Evaluation Tools for GEM-MACH

Alex Lupu

Environment and Climate Change Canada
Environnement et Changement climatique Canada

alexandru.lupu@canada.ca

December 15, 2015

Overview

Unix Environment Setup

IDL 8.4

```
. ssmuse-sh -d /ssm/net/hpcs/ext/idl-8.4
```

Fortran 90 compilers PGI 14.1 and IBM XL 13.1

```
. ssmuse-sh -d hpcs/201402/02/base -d hpcs/201402/02/pgi1401
. ssmuse-sh -d hpcs/201402/02/base -d hpcs/ext/xlf_13.1.0.10
```

RPN libraries and utilities 15.2

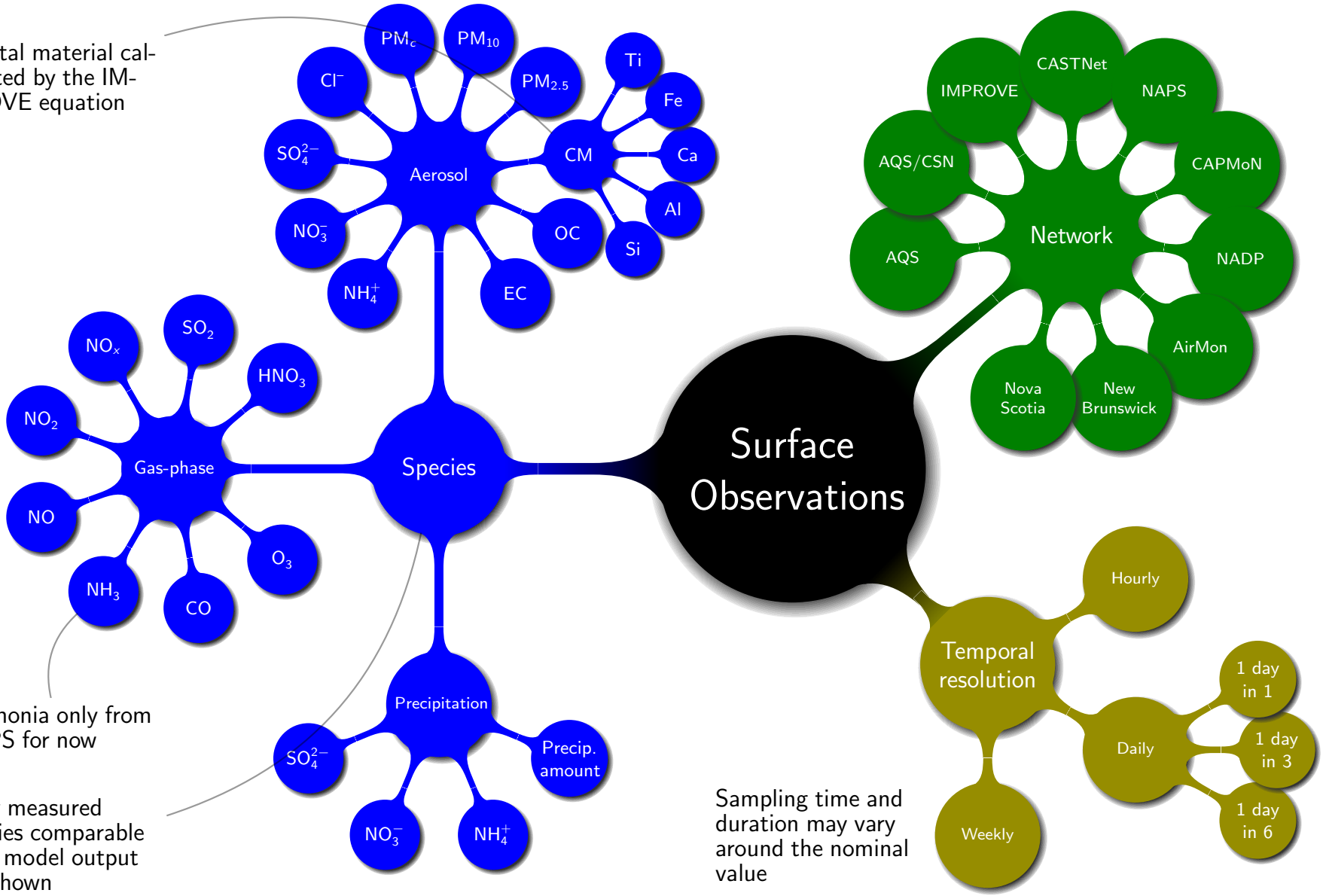
```
. ssmuse-sh -d rpn/libs/15.2
. ssmuse-sh -d rpn/utils/15.2
```

netCDF libraries 3.6.1 for PGI 14.1 and IBM XL 13.1

SSM_TOOLS=/home/ordenv/ssm-domains/ssm-tools

```
NCDF_INC=$SSM_TOOLS/netcdf_3.6.1_multi/include/Linux_x86-64/pgi9xx
NCDF_LIB=$SSM_TOOLS/netcdf_3.6.1_multi/lib/Linux_x86-64/pgi9xx
NCDF_INC=$SSM_TOOLS/netcdf_3.6.1_multi/include/AIX-powerpc7/xlf13
NCDF_LIB=$SSM_TOOLS/netcdf_3.6.1_multi/lib/AIX-powerpc7/xlf13
```

Surface Observation Datasets for Year 2010



Format of Observation Data Files

NAtChem data exchange standard

- Self-describing, csv dataset (metadata included in the file)
- Based on the NARSTO format
- All species are provided in this format except for the NAPS speciated aerosol measurements
- One file per network
- Reference: `http://cdiac.ornl.gov/programs/NARSTO`

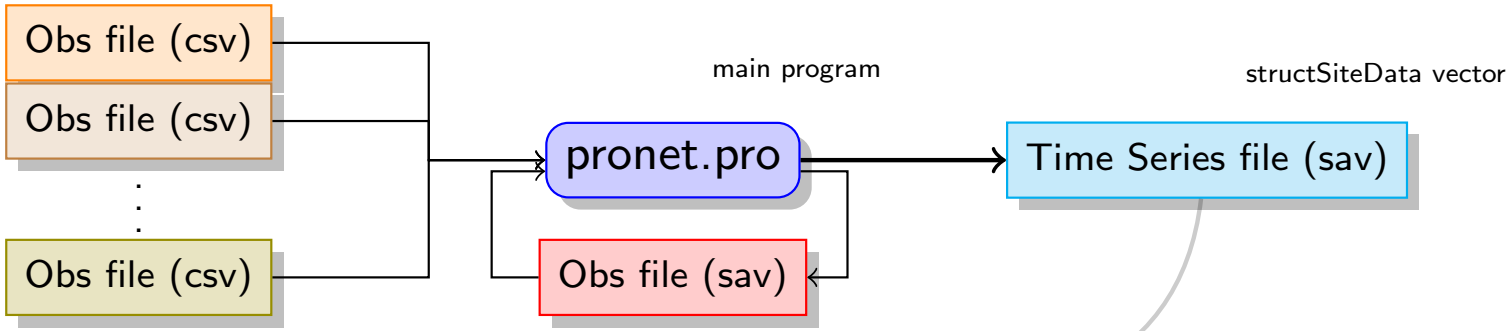
NARSTO: North American Research Strategy for Tropospheric Ozone

