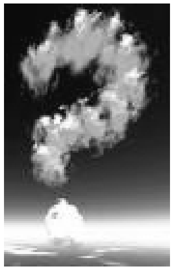
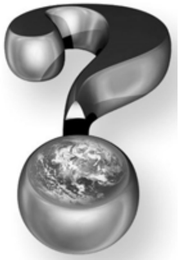
Model Evaluation Tools (MET) tutorial

4 – 5 February, 2009



- What?
  - Set of verification tools for evaluating forecasts via
    - standard statistics
    - object-based methods
    - scale decompositions

- Why?
  - Make verifiying easy.
  - Encourage verification.
  - Promote consistency across users.

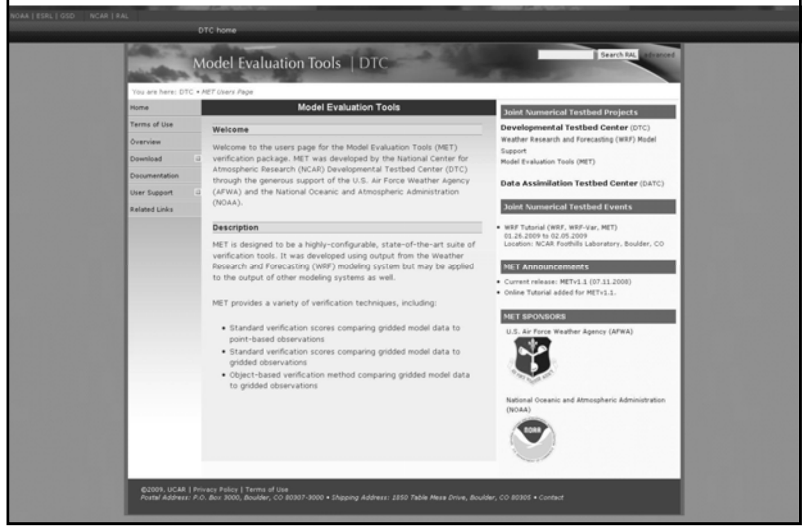


- How?
  - A (unix like) package of software tools and scripts.
  - Community contributed methods, graphics, etc.

Downloading MET

- MET distributed as a tarball to be downloaded and compiled locally.
  - METv1.1 released on July 11, 2008.
  - METv2.0 scheduled for February 2009.
  - METv2.0 BETA pre-installed on tutorial machines.
  - Register and download: **www.dtcenter.org/met/users**
- Language:
  - Written primarily in C++ with calls to a Fortran library
- Supported Platforms and Compilers:
  - Linux with GNU compilers
  - Linux with Portland Group (PGI) compilers
  - IBM machines with IBM compilers

www.dtcenter.org/met/users



Dependencies

- REQUIRED:
  - GNU Make Utility
  - C++ and Fortran compilers (GNU, PGI, or IBM)
  - NetCDF version 3 Library
  - BUFRLIB Library
  - GNU Scientific Library (GSL)
  - F2C Library (f2c or g2c, for some compilers)
- RECOMMENDED:
  - WRF Post-Processor
  - COPYGB (included with WRF-Post)
  - R statistics and graphics package
  - *CWORDSH (not needed for v2.0)*

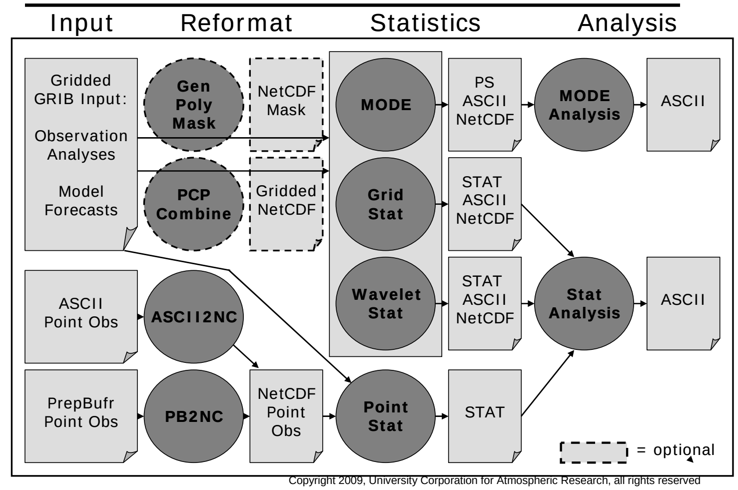
Directory Structure

File or Directory	Contents
README	Installation instructions and release notes.
Makefile_gnu (pgi, ibm)	Top-level Makefile to be configured prior to building MET.
lib/	Source code for internal MET libraries.
src/	Source code for the MET applications.
doc/	MET User's Guide.
bin/	Built MET executables.
scripts/	Test scripts to be run after building MET.
data/	Sample data used by the test scripts.
out/	Output generated by the test scripts.

Building MET

- Steps for building MET:
  1. Build the required libraries with the same family of compilers to be used with MET.
  2. Select the appropriate Makefile.
     - GNU, PGI, or IBM
  3. Configure the Makefile.
     - C++ and Fortran compilers
     - Paths for NetCDF, BUFRLIB, GSL, and F2C libraries
  4. Execute the GNU Make utility to build all of the MET utilities.
  5. Run the test script and check for runtime errors.
     - Runs each of the MET tools at least once.
     - Uses sample data distributed with the tarball.

MET v2.0 Flowchart



Graphics

- Limited graphics incorporated into MET.
- Several options for plotting MET statistical output - including R, NCL, IDL, GNUPlot, and many others.
- Users are encouraged to submit their own plotting and/or analysis scripts using any tool for posting to the MET website.

R Statistics and Graphics

- The R Project for Statistical Computing (www.r-project.org)
  - Powerful statistical analysis and plotting tools.
  - Freely available and well supported for Linux/Windows/Mac.
- Sample R plotting and analysis scripts posted on the MET website.
- Recent praise for R in the technology section of the NY Times:
  - http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=2

Configuration Files

- Set of command line tools controlled using command line options and ASCII configuration files.
  - Well commented and documented in MET User’s Guide.
  - Easy to modify.
  - Distributed with the tarball.
- Configuration files control things such as:
  - Fields/levels to be verified.
  - Thresholds to be applied.
  - Interpolation methods to be used.
  - Verification methods to be applied.
  - Regions over which to accumulate statistics.

Data Formats and Types

Tressa L. Fowler

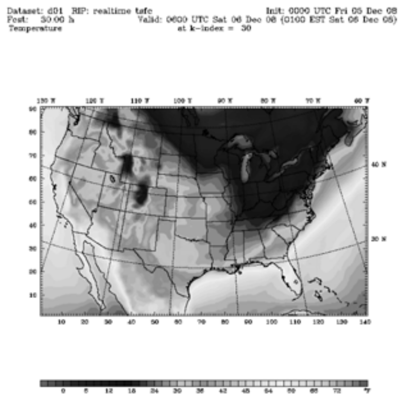
What you can do with MET depends on what type of data you have.

The **format** (grid, point) of your data determines your MET tool(s).

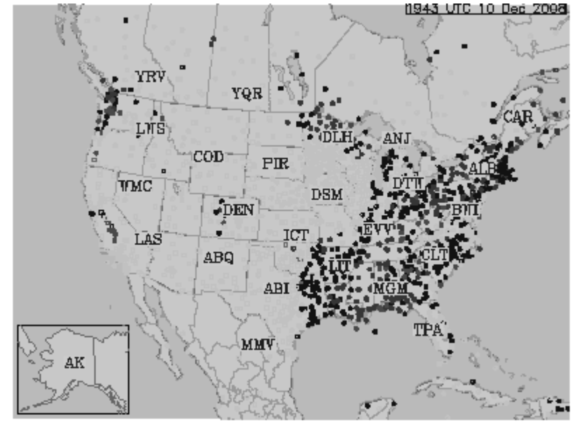
The **type** (continuous, binary) of your data determines the analyses to use within each tool.

Data Formats

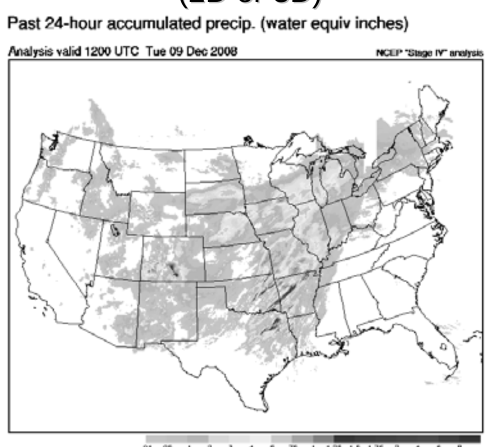
Gridded Forecasts
(2D or 3D)



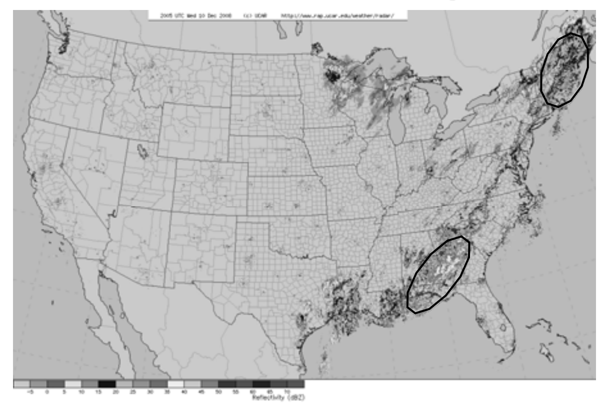
Point Observations
(2D or 3D)



Gridded Observations
(2D or 3D)



Gridded data to transform into **Objects**



Data	MET Tool
Gridded Forecasts Gridded Observations	Grid stat Wavelet Stat MODE
Gridded Forecasts Point Observations	Point Stat

Data Types

Types of Forecasts

- Continuous
  - Wind speed
  - Temperature
- Categorical (includes Binary)
  - Rain / No Rain
  - Hurricane Category 1 - 5
- Probabilistic
  - Prob of freezing precip
- Ensembles

Types of Observations

- Continuous
  - Wind speed
  - Temperature
- Categorical (includes binary)
  - Rain / No Rain
  - No tornado, F1, F2, F3, F4, F5

copyright 2009, UCAR, all rights reserved.

Data type	Analyses
Continuous forecasts Continuous observations	Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Bias
Continuous forecasts, Categorical observations	Receiver Operating Characteristic (ROC) curve, Kolmogorov-Smirnov
Categorical forecasts, Categorical observations	Contingency table statistics and skill scores
Probabilistic forecasts, Categorical observations	Brier score, ranked probability score (RPS), reliability diagram

Introduction to Standard Verification

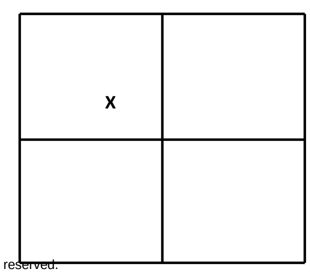
Tressa L. Fowler

Basics

- Match up forecasts and observations at points.
- Calculate differences, sums, or counts over all the points.
- Summarize these things as statistics.

Matching Points to Grids

- Observation points are unlikely to fall exactly on forecast grid points.
- Match in horizontal space via choice of methods:
  - Closest
  - Interpolate
  - Function of surrounding points, *e.g.*
    - Min of closest 4
    - Median of closest 25
- Match in vertical by interpolating between level above and below.

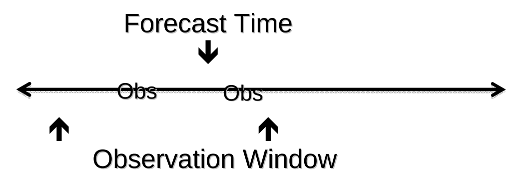


Matching Grids to Grids

- Must use some converter to put forecasts and observations on the same grid.
  - – Example: copygb

Time

- If your forecasts and observations are not at the same time, you may need to define a time window for your observations.

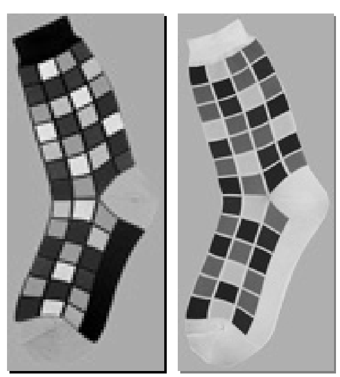


Now you have a bunch of forecast / observation pairs

How well do they match?

F	O
0.32	0.03
0. 51	0.0
0.42	0.48
0.08	0.14
0.20	0.23
0.75	0.33

F	O
1	1
0	0
0	1
1	0
0	1
1	1



From those pairs, you can calculate a variety of statistics . . . like these

http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html

$$POD = \frac{hits}{hits + misses}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2}$$

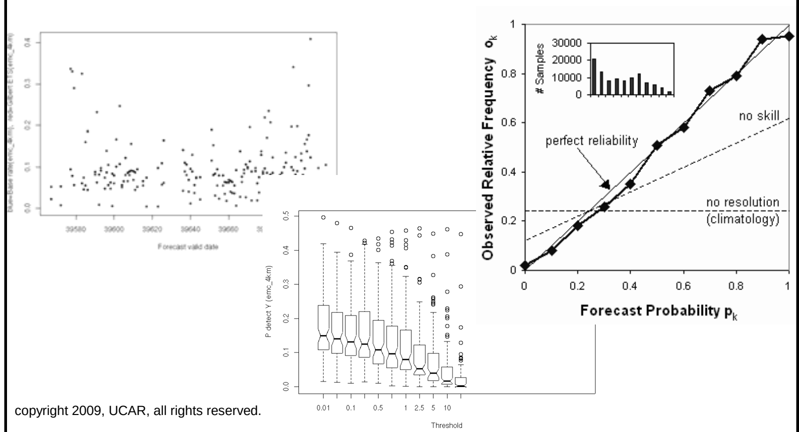
$$ETS = \frac{hits - hits_{random}}{hits + misses + false\ alarms - hits_{random}}$$

Correlation *r*

$$r = \frac{\sum (F-\bar{F})(O-\bar{O})}{\sqrt{\sum (F-\bar{F})^2}\sqrt{\sum (O-\bar{O})^2}}$$

	Observed Event	Observed Non-event
Forecast Event	Count = 532 (Hits)	Count = 219 (False Alarm)
Forecast Non-event	Count = 393 (Miss)	Count =1,627 (Correct No's)

Example Plots



References for information about verification

- http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html
- Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences. 2nd Ed, Academic Press.
- Jolliffe, Stephenson (eds.), 2003. Forecast Verification: A Practitioner's Guide in Atmospheric Science. Wiley.

