SAMURAI Users Manual 1.0

Spline Analysis at Mesoscale Utilizing Radar and Aircraft Instrumentation

SAMURAI is a variational analysis technique developed based primarily on the work of Ooyama (1987), Ooyama (2002) and Gao et al.(2004). The SAMURAI analysis yields a maximum likelihood estimate of the atmospheric state for a given set of observations and error estimates by minimizing a variational cost function.

The technique has several advantages over traditional objective analysis techniques, including:

- observational error specifications for different instrumentation
- more complex observation operators for remote sensing data
- the addition of balance constraints such as mass continuity
- including a priori background estimates of the atmospheric state when available.

A distinguishing characteristic of the SAMURAI technique compared to other variational solvers is that the analysis can be performed directly in an axisymmetric cylindrical coordinate system or in a 3D Cartesian coordinate system. The two-dimensional solver improves the computational efficiency and minimizes potential errors in mass conservation that arise when interpolating from a three-dimensional domain.

Another distinguishing characteristic from other variational solvers is the use of a Galerkin approach, which is similar to the Fourier spectral transform but uses the cubic B-spline as a basis Ooyama (2002). The disadvantage of the B-spline basis is that it is not orthogonal and therefore requires an extra matrix to obtain the spline coefficients, but this is a fair trade-off with its other desirable characteristics. The basis is computationally efficient and continuously differentiable to second order, allowing for efficient, accurate interpolation to observation locations, flexible incorporation of boundary conditions, and high numerical accuracy of kinematic derivatives.

The analysis is performed in a manner similar to the spectral transform method Machenhauer (1979), transforming to and from the spline coefficients and physical space at each step of the cost function minimization.

A more technical description of SAMURAI is given in the appendix of Bell et al. (2012).

Please reference Bell et al. (2012) if you use SAMURAI in your research. A description of the 3D version including analytic tests is currently in preparation, and will serve as a useful citation for v1.x shortly.

Contributing to SAMURAI

- Check out the latest master to make sure the feature hasn't been implemented or the bug hasn't been fixed yet
- Check out the issue tracker to make sure someone already hasn't requested it and/or contributed it
- Fork the project
- Start a feature/bugfix branch
- Commit and push until you are happy with your contribution, then open a Pull Request.
- Make sure to add tests for the feature/bugfix. This is important so I don't break it in a future version unintentionally.

Copyright

References

Bell, M. M., M. T. Montgomery, and K. A. Emanuel, 2012: Air-sea enthalpy and momentum exchange at major hurricane wind speeds observed during CBLAST. *J. Atmos. Sci.*, *in press*.

Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, **21**, 457–469.

Machenhauer, B., 1979: The spectral method. *Numerical methods used in atmospheric models*, A. Kasahara, Ed., GARP Publications Series No 17, WMO and ICSU, Vol. 2, pp. 121–275.

Ooyama, K. V., 1987: Scale controlled objective analysis. *Mon. Wea. Rev.*, **115**, 2479–2506.

Ooyama, K. V., 2002: The cubic-spline transform method: Basic definitions and tests in a 1d single domain. *Mon. Wea. Rev.*, **130**, 2392–2415.

Compilation

To compile, use [CMake](#) from the top-level directory:

```
$ cmake .
```

to create a Makefile or Xcode project for your machine. Run `make` or build via Xcode to create the `samurai` binary. Run `make install` to install the binary to /usr/local/bin

To install samurai in an alternate location:

```
$ cmake -DCMAKE_INSTALL_PREFIX=/some/other/location .
```

To use an alternate compiler (for example the Intel C++ compiler):

```
$ cmake -DCMAKE_CXX_COMPILER="icpc" .
```

The program is known to work with GCC and Intel compilers, but has not been tested with other compilers.

A few external libraries are required:

[Geographiclib](#) is used for geolocation of data and map projection.

[Qt](#) is used for helper classes including XML parsing. A graphical user interface based on Qt will be available in a future release.

[NetCDF](#) is used for output of the gridded analysis results.

[cURL](#) and [HDF5](#) are prerequisites for NetCDF4.

Installation

The pre-compiled binaries for Mac and Linux contain the necessary libraries in the tarfile. You can untar the binary and libraries into your favorite directory and it should work from there. On Linux, run the shell script `samurai.sh` instead of `samurai` in the bin directory to use the included libraries. The Mac version has the library paths embedded in the binary. If you compiled from source, then you can run the binary from anywhere as long as the system libraries are in your path.

Running SAMURAI

The `samurai` binary takes a single argument in the form of an XML configuration file describing the analysis. Observational data and reference frame information should be placed in a subdirectory specified in the configuration.