NAPS Excel tables

- NAPS speciated aerosol measurements
- No standard (inconsistent format between years)
- No validity flags, but 'quality assured'
- One file per station and instrument
- Files were converted to csv with a VisualBasic script
- Reference: `http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx`

NAPS: National Air Pollution Surveillance Program

Pre-processing of Surface Observations: pronet.pro



IDL program pronet.pro reads in the csv data files and saves the observation time series filtered by validity flags (if present) to IDL binary sav file(s)

Output

- One time series file per species, instrument/method and network

Location of files

Warning Location of files will most likely change.

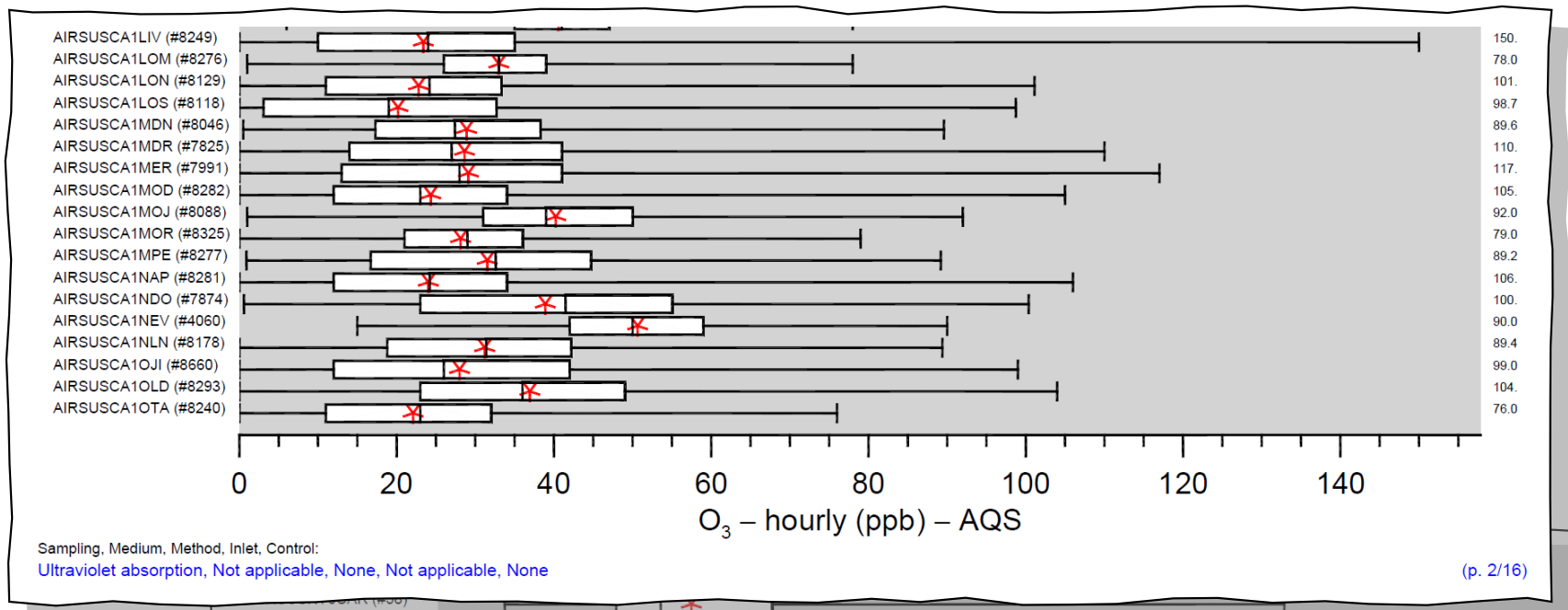
- IDL code: `~arqpalu/work/eval/dev/`
- Time series files (IDL binary sav): `~arqpalu/cnfs/eval/ts/`
- Original csv data files: `~arqpalu/cnfs/2010_dat/`

Time Series IDL structure

```
1  structSiteData = { ssd, $
2     siteID : '', $
3     siteDescription : '', $
4     lat : 0., lon : 0., elev : 0., lu : '', tz : 0., $
5     jdStart : ptr_new(), jdEnd : ptr_new(), $
6     c : ptr_new(), cx : ptr_new(), units : '', $
7     cntFlags : ptr_new(), $
8     sampling : '', medium : '', control : '', $
9     method : '', inlet : '', low : '', upp : '' $
10 }
```

- `siteID` – site identification code
- `lat`, `lon` – latitude and longitude
- `jdStart`, `jdEnd` – pointers to vectors of start and end of sampling times (Julian dates)
- `tz` – time offset from UTC
- `c` – pointer to vector of mixing ratios or concentrations
- `cx` – pointer to auxiliary vector, e.g., used to store precipitation amount for precipitation chemistry measurements