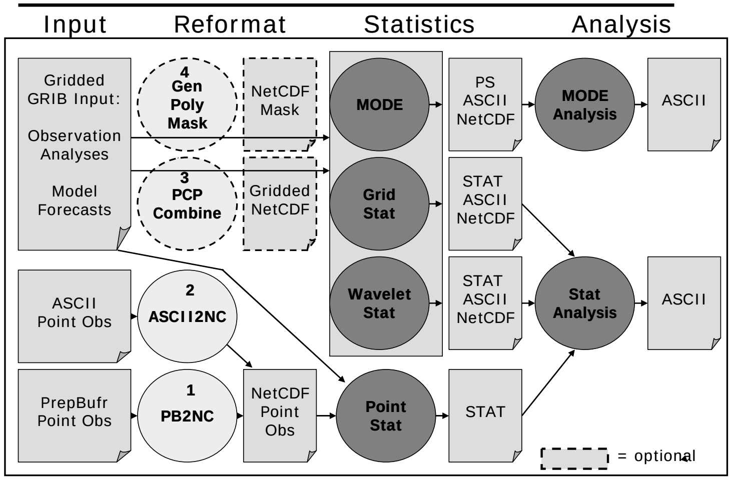
File Formats and Pre-Processing

- Data Formats
- Pre-processing Tools
- Useful Links

Supported Data Formats

- **Forecasts**
  - **GRIB** – GRIdded Binary data format (*version 1)
  - **NetCDF** – network Common Data Format
- **Observations**
  - **PREPBUFR** – binary dataset prepared by NCEP from varied data sources.
  - **ASCII** – “Met_Point” format (10-cols x n-rows)
  - gridded **GRIB** – i.e. NEXRAD Level II or IV
  - point **NetCDF** – i.e. aircraft observations

Pre-Processing / Reformatting



Data Reformating Tools

- **PB2NC and ASCII2NC**
  - Arrange observational data into the NetCDF point format expected by Point-Stat.
- **PCP_Combine (optional)**
  - Sum precipitation values across two or more time periods.
  - Subtract precipitation values to create values for finer subperiods.
  - Produces gridded NetCDF file that can be used as input grid for any Statistics tool.
- **Gen_Poly_Mask (optional)**
  - Used when more complex masking is needed.
  - Produces a NetCDF file of pre-defined mask.
  - May be used for masking in any Statistics tools.

1. PB2NC Tool

- **Stands for "PREPBUFR to NetCDF"**
- **Functionality:**
  - Filters and reformats PREPBUFR point observations into intermediate NetCDF format.
  - Configuration file specifies:
    - Observation types, variables, locations, elevations, quality marks, and times to retain or derive for use in Point-Stat.
- **Data formats:**
  - Reads PREPBUFR using NCEP's BUFRLIB.
  - Writes point NetCDF as input to Point-Stat.

  Note: v2.0 no longer requires CWORDSH to pre-process PREPBUFR files.

PB2NC: Usage

Usage: pb2nc
prepbufr_file
netcdf_file
config_file
[-pbfile prepbufr_file]
[-valid_beg time]
[-valid_end time]
[-nmsg n]
[-dump path]
[-v level]

prepbufr_file	**Input obs file in PrepBufr format**
netcdf_file	**Output name for NetCDF file**
config_file	**PB2NC configuration file**
-pbfile	**Additional input PrepBufr files**
-valid_beg	**Beginning of valid time window** [YYYYMMDD_[HH[MMSS]]
-valid_end	**End of valid time window** [YYYYMMDD_[HH[MMSS]]
-nmsg	**Number of PrepBufr messages to process**
-dump	**Dump entire contents o PrepBufr file to file in path**
-v	**Level of logging**

PB2NC: Run

- **METv2.0/bin/pb2nc** \
  ndas.t00z.prepbufr.tm12.20070401.nr \
  tutorial/out/pb2nc/tutorial_pb.nc \
  tutorial/config/PB2NCConfig_tutorial \
  -v 2

```
==> append : to filename to view the  data source
BUFR
230ADPUPA   UPPER-AIR (RAOB, PIBAL, RECCO, DROPS) REPORTS
231AIRCAR   MDCRS ACARS AIRCRAFT REPORTS                    232AIRCFT
AIREP/PIREP, AMDAR(ASDAR/ACARS), E-ADAS(AMDAR BUFR) ACF233SATWND
SATELLITE-DERIVED WIND REPORTS                 234PROFLR   WIND
PROFILER REPORTS                    235VADWND   VAD (NEXRAD) WIND
REPORTS                   236SATEMP   TOVS SATELLITE DATA (SOUNDINGS,
RETRIEVALS, RADIANCES) 237ADPSFC   SURFACE LAND (SYNOPTIC, METAR)
REPORTS              238SFCSHP   SURFACE MARINE (SHIP, BUOY, C-MAN
PLATFORM) REPORTS    239SFCBOG   MEAN SEA-LEVEL PRESSURE BOGUS
REPORTS              240SPSSMI   SSM/I RETRIEVAL PRODUCTS (REPROCESSED
WIND SPEED, TPW) 241SYNDAT   SYNTHETIC TROPICAL CYCLONE BOGUS
REPORTS             242ERS1DA   ERS SCATTEROMETER DATA (REPROCESSED
WIND SPEED)        243GOESND   GOES SATELLITE DATA (SOUNDINGS,
RETRIEVALS, RADIANCES) 244QKSWND   QUIKSCAT SCATTEROMETER DATA
(REPROCESSED WIND SPEED)   245MSONET   MESONET SURFACE REPORTS
(COOPERATIVE NETWORKS)        246GPSIPW   GLOBAL POSITIONING
SATELLITE-INTEGRATED PRECIP. WATER  247RASSDA   RADIO ACOUSTIC
SOUNDING SYSTEM (RASS) TEMP PROFILE RPTSM063000BYTCNT
...
```

Result of running linux "less" command:

**>less \
ndas.t00z.prepbufr.tm12.20070401.nr**

PREPBUFR

- **BUFR** is the World Meteorological Organization (WMO) standard binary code for the representation and exchange of observational data.
  - **http://www.nco.ncep.noaa.gov/sib/decoders/BUFRLIB/**
  - **http://www.ecmwf.int/products/data/software/**
- The **PREPBUFR** format is produced by NCEP for analyses and data assimilation. The system that produces this format:
  - Assembles observations dumped from a number of sources
  - Encodes
    - information about the observational error for each data type
    - background (first guess) interpolated to each data location
  - Performs both rudimentary multi-platform quality control and more complex platform-specific quality control
- **MET currently only supports PREPBUFR.** Let us know if you need to read in other BUFR format.

PREPBUFR - Known Issues

- CWORDSH blocking
  - In v1.0 and v1.1 it was needed to structure to data properly for reading by the MET fortran code.
  - Need eliminated in v2.0.
- Compilation 64-bit OS
  - Will work if you compile using 32-bit flags.
  - Investigation of this problem in ongoing.

2. ASCII2NC Tool

- **Stands for "ASCII to NetCDF"**
- **Functionality:**
  - Reformat ASCII point observations into intermediate NetCDF format.
  - One input ASCII format supported (10 columns):
  - No configuration file.
- **Data formats:**
  - Reads Met_Point ASCII.
  - Writes point NetCDF as input to Point-Stat.
  - ***Future: support additional standard ASCII formats.***

ASCII2NC: Usage

Usage: ascii2nc
ascii_file
netcdf_file
[-format ascii_format]
[-v level]

ascii_file	**Input obs file in ASCII format**
netcdf_file	**Output name for NetCDF file**
-format	**Override default MET_Point format** (*Future Option)*
-v	**Level of logging**

MET_Point Format

```
Msg    STID  ValidTime       Lat   Lon     Elev   GC Lvl  Hgt  Ob
ADPUPA 72365 20070331_120000 35.03 -106.62 1618.0 7 837.0 1618 1618
ADPUPA 72365 20070331_120000 35.03 -106.62 1618.0 11 837.0 1618 273.05
ADPUPA 72365 20070331_120000 35.03 -106.62 1618.0 17 837.0 1618 271.85
ADPUPA 72365 20070331_120000 35.03 -106.62 1618.0 52 837.0 1618 92
ADPUPA 72365 20070331_120000 35.03 -106.62 1618.0 53 837.0 1618 0.00417
ADPUPA 72365 20070331_120000 35.03 -106.62 1618.0 7 826.0 1724 1724
ADPUPA 72365 20070331_120000 35.03 -106.62 1618.0 11 826.0 1724 274.55
```

* *Use a value of "-9999" to indicate missing data*

Msg	**Message type**
STID	**WMO Station ID**
Lat	**Latitude** [North]
Lon	**Longitude** [East]
Elev	**Elevation** [m] (Note: currently not used by MET code so can be filled with -9999.)
GC	**GRIB code for variable *(i.e. AccPrecip = 61; MSLP = 2; Temp = 11, etc…)*** **http://www.cpc.ncep.noaa.gov/products/wesley/opn_gribtable.html**
Lvl	**Pressure** [mb] or **Accumulation Interval** [hr]
Hgt	**Height above Mean Sea Level** [m – MSL] (Note: currently not used by MET code so can be filled with -9999.)
Ob	**Observed value**

ASCII2NC: Run

- **METv2.0/bin/ascii2nc** sample_obs.txt sample_ascii.nc -v 2

Result of ncdump –h

```
netcdf sample_ascii {
dimensions:
        mxstr = 15 ;
        hdr_arr_len = 3 ;
        obs_arr_len = 5 ;
        nhdr = 5 ;
        nobs = UNLIMITED ; // (2140 currently)
variables:
        char hdr_typ(nhdr, mxstr) ;
                hdr_typ:long_name = "message type" ;
        char hdr_sid(nhdr, mxstr) ;
                hdr_sid:long_name = "station identification" ;
        char hdr_vld(nhdr, mxstr) ;
                hdr_vld:long_name = "valid time" ;
                hdr_vld:units = "YYYYMMDD_HHMMSS UTC" ;
        float hdr_arr(nhdr, hdr_arr_len) ;
                hdr_arr:long_name = "array of observation station header values" ;
                hdr_arr:_fill_value = -9999.f ;
                hdr_arr:columns = "lat lon elv" ;
                ... ;
        float obs_arr(nobs, obs_arr_len) ;
                obs_arr:long_name = "array of observation values" ;
                obs_arr:_fill_value = -9999.f ;
                obs_arr:columns = "hdr_id gc lvl hgt ob" ;
                obs_arr:hdr_id_long_name = "index of matching header data" ;
                ... ;
```

Result of ncdump –v obs_arr

```
obs_arr =
 1, 7, 837, 1618, 1618,
 1, 11, 837, 1618, 273.05,
 1, 17, 837, 1618, 271.85,
 1, 52, 837, 1618, 92,
 1, 53, 837, 1618, 0.00417,
 1, 7, 826, 1724, 1724,
 1, 11, 826, 1724, 274.55,
 1, 17, 826, 1724, 272.15,
 1, 52, 826, 1724, 84,
 1, 53, 826, 1724, 0.00432,
 1, 7, 815.3, 1829, 1829,
 1, 11, 815.3, 1829, 276.45,
 1, 17, 815.3, 1829, 265.75,
 1, 52, 815.3, 1829, 45,
 1, 53, 815.3, 1829, 0.0027,
 1, 7, 815, 1832, 1832,
 1, 11, 815, 1832, 276.55,
 1, 17, 815, 1832, 265.55,
 1, 52, 815, 1832, 44,
 1, 53, 815, 1832, 0.00266,
 1, 7, 784.7, 2134, 2134,
 1, 11, 784.7, 2134, 274.05,
 1, 17, 784.7, 2134, 264.15,
 1, 52, 784.7, 2134, 47,
```


3. PCP-Combine Tool

- **Stands for “Precip-Combine”**
- **Functionality:**
  - Mathematically combines precipitation fields across multiple files.
    - Add precipitation over 2 files
      - *2 NMM output files to go from 3-hr to 6-hr accumulation.*
    - Sum precipitation over more than 2 files
      - *12 ARW output files to go from 1-hr to 12 hr accumulation.*
    - Subtract precipitation in 2 files
      - *2 Level II radar files).*
  - Specify field name on the command line.
  - No configuration file.
- **Data formats:**
  - Reads GRIB format.
  - Writes gridded NetCDF as input to stats tools.

PCP-Combine: Usage

Usage: pcp_combine
[-sum sum_args]
or [-add add_args]
or [-subtract sub_args]
[-gc code]
[-v level]

-sum	**Accumulates data over multiple files.** ***Sum_args***: (init_time, in_accum, valid_time, out_accum, out_file, -pcpdir path, -pcprx reg_exp)
-add	**Accumulates data over two files.** ***Add_args:*** (in_file1, Accum1, in_file2, Accum2, out_file).
-subtract	**Subtracts data over two files.** ***Sub_args:*** (in_file1, Accum1, in_file2, Accum2, out_file).
-gc	**GRIB code for variable** *(i.e. AccPrecip = 61; MSLP = 2; Temp = 11, etc…).*
-v	**Level of logging**

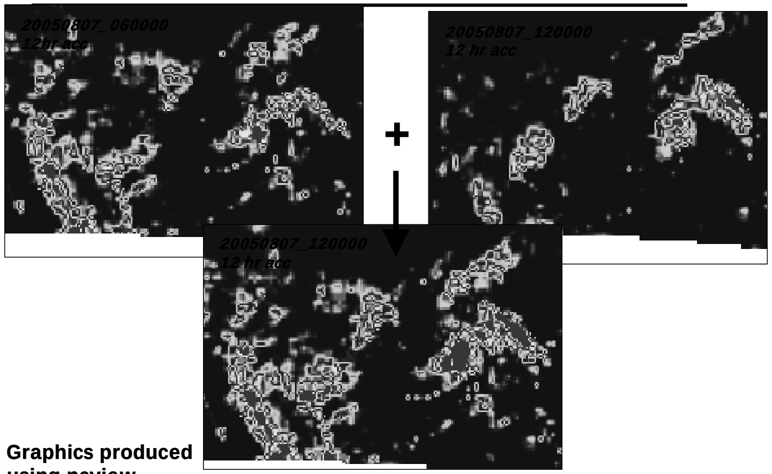
PCP-Combine: Run

- **Two example command lines**

1) Summing two 6-hourly accumulation forecast files into a single 12-hour accumulation forecast.
   - ***METv2.0/bin/pcp_combine ***
     *-add 20050807_060000 6 20050807_120000 6 *
     *sample_fcst.nc *
     -pcpdir data/2005080700

2) Summing 12 1-hourly accumulation observation files into a single 12-hour accumulated observation.
   - ***METv2.0/bin/pcp_combine ***
     *-sum 00000000_000000 1 20050807_120000 12 *
     *sample_obs.nc *
     -pcpdir data/ST2ml

PCP-Combine: Example #1



Graphics produced using ncview

GRIB vs GRIB2

- **GRIB (or GRIB1): WRF postprocessor (WPP) produces GRIB1 format using copyGB**
  - Performs horizontal interpolation and destaggering (in the case of WRF-NMM) onto a defined grid.
  - Useful for both cores in creating an output grid not fixed by the model integration domain.
    - **http://www.dtcenter.org/wrf-nmm/users/downloads/**
- **GRIB2: NCEP and other WMO organizations have historical data in GRIB1 but now use GRIB 2 as the standard for gridded binary data.**
- **At least two GRIB2toGRIB1 converters available**
  - **NCEP cnvgrib**
    - **http://www.nco.ncep.noaa.gov/pmb/codes/GRIB2/**
  - **UCAR/CISL Grib Converter**
    - **http://dss.ucar.edu/libraries/grib/c.html**

Data Inventory Tools

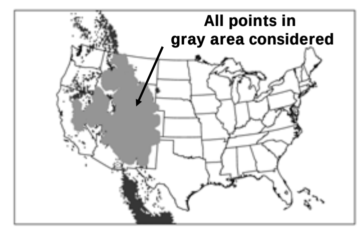
- **wgrib** – dumps GRIB1 headers and data.
  - **http://www.cpc.ncep.noaa.gov/products/wesley/wgrib.html**
- **wgrib2** – dumps GRIB2 headers and data.
  - **http://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/**
- **ncdump -** dumps NetCDF headers and data.
  ncview – plots gridded NetCDF data.
  - **http://www.unidata.ucar.edu/software/netcdf/**
- **GrADS** – command line interface to produce plots.
  - **http://www.iges.org/grads/downloads.html**
- **NCL** – command line interface to produce plots.
  - **http://www.ncl.ucar.edu/**
- **IDV** – gui-driven visualization of many gridded and point datasets.
  - **http://www.unidata.ucar.edu/software/idv/**

4. Gen Poly Mask Tool

- **Stands for “General Polyline Mask”**
- **Functionality:**
  - Pre-computes a lat/lon polyline to generate a 0/1 mask field to be applied to your data.
  - Applies this mask once – prior to running Point-Stat or Grid-Stat
  - No configuration file.
- **Data formats:**
  - Reads Ascii formatted polyline file.
  - Reads GRIB file.
  - Reads NetCDF files from PCP-Combine.
  - Writes gridded NetCDF file of 0/1 mask.