The current interface is available for 3D Cartesian and 3D cylindrical analysis. The 2D Cylindrical interface will be added soon.

Several utility scripts are included in the `util` subdirectory, and more will be added as available.

This documentation covers the configuration and interface for the 1.x release series. There are no plans to change the mandatory elements of the XML configuration structure through the 1.x releases. Optional elements may be added, but no changes would be required for existing analysis configurations.

Example XML files for Cartesian and cylindrical analyses are included in the `bin` directory and are described below. During initialization the program looks for data in the specified `data_directory` subdirectory.

The only requirement for doing an analysis is a frame-of-reference file in the `data_directory`. This file contains 1 second positions of a lat/lon reference point for the domain, and the motion of the frame. For tropical cyclone (TC) analyses this is usually the center of circulation, but can be any arbitrary reference point. The filename is *YYYYMMDD.cen* where the date refers to the first line of the file. The file structure is multiple lines consisting of *HHMMSS Latitude Longitude V-speed U-speed*. Times are in UTC, and exceed 240000 when crossing over to the following day. Note that V and U follow latitude and longitude (not U then V). Only data from the first time to last time in the file will be used in the analysis, so this file sets both the time limits and analysis location. A simple linear track can be generated using the supplied `samurai_lineartrack.pl` script in the util directory.

If a background state is used, the software looks for a file called `samurai_Background.in`. This is a textfile consisting of kinematic and thermodynamic data for an *a priori* estimate of the atmosphere. Numerical model analyses, idealized states, or other estimates can be used in this file. The file format is:

Time (UTC unix seconds past 1-1-1970) Latitude Longitude (deg) Altitude (m) U V W (m/s) Temperature (K) Qv (g/kg) Dry air density (kg/m-3)

The time has to be in the limits specified by the frame-of-reference file, and the data should extend beyond the analysis domain boundaries. The data does not have to be evenly spaced, but the ordering of the positions is important. Data should be sorted by altitude, such that each successive column of data is grouped together with altitude increasing. The data from the background file is objectively analyzed and then transformed to spline coefficients. If greater fidelity and mass-conservation of the background is desired, the background state can be "adjusted" to closely match the pseudo-observations given in the file and mass-balanced.

SAMURAI can read several text data formats commonly used for aircraft observations. All datafiles in the `data_directory` will be included in the analysis. The read subroutine for each datatype is determined by the file suffix.

- frd = Dropsonde format used by HRD
- cls = Dropsonde format used by older versions of ASPEN
- sec = 1 second flight level data format used by HRD
- ten = 10 second flight level data format used by USAF
- swp = Doppler radar format used by NCAR/EOL
- sfmr = Custom format for SFMR data
- Wwind = Dropsonde format used by some versions of ASPEN (modified CLASS)
- eol = Dropsonde format used by most recent version of ASPEN (modified CLASS)
- qscat = Custom format for QUIKSCAT data
- ascat = Custom format for ASCAT data
- nopp = Custom format for NOPP scatterometer products
- cimss = Atmospheric motion vector format used by CIMSS/Wisconsin
- dwl = Doppler lidar format used by Simpsons Weather Associates
- insitu = Custom format for insitu observations

Native formats should work without too much trouble. Please refer to the VarDriver.cpp code for details on the custom formats. Instrument errors are configurable in the XML configuration, but are currently limited to fixed errors per instrument except for Doppler radar.

Configuration

The configuration file is a standard XML format that allows the user to customize the analysis. Details on the specific options are given below.

grid

These options define the analysis grid.

i*min,* i*max,* i*incr,* j*min,* j*max,* j*incr,* k*min,* k*max, k_incr*

The minimum and maximum horizontal distances are relative to the moving frame-of-reference. The vertical distance is relative to the surface. Topography is not currently supported, but will be in a future release. All distances are in kilometers, and the number of gridpoints is calculated from the given parameters. For the RTZ mode, the j dimension is in degrees. For periodic boundary conditions, the min and max should be the points where the domain matches (e.g. 0 and 360).

operation

These options determine how the program runs.

mode

The mode of operation. Two modes are currently supported. The "XYZ" option performs a 3D Cartesian analysis. The "RTZ" option performs a 3D cylindrical analysis. A 2D cylindrical mode is available in the code, but is not currently exposed to the interface. An axisymmetric version of the 3D cylindrical analysis mode will be available later in the 1.x series.

load_background

If this option is set to 'true' then the `samurai_Background.in` file will be loaded and objectively analyzed.