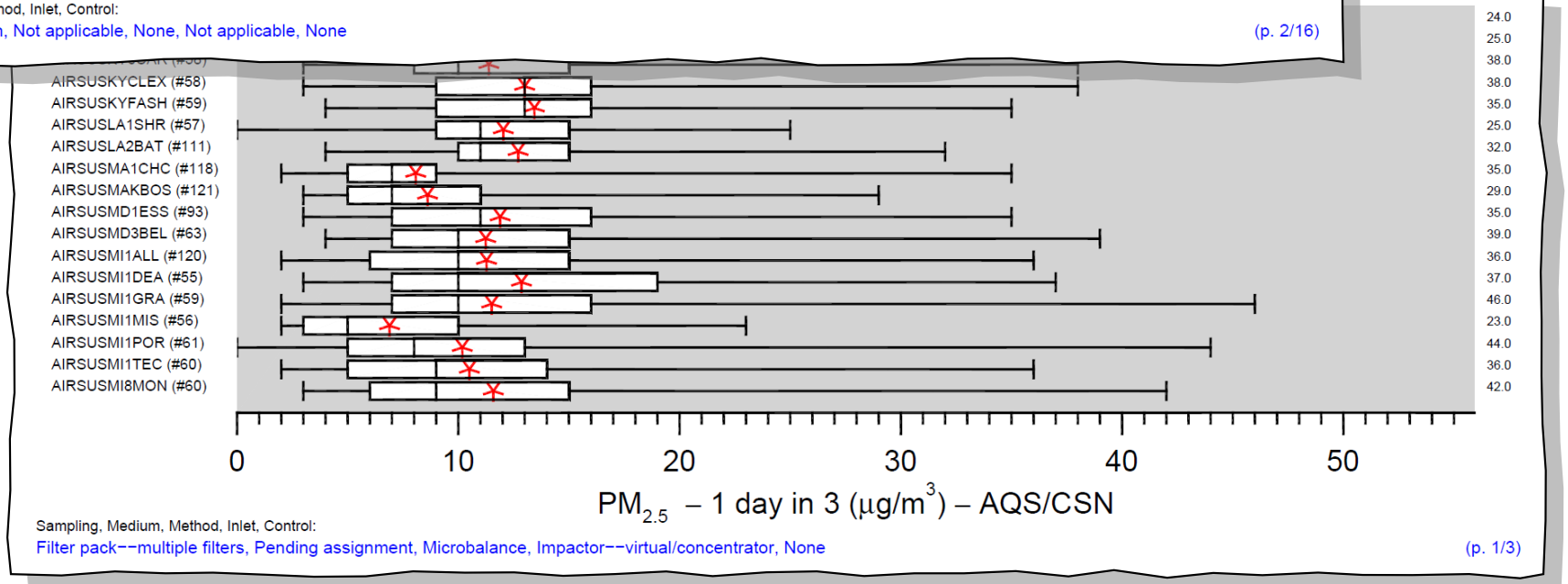
Data Summary by Station – AQS O_3 and AQS/CSN $PM_{2.5}$