Example polyline mask

Coordinates generated from underlying points of Elevation > 5000 ft.

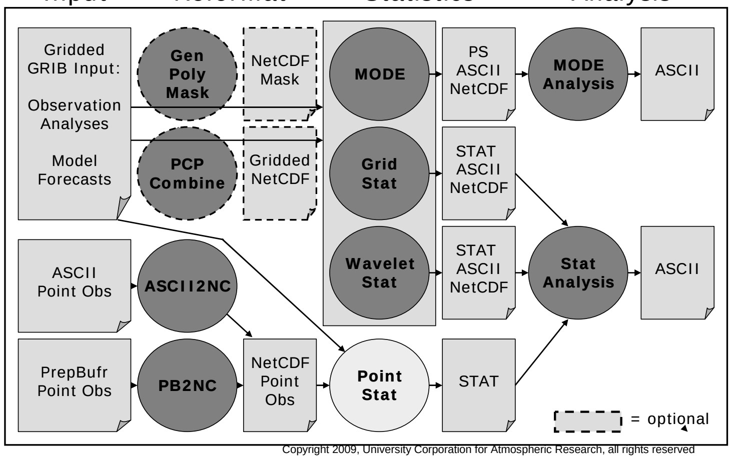


Example polyline file:
data/poly/CONUS.poly

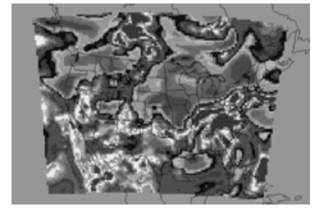
```
CONUS
35.81352 -122.33439
36.01565 -121.51420
35.35092 -121.27343
34.68935 -121.03940
34.03100 -120.81184
34.21258 -120.01634
33.55392 -119.80508
...
```

Point-Stat Tool

Input | Reformat | Statistics | Analysis

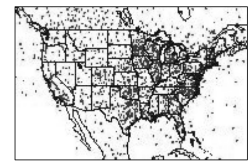


Point-Stat: Overview



- Compare **gridded forecasts** to **point observations.**
- Accumulate matched pairs over a defined area at a *single* point in time.
- Verify one or more variables/levels.
- Analysis tool provided to aggregate through time.

- Verification methods:
  - Continuous statistics for raw fields.
  - Categorical counts and statistics for thresholded fields.
  - Parametric and non-parametric confidence intervals for statistics.
  - Compute partial sums for raw fields and/or the raw matched pair values.
  - Probabilistic methods.



Point-Stat: Input/Output

- Input Files
  - Gridded forecast file
    - GRIB output of WRF Post-Processor (or other)
    - NetCDF output of PCP-Combine
  - Point observation file
    - NetCDF output of PB2NC
    - NetCDF output of ASCII2NC
  - ASCII configuration file
- Output Files
  - ASCII statistics file with all output lines (end with ".stat")
  - Optional ASCII files sorted by line type with a header row (ends with "_TYPE.txt")

Point-Stat: Usage

Usage: point_stat
fcst_file
obs_file
config_file
[-climo climo_file]
[-ncfile netcdf_file]
[-valid_beg time]
[-valid_end time]
[-outdir path]
[-v level]

fcst_file	Forecast file in GRIB or NetCDF
obs_file	Point observation file in NetCDF (PB2NC or ASCII2NC)
config_file	ASCII configuration file
-climo	Climatological file for computing anomaly partial sums
-ncfile	Additional point observation files in NetCDF
-valid_beg	Beginning of valid time window for matching
-valid_end	End of valid time window for matching
-outdir	Output directory to be used
-v	Level of logging

Point-Stat: Configuration

- 25 configurable parameters – only set a few:
  - Temperature at the surface (2-meter).
    - **fcst_field[] = ["TMP/Z2“];**
  - Temperature below freezing.
    - **fcst_thresh[] = [“le273“];**
  - Match to observations at the surface.
    - **message_type[] = ["ADPSFC"];**
  - Look at all the points in my domain.
    - **mask_grid[] = ["FULL"];**
  - Match observation to the nearest forecast value.
    - **interp_wdth[] = [1];**
  - Generate all possible statistic types.
    - **output_flag[] = [2, 2, 2, 2, 2, 2, 2, 2, 2];**

Point-Stat: Run

- METv2.0/bin/point_stat \
  sample_fcst.grb sample_pb.nc \
  PointStatConfig_TMPZ2 -outdir out -v 2

```
Forecast File: sample_fcst.grb
Climatology File: none
Configuration File: PointStatConfig_TMPZ2
Observation File: sample_pb.nc
----------------------------------------
Reading records for TMP/Z2.
For TMP/Z2 found 1 forecast levels and 0 climatology levels.
----------------------------------------
Searching 179772 observations from 44076 PrepBufr messages.
----------------------------------------
Processing TMP/Z2 versus TMP/Z2, for observation type ADPSFC, over region FULL, for interpolation method
UW_MEAN(1), using 11370 pairs.
Computing Categorical Statistics.
Computing Continuous Statistics.
Computing Scalar Partial Sums.
----------------------------------------
Output file: out/point_stat_360000L_20070331_120000V.stat
Output file: out/point_stat_360000L_20070331_120000V_fho.txt
Output file: out/point_stat_360000L_20070331_120000V_ctc.txt
Output file: out/point_stat_360000L_20070331_120000V_cts.txt
Output file: out/point_stat_360000L_20070331_120000V_cnt.txt
Output file: out/point_stat_360000L_20070331_120000V_sl1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_sal1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_vl1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_val1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_mpr.txt
```

Point-Stat: ASCII Output Types

- Statistics line types: 9 possible
  - Categorical - apply threshold
    - Contingency table counts (**FHO**, **CTC**)
    - Contingency table statistics (**CTS**)
  - Continuous - raw fields
    - Continuous statistics (**CNT**)
    - Partial Sums (**SL1L2**, **SAL1L2**, **VL1L2**, **VAL1L2**)
  - Matched Pairs
    - Raw matched pairs – a lot of data! (**MPR**)
  - Additional line type(s) for probabilistic methods
- 21 header columns common to all line types
- Remaining columns specific to each line type

Point-Stat: Sample Output

- STAT file output for sample run:
  - 1 line each for **FHO**, **CTC**, **CTS**, **CNT**, and **SL1L2**
  - 11,370 lines for **MPR**!
- 6 additional TXT files with lines sorted by type

Point-Stat: CTC Output Line

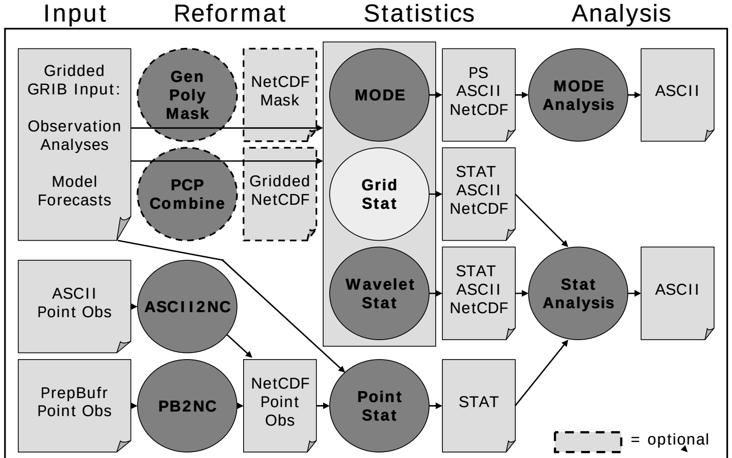
VERSION	V2.0	**VX_MASK**	FULL
MODEL	WRF	**INTERP_MTHD**	UW_MEAN
FCST_LEAD	360000	**INTERP_PNTS**	1
FCST_VALID_BEG	20070331_120000	**FCST_THRESH**	<273.000
FCST_VALID_END	20070331_120000	**OBS_THRESH**	<273.000
OBS_LEAD	000000	**COV_THRESH**	NA
OBS_VALID_BEG	20070331_103000	**ALPHA**	NA
OBS_VALID_END	20070331_133000	**LINE_TYPE**	CTC
FCST_VAR	TMP	**TOTAL**	11370
FCST_LEV	Z2	**FY_OY**	1635
OBS_VAR	TMP	**FY_ON**	380
OBS_LEV	Z2	**FN_OY**	438
OBTYPE	ADPSFC	**FN_ON**	8917

Point-Stat: Matched Pairs

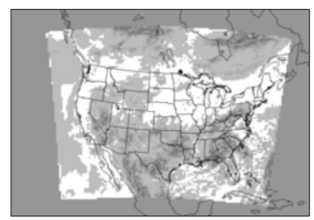
- **Matched Pair (MPR) line type contains raw matched pairs.**
  - **Data overload!**

TOTAL	INDEX	OBS_LAT	OBS_LON	OBS_LVL	OBS_ELV	FCST	OBS	CLIMO
11370	1	43.93000	-60.01000	1010.79999	4.01053	271.87788	271.54999	NA
11370	2	46.43000	-71.93000	1016.09998	102.04903	268.50255	269.45001	NA
11370	3	44.23000	-78.36000	1004.50000	191.44466	272.94013	272.35001	NA
11370	4	51.67000	-124.40000	916.50000	872.82202	263.69020	264.95001	NA
11370	5	58.61000	-117.16000	973.90002	337.50449	272.37757	270.95001	NA
11370	6	52.18000	-122.04000	906.50000	938.08594	271.37738	264.35001	NA
11370	7	50.68000	-127.36000	1020.20001	22.03931	275.44020	275.04999	NA
11370	8	50.45000	-100.59000	949.09998	562.38477	272.18978	271.75000	NA
11370	9	57.13000	-61.47000	899.70001	834.87476	258.06464	254.64999	NA
11370	10	47.56000	-59.16000	1000.90002	40.06803	272.06486	269.54999	NA
11370	11	47.41000	-72.79000	1006.90002	169.37592	266.37724	265.95001	NA
11370	12	45.76000	-62.68000	1014.00000	1.99518	268.94018	268.64999	NA
11370	13	49.24000	-65.33000	1014.90002	28.96468	264.25276	267.25000	NA
11370	14	43.29000	-79.79000	1017.79999	77.03765	273.56474	275.85001	NA
11370	15	48.78000	-123.04000	1015.70001	23.93772	278.12724	280.25000	NA

Grid-Stat Tool

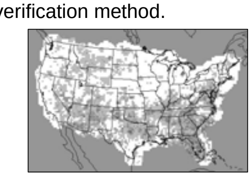


Grid-Stat: Overview



- Compare **gridded forecasts** to **gridded observations** on the *same grid.*
- Accumulate matched pairs over a defined area at a *single* point in time.
- Verify one or more variables/levels.
- Analysis tool provided to aggregate through time.

- Verification methods:
  - Continuous statistics for raw fields.
  - Categorical counts and statistics for thresholded fields.
  - Parametric and non-parametric confidence intervals for statistics.
  - Compute partial sums for raw fields.
  - Probabilistic methods.
  - Continuous statistics and categorical counts/statistics using neighborhood verification method.



Grid-Stat: Input/Output

- Input Files
  - Gridded forecast and observation files
    - GRIB output of WRF Post-Processor (or other)
    - NetCDF output of PCP-Combine
  - ASCII configuration file
- Output Files
  - ASCII statistics file with all output lines (end with ".stat")
  - Optional ASCII files sorted by line type with a header row (ends with "_TYPE.txt")
  - Optional NetCDF matched pairs file

Grid-Stat: Usage

Usage: point_stat
fcst_file
obs_file
config_file
[-outdir path]
[-v level]

fcst_file	Forecast file in GRIB or NetCDF
obs_file	Observation file in GRIB or NetCDF
config_file	ASCII configuration file
-outdir	Output directory to be used
-v	Level of logging

Grid-Stat: Configuration

- 24 configurable parameters – only set a few:
  - Precipitation accumulated over 24 hours.
    - **fcst_field[] = ["APCP/A24"];**
  - Any rain and moderate rain.
    - **fcst_thresh[] = ["gt0.00, gt20.00"]; (mm)**
  - Look at all the points and only the eastern United States.
    - **mask_grid[] = ["FULL"];**
    - **mask_poly[] = ["EAST.poly"];**
  - Compute neighborhood statistics with two sizes. **nbr_width[] = [3, 5];**
  - Generate all possible statistic types.
    - **output_flag[] = [2, 2, 2, 2, 2, 2, 2, 2, 2];**

Grid-Stat: Run

- METv2.0/bin/grid_stat \
  sample_fcst.grb sample_obs.nc \
  GridStatConfig_APCP24 -outdir out -v 2

```
Forecast File: sample_fcst.grb
Observation File: sample_obs.nc
Configuration File: GridStatConfig_APCP24
----------------------------------------
Processing APCP/A24 versus APCP/A24, for interpolation method UW_MEAN(1), over region FULL, using 6412 pairs.
Computing Categorical Statistics.
Computing Continuous Statistics.
Processing APCP/A24 versus APCP/A24, for interpolation method UW_MEAN(1), over region EAST, using 2586 pairs.
Computing Categorical Statistics.
Computing Continuous Statistics.
Processing APCP/A24 versus APCP/A24, for interpolation method NBRHD(9), raw thresholds of >0.000 and >0.000,
over region EAST, using 5829 pairs.
... MORE NEIGHBORHOOD COMPUTATIONS ...
----------------------------------------
Output file: out/grid_stat_240000L_20050808_000000V.stat
Output file: out/grid_stat_240000L_20050808_000000V_fho.txt
Output file: out/grid_stat_240000L_20050808_000000V_ctc.txt
Output file: out/grid_stat_240000L_20050808_000000V_cts.txt
Output file: out/grid_stat_240000L_20050808_000000V_cnt.txt
Output file: out/grid_stat_240000L_20050808_000000V_sl1l2.txt
Output file: out/grid_stat_240000L_20050808_000000V_nbrctc.txt
Output file: out/grid_stat_240000L_20050808_000000V_nbrcts.txt
Output file: out/grid_stat_240000L_20050808_000000V_nbrcnt.txt
Output file: out/grid_stat_240000L_20050808_000000V_pairs.nc
```