adjust_background

If this option is set to 'true' then a variational analysis will be conducted using the pseudo-observations. The analysis parameters will be set to those from iteration 1 below.

preprocess_obs

If this option is set to 'true' then the raw data files will be preprocessed according to their file suffixes. The output of the preprocessor is a text file called `samurai_Observations.in`. If this option is set to 'false' then the program will load the observations directly from the Observations.in file, located in the `data_directory`. Setting this to 'false' is useful for synthetic observations, but is not recommended for restarts at the current time.

num_iterations

Multiple iterations of the analysis can be performed by incrementing this number. The parameters for each iteration can change, which is useful for multi-pass (i.e. coarse to fine) analyses. The analysis from the previous iteration is used as the background for the next iteration.

output_mish

The analysis is performed computationally on a Gaussian quadrature "mish" that is finer resolution than the analysis grid defined above. Normally you just want the nodal values, but you can also output the mish values if you like. This would be useful if you are planning to integrate any of the quantities. The mish values are only output to the text file, not to the NetCDF or Asi files.

output_txt

The data can be output in a tab-delimited text file for easy ingest into Matlab or other analysis software. Columns are labeled with the variable names, but see the NetCDF output for units and long names. The output file name is samurai__analysis.out (e.g. samurai_XYZ_analysis.out).

output_qc

A summary text file of all the observations and their weights that went into the analysis can be output. This file is useful for comparing the analysis versus the observations. The output file name is samurai_QC_analysis.out

output_netcdf

The most compact and complete output format is NetCDF. Metadata follows the COARDS conventions for use with GRaDs. The output file name is samurai__analysis.nc

output_asi

This output format is an ASCII representation of the CEDRIC data format used in radar analysis. It can be used with the grid2ps graphics package.

background

These options set information about the background state.

ref_state

This is a filename containing parameters that define the hydrostatic reference state used in the analysis. The default is the Dunion (2011) moist tropical sounding (dunion_mt.snd) which will be used if the specified sounding file cannot be found. The sounding must be given in terms of a surface pressure value in hPa, and 4th order polynomial curvefit coefficients for vertical profiles of qv, dry air density, and vertical pressure gradient. Note that qv is given in terms of a biased hyperbolic transform, which is approximately 0.5*qv for values above 10^-7 g/kg. The pressure gradient is in Pa/m. Temperature is calculated from the other quantities. The 4th order fit allows for easy differentiation and integration to ensure hydrostatic balance and accurate vertical derivatives. The drawback is that complex reference states are not possible. Since the main purpose of the reference state is to remove strong vertical gradients from the analysis, not necessarily provide a perfectly realistic sounding for numerical weather prediction, the 4th order representation is believed sufficient. Different (better?) sounding specifications may be available in a later release as requested.

ref_time

This defines the reference time HH:MM:SS when outputting to netCDF. This time needs to be within the defined frame-of-reference time limits, and will be used to convert the Cartesian analysis to lat/lon in the netCDF file.

i_background_roi, j_background_roi

The background radius of influence in km, or degrees in j dimension of RTZ mode. This sets the radius of influence for the exponential weighting function used to objectively analyze the background estimates. The ROI should be set to approximately the resolution of your background field, or slightly longer for a smoother background. The exponential weighting function is given by $\exp(-2.302585092994045*r^2/R^2)$ where r is the distance from the background estimate to the mish location, and R is the magnitude of the radius of influence vector. The multiplier is used such that the weight drops to 0.1 for one node increment, and 0.01 for two node increments.

The vertical interpolation is performed by cubic B-splines in a logarithmic height coordinate. This allows for better interpolation from pressure or sigma coordinate background fields and maintains high resolution in the vertical.

"Adjustment" of the background field is recommended for large departures from the original background geometry to the regular SAMURAI height grid.

radar

These options set radar analysis parameters.

radar_skip

This option can be used to skip beams in the radar data. When set to '1', all beams are used. This is primarily used to thin the data to decrease the computational burden when using a lot of radar data.

radar_stride

This option sets the number of gates over which averaging occurs along the beam. A stride of '1' uses all data, and higher numbers average multiple gates of the given stride. This is used to thin the data for computational reasons, to reduce noise, or to reduce the spatial scale of the radar observations.

dynamic_stride

When set to '1' the stride is dynamically determined based on the range. At short range, the minimum stride is set to the radarstride value. As the range increases, the stride increases to try and approximate spherical pulse volumes.

qr_variable

In the default 'dbz' mode, the reflectivity is objectively analyzed and not included in the cost function. If this is set to 'qr', then reflectivity is converted to liquid water using Z-M relationships defined in Gamache et al. (1993) and used as an additional variable in the cost function minimization. This could be used to combine reflectivity with other liquid water measurements, but does increase the computation time.