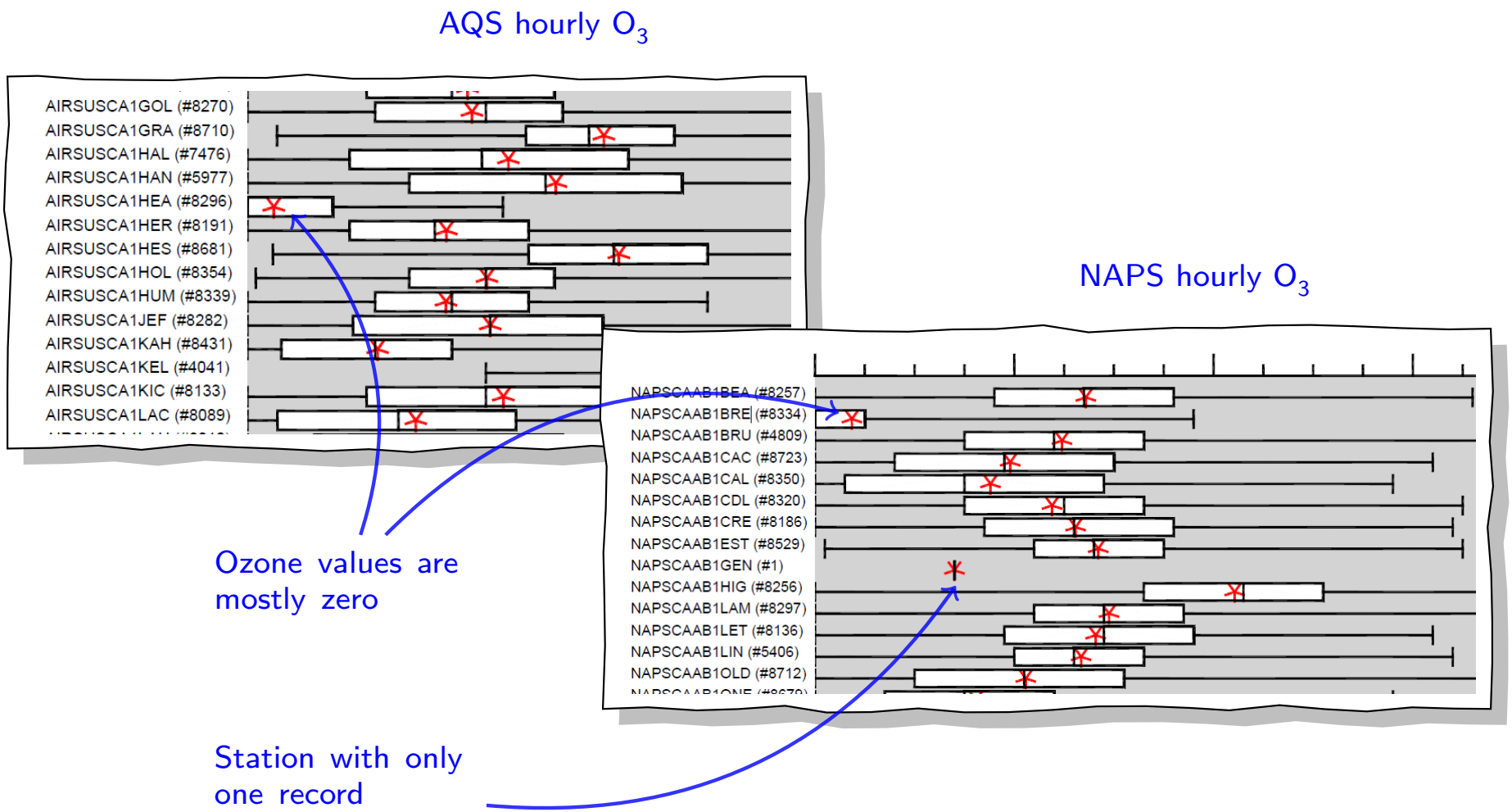
AQS O_3

plots produced by airchemplotbox.pro

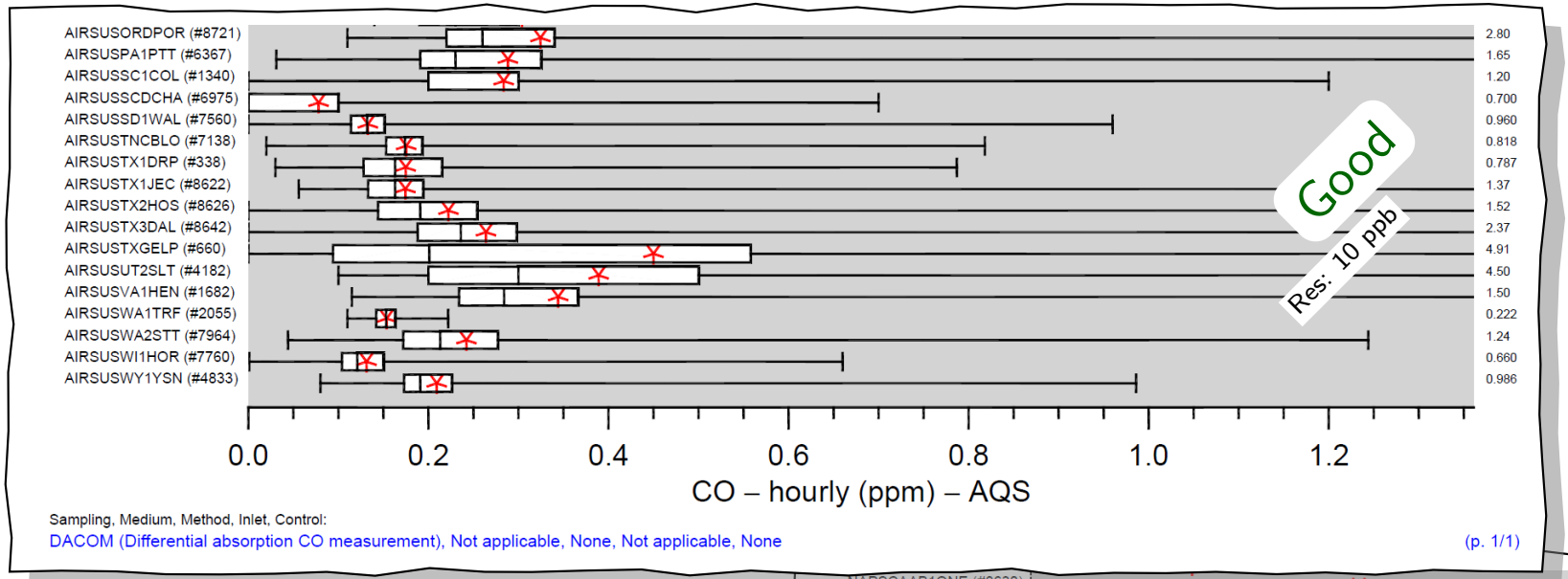
AQS/CSN $PM_{2.5}$

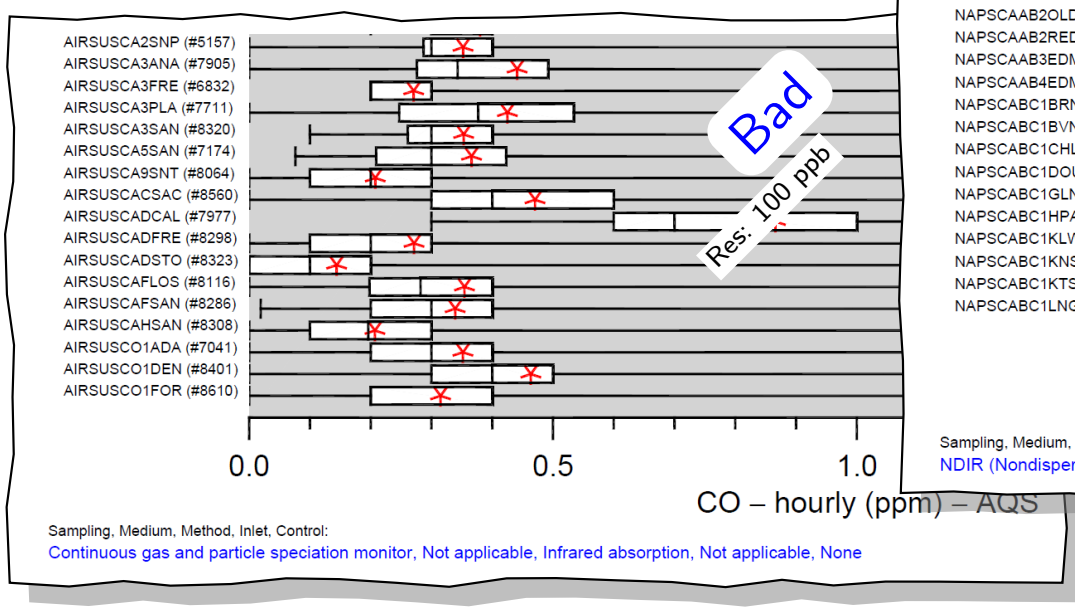


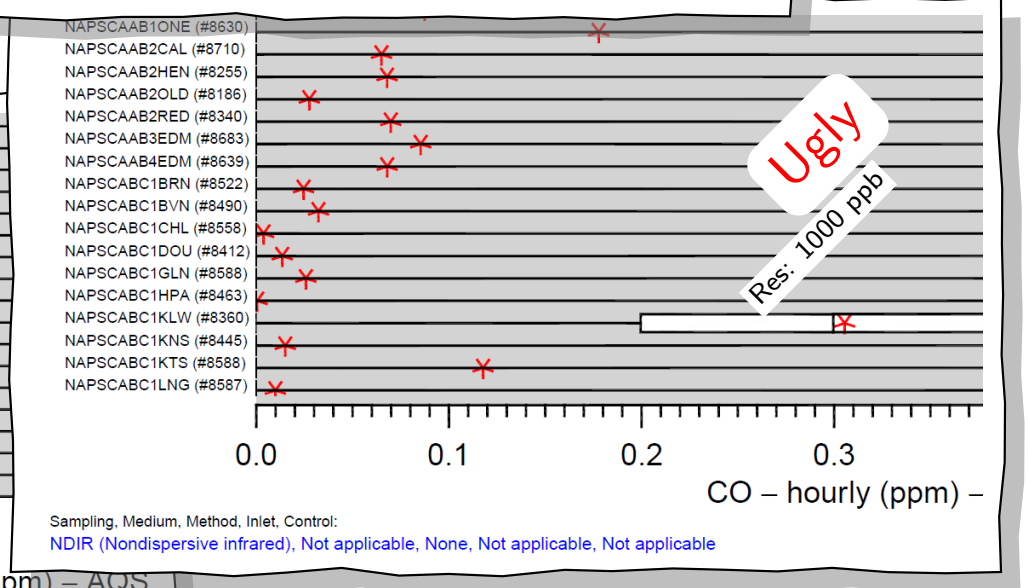
Data Summary by Station – Identifying problems



Data Summary – CO measured by different methods







Offline Extraction of Model Time Series: xtr.F90

Fortran program xtr.F90 performs bilinear interpolation at site locations for air chemistry and nearest neighbour for precipitation chemistry