Grid-Stat: ASCII Output Types

- Statistics line types: 8 possible
  - Same as Point-Stat
    - **FHO**, **CTC**, **CTS**, **CNT**, and **SL1L2**
  - Omitted for Grid-Stat
    - ***SAL1L2***, ***VL1L2***, or ***VAL1L2***
  - Neighborhood – define neighborhood, apply threshold
    - Neighborhood continuous statistics (**NBRCNT**)
    - Neighborhood contingency table counts (**NBRCTC**)
    - Neighborhood contingency table statistics (**NBRCTS**)
  - Additional line type(s) for probabilistic methods
- 21 header columns common to all line types
- Remaining columns specific to each line type

Grid-Stat: Sample Output

- STAT file output for sample run:
  - 2 lines each for **CNT** and **SL1L2**
    - = 2 verification regions (FULL and EAST)
  - 4 lines each for **FHO**, **CTC**, and **CTS**
    - = 2 regions * 2 thresholds
  - 8 lines each for **NBRCNT**, **NBRCTC**, **NBRCTS**
    - = 2 regions * 2 thresholds * 2 neighborhood sizes
- 8 additional TXT files with lines sorted by type
- NetCDF file containing matched pairs

Copyright 2009, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: CTC Output Line

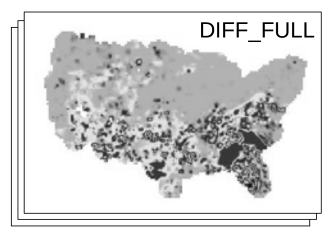
VERSION	V2.0
MODEL	WRF
FCST_LEAD	240000
FCST_VALID_BEG	20050808_000000
FCST_VALID_END	20050808_000000
OBS_LEAD	000000
OBS_VALID_BEG	20050808_000000
OBS_VALID_END	20050808_000000
FCST_VAR	APCP
FCST_LEV	A24
OBS_VAR	APCP
OBS_LEV	A24
OBTYPE	MC_PCP

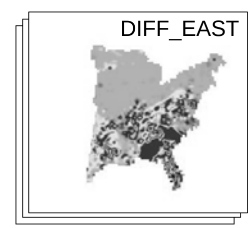
VX_MASK	EAST
INTERP_MTHD	UW_MEAN
INTERP_PNTS	1
FCST_THRESH	>=20.000
OBS_THRESH	>=20.000
COV_THRESH	NA
ALPHA	NA
LINE_TYPE	CTC
TOTAL	2586
FY_OY	5
FY_ON	104
FN_OY	70
FN_ON	2407

Copyright 2009, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: NetCDF Matched Pairs

- Forecast, observation, and difference fields for each combination of…
  - Variable, level, masking region, and interpolation method (smoothing)
- Sample output contains 6 variables:
  - FCST, OBS, and DIFF for **FULL** and **EAST**





Copyright 2009, University Corporation for Atmospheric Research, all rights reserved

Variations

- **Grid-Stat** and **Point-Stat** may be used to compare two *different* variables.
  - Not strictly verification.
  - Leave observation blank to use forecast setting.
  - Selecting variable/levels:
    - **fcst_field[] = ["61/A24"];**
    - **obs_field[] = [];**
  - Selecting thresholds:
    - **fcst_thresh[] = ["gt0.0 ge20.0"];**
    - **obs_thresh[] = [];**

Stat Analysis Tool

- Filtering
- Summarizing
- Aggregating

of Grid Stat, Point Stat,
& Wavelet output

What can Stat Analysis do for you?

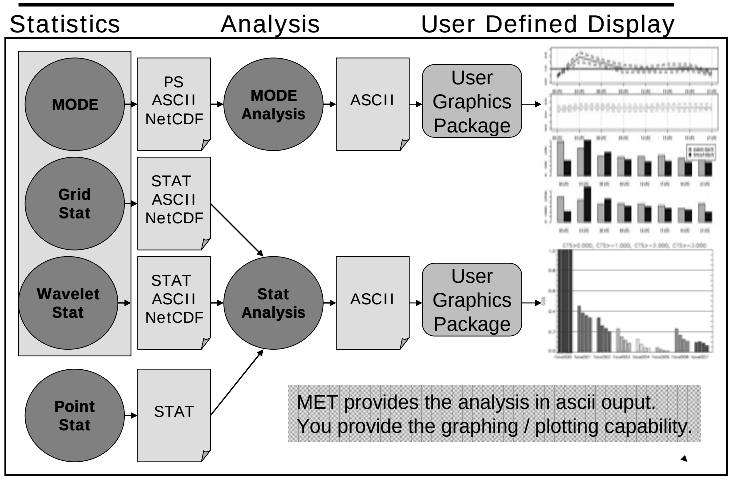
✓ **Q:** If I wanted the overall statistics for all the gridded observations compared to the forecasts for hours 0 through 24 together, can MET do this?
A: Yes – using Stat Analysis Tool on Grid-Stat output

✓ **Q:** Can I compute long-term statistics at individual sites (eg, mean absolute error or RMS error for daily forecasts for a month).?
A: Yes – using Stat Analysis Tool on Point-Stat output

✓ **Q:** Can I aggregate my contingency table statistics over multiple runs?
A: Yes – using Stat Analysis Tool on any output

❑ **Q:** Can I aggregate statistics for a large number of individual stations in one simultaneous run?
A: No – but you could get the aggregate statistics for each individual station and run Stat Analysis Tool n-times.

Stat Analysis Tool



Stat Analysis Jobs

- **Filtering**
  - **stat_job_filter -** filters out lines from one or more stat files based on user-specified filtering options.
- **Summarizing**
  - **stat_job_summary -** produces summary information from a single column of data including:
    - mean, standard deviation, min, max, and the 10th, 25th, 50th, 75th, and 90th percentiles.
- **Customized tool for AFWA**
  - **stat_job_go_index -** computes the GO Index, a performance Statric used primarily by the United States Air Force.

Stat Analysis Jobs

- **Aggregation**
  - **stat_job_aggr -** aggregates stat data across multiple time steps or masking regions to sum contingency table data or partial sums across multiple lines of data.
  - **stat_job_aggr_cts -** aggregates contingency table to produce aggregated contingency table statistics.
  - **stat_job_aggr_cnt -** computes as many continuous statistics from aggregated partial sums as possible.
  - **stat_job_aggr_mpr -** aggregates the matched pair output of Point-Stat and recomputes the requested stat line type.

PB2NC: Usage

Usage: stat_analysis
-lookin path
-out filename
[-v level]
[-config config_file]
or command line options with associated arguments
[stat_job_filter]
[stat_job_summary]
[stat_job_aggr]
[stat_job_cts]
[stat_job_aggr_cnt]
[stat_job_aggr_mpr]
[stat_job_go_index]

-lookin	Directory where *.stat files reside
-out	Output name for NetCDF file
-v	Level of logging
-config	StatAnalysisConfig file
stat_job_filter	See previous 2 slides
stat_job_summary	See previous 2 slides
stat_job_aggr	See previous 2 slides
stat_job_cts	See previous 2 slides
stat_job_aggr_cnt	See previous 2 slides
stat_job_aggr_mpr	See previous 2 slides
stat_job_go_index	See previous 2 slides

Point-Stat: Configuration

- 22 configurable parameters – only set a few:
  - Apply NAM G212 mask
    - **vx_mask[ ] = ["G212"];**
  - Using only the Temperature variable
    - **var[ ] = ["TMP"];**
  - Filter on CTC lines that have been threshholded var[ ]>278
    - **line_type[ ] = ["CTC"];**
    - **fcst_thres[ ] = [">278"];**
  - Dump the filtered stat data to a file AND sum contingency table count (CTC) lines of data for pressure levels between 850 and 750
    - **joblist[ ] = *["-job stat_job_filter -dump_row out/filter_job.stat", "-job stat_job_aggr -dump_row out/aggr_ctc_job.stat -level P850-750"*];**

Stat Analysis Tool: Run stat_job_aggr

"-job stat_job_aggr -dump_row out/aggr_ctc_job.stat -level P850-750"

Point Stat Output *(i.e. point_stat_out.stat)*

```
V2.0    WRF   … ADPUPA G212 … TMP
   P850-750 … >278.00 CTC
   = 401      192              11
   24               174
   UW_MEAN          1
```

		OBS		
FCST		Y	N	
	Y	192	11	*203*
	N	24	174	*198*
		216	*185*	**401**

```
V2.0    WRF   … ADPSFC G212 … TMP
   P850-750 … >278.00 CTC
   = 167      25               23
   0                119
   UW_MEAN          1
```

		OBS		
FCST		Y	N	
	Y	25	23	*48*
	N	0	119	*119*
		25	*142*	**167**

(NOTE: header modified to show only pertinent info)

Stat Analysis Tool: Run stat_job_aggr

Stat Analysis Output *(i.e. stat_analysis.out)*

```
FILTER:  -job vsdb_job_filter
   -vx_mask G212 -line_type CTC
   –fcst_thresh >278.000 -var TMP
   -dump_row out/filter_job.vsdb
COL_NAME:         TOTAL
   FY_OY            FY_ON
   FN_OY          FN_ON
   INTERP_MTHD      INTERP_PNTS
CTC:              568              217
   34               24          293

   -9999                        -9999
JOB_COMMAND: -job vsdb_job_aggr
   -vx_mask G212 -line_type CTC
   -fcst_thres >278.000 -var TMP
   -level P850-750 -dump_row
   out/aggr_ctc_job.vsdb
```

		OBS		
FCST		Y	N	
	Y	217	34	*251*
	N	24	293	*317*
		241	*327*	**568**

Identifying JOB_COMMAND Line comes after data

Stat Analysis Tool: Run stat_job_summary

"-job stat_job_summary –line_type CNT –alpha = 0.50 -var TMP \ –dump_row ../out/job_summary_RMSE.vsdb –column RMSE"

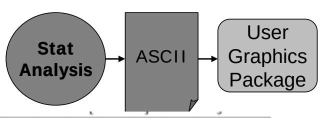
Column Number	Description
1	Summary (job type)
2	Total
3-7	Mean* *Includes normal and bootstrap upper and lower confidence limits*
8-10	Standard deviation** *Includes bootstrap upper and lower confidence limits*
11	Minimum value
12	10th percentile
13	25th percentile
14	Median (50th percentile)
15	75th percentile
16	90th percentile
17	Maximum value

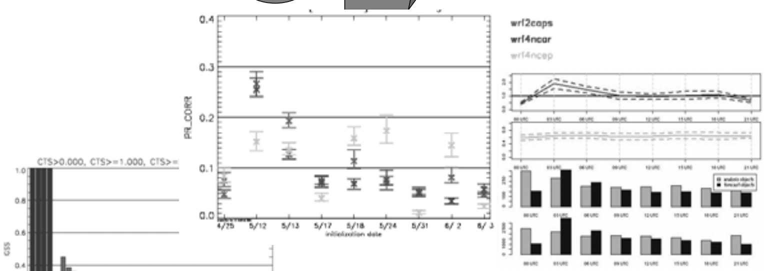
(stat_analysis.out cont.)

```
COL_NAME: TOTAL MEAN
MEAN_NCL MEAN_NCU MEAN_BCL
MEAN_BCU STDEV STDEV_BCL
STDEV_BCU MIN  P10
P25 P50 P75
P90 MAX
SUMMARY: 4  1.98438
1.33219 2.63656  1.58837
2.29289 0.40986 0.04574
0.55950 1.41291 1.59671
1.87241 2.07130 2.18328
2.18328 2.30251
JOB_COMMAND: -job
vsdb_job_summary -line_type
CNT ….
```

Identifying JOB_COMMAND Line comes after data

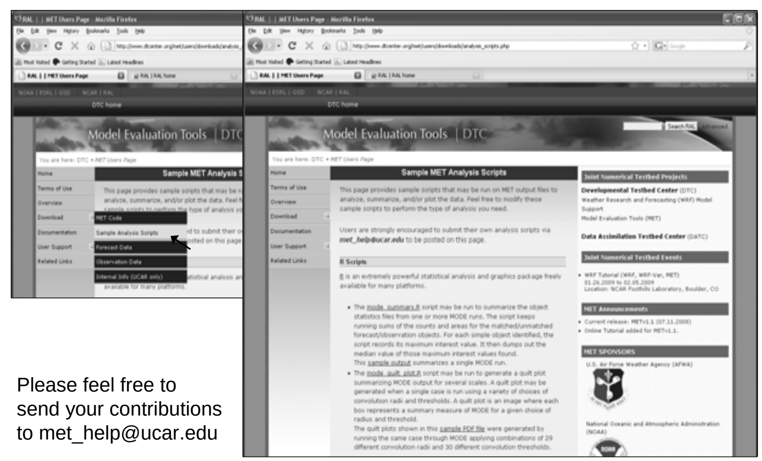
Use your favorite plotting software





MET provides the analysis in ascii ouput.
You provide the graphing / plotting capability.

User Contributed Plotting Scripts

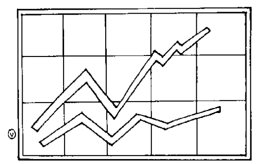


Please feel free to send your contributions to met_help@ucar.edu

Thanks - Any Questions?

Now you know how to:
Filter it.
Aggregate it.
and
Plot it...

Next up is:
Analyzing it.
and
Customizing it.



Interpretation of Output – Grid and Point Stat

Tressa L. Fowler

Point and Grid Stat Output Lines

- CTC - Contingency Table Counts
- CTS - Contingency Table Statistics
- CNT - Continuous Statistics
- *FHO - Forecast, Hit, Observation Rates*
- *SL1L2 - Scalar L1L2 Partial Sums*
- *SAL1L2 - Scalar Anomaly L1L2 Partial Sums when climatological data is supplied*
- *VL1L2 - Vector L1L2 Partial Sums*
- *VAL1L2 - Vector Anomaly L1L2 Partial Sums when climatological data is supplied*
- MPR - Matched Pair data

Available in the newest release MET 2.0, line types for probability forecasts.

