15 September 2016

TO: ARISTO-LAMS archive
FROM: William Cooper
SUBJECT: workflow for generating this report

1 Purpose

This workflow description documents the steps leading to the code in “ARISTO-LAMS.Rnw” and provides additional detail not in the associated report “ARISTO-LAMS.pdf.” ARISTO-LAMS.Rnw contains both text (in LaTeX format) and R processing code for the analyses in the resulting report that describes some studies of the measurements from the LAMS on the C-130 in the 2015 and 2016 ARISTO project. The description of workflow provided here includes the process of collecting the observations and processing them to data files, the data archives used, the steps required to generate the plots and other results including the instances where manual intervention is required to identify appropriate subsets of the data, the relevant R code and LaTeX documents, and all the steps leading to the generation of the text in the manuscript. "ARISTO-LAMS.Rnw" is the authoritative reference, but this overview may help explain the workflow at a general level and so could substitute for or supplement reading the R and LaTeX code in most cases. The intent is to describe the workflow in sufficient detail to support replication of the analysis and figures presented in the report, to facilitate changes based on new data or new analysis approaches, and to make it practical to apply the proposed algorithms to data collected by research aircraft used in atmospheric studies.

The description provided here is not complete because many aspects of the workflow have already been included in the main document. This is indended to supplement those descriptions but not provide a complete substitute for them.

2 Acquisition of the primary data

The measurements used in this report were collected using the NSF/NCAR C-130 research aircraft during the ARISTO projects of 2015 and 2016. The onboard data-acquisition program ’aeros’ recorded the data in digital format, and those data files were then processed by the program ’nimbus’ to produce an archive in NetCDF format. The software management group of NCAR/EOL maintains a version-controlled archive of these programs, so if they are of interest they can be obtained by contacting the data-management group of EOL (at this or this address). The data files available from NCAR/EOL can be found at links on this URL. The details of the processing algorithms are documented here: Processing Algorithms. Some procedures as they pertain to the measurement of wind are also documented in Cooper et al. (2016), and there is supplemental material on the LAMS there also. The resulting data files contain measurements in scientific units and brief descriptions of each measurement, included as netCDF attributes for the files and for each variable.

One special file was generated and used in ARISTO-LAMS.Rnw, a netCDF file named AR16rf06hr.nc. It was generated by the program nimbus (version of June 2016) by specifying a 25-Hz data rate, with LAMS histograms at 50 Hz. At the time of this run, the 25-Hz option for the beam histogram variables (BEAM[1-4]_LAMS) did not work so it was necessary to produce those variables at 50-Hz and then average to 25 Hz during processing. That netCDF file is not part of the standard project archives and so must be obtained separately, as described in the "Reproducibility" appendix to the report.

3 The ARISTO-LAMS.Rnw file

The .Rnw file is basically LaTeX text, generated for simplicity using LyX and exported to .Rnw format and then processed in RStudio (RStudio (2009)). The .lyx file (ARISTO-LAMS.lyx) will run equivalently and produce a PDF-format version of the manuscript. The .Rnw file was produced by exporting in 'Rnw (knitr)' format from LyX. Within the .Rnw file or within the .lyx file there are "chunks" of R code (R Core Team (2013)), delineated by a header having this format:

```
<<title, var=setting, ...>>=
...R code...
@
```

These chunks generate plots and other results of analyses that are incorporated into the manuscript using 'knitr' (Xie (2013, 2014)). In RStudio, the chunks appear as gray sections in the file when it is edited. Where tasks involve execution of R code, the chunk containing the code is referenced in the discussion below. Any results from the processing can be incorporated into the LaTeX text via "\Sexpr{}" calls embedded in the LaTeX portion of the file.

Two "switches" serve to speed execution of the code: (1) "ReloadData," which when TRUE causes the original archive data files to be read; and the option "cache" which, when true, causes previously generated subsets of the results to be re-used as saved in a subdirectory "cache." Both are provided to speed execution in cases where small changes are made. When ReloadData is FALSE subsets of data files previously saved as "Rdata" files are restored to R data.frames, a process that is much faster than re-reading the original archive data file. If cache is FALSE all calculations are performed, but using "cache=TRUE" is usually much faster after the first run.

The file and the report are organized into seven sections, each of which is discussed in this workflow document. This document was in turn generated by a LyX file ("WorkflowARISTO-LAMS.lyx").