Shell script: xtr_loop.sh

```
1  dhome="/home/pxarqd/arqp/alu/work/dip/oxtr"
2  MDIR0="gmv2r_r1922_05a"
3  DIN="/cnfs/dev/aq/aq02/arqp/alu/dip/xtrin/"
4  DOUT="/cnfs/dev/aq/aq02/arqp/alu/dip/$MDIR0/ts/"
5  VARLIST="ts.varlist"
6  sdat='2010010100'
7  fdat='2010123112'
8  cdat=$sdat # current date
9  incr=12
10 tini=1
11 tfin=24
12 maxnproc=8 # max number of processors
13 cd $dhome
14 nproc=0
15 while [ "$cdat" -le "$fdat" ]; do
16    nproc=$(($nproc+1))
17    MDIR=$MDIR0/${cdat}0000/model/
18    xtr.Abs $cdat $tini $tfin $MDIR $DIN $DOUT $VARLIST &
19    if [ "$nproc" -ge "$maxnproc" ]; then
20       wait
21       nproc=0
22    fi
23    cdat=`r.date $cdat +$incr`
24    cdat=$(echo $cdat | cut -c1-10)
25 done
26 wait
```

Script variables

- MDIR0 – path to *hourly surface* model output
- DIN – path to input files containing list of stations and model fields
- VARLIST – name of file containing list of model fields
- DOUT – path to time series output (netCDF)
- sdat, fdat – start and end date (yyyymmddhh)
- incr – time interval between forecasts in hours
- tini, tfin – first and last forecast hour to be saved

List of stations

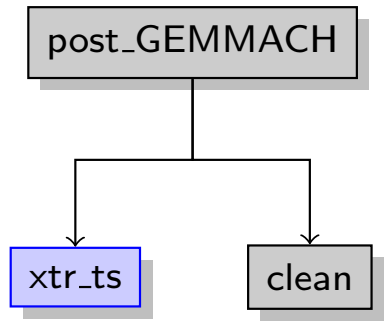
```
1 'NAPSCAYT1HER'   60.718601 -135.050003
2 'CASTUSAK1DEN'   63.725800 -148.960007
```

List of model fields

```
1 'TO3'    'o3_air.stations'       'surf'
2 'TO3'    'o3_woudc.stations'     'prof'
3 'TJ'     'o3_air.stations'       'diag'
4 'WNO3'   'no3m_precip.stations'  'wdep'
```

Online Extraction of Model Time Series: xtr.F90

Program xtr.F90 can be run online as a Maestro task. Useful if vertical profile extraction is desired. After extraction, unnecessary 3D fields that can take up a lot of disk space can be removed from the output by running editout.F90.



Maestro task: xtr_ts.tsk

```
1  if [ -f ${TASK_INPUT}/stations/${VARLIST} ];
2  then
3    . ${TASK_INPUT}/date.cfg
4    # extract time series; *** assuming hourly output
5    ${TASK_BIN}/xtr $(echo ${start_date} | cut -c1-10) \
6         0 $(( ${Fcst_end_S} / 3600 )) \
7         ${TASK_INPUT}/model_hyb_in/ \
8         ${TASK_INPUT}/stations/ \
9         ${TASK_OUTPUT}/model_hyb_ts/ \
10        ${VARLIST}
11   # parse outcfg.arch
12   python ${TASK_BIN}/readoutcfg.py \
13       -g ${TASK_INPUT}/gem_settings.nml \
14       -o ${TASK_INPUT}/outcfg.arch \
15       -v ${TASK_WORK}/varlist.arch
16   # remove output fields not required for archiving
17   ${TASK_BIN}/editout $(echo ${start_date} | cut -c1-10) \
18        ${TASK_INPUT}/model_hyb_in/ \
19        ${TASK_WORK}/varlist.arch
20 else
21   echo "[xtr_ts] no varlist file: task skipped"
22 fi
```

Sub-tasks

1. xtr.Abs – extract surface and/or vertical profile time series at specific locations (see slide 10)
2. readoutcfg.py – parse output configuration file (outcfg.arch) and create file with list of variables (varlist.arch) to be passed to editout.Abs
3. editout.Abs – remove output fields not required for archiving

Code location

- Time series extraction code: ~arqpalu/work/out/oxtr/dev/
- Outcfg.out Python parser: ~arqpalu/work/out/opars/dev/
- G-M output archiving code: ~arqpalu/work/out/oedit/dev/
- Maestro project: ~arqpalu/.suites/tjdev/

Structure of Model Time Series File

- Files are in netCDF format
- One file per field (species) and forecast
- The model field is saved as a 3D float array. Its dimensions are: the station index, the vertical level index, and the time since start of integration (in hours)
- The name of the file contains the forecast start time, the name of the variable (as in the model output), and the name of the file with list of stations

netCDF file structure

```
1  netcdf 2010123112_TO3_o3_air.stations_ts {
2  dimensions:
3     station = 1555 ;
4     time = UNLIMITED ; // (24 currently)
5     level = 1 ;
6
7  variables:
8     float time(time) ;
9        time:standard_name = "time" ;
10       time:units = "hours since 2010-12-31 12:00:00" ;
11    float TO3(time, level, station) ;
12
13 // global attributes:
14       :featureType = "timeSeries" ;
15 }
```

IDL Matching Code: matchts.pro

Motivation

- Observations come in various temporal resolutions. While hourly data can be matched directly with the model output, for consistent comparison with observations that have larger sampling periods (days, weeks), the model output has to be averaged or integrated accordingly

- The IDL program matchts.pro pairs observations of any temporal resolution with corresponding model output averages or integrals.
- The code takes as input the **observation time series IDL sav** file(s) produced by pronet.pro and the **model time series netCDF** files generated by xtr.F90.