Point and Grid stat output example

Header Line:

- TOTAL BASER BASER_NCL BASER_NCU BASER_BCL BASER_BCU FMEAN FMEAN_NCL FMEAN_NCU FMEAN_BCL FMEAN_BCU ACC ACC_NCL ACC_NCU ACC_BCL ACC_BCU FBIAS FBIAS_BCL FBIAS_BCU PODY PODY_NCL PODY_NCU PODY_BCL PODY_BCU PODN PODN_NCL PODN_NCU PODN_BCL PODN_BCU POFD POFD_NCL POFD_NCU POFD_BCL POFD_BCU FAR FAR_NCL FAR_NCU FAR_BCL FAR_BCU CSI CSI_NCL CSI_NCU CSI_BCL CSI_BCU GSS GSS_BCL GSS_BCU HK HK_NCL HK_NCU HK_BCL HK_BCU HSS HSS_BCL HSS_BCU ODDS ODDS_NCL ODDS_NCU ODDS_BCL ODDS_BCU

Data Line:

- CTS 5 1.00000 0.56552 1.00000 NA NA 0.40000 0.11762 0.76928 NA NA 0.40000 0.11762 0.76928 NA NA 0.40000 NA NA 0.40000 0.11762 0.76928 NA NA NA NA NA NA NA NA NA NA NA NA 0.00000 0.00000 0.43448 NA NA 0.40000 0.11762 0.76928 NA NA 0.00000 NA NA NA NA NA NA NA 0.00000 NA NA NA NA NA NA NA

See Appendix C of MET documentation for equations and details

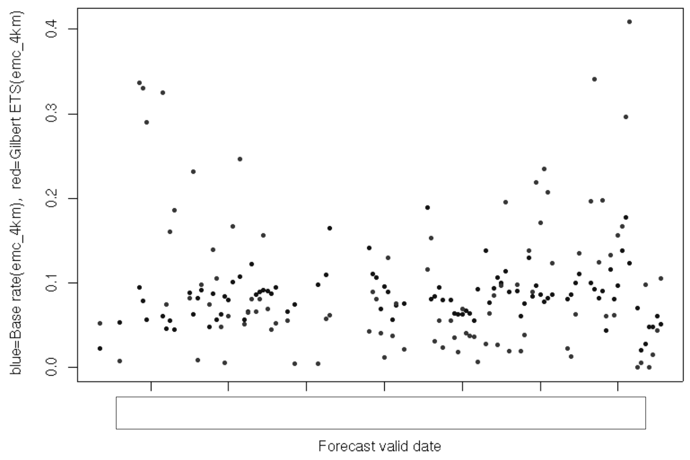
- TOTAL – Count of total pairs included in the stats for this line
- BASER – base rate (e.g. sample climatological rate of event)
- FMEAN – mean of forecast values
- ACC - accuracy
- FBIAS – frequency bias
- PODY – probability of detection (events)
- PODN – probability of detection (non-events)
- POFD – probability of false detection
- FAR – False Alarm Ratio
- CSI – Critical success index
- GSS – Gilbert Skill Score
- HK - Hanssen-Kuipers Discriminant
- HSS - Heidke Skill Score
- ODDS – Odds Ratio
- Etc., etc., etc. . .

Confidence Intervals for some measures are included with the following suffixes:
_NCL = Normal Confidence Lower
_NCU = Normal Confidence Upper
_BCL = Bootstrap Confidence Lower
_BCU = Bootstrap Confidence Upper

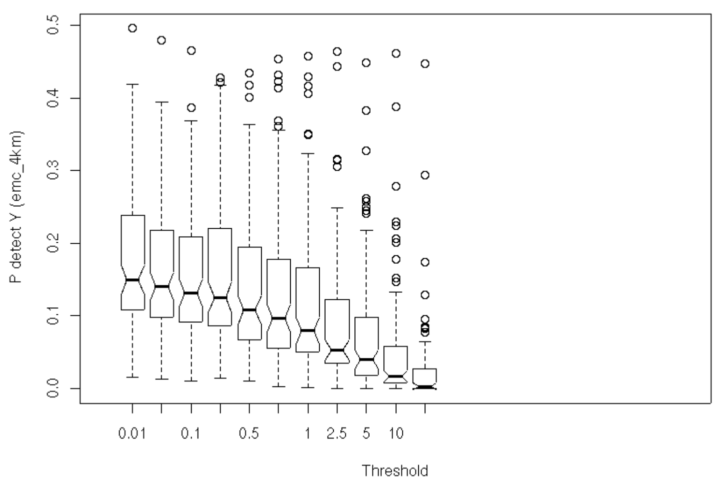
Example of Appendix C info

- **Accuracy -** for a 2x2 contingency table is defined as $\frac{n_{11} + n_{00}}{T}$
- That is, it is the proportion of forecasts that were either hits or correct rejections – the fraction that were correct.
- Accuracy ranges from 0 to 1; a perfect forecast would have an accuracy value of 1.
- Accuracy should be used with caution, especially for rare events, because it can be strongly influenced by large values of n_{00}.

Example Scatter Plot



Example Box Plot



Customizing Point-Stat and Grid-Stat Output

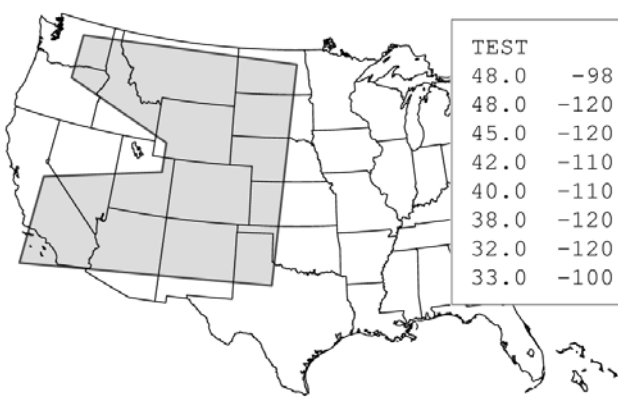
We'll restrict this discussion to an explanation of masking and interpolation.

Masking Methods

	Polyline	Grid	Stations	On/Off Bit Map
Point Stat	✓	✓	✓	✓
Grid Stat	✓	✓	N/A	✓

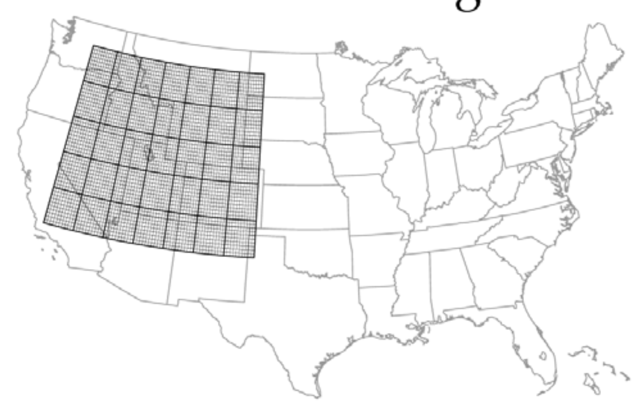
Masking by stations doesn't apply to Grid Stat.

Polyline Masking



```
TEST
48.0   -98.0
48.0  -120.0
45.0  -120.0
42.0  -110.0
40.0  -110.0
38.0  -120.0
32.0  -120.0
33.0  -100.0
```

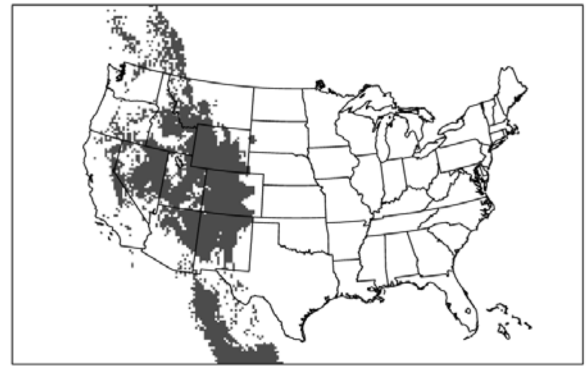

Grid Masking



NCEP grids link: www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html

Station Masking

Data Threshold Mask



Topography > 5000 feet

Interpolation Methods

	Min	Max	Median	UW Mean	DW Mean	Nearest Nbr	Least Squares
Point Stat	✓	✓	✓	✓	✓	✓	✓
Grid Stat	✓	✓	✓	✓	N/A	N/A	N/A

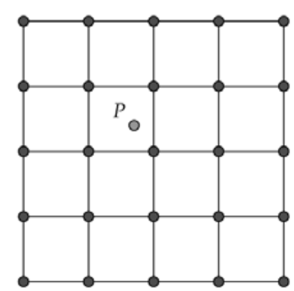
For Grid Stat, these are smoothing methods.

Interpolation

Need to Choose:

(1) Method

(2) Width



Min, Max, Median

Takes minimum, maximum or median of values in interpolation square.

Median separates the upper half of data values from the lower half. This is different from the mean, which is an average.

Nearest Neighbor

Essentially, no interpolation is performed.

Value at interpolation point is simply the data value at the closest grid point.

Unweighted Mean
Distance-Weighted Mean

Unweighted Mean is the average.

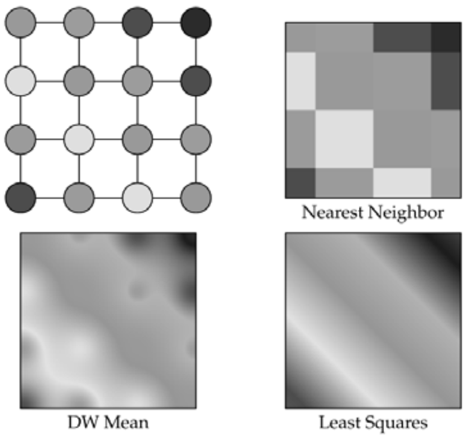
Distance-Weighted Mean is an average weighted according to distance from nearby grid points.

Least Squares

Performs a local Least-Squares linear fit in interpolation square.

$$z = Ax + By + C$$

Interpolation Examples



Giving meaning to your forecast verification results

What is the answer to life, the universe and everything?

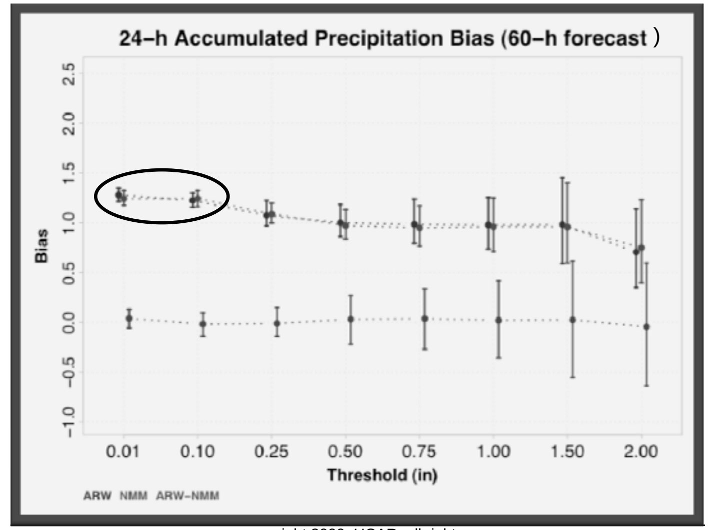
42.

Eric Gilleland

Research Applications Laboratory,

National Center for Atmospheric Research

Boulder, Colorado Email: EricG @ ucar.edu



Accounting for Uncertainty

- Observational
- Model
  - Model parameters
  - Physics
  - Verification scores
- Sampling
  - Verification statistic is a realization of a random process.
  - What if the experiment were re-run under identical conditions?

Hypothesis Testing and Confidence Intervals

- Hypothesis testing
  - Given a null hypothesis (e.g., no bias), is there enough evidence to reject it?
  - One- or two-sided, but test is against a single null hypothesis.
- Confidence intervals
  - Related to hypothesis tests, but more useful.
  - How confident are we that the true value of the statistic (e.g., bias) is different from a particular value?
  - Interpretation for most *frequentist* intervals is a bit awkward.

Hypothesis Testing and Confidence Intervals

"If we re-run the experiment 100 times, and create 100 (1-α)100% CI's, then we expect the true value of the parameter to fall inside (1-α)100 of the intervals."

Hypothesis Testing and Confidence Intervals

Example: The difference in bias between two models is 0.01.

Hypothesis test: Is this different from zero?

Confidence interval: Does zero fall within the interval? Does 0.5 fall within the interval?

Confidence Intervals (CI's)

- Parametric
  - Assume the observed sample is a realization from a known *population* distribution with possibly unknown parameters (e.g., normal).
  - Normal approximation CI's are most common.
  - Quick and easy.
- Nonparametric
  - Assume the distribution of the observed sample is representative of the *population* distribution.
  - Bootstrap CI's are most common.
  - Can be computationally intensive, but easy enough.

Normal Approximation CI's

$$\hat{\theta} \pm z_{\alpha/2} se(\theta)$$

Is a (1-α)100% Normal CI for Θ, where Θ is the statistic of interest (e.g., the forecast mean), se(Θ) is the standard error for the statistic, and z_v is the v-th quantile of the standard normal distribution.

Normal Approximation CI's

Example: Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) sample from a normal distribution with variance σ_X^2.

Then, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an estimate of the mean

of the sample. And a (1-α)100% CI is given by

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

Normal Approximation CI's

- Numerous verification statistics can take this approximation in some form or another. In other cases (e.g., forecast/observation variance, linear correlation), different parametric CI's can be used that still rely on the underlying sample's being iid normal.
- Contingency table verification scores such as probability of detection (POD) and false alarm ratio (FAR) also have normal approximation CI's (for large enough sample sizes).

Normal Approximation CI's

- Check the validity of the independence assumption.
- Check the validity of the normal distribution (e.g., qq-plots, automatic tests exist too).

(cf. Gilleland, 2008)

Bootstrap CI's

IID Bootstrap Algorithm

1. Resample *with replacement* from the sample, $X_1, \ldots, X_n$.
2. Calculate the verification statistic(s) of interest from the resample in step 1.
3. Repeat steps 1 and 2 many times, say B times, to obtain a sample of the verification statistic(s).
4. Estimate (1-α)100% CI's from the sample in step 3.

Bootstrap CI's

IID Bootstrap Algorithm: Types of CI's

1. Percentile Method CI's*
2. Bias-corrected and adjusted (BCa)*
3. ABC
4. Basic bootstrap CI's
5. Normal approximation
6. Bootstrap-t

*1 and 2 are available In MET

Bootstrap CI's

Block Bootstrap Algorithm

- Block bootstrapping is one way to obtain bootstrap CI's for dependent samples.
- Same as IID Bootstrap, but resample blocks of contiguous data points.
- Use percentile method for CI's.
- Block sizes should be substantially larger than the correlation length, but substantially smaller than the sample size. Usually, one takes the greatest integer below the square root of the sample size.

Bootstrap CI's

Block Bootstrap Algorithm: Blocks

Non-overlapping (NBB)

$$Y_1 = \{X_1, \ldots, X_l\}, \ldots, Y_k = \{X_{(k-1)l+1}, \ldots, X_{kl}\}, \ldots, Y_b = \{X_{(b-1)l+1}, \ldots, X_n\}$$

Moving (MBB)

$$Y_i = \{X_i, \ldots, X_{i+l-1}\}$$

Circular (CBB)

Bootstrap CI's

Parametric Bootstrap Algorithm:

- Same as IID Bootstrap algorithm, but first model the dependence in the data, then in step 1, take random samples from the model instead of the data.
- Generally preferable to block bootstrap.
- Stronger assumptions about the sample than block bootstrap.
- Use percentile method for CI's.

Bootstrap CI's

Configurable parameters for the bootstrap in MET. For example, usually it is appropriate to take resamples of size *n*, where *n* is the size of the original sample. However, there are cases where it is better to take samples of size, *m*<*n* (e.g., if the population distribution is heavy-tailed; see Gilleland, 2008).

Practical Considerations

- Point-stat is quicker than Grid-stat, so bootstrap is quicker with Point-stat.
- May be prohibitively computationally inefficient to bootstrap over an entire field (i.e., over several thousand points), but can also bootstrap the statistics for each field over time. Measures the (between-field) uncertainty of the estimates over time, rather than the within field uncertainty.
- Normal approximation intervals are quick, and generally accurate. Check the normality assumption!
- Check whether the samples (for either type of interval) are independent or not.

