a function of the sequence-index;
- the number of retrieval iterations;
- a flag indicating the result of retrieval.

Quick View details will be displayed in a suitable plot.

[4] Working Selection

Logic level to select specific sequences that will be analysed.

The user is allowed to select the sequences and retrieval that will be processed.
The selection of specific retrieval type will be possible by the way of a button bar. Only one retrieval is selectable at a time.
The selection of the sequences will be possible by the way of a table containing data referred to the sequence. For the selected retrieval, each row contains:

- the index of the sequence;
- the latidude;
- the sun elevation;
- χ^2

By default, all sequences will be selected but the ones whose retrieval failed.

[5] Product Analysis

Logic level to choose specific information about selected retrieval.

The following menu is available to the user:

- The averaged p,T or VMR profile as a function of the altitude
- The averaged corrections to the tangent altitude
- Eigenvalues of the V-C Matrix
- Correlation graphs related to the gradient

[6] Display of Product Analysis

Logic level to display specific information previously chosen.

The result of selections listed in Product Analysis (level 5) will be displayed in a suitable plot.

[7] Occupation Matrix *Quick View*

Logic level to quickly analyse the results of retrieval over MicroWindows.

The *Quick View* of the occupation matrix displays a [18 rows x 21 columns] matrix. Rows refer to the sweeps, while columns refer to the MicroWindows. Each matrix element corresponds to a MicroWindow and displays alternatively one of the following parameters:

- χ^2_{orm} : χ square calculated in ORM by way of the variance-covariance matrix;
- χ^2_e : χ square expected;
- χ^2_c : χ square calculated in the ST using only the selected elements of the MW;
- χ^2_a : χ square calculated in the ST using all the elements of the MW;
- $R_{orm/e}$: Ratio between χ^2_{orm} and χ^2_e ;
- $R_{c/e}$: Ratio between χ^2_c and χ^2_e ;
- $R_{a/e}$: Ratio between χ^2_a and χ^2_e ;
- $R_{orm/c}$: Ratio between χ^2_{orm} and χ^2_c ;
- Res_{av} Average value of the residual;
- N_{points} MW population;

The selection of parameter to be displayed is available on a button bar.
The bottom row as well as the last column are dedicated to the average value of the displayed parameter.
The average is calculated over each column and row respectively.
Global average all over MicroWindows is also available.
The user is also allowed to display:

- The *p,T* or *VM*R profile as a function of the altitude
- Theoretical *p,T* or *VMR* profile
- The *VMR* profiles of "initial guess "
- Correlation graph related to the gradient

[8] MicroWindows selection

Logic level to select specific MicroWindows that will be analysed.

User is allowed to select one or more MWs.
On the chosen selection Statistical Tool will operate on averaged data.
User is also allowed to choose among the graphs listed below:

- The average of the residuals
- Continuum
- Spectral data (measured, calculated, residual, logic vector)

[9] Display of MicroWindows data

Logic level to display specific information previously chosen.

The result of selections listed in MicroWindows selection (level 8) will be displayed in suitable graphs.

[10] Historical Analysis of retrieval

Logic level to display detailed information over the selected retrieval.

The following details referred to the selected retrieval are available to the user:

:eigenvalues of V-C matrix
:the number of iterations needed to reach the convergence of retrieval
and for each iteration:
:the number of microiterations
:intermediate χ^2
:λ of Marquardt constant
:intermediate calculation time

[11] Display of Historical Data

Logic level to display specific information related to retrieval.

Details related to the retrieval (level 10) will be displayed in suitable report.

8.1 Side by side with the user

A Macro selection, overview, and micro selection: a preliminary work.

Step 0

To start working with Statistical Tool a text-format list is necessary to be present and available to the procedure. This list shows the entire path of files the user wants to analyse. This file normally uses *lst* extension. Other extensions are allowed.

Step 1

At the very beginning the user will encounter a dialog window to enter selection.
To start selection the user has to press “Preliminary” menu.
The user can decide to select either a single-sequence file or few files related to different sequences or the total amount of sequences collected in an orbit. The selection of sequences is possible by using the file **.lst*.

Step 2

The list of sequences is shown to the user. An useful instrument to overview general trends of retrievals in different sequences is the so called Overall Quick View.
This table is retrieval dependent (pT, h2o, o3, hno3, ch4, n2o, no2); it means that the user has to choose a specific retrieval, by pressing one of the button presented on the top of the table. Of course a retrieval selection is auto-exclusive, and analysis over each retrieval has to be carried out alone.
At this point the user is allowed to change and improve selection by marking or unmarking specific files from the list.
Fast selection or deselection of all sequences listed is available on Utility menu. Through the same menu, the new list of sequences, obtained by deselecting specific sequences can be saved on the storage device for further analysis.

Step 3

When the group of sequences and the retrieval are defined, the user finally enter inside the process.
A number of graphical information are available in Analysis menu, such as:

Chi square as a function of sequence
And others

Step 4

The user is now allowed to choose between Products Visibility and analysis of Housekeeping (i.e. auxiliary data).

B Pressure, Temperature, Volume Mixing Ratio and Continuum: Products Visibility

Step 0

A group of sequences is selected in previous section; at any time the user is allowed to change retrieval, without coming out of Product analysis;

Step 1

The user can decide to display the results of retrieval.
First of all the correction to the tangent altitude can be shown versus "a priori" estimated values.
Then the retrieved values of pressure, temperature, or volume mixing ratio for a specific gas are available. All values displayed are obtained by averaging over all sequences selected and presented versus tangent altitude.
Finally, continuum values are available to display,

Step 2

The user can save all plots displayed in ps format, for further application.

C The Occupation Matrix and Historical details: Housekeeping for Statistical analysis

Step 0

A group of sequences is selected in previous section. Now the user picks up a reference Occupation Matrix, here called Nominal, by entering in Utility menu. Nominal Occupation Matrix is coded in a separate file, and generated outside of Statistical Tool.

Step 1

The user can enter inside of Housekeeping section by pressing third menu of Overall Quick View. At any time the user is allowed to change retrieval, without coming out of Housekeeping.

Step 2

An internal control occurs between Nominal Occupation Matrix and sequence Occupation Matrix related to specific retrieval.
The user is informed about how many sequences take part to statistical population thanks to a numerical indication at the end of each line of sight, an shown in the picture.
The process analysis will be carried out only over matching line of sight of Occupation Matrix.

9 Data exchanged with ORM procedure.

Inside of ORM procedure a routine has been added in order to extract data useful to Statistica Tool. See [RD2] for further details.

10 Test of Actual Version of Statistical Tool

Present version has been tested on Linux machines, with the following Operative Systems:

RedHat release 6.01
Debian