IDL Matching Code: matchts.pro

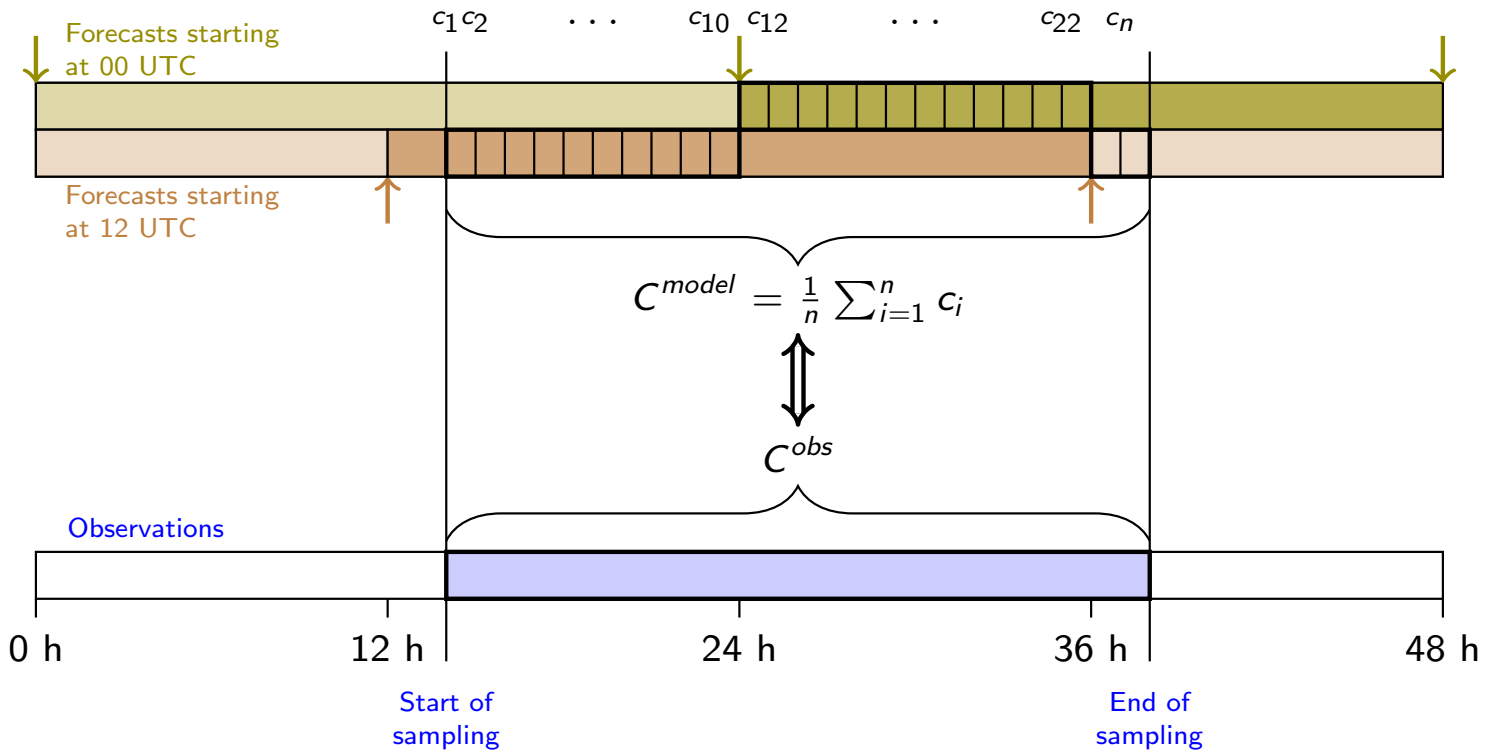


Figure: Diagram illustrating the matching algorithm. For each observation, the model output that falls within the sampling period is identified ($c_1, c_2, \ldots, c_n$), an averaging (for air chemistry variables – shown) or accumulation (for wet deposition fluxes) of these values is performed, and the corresponding pair (C^{obs}, C^{model}) is saved to an ascii file. If two (or more) forecast values are available for the same point in time, the one that comes from the most recent forecast is used. The figure shows the case with two 24-h forecasts starting at 00 and 12 UTC, respectively, and hourly output.

Pairing of Observations with Model Output

Shell script: `matchts.sh`

```
dhome="/home/pxarqd/arqp/alu/work/eval/dev/"
obsPath="/cnfs/dev/aq/aq02/arqp/alu/eval/ts/"
stationPath="/cnfs/dev/aq/aq02/arqp/alu/xtrin/"
modPathBase="/cnfs/dev/aq/aq02/arqp/alu/xtrout_r1864/"
outPath="/cnfs/dev/aq/aq02/arqp/alu/mots_r1864_v02/"
params="dtStartHr = 12, dtModHr = 1, tOffHr = 1"
species="$1"
pathList="obsPath='$obsPath', stationPath='$stationPath', \
   modPathBase='$modPathBase', outPath='$outPath'"
cd $dhome # working directory
for s in $species ; do
   cat > batch_$s.job <<-EOF
      cd $dhome
      idl -e "matchts, species = '$s', $pathList, $params"
   EOF
   ord_soumet batch_$s.job -mach pollux  -cm 8G \
      -t 10800 -listing ~/listings/pollux/
done
```

Code location

`~arqpalu/work/eval/dev/matchts.pro`

Script variables

- `species` – species name, passed as command-line argument (air: `O3`, `NOX`, `CO`, `SO2`, `HNO3`, `SO4`, `NO3`, `NH4`, `CM`, `PM25`, `PM10`, `PMC`, `OC`, `EC`, `CL`; precipitation: `SO4_PR`, `NO3_PR`, `NH4_PR` (deposition fluxes), `SO4_PR_CONC`, `NO3_PR_CONC`, `NH4_PR_CONC` (concentration), `PRECIP`)
- `dhome` – location of IDL code (main program: `matchts.pro`)
- `obsPath` – path to IDL sav files containing observed time series (see slide 6)
- `modPathBase` – path to netCDF files containing model time series (see slide 10)
- `stationPath` – path to files containing list of stations (for now, station files must be the same as those used for extracting the model time series – see slide 10)
- `outPath` – output location (model–observation pairs)
- `dtStartHr` – time interval between forecasts (must be equal to `incr` – see slide 10)
- `dtModHr` – model output time interval in hours
- `tOffHr` – time offset in hours (for future use; for now, must be equal to `tini` – see slide 10)

Structure of Model-Observation Files

CASTNet weekly HNO_3 air concentration

```
#-------------------------------------------------------------------------------------------------
#SITE ID         SITE DESCRIPTION        LAT       LONG     ALT      LAND USE   TZ OFF       UNITS
#CASTCAON1EGB             Egbert    44.2317   -79.7840   251.0         Rural    -5.00      ug sm-3
#-------------------------------------------------------------------------------------------------
#SITE ID             TIME START               TIME END                    OBS         MODEL
CASTCAON1EGB         2010-01-05T13:26-05:00  2010-01-12T16:58-05:00       0.29100      0.40304
CASTCAON1EGB         2010-01-12T16:58-05:00  2010-01-19T12:19-05:00       1.06700      1.38875
CASTCAON1EGB         2010-01-19T12:26-05:00  2010-01-26T11:32-05:00       0.23800      0.49319
CASTCAON1EGB         2010-01-26T11:50-05:00  2010-02-02T13:44-05:00       0.26500      0.25621
CASTCAON1EGB         2010-02-02T13:55-05:00  2010-02-09T11:09-05:00       0.36000      0.08407
CASTCAON1EGB         2010-02-09T11:24-05:00  2010-02-16T11:26-05:00       0.16800      0.05477
CASTCAON1EGB         2010-02-16T11:50-05:00  2010-02-23T12:40-05:00       0.12200      0.13647
CASTCAON1EGB         2010-02-23T12:56-05:00  2010-03-02T11:14-05:00       0.21900      0.11606
# ...
CASTCAON1EGB         2010-12-18T12:17-05:00  2010-12-25T10:40-05:00       0.49600      0.31223
CASTCAON1EGB         2010-12-25T10:51-05:00  2011-01-01T11:00-05:00       0.26100      1.11110
#-------------------------------------------------------------------------------------------------
#SITE ID         SITE DESCRIPTION        LAT       LONG     ALT      LAND USE   TZ OFF       UNITS
#CASTUSAK1DEN           Denali Np    63.7231  -148.9700   661.0         Rural    -9.00      ug sm-3
#-------------------------------------------------------------------------------------------------
#SITE ID             TIME START               TIME END                    OBS         MODEL
CASTUSAK1DEN         2010-01-05T11:35-09:00  2010-01-12T18:02-09:00       0.18000      0.35415
CASTUSAK1DEN         2010-01-12T18:10-09:00  2010-01-19T16:08-09:00       0.27700      0.29123
CASTUSAK1DEN         2010-01-19T16:14-09:00  2010-01-26T18:03-09:00       0.12500      0.29466
CASTUSAK1DEN         2010-02-02T16:59-09:00  2010-02-09T17:53-09:00       0.09100      0.54359
CASTUSAK1DEN         2010-02-09T17:57-09:00  2010-02-16T17:54-09:00       0.03300      0.10272
CASTUSAK1DEN         2010-02-16T18:01-09:00  2010-02-24T15:24-09:00       0.04700      0.24105
CASTUSAK1DEN         2010-02-24T15:29-09:00  2010-03-02T19:51-09:00       0.08300      0.34160
CASTUSAK1DEN         2010-12-21T19:29-09:00  2010-12-28T15:22-09:00       0.05900      0.22825
CASTUSAK1DEN         2010-12-28T15:35-09:00  2011-01-04T10:43-09:00       0.03900      0.13542
# ...
```