Thank you. Questions?

For more information, see:

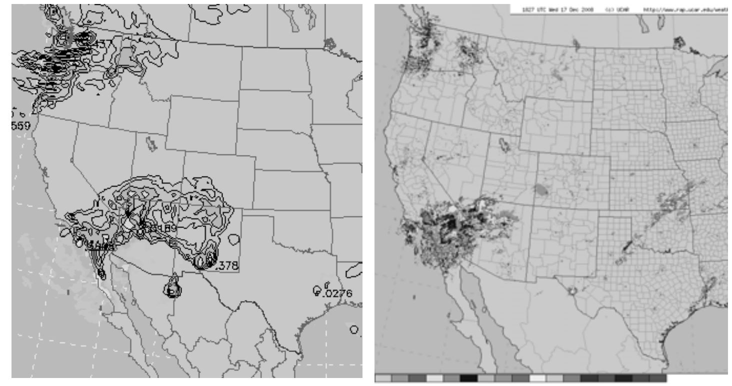
Developmental Testbed Center, 2008. Model Evaluation Tools Version 1.1 (METv1.1) User's Guide. Available at: http://www.dtcenter.org/met/

Gilleland E, 2008. Confidence intervals for forecast verification. *Submitted* as an NCAR Technical Note. Available at: http://www.ral.ucar.edu/~ericg/Gilleland2008.pdf

Why Use Objects?

Tressa L. Fowler

Typical situation

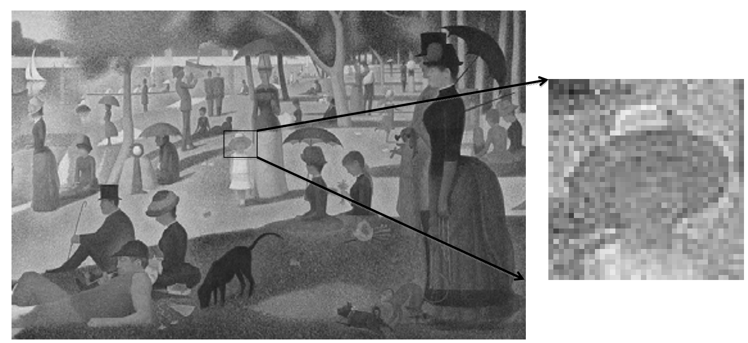


Forecast Observation

Traditional verification matches up points, then sums them up.

Many forecasts are more than the sums of their parts.

Pixels or Pictures?

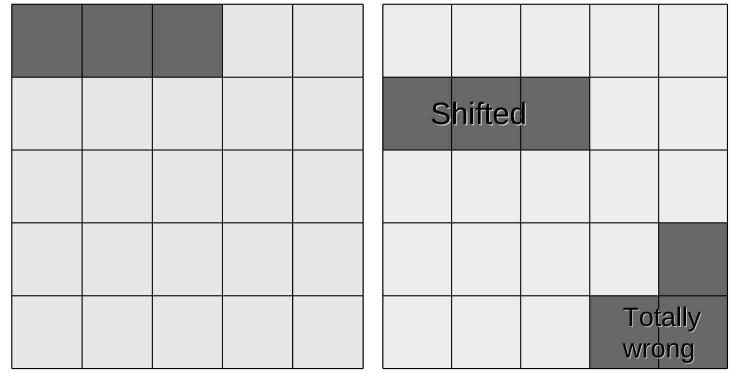


Object verification
is more like what
humans
do.

Objects recognize
the spatial relationship
between points.

Simple example

Observed Forecast

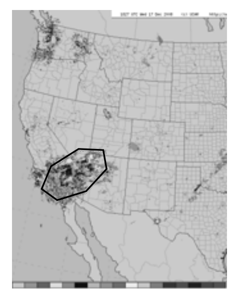


REAL - observed

Forecast 1 –
Distorted view of reality

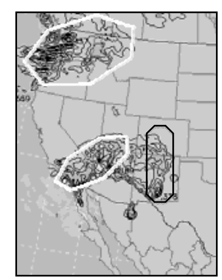
Forecast 2 –
Another distorted
view of reality

REAL - observed



Forecast 1 –
Distorted view of reality

Forecast 2 –
Another distorted
view of reality



This is not really a new idea . . .

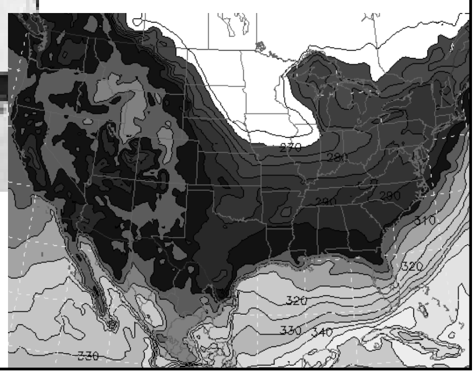
Analytic cubists "analyzed" natural forms and reduced the forms into basic geometric parts on the two-dimensional picture plane.

Analytic cubism was developed between 1908 and 1912 . . .

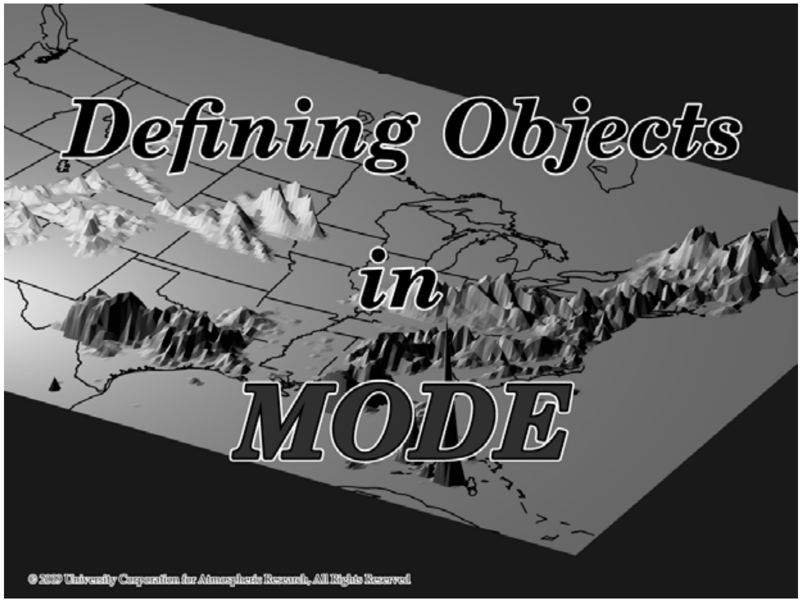
Comparing objects can tell you things about your forecast like . . .

This:	Instead of this:
30% Too Big	POD = 0.35
Shifted west 1 km	FAR = 0.7235
Rotated 15°	CSI = 0.1587
Peak Rain 1/2” too much	

Verifying with objects doesn’t always make sense . . .



- In MET, object based verification is done using the MODE (Method for Object-Based Diagnostic Evaluation) tool.
  - Define objects
  - Compute attributes (e.g. area, centroid, axis angle, intensity)
  - Merge objects (e.g. thunderstorm cells merge into line)
  - Match forecast and observed objects
  - Compare attributes between matches
  - Output summary statistics

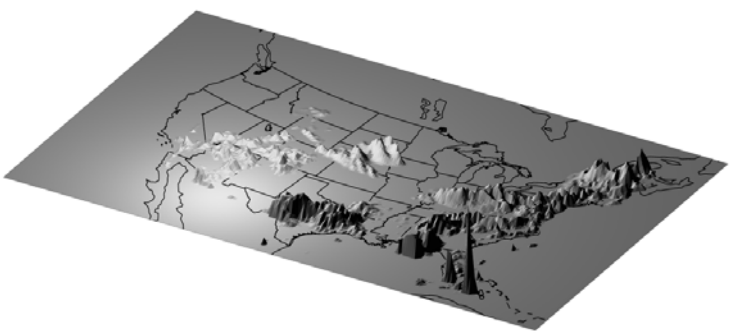


What are Objects?

Raw Field

Object Field

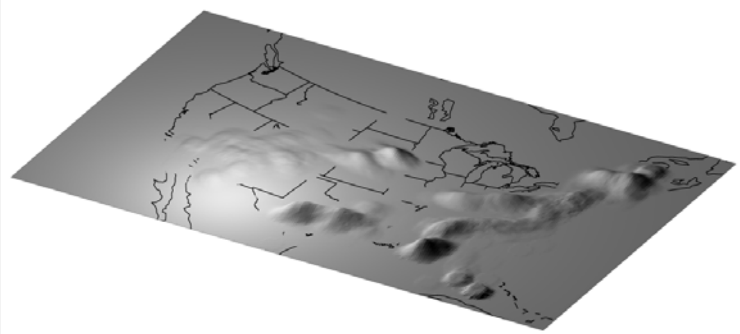
Objects are Regions of Interest



Step #1

Start with the raw data field.

In this case, a precipitation field.

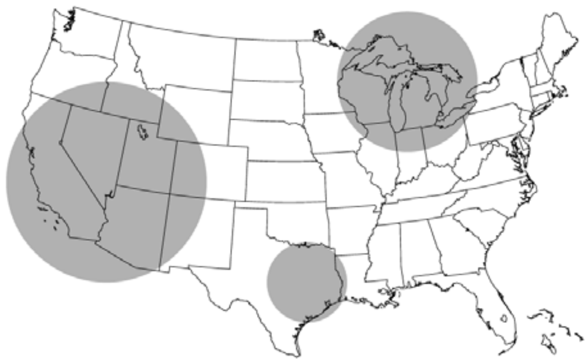


Step #2

Apply convolution operator.

This is basically a smoothing operation.

Convolution Radius



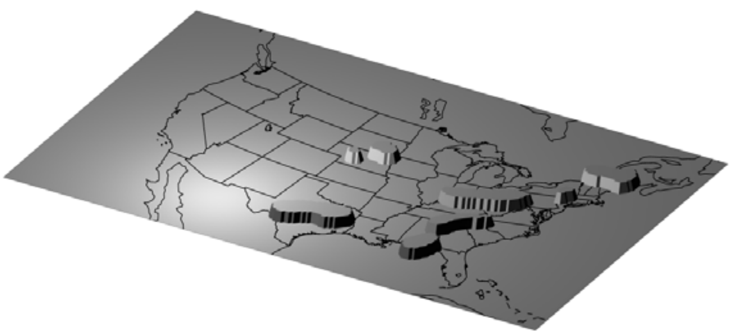
Radius of Influence

Convolution

Requires raw field $f(x,y)$ *and*

filter function $\phi(x,y)$

$$C(x,y) \; = \sum_{(\hat{x},\hat{y})\in G} \phi(\hat{x},\hat{y})\, f(x-\hat{x}, y-\hat{y})$$



Step #3

Threshold the smoothed field.

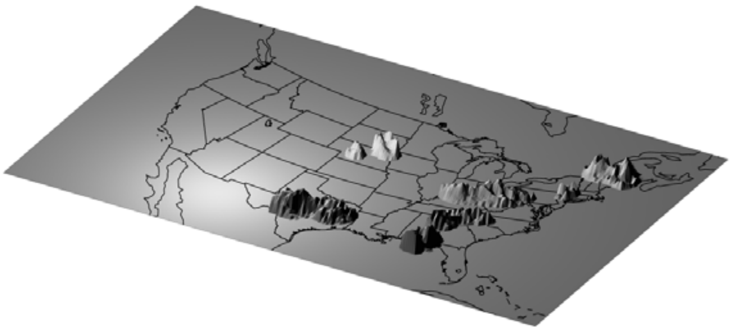
This produces an on/off mask field.

Masking

Uses convolved field $C(x,y)$

and threshold T

$$M(x,y) \; = \begin{cases} 1 & \text{if } C(x,y) \; \geq \; T \\ 0 & \text{else} \end{cases}$$

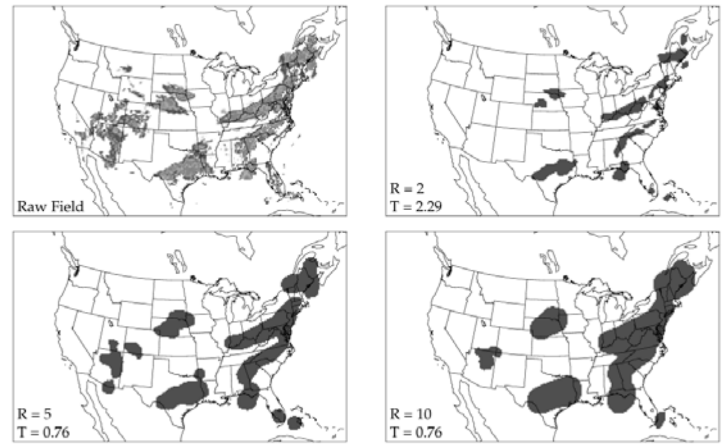


Step #4

Restore original data to object interiors.

This gives us our objects.

Changing Object-Definition Parameters



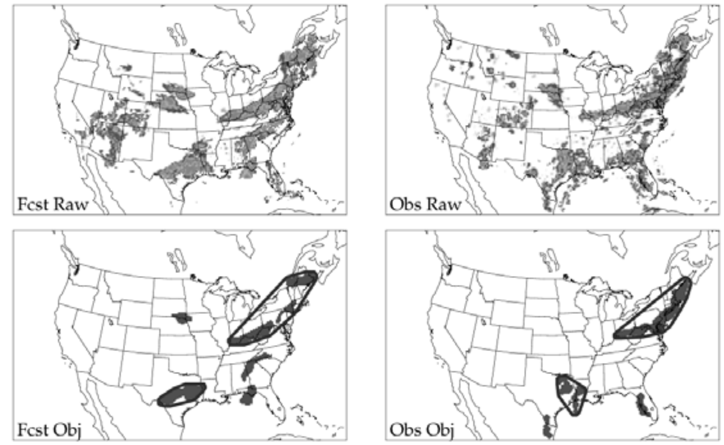
Matching & Merging

Merging: Associating objects in the same field.

Matching: Associating objects in different fields.

MODE does this using a Fuzzy-Logic engine.

