





Requirements for model output and observation data formats for Verification Tools

Some background info.....

- Long-term strategy on COSMO verification software was approved (GM2015)
- Further to that, STC decided that with regard to the future strategy for verification tools, support should be provided for the COSMO Additional Verification Tools (AVT) using a **common data format**
- WG5 and TAG was asked to prepare a policy to implement the requirements addressed in the verification tools strategy document. Also, this is linked with the request for PP INSPECT to prepare the project document for a policy on model output and measurement data formats used within the project and, subsequently, on formats used by different spatial verification tools.

Additional Verification Tools (AVT) are targeted to provide code for the implementation of any “useful” verification methods (in addition to the ones included in the CVS).

Each software or code has to be well documented but no technical support is necessary to be provided.

The adaptation to input datasets, the connection to local Databases, installation specifics and any maintenance issue will be the responsibility of each member that is interested to use them.

AVT, **in relation with INSPECT**, includes spatial methods software (VAST for fuzzy verification, SpatialVX various libraries for displacement or feature based methods)

Common agreement

When estimating the effort required for the adjustment of verification tools to local DB and I/O formats, it should be noted that:

- **Any data format adaptation required (by the verification software) will have to be performed externally**, separating in this way the main focus of the verification tools, which is the correct application of statistical methods, from the preprocessing of data or the graphical representation, which is the focus of other tools.
- **Fieldextra is official COSMO software and its use is suggested for the postprocessing of any gridded data** (forecast or observation field), while LIBSIM and other tools can be utilized for other forms of observations (e.g. BUFR).
- **AVT software should be followed by input data format standards**, in this way making the adaptation of each tool easier for most users and facilitating the exchange of data for common experiments.

Input file for Additional Verification Tools (AVT) - INSPECT related

VAST software - spatial verification methods for neighborhood approach.

- ✓ Main code uses CSV (Comma Separated Values) files as precipitation input data, identical in dimension and in a regular lat-lon horizontal spacing (obs and fct)
- ✓ Actions: regridding, production of precipitation accumulations, GRIB to CSV conversion
- ✓ offers the possibility (optional) to perform data input preprocessing (LIBSIM routines)
Obs (GRIB or BUFR) and Fct (GRIB) format can be converted into CSV VAST input files
Prerequisite that initial files must be in higher resolution and cover a larger spatial area
- ✓ Alternatively, data preparation could be performed with FieldExtra

Longitude	Latitude	Value
6.315	43.010	-999.000
6.352	43.010	-999.000
6.389	43.010	-999.000
6.426	43.010	-999.000
6.463	43.010	-999.000
6.500	43.010	-999.000
6.538	43.010	-999.000
6.575	43.010	-999.000
6.612	43.010	0.000
6.649	43.010	0.000
6.686	43.010	0.000
6.723	43.010	0.000
6.760	43.010	0.000
6.797	43.010	0.000

Fieldextra is unable to process non-gridded data (BUFR) and gridded data other than that in GRIB(1/2) f (i.e. it is used primarily for forecast model output processing).

Input file for Additional Verification Tools (AVT) - INSPECT related

SpatialVx software - assortment of R libraries for applying a large variety of spatial verification methods (neighborhood, scale separation, feature-based, field deformation).

- ✓ Most of the libraries call for gridded data in an arbitrary textual format
- ✓ Functions to manipulate the data are provided if required by a certain method: e.g. create field accumulations and identify important features (e.g. to determine a precipitation object)
- ✓ Observation and forecast data must be given in identical format, but not necessarily on a regular lat-lon grid
- ✓ Preprocessing (e.g. decoding, regridding, interpolation) of model output (GRIB1/2), surface observations (BUFR) or radar/satellite data (NetCDF, UF, HDF, XML) is usually performed outside SpatialVx, but R packages are available to perform part of the preprocessing tasks.

Input file for Additional Verification Tools (AVT) – non INSPECT related

DWD Verification Software (Rfdbk package) - structured on an R-based environment,

- taking advantage of “feedback files” that are produced during the assimilation cycle
- It can perform verification using both SYNOP and TEMP for all parameters that are included in the DA files
- input files required by the Rfdbk software have to be produced with MEC
- MEC requires model data in GRIB format while for observations:
 - fof/mon/cof/ekf/ver-files (existing fdbkfiles extracted from nudging analysis)\
 - CDFIN (BUFR converted by buf2netcdf to NetCDF, BUFR in WMO-templates as used by DWD)

The MEC output files used as input to the Rfdbk software are in **NetCDF** format

Data preprocessing of FORECAST data

Forecast model output is used in either GRIB1/2 (e.g. VERSUS), NetCDF (e.g. Rfdbk) or textual format (e.g. CSV)

Processing as AVT input

- Fieldextra used for GRIB1/2 to text (CSV) conversion and for various processing of GRIB1/2 files
- GrADS (Grid Analysis and Display System) to convert and/or interpolate GRIB1/2 to ASCII (also to NetCDF for one variable). PPINSPECT: convert COSMO grib data into ASCII input for R SpatialVx (examples with templates available in WG5 repository).
- R-packages [RNetCDF](#) and NCDF to import the NetCDF data in R. For data in netCDF4, an R-package ncdf4 is also available.
- Other R libraries exist for some of the preprocessing tasks. *rNOMADS* used for reading GRIB files in R tuned to NOAA model that allows R users to download global and regional weather model data for processing.
- NetCDF Java library can be used to convert a GRIB file into a NetCDF format on Windows platform. NCL (NCAR command language) can read and write netCDF-3, netCDF-4 classic, netCDF-4, HDF4, binary, and ASCII data. It can read HDF-EOS2, HDF-EOS5, GRIB1, GRIB2, and OGR files