Temporal Aggregation Code: aggregate.pro

- The IDL program aggregate.pro calculates weekly ('w'), monthly ('m'), quarterly ('q'), seasonal ('s') and annual ('a') averages from the model–observation pairs produced by matchts.pro
- A **completeness criterion** can be applied to the data. Averages produced from a fraction of valid data that is less than a specified threshold are discarded.

  For precipitation chemistry data, the total precipitation associated with valid chemistry data has to be at least the specified fraction of the total precipitation over the averaging period.

Running the aggregation program (examples)

- Ozone weekly average

```
idl -e "aggregate, 'o3_uv-abs_aqs.asc' , 'w', factComp = 0.75"
```

- Ozone monthly average

```
idl -e "aggregate, 'o3_uv-abs_aqs.asc', 'm', factComp = 0.75"
```

- Sulphate seasonal precipitation-weighted mean concentration in rain water

```
idl -e "aggregate, ['so4prconc_capmon_precip.asc', 'pr_capmon_precip.asc'],
's', factComp = 0.85, /conc"
```

- Sulphate seasonal wet deposition

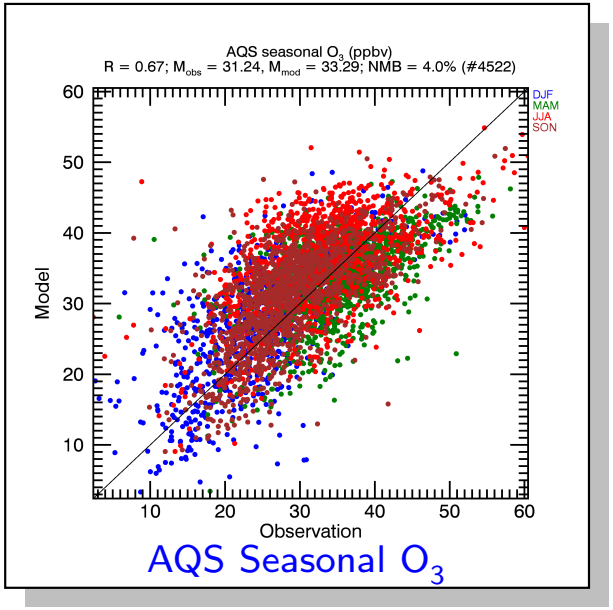
```
idl -e "aggregate, ['so4pr_capmon_precip.asc', 'pr_capmon_precip.asc'], 's',
factComp = 0.85, /accum"
```

Structure of Output File

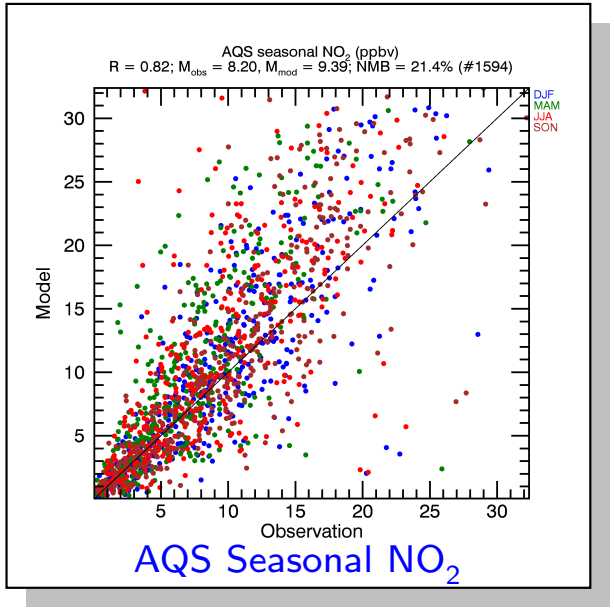
NAPS monthly mean SO_4^{2-} air concentration

```
#-------------------------------------------------------------------------------------------------------------------
#SITE ID         SITE DESCRIPTION        LAT       LONG     ALT      LAND USE   TZ OFF       UNITS      INTV      FREQ
#NAPSCAON1TOR  Toronto, Centennia    43.6485   -79.5914   137.0             R    -5.00      ug m-3        24       144
#-------------------------------------------------------------------------------------------------------------------
#SITE ID             TIME PERIOD                      OBS          MODEL          COMPL        N
NAPSCAON1TOR         01/02/2010 - 01/03/2010       0.74640        0.50832           0.18        5
NAPSCAON1TOR         01/03/2010 - 01/04/2010       1.91460        1.48467           0.16        5
NAPSCAON1TOR         01/04/2010 - 01/05/2010       2.17820        2.27218           0.17        5
NAPSCAON1TOR         01/05/2010 - 01/06/2010       2.30660        2.48560           0.16        5
NAPSCAON1TOR         01/06/2010 - 01/07/2010       1.58060        3.01655           0.17        5
# ...
NAPSCAON1TOR         01/11/2010 - 01/12/2010       1.94980        2.17894           0.17        5
NAPSCAON1TOR         01/12/2010 - 01/01/2011       1.65340        0.94683           0.16        5
#-------------------------------------------------------------------------------------------------------------------
#SITE ID         SITE DESCRIPTION        LAT       LONG     ALT      LAND USE   TZ OFF       UNITS      INTV      FREQ
#NAPSCABC3MEA  Whistler, Meadow P    50.1439  -122.9611   638.0             R    -8.00      ug m-3        24        72
#-------------------------------------------------------------------------------------------------------------------
#SITE ID             TIME PERIOD                      OBS          MODEL          COMPL        N
NAPSCABC3MEA         01/01/2010 - 01/02/2010       0.16912        0.04810           0.26        8
NAPSCABC3MEA         01/02/2010 - 01/03/2010       0.19560        0.17702           0.36       10
NAPSCABC3MEA         01/03/2010 - 01/04/2010       0.34060        0.23700           0.32       10
#-------------------------------------------------------------------------------------------------------------------
#SITE ID         SITE DESCRIPTION        LAT       LONG     ALT      LAND USE   TZ OFF       UNITS      INTV      FREQ
#NAPSCANU3IQA  Iqaluit, Wildlife     63.7500   -68.5170    33.0             U    -5.00      ug m-3        24       144
#-------------------------------------------------------------------------------------------------------------------
#SITE ID             TIME PERIOD                      OBS          MODEL          COMPL        N
NAPSCANU3IQA         01/01/2010 - 01/02/2010       0.52575        0.40232           0.13        4
NAPSCANU3IQA         01/02/2010 - 01/03/2010       0.33160        0.30879           0.18        5
NAPSCANU3IQA         01/03/2010 - 01/04/2010       0.43240        0.25816           0.16        5
NAPSCANU3IQA         01/04/2010 - 01/05/2010       0.50420        0.27493           0.17        5
# ...
```