Example of Matching & Merging



Object Attributes

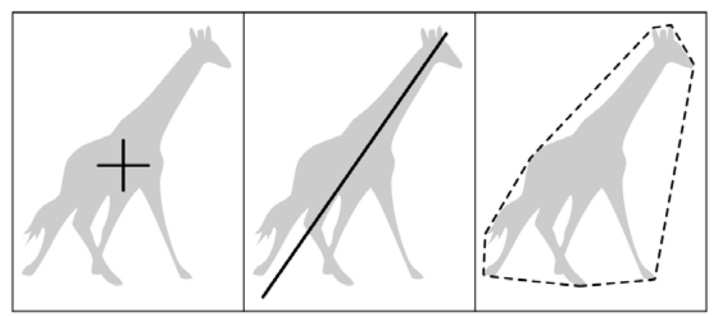
Single:

- Area
- Centroid
- Axis Angle
- Median Intensity
- Complexity
- Aspect Ratio
- Curvature

Pair:

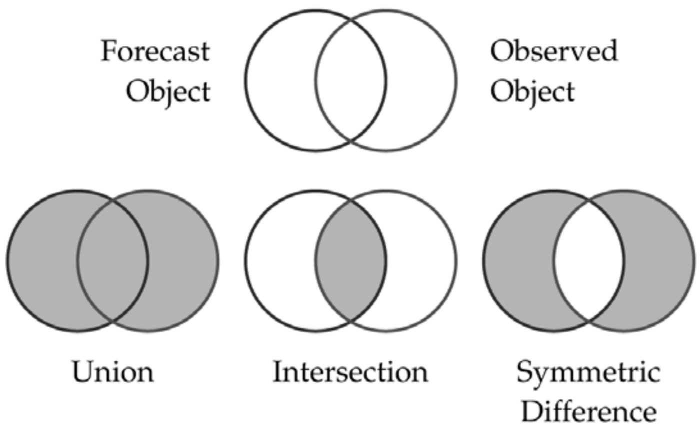
- Centroid Distance
- Angle Difference
- Median Intensity Ratio
- Intersection Area
- Convex Hull Distance
- Boundary Distance
- Area Ratio

Example Single Attributes



Centroid Axis Convex Hull

Example Pair Attributes



Interest Maps

Map attributes to interest values.

Example: Centroid Distance

All interest maps can be changed in the config file.

Weights

Express relative importance of different attributes in matching and merging.

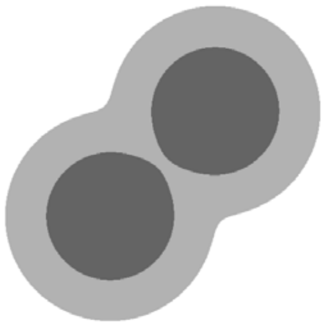
All weights can be changed in the config file.

Total Interest

Calculated from weights, attributes, and interest maps.

$$T(\alpha) = \frac{\sum_i w_i\, C_i(\alpha)\, I_i(\alpha)}{\sum_i w_i\, C_i(\alpha)}$$

Alternative Merging Method



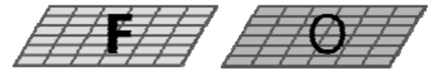
Double
Thresholding

Verifying WRF with Objects

Running MODE and MODE analysis tool

Running MODE - Input

- 2 gridded fields, same grid, possibly different variables
  - forecast (fcst_file) and observation (obs_file)



- configuration settings (config_file)
- syntax
  - `>mode fcst_file obs_file config_file`

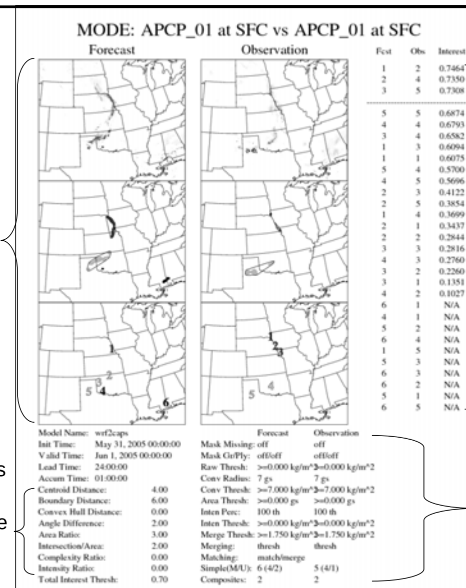
Running MODE - Output

- PostScript
  - object pictures
  - parameter summary
  - total interest for each object pair

object pictures

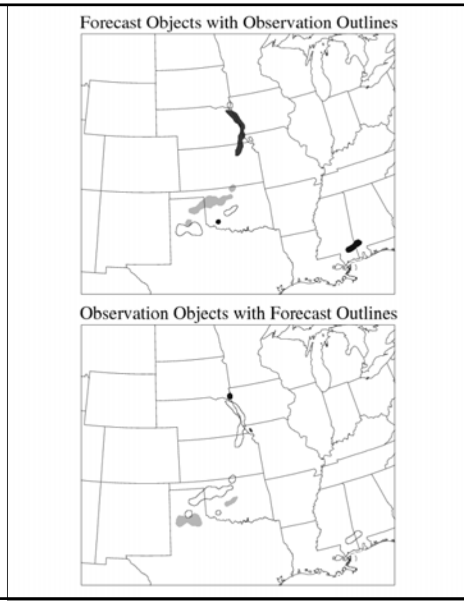
model info
name, times

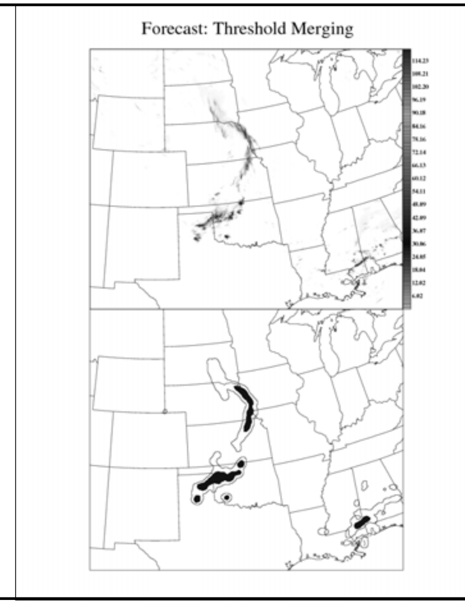
object-attribute weights



object-pairs total interest

parameter summary
threshold
smoothing radius
match/merging





Running MODE - Output

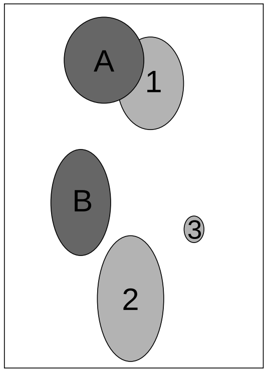
- PostScript
  - object pictures
  - parameter summary
  - total interest for each object pair
- ASCII
  - object sizes, shapes, positions
  - stats for simple, paired objects and clusters
  - standard contingency table stats based on objects
- netCDF
  - gridded object fields

Interpretation of MODE output

- How well do objects match?
  - median of the maximum interest (MMI)*
- What % of objects are matched?
- Displacement of matched objects
- mode_analysis tool

* Davis et al., 2009: The Method for Object-based Diagnostic Evaluation (MODE) Applied to WRF Forecasts from the 2005 SPC Spring Program. Submitted to Weather and Forecasting

Median of the Max. Interest (MMI)



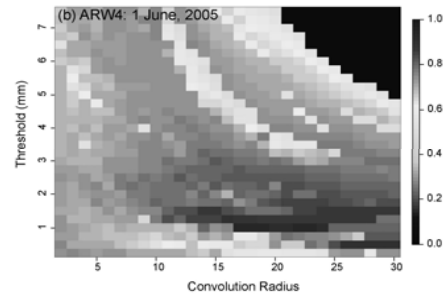
Interest Matrix

observed

forecast	A	B
1	0.90	0.65
2	0.50	0.80
3	0.40	0.55

MMI = median { 0.55, 0.80, 0.80, 0.90, 0.90 } = 0.80

Median of the Max. Interest (MMI) Quilt Plot



MMI as a function of convolution radius (grid squares) and threshold (mm) for 24-h forecast of 1-h rainfall

What % of Objects are Matched?

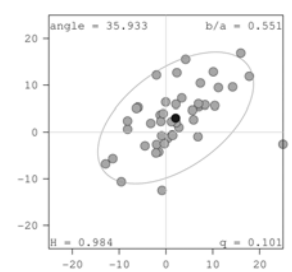
- Hit rate, Threat Score, Heidke Skill Score
- Other standard contingency table scores

	Forecasted object	Unforecasted object
Observed object	Matched	Missed
Unobserved object	False alarm	Correct null

MODE Analysis

- displacement of forecast clusters from matched observed clusters

Composite Object Centroid Differences over the Appalachian Mountains



MODE Analysis Tool
SUMMARY Example

Command Line

```
mode_analysis -summary
              -mask_file ttt -config config/mode_test_config    \
              -dump_lines out -lookin /d1/score/mode_files  \
              -fcst -composite -area_min 3000    \
              -centroid_x_min 600
              -centroid_x_max 1100    \
              -column CENTroid_x   \
              -column centroid_y    \
              -column centroid_lat    \
              -column centroid_lon    \
              -column area   \
              -column axis_ang   \
              -column length
```

Output

```
Total mode lines read =  73,330
Total mode lines kept =     539
```

Field	Min	Max	Mean	StdDev	P10	P25	P50	P75	P90
centroid_x	600.23	914.61	779.36	97.98	626.36	687.96	804.30	866.75	894.24
centroid_y	55.22	560.08	335.55	113.08	189.48	240.58	333.51	421.06	496.82
centroid_lat	22.00	40.14	32.08	4.06	26.83	28.66	32.00	35.15	37.87
centroid_lon	-107.03	-95.01	-100.18	3.75	-106.03	-103.67	-99.23	-96.84	-95.79
area	3210.00	85486.00	12680.96	9931.67	4935.00	6256.00	9445.00	16106.00	23678.00
axis_ang	-88.84	89.90	13.54	44.82	-55.80	-16.36	17.28	48.23	71.27
length	100.57	494.54	200.08	82.11	112.44	133.62	179.96	249.06	315.33

MODE Analysis Tool
By Case Example

Command Line

```
mode_analysis -bycase
              -dump_lines out
              -mask_file ttt -config config/mode_test_config
              -fcst_valid_min 20070702 -fcst_valid_max 20070702_12
              -area_min 3000
              -centroid_x_min 600 -centroid_x_max 1100
               /d1/score/mode_files/ncwf2_vs_ncwdp/* /d1/score/mode_files/rcpf_vs_ncwdp/*
```

Output

Fcst Valid Time	Area Matched	Area Unmatched	# Fcst Matched	# Fcst Unmatched	# Obs Matched	# Obs Unmatched
Jul 2, 2007 00:00:00	12392	20786	0	1	1	1
Jul 2, 2007 01:00:00	6706	11038	0	0	1	2
Jul 2, 2007 02:00:00	7507	18696	0	0	1	3
Jul 2, 2007 03:00:00	19401	32268	2	3	1	2
Jul 2, 2007 04:00:00	0	16551	0	2	0	1
Jul 2, 2007 05:00:00	15311	29730	1	2	1	2
Jul 2, 2007 06:00:00	4730	8182	0	0	1	2
Jul 2, 2007 07:00:00	3733	13285	0	1	1	2
Jul 2, 2007 08:00:00	6994	6994	0	0	1	1
Jul 2, 2007 09:00:00	15981	15981	0	0	2	2
Jul 2, 2007 10:00:00	51501	53427	2	2	4	4
Jul 2, 2007 11:00:00	15779	21089	1	1	1	2
Jul 2, 2007 12:00:00	31339	40665	1	2	2	2

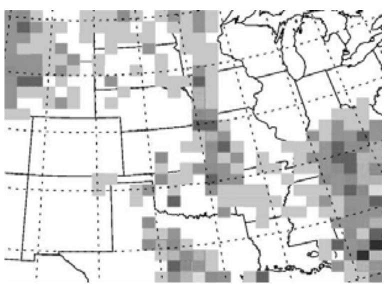
Thurs AM Lecture Wrap-Up

- Confidence intervals
- Verifying WRF with objects
  - matching and merging
- Running MODE
  - MODE analysis
- Questions?

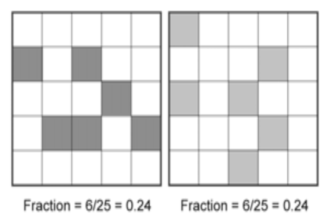
Filter Methods

- Smoothing (e.g., FSS, Upscaling)
  - Neighborhood or “fuzzy” (Ebert 2008)
  - At what resolution does the model have skill?
  - Credit a forecast for being partially correct.
- Band-pass (e.g., Fourier, Wavelets)
  - How well does a model reproduce the spatial structure of the verification field?
  - Examine performance at different scales separately.

Filter Methods



Filter Methods



Roberts and Lean (2008)

Numerator is Fractions Brier Score (FBS)

$$FSS = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(P_{fcst} - P_{obs})^2}{\frac{1}{N}\sum_{i=1}^{N}P_{fcst}^{\;2} + \frac{1}{N}\sum_{i=1}^{N}P_{obs}^{\;2}}$$

Fractions Skill Score (FSS)

Filter Methods

Slide from E Ebert. See Ebert (2008) for full references.

Fuzzy method	Matching strategy*	Decision model for useful forecast
Upscaling (Zepeda-Arce et al. 2000; Weygandt et al. 2004)	NO-NF	**Resembles obs when averaged to coarser scales**
Minimum coverage (Damrath 2004)	NO-NF	**Predicts event over minimum fraction of region**
Fuzzy logic (Damrath 2004), joint probability (Ebert 2002)	NO-NF	**More correct than incorrect**
Fractions skill score (Roberts and Lean 2007)	NO-NF	**Similar frequency of forecast and observed events**
Area-related RMSE (Rezacova et al. 2006)	NO-NF	**Similar intensity distribution as observed**
Practically perfect hindcast (Brooks et al. 1998)	NO-NF	**Resembles a forecast based on perfect knowledge of observations**
Pragmatic (Theis et al. 2005)	SO-NF	**Can distinguish events and non-events**
CSRR (Germann and Zawadzki 2004)	SO-NF	**High probability of matching observed value**
Multi-event contingency table (Atger 2001)	SO-NF	**Predicts at least one event close to observed event**

*NO-NF = neighborhood observation-neighborhood forecast, SO-NF = single observation-neighborhood forecast

Filter Methods

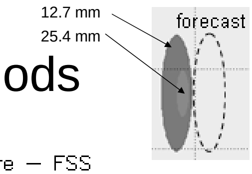
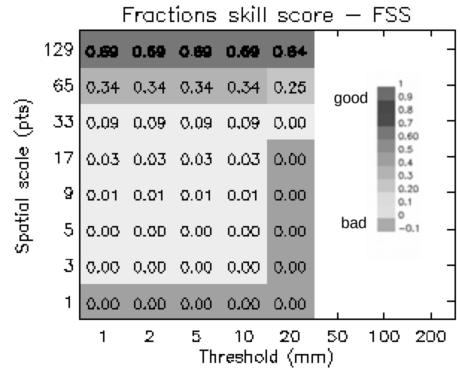


Fig. courtesy of E Ebert.



Filter Methods

- In MET, Neighborhood methods are in grid-Stat tool. See MET documentation for specific filters.
- Can be cumbersome (many numbers for one filter).
- Choose filter that addresses the verification question of interest.
- See Ebert (2008) for a good summary and comparison of these techniques (and references).