i_reflectivity_roi, j_reflectivity_roi, k_reflectivity_roi

The reflectivity radius of influence in km, or degrees in j dimension of RTZ mode. This sets the radius of influence for the exponential weighting function used to objectively analyze the reflectivity. A vertical radius of influence is included here, as opposed to the background ROI, to account for the interpolation from a spherical coordinate system used in radar scanning as opposed to (generally) from pressure or sigma coordinates for background fields. See background_roi description above for details on the exponential weighting function.

dbz_pseudow_weight

Setting this greater than '0' will create pseudo-observations of the vertical velocity boundary condition based on the reflectivity. A non-zero value sets a weak constraint for zero vertical velocity at echo-top and at the surface. The number specifies the pseudo-error in m/s. Note that larger values set a weaker constraint, and smaller numbers greater than zero set a stronger constraint. Strong vertical velocity constraints at the surface and domain top can be set using boundary conditions, but this is provides a weak constraint at the surface and a variable height in the middle of the domain.

radar_dbz

The name of the reflectivity field in the radar data.

radar_vel

The name of the Doppler velocity field in the radar data.

radar_sw

The name of the spectrum width field in the radar data.

mask_reflectivity

The analysis can be set to missing data where there is no reflectivity. If set to 'None' then no masking is performed. A

numerical value will be used as a threshold for the masking, with all data at nodes having less than the given reflectivity value removed.

parameters

These parameters set the errors, filtering, and weights for the cost function for each iteration of the outer loop. The iteration is set via the "iter" attribute. If the number of iterations is less than "iter" then the parameter values are ignored.

bg_rhou_error, bg_rhov_error, bg_rhow_error, bg_tempk_error, bg_qv_error, bg_rhoa_error, bg_qr_error

The error numbers represent the standard deviation of the background error for each variable. If no background field is used, these should be set relatively high.

mc_weight

This number specifies the weight given to the mass continuity constraint. Mass continuity is enforced by pseudo-observations of the momentum gradients at the spline nodes. Larger numbers indicate a stronger constraint, zero means no constraint.

i_filter_length, j_filter_length, k_filter_length

These numbers set the Gaussian recursive filter length scale in gridpoints. If set to '-1', the recursive filter is turned off for the given dimension. The minimum recommended value is 2.0, with higher values corresponding to larger spatial influence of the observations. Higher values also result in a smoother analysis. A journal article and Wiki page describing the available filters in SAMURAI are in preparation.

i_spline_cutoff, j_spline_cutoff, k_spline_cutoff

These numbers set the spline cutoff length scale in gridpoints. See Ooyama (2002) for a discussion of the spline cutoff length. The recommended value is 2.0.

i_max_wavenumber, j_max_wavenumber, k_max_wavenumber

These numbers act as a direct filter in frequency space. They are currently NOT implemented, but are included as placeholder for future 1.x versions.

boundary_conditions

Homogeneous spline boundary conditions can be enforced in either or both the horizontal and vertical. Available options are R0, R1T0, R1T1, R1T2, R2T10, R2T20, R3, and PERIODIC following Ooyama (2002). The default "non"-boundary condition (R0) adds a buffer set of gridpoints that are used to calculate the solution but are discarded for output. Different boundary conditions can be set on the left (L) or right (R) side of the domain for each variable and dimension. The most common option other than R0 would be R1T0 for vertical velocity (rhow = 0) at the surface and/or domain top. Periodic domains are only valid for the i and j dimension, but are available in both the XYZ and RTZ mode. R1T10 and non-homogeneous boundary conditions should be available in a future release.

observation_errors

Errors for a variety of standard instrumentation are included in the example configurations. Note that the specified error is given in terms of a standard deviation, and includes both instrument and representation error. Errors are fixed for all observations from a particular instrument except radar. In the radar case, the spectrum width and elevation angle are used to define the error for each radar gate. A minimum error (radarminerror) is also enforced.

options

This group is reserved for future options that are not mandatory. Users wishing to implement new options are

encouraged to add them here for inclusion in a future release.

spline_approximation

The cubic B-spline used in the analysis requires several mathematical evaluations to compute, and is called many times during an analysis. To reduce computation time, the spline can be approximated by a pre-calculated look-up table with ~1e-12 or ~1e-6 precision. A value of '0' here uses the full, unapproximated spline calculation, with '1' or '2' increasing the approximation and decreasing the precision. Tests suggest that the results are minimally changed with the '2' approximation, but computation time can be decreased significantly.