Data preprocessing of OBSERVATION data

BUFR is the WMO standard binary code for the representation and exchange of observational data (point observations)

Processing as AVT input

BUFR files can be decoded to text format with `grib_api` functions (ECMWF library) or processed with libraries included in LIBSIM. In addition with the use of LIBSIM, it is possible to aggregate observations into low-resolution regular boxes or aggregate them into arbitrary polygons

Radar data (HDF4-5, XML)

OPERA Data Center creates 3 composite products: instantaneous surface rain rate, instantaneous max reflectivity and 1 hour rainfall accumulation. Data are available in both **BUFR** and **HDF5** formats.

Like all observational data, many issues (e.g. problems with parallax, radar beams above the precipitation at long ranges, evaporation of rainfall at lower levels beneath the beam, anomalous propagation, etc.) have to be considered when using such data for verification purposes.

Data preprocessing of OBSERVATION data

Satellite data sources (HDF4, GRIB, BUFR)

For PP INSPECT and other WG5 applications, satellite data from the following sources/formats are used:

- SEVIRI/Meteosat (EUMETSAT) data: cloud cover, rain intensity, surface temperature and albedo. Formats: HDF4, GRIB, and ASCII.
- Polar-orbiting satellites (NOAA, Terra, Aqua, Suomi NPP): vertical profiles, Total Precipitable Water (TPW), and ozone. Formats: NetCDF, HDF4, and ASCII.
- EUMETCast products (SAF EUMETSAT): rain intensity, precipitation accumulations, snow cover, and surface humidity. Formats: GRIB, BUFR, and HDF4

Processing as AVT input

HDF5 files can be read directly into the R based SpatialVx software by using “HDF5 interface to R”: Bioconductor <https://www.bioconductor.org/>.

Examples of the procedures that have been followed for the decoding of binary observations (radar and satellite) and their conversion to input files for SpatialVx applications as required for PP INSPECT are available in the WG5 repository: <ftp.meteoam.it/WG5-Repository>.

Suggestions

- Due to the **wide variety of formats** and other issues that must be considered depending on the instrument, measured parameter/index, volume of data, geographic projection, etc., many approaches can be followed to prepare information as input for AVT
- Such observation and forecast data format conversion is usually performed prior to use of AVT
- While it would be **desirable to have a unique COSMO tool dedicated to preparation of verification data**, it is currently not deemed possible given the diversity of database structures, types of observational data used, requirements to interface with models other than COSMO, etc.
- Nevertheless, **any future COSMO data processing software development should also aim to fulfill the requirements of verification applications**. Treatment of non-gridded binary observations (BUFR) as well as the handling of data formats such as NetCDF or HDF4/5 are considered to be desirable features of any such preprocessing software.
- NetCDF format is an alternative format for int2lm input and COSMO model output, and in general it has the potential of becoming a standard data exchange format. **It is suggested that NetCDF support should be included in existing COSMO preprocessing software**

Remarks from TAC (JMB):

- * Good overview document !*
- * It is planned to add NetCDF import (CF standard) in the next major release of fieldextra, available at the end of this year*
- * It is envisaged to add a 'friendly' R export in fieldextra, to facilitate the usage of fieldextra as a pre-processing software for specialized R applications*
- * With these two additions, and some code tuning, fieldextra should be able to fulfill many of the AVT pre-processing needs **for gridded fields** (model, satellite, radar).*
- * One advantage of using fieldextra over free tools, is that it is well tested, fully compatible with COSMO/ICON/IFS, robust, and help the user avoid mistakes thanks to the many build-in tests*
- * One could maybe define a COSMO PT 'fieldextra as AVT pre-processing tool'. MeteoSwiss would appreciate additional help.*

Ideas for discussion.....

	Common formats	AVT format	Actions	Code	Resources
Forecast	GRIB1/2 NetCDF	GRIB1/2 Textual NetCDF	QC Decode Interpolate Extremes Change of projection	R libraries Fieldextra LIBSIM	COSMO
Obs	GRIB1/2 NetCDF	Textual	Aggregation (spatially/timely)	R libraries Fieldextra LIBSIM GRADS	

Opinions?

Decide on the possibility to expand current pre-processing tools (Fieldextra, LIBSIM, etc.) through COSMO PT, to include as input netCDF files. Fieldextra does not handle non-gridded data.

To an ambitious extend, it can be examined the possibility of the development of a common observation pre-processing system to provide observations in an appropriate format for verification. Aim to avoid duplication of work on observation pre-processing (serve also DA purposes). Such tool could comprise mainly decoding, conversion to local DBs, simple QC and conversion to a format for AVT with special attention to remote sensing data.

Exchange of “community code” for verification input data processing. Expand WG5 repository with code and/or instructions for AVT input data preparation (currently restricted only through PPINSPECT Task2b)