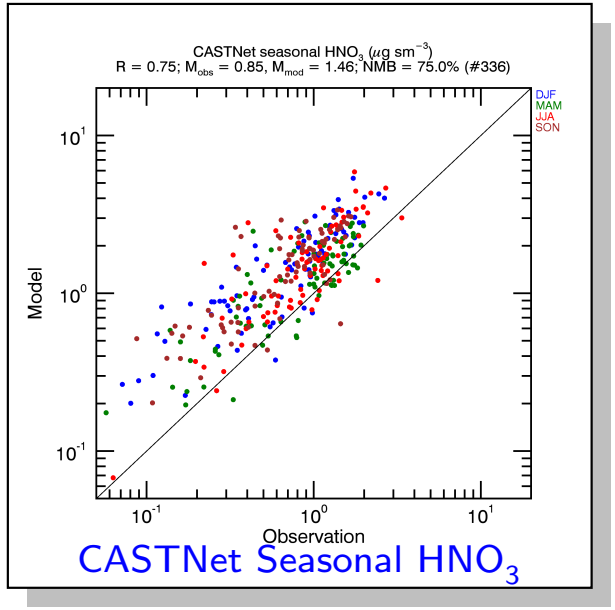
Scatter Plots



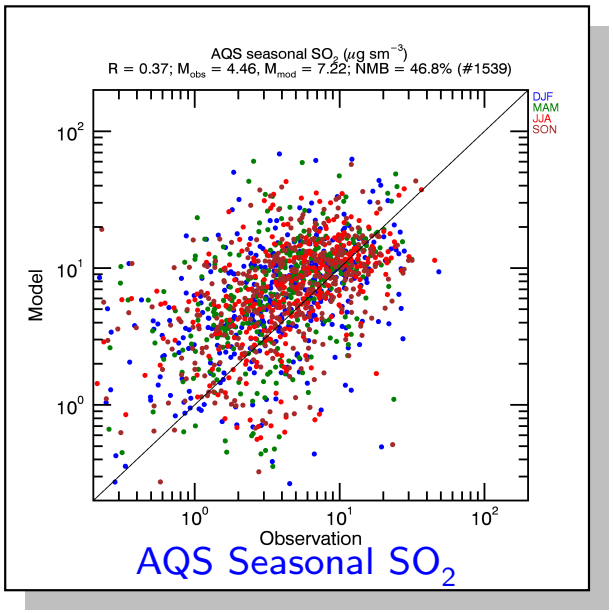
AQS Seasonal O_3



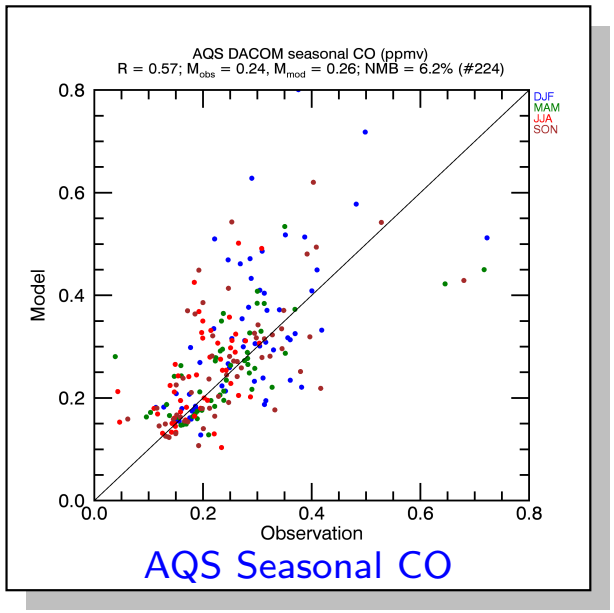
AQS Seasonal NO_2



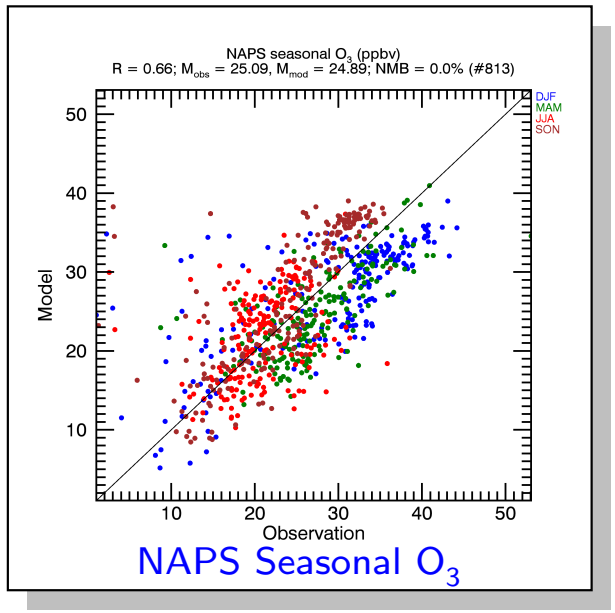
CASTNet Seasonal HNO_3



AQS Seasonal SO_2



AQS Seasonal CO



NAPS Seasonal O_3

plots produced by plotmodobs.pro