Band-pass Methods

- Fourier

$$Z(\mathbf{x}) = \sum_{k=1}^{n} Z_k \exp\left(i\omega_k^T \mathbf{x}\right)$$

- Wavelets

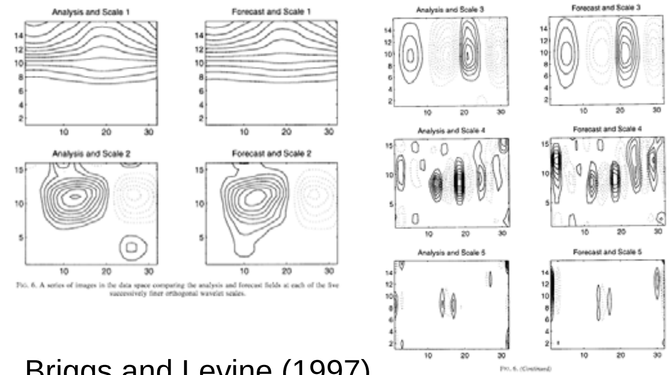
$$Z(\mathbf{x}) = \sum_{j,k} c_{j,k}\varphi_k\left(2^{j-1}\mathbf{x}\right) + \sum_{j,k} d_{j,k}\psi_k\left(2^{j-1}\mathbf{x}\right)$$

Band-pass Methods

- Fourier
  - Skamarock (2004), MWR **132**:3019-3032
  - Harris *et al*. (2001), J Hydrometeorol. **2**:406-418
  - Tustison *et al*. (2001), JGR **106**(D11):11775-11784
  - *and many more…*
- Wavelet
  - Briggs and Levine (1997), MWR **125**:1329-1341
  - Casati *et al*. (2004). [In MET Wavelet tool]

Band-pass Methods

Wavelets

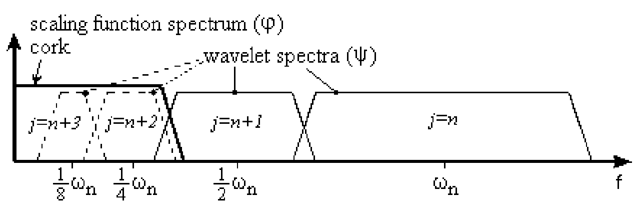


Briggs and Levine (1997)

Band-pass Methods

Wavelets

$$Z(\mathbf{x}) = \sum_{j,k} c_{j,k}\varphi_k\left(2^{j-1}\mathbf{x}\right) + \sum_{j,k} d_{j,k}\psi_k\left(2^{j-1}\mathbf{x}\right)$$

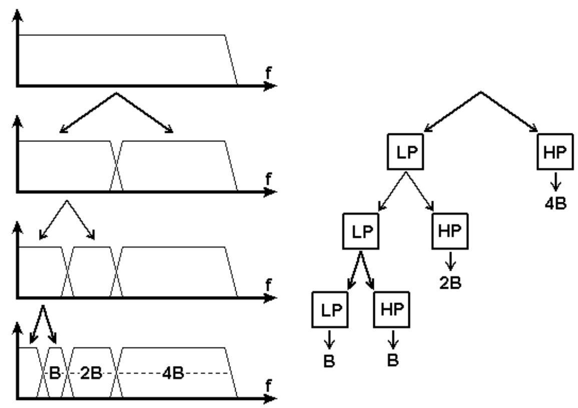


Lifted from:

http://pagesperso-orange.fr/polyvalens/clemens/wavelets/

Band-pass Methods

Wavelets



Lifted from:

http://pagesperso-orange.fr/polyvalens/clemens/wavelets/

Band-pass Methods

Wavelets: Practical considerations

- Applicable only on square grids of size, 2^n.
- If grid does not meet this, then must alter the grid in some way. For example,
  - Pad with zeros,
  - Cut the domain down.
- Several choices for wavelets. Haar Wavelets are a good choice for precipitation (because of their discrete nature), but other choices might be better for other fields.

Quick Contingency Table Review

	obs > threshold	obs < threshold	
forecast > threshold	a hits	b false alarms	a+b
forecast > threshold	c misses	d correct rejections	c+d
	a+c	b+d	n=a+b+c

Band-pass Methods

Intensity Scale (IS) (Casati *et al*., 2004)

1. Create binary fields for a threshold, and calculate MSE (denote, MSE(t)).
2. Apply wavelet decomposition to binary fields.
3. Calculate MSE for each threshold (t) and scale (j) (denote, MSE(t,j)), and find MSE%(t,j)=MSE(t,j)/MSE(t).
4. Find $MSE(t)_{random}$ = FBI*Br*(1-Br) + Br*(1-FBI*Br), where FBI=(a+b)/(a+c) is the frequency bias index and Br=(a+c)/n is the sample climatology from the contingency table.
5. Calculate SS(t,j)=1-MSE(t,j)*(n+1)/$MSE(t)_{random}$.

Band-pass Methods

In addition to MSE and SS, the energy squared (denote En2) is calculated. That is,

$$En2(X) = \sum_{i=1}^{n} X_i^2$$

Band-pass Methods

Also find the relative difference in En2 between the forecast and verification fields. That is,

[En2(F)-En2(O)]/[En2(F)+En2(O)]

Similar to MSE, let En2(X,t) be the energy for the raw binary field X (where X is either the verification or forecast field), and En2(X,t,j) is the same, but taken only at scale j. Then

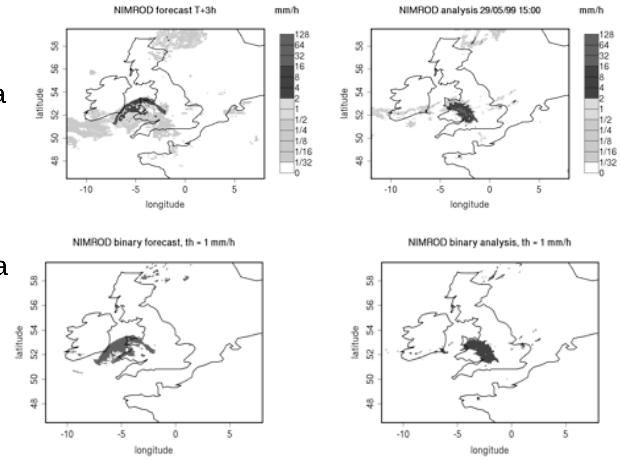
En2%(X,t,j) = En2(X,t,j)/En2(X,t).

Casati *et al*.(2004)

Forecast   Observation

Raw Data

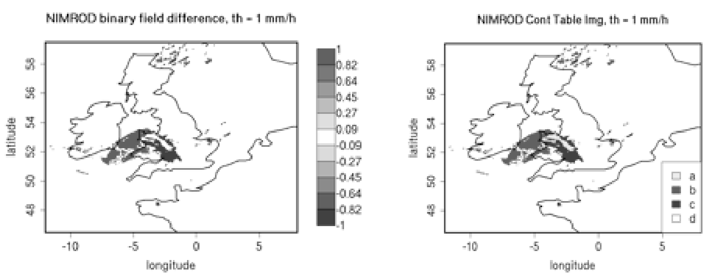
Binary Data



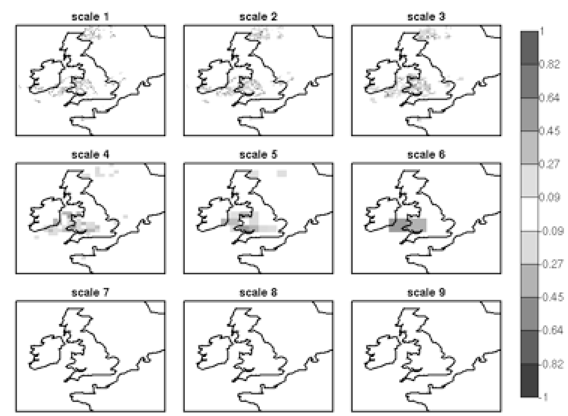
Casati *et al*.(2004)

Binary Difference Field

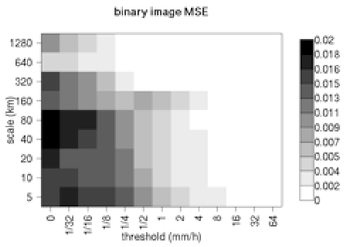
Contingency Table Contributions

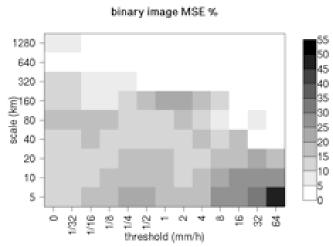


Casati *et al*.(2004)



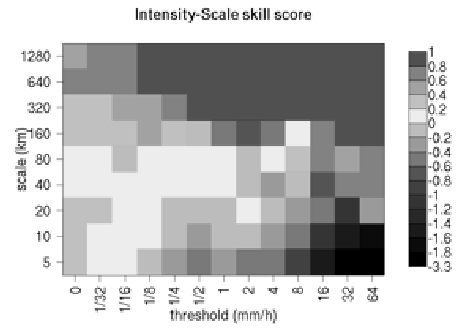
Casati *et al*.(2004)





Casati *et al*.(2004)

Intensity Skill Score (ISS) is the Heidke Skill Score (HSS), but calculated at each threshold and wave number.



HSS is an accuracy score corrected for random chance.

Thank you…Questions?

References

Casati B, G Ross, and DB Stephenson, 2004. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl*. **11**:141--154.

Ebert EE, 2008. Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteorol. Appl*., **15**:51--64. DOI: 10.1002/met.25.

See also, http://www.ral.ucar.edu/projects/icp

Customizing and Understanding Wavelet Stats

Tressa L. Fowler

Options for Handling Missing data

- Points with valid forecasts. F
- Points with valid observations. O
- Points with both valid forecasts and observations. $F \cap C$
- All points. $F \cup C$

Thresholds

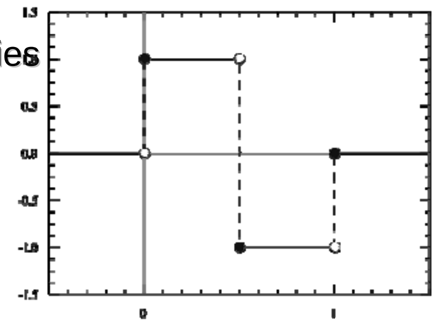
Forecast		Threshold	Event
0.05	<	0.2	0
0.17	<	0.2	0
0.45	>	0.2	1
2.15	>	0.2	1
0.05	<	1	0
0.17	<	1	0
0.45	<	1	0
2.15	>	1	1

Tiles – Grid must be 2^n x 2^n

- Cut down – user selected subset (square)
- Tiles – automated selection of subset(s)
- Pad with zeros – not recommended unless adds very small number of points.

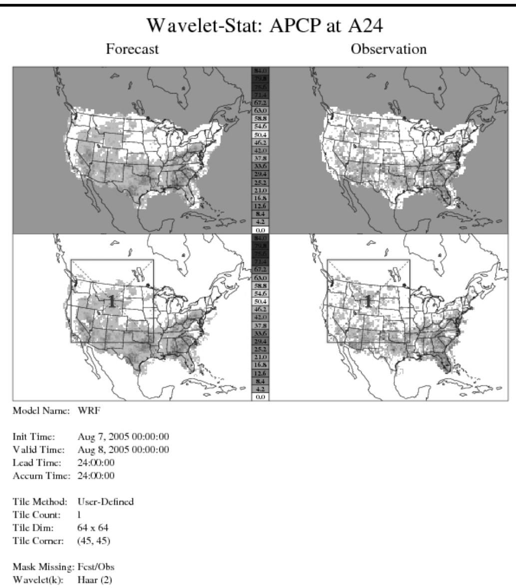
Wavelets

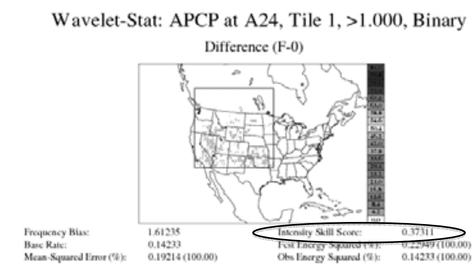
- Haar
- Centered Haar
- Daubechies
- Centered Daubechies
- B spline
- Centered B spline



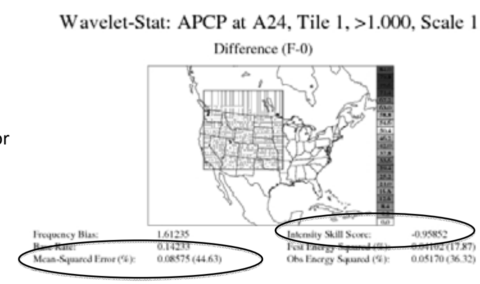
Output

- Text files
  - Options
  - Statistics
  - Can be read into stat analysis tool.
- Postscript files
  - Graphics
  - Options
  - Statistics



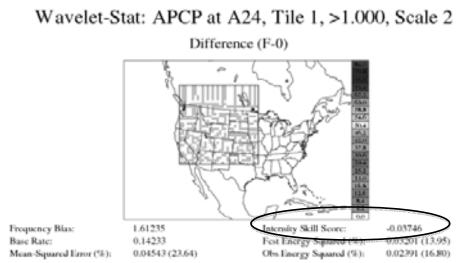


Overall forecast has skill (ISS > 0)

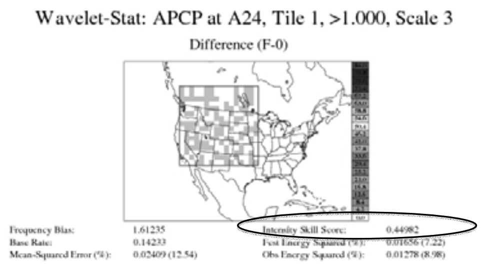


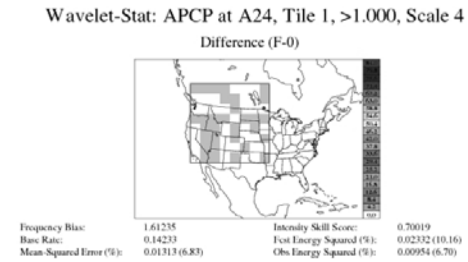
Errors at this scale account for nearly half of the MSE.

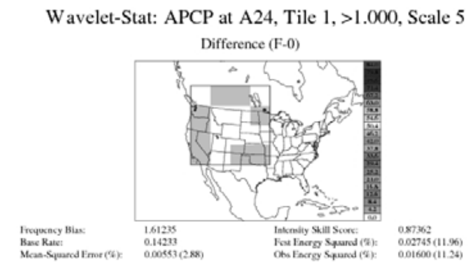
At this scale, forecast does not have skill (ISS < 0)

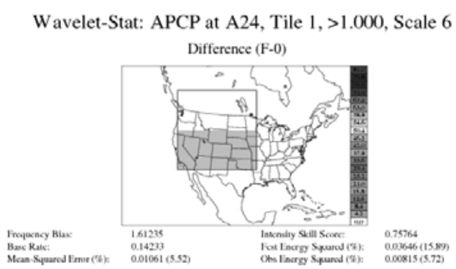


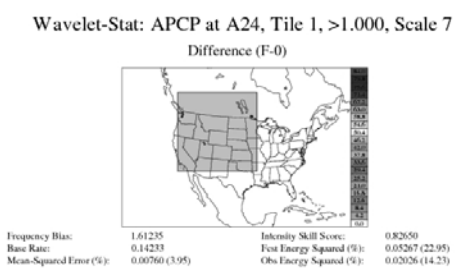
Forecast transitions from no skill to skill at scale 3.











Summary

- Wavelet tool provides a flexible method for decomposing spatial fields into different scales.
- Once decomposed, verification measures at each *physical* scale can be examined and compared.