4 Required R packages including Ranadu

The R code used for analysis reported in this paper relies heavily on a package of routines for R called "Ranadu." This is a set of R scripts for working with the NetCDF archive data files produced by NCAR/EOL/RAF, and it includes some convenience routines for generating plots and performing other data-analysis tasks with the archived data. The Ranadu package is available at this GitHub address. To run the R code referenced here, that package should be included in the R installation. The KalmanFilter.Rnw routine requires that package and also some others referenced in the file, including “knitr”, “ggplot2”, “grid”, “ggthemes”, “zoo” and “signal”. In addition, Ranadu requires “ncdf4”, “tcltk”, “stats”, and “reshape2”.Some parts of “Ranadu” reference additional packages as needed, but they are not used in ARISTO-LAMS.Rnw so do not need to be available for this routine to run.

The netCDF files sometimes have mixed rates, e.g., 5_Hz for GPS-provided measurements and 25 Hz for some others including interpolation to 25 Hz from 13 Hz for IRS-provided measurements. In “Ranadu”, this is handled appropriately in the “getNetCDF()” routine, sometimes with the “ShiftInTime()” routine to adjust for delays among the variables and the “SmoothInterp()” routine to interpolate for missing values and apply smoothing of signals. Most of the plots are generated using “plotWAC()”, which is a convenience routine that sets various options preferred by the author when using the R “plot()” routine.

5 Discussion of Sections 1 and 2

Section 1 includes a table listing the flights used in this study. The data files are netCDF-format files located in the directories /scr/raf_data/ARISTO2015 and /scr/raf_data/ARISTO2016 as of Sept. 2016. These files may change if the project is reprocessed, so they are not guaranteed to remain as they were used. To ensure reproducibility, the R data.frames containing the data as read from these files and used in this program were saved separately in a set of files contained in SaveRData.zip, which is part of the archive for this document. This set of files should be restored to the working directory containing ARISTO-LAMS.Rnw before running that program.

Section 2 is a description of the data-processing required to run this program. That section describes the workflow for that step in the process, so it does not need further discussion here. It also includes links to further discussion of the processing algorithm incorporated into the Python program LAMS-ARISTO.py that is included as part of the archive named ARISTO-LAMS.zip for this project.

6 Discussion of Section 3

There is some further discussion of the “standard” approach to determining the sensitivity coefficients in past analysis-project archives such as discussed in this link, a workflow description for a previous program. The key change used here is to use the LAMS-measured angle of attack as the reference for the fits, as described in the text. The discussion of angle of attack is unfortunately separated into two sections because of the history of how the analysis was performed, and the second section (Sect. 4.3 in the report) should be considered the preferable determination of sensitivity coefficients for angle of attack because the data used span a larger range of flight conditions.

7 Section 4: Corrections to ambient and dynamic pressure

The original reference for this approach is Cooper et al. (2014). The standard discussion of processing algorithms is in this technical note. Because the parameterization relies on the measured angle of attack, the sensitivity coefficients for this measurement were revised based on the extensive range of flight conditions present in flight 6 of ARISTO-2016, and that discussion is presented in Sect. 4.3 of the report. Workflow for that section follows the same approach as in Sect. 3, so the references there can be consulted for additional detail.

The change in approach from previous work is in how the empirical terms in the fitted equation were selected. Candidate terms expected to contribute were the ratio of dynamic to static pressure, the Mach number (also dependent on dynamic and static pressure but not in a linear way), the angle of attack (potentially changing the static defect), the rate of climb, the pitch angle, and the sideslip angle. Several of these, including sideslip angle, rate of climb, and pitch angle (separate from angle of attack) were eliminated in preliminary checks. Then a combination of these factors was defined by this R equation in ARISTO-LAMS.Rnw line 1279:

$$\mathrm{lm(DQP \sim AKRD * M * QP * I(AKRD\hat{\ }2) * I(QP\hat{\ }2) * I(M\hat{\ }2), data = DW)}$$

Here DQP is the ratio of the error in dynamic pressure to the measured (uncorrected) ambient pressure, AKRD is the angle of attack, QP is the ratio of uncorrected dynamic to uncorrected ambient pressure, M is the Mach number calculated from the uncorrected pressures, and DW is a subset of the measurements qualified by meeting the defined chisquare test with some additional retrictions on airspeed and roll angle to eliminate turns and periods near takeoff and landing. When defined in this way, the R routine 'lm()' constructs terms involving all the factors multiplied together in the argument and also all possible products of those terms; i.e., AKRD, M, and QP but also AKRD*M, AKRD*M_QP^2, etc., for a total of 64 candidate terms. Next, those that did not meet a significance test of <0.001 in the t-test for that coefficient in the fit were eliminated, leaving 23 coefficients. From that point, analysis of variance (provided by the R routine 'anova()') was used to eliminate the least significant contributing terms repeatedly while monitoring the change in the standard deviation of the fit about the reference values DQP. This led to the terms listed in Eq. (10) in the primary document, at which point elimination of any additional terms caused the standard deviation to increase by 10% above that for the 23-coefficient fit. The only surprise in this approach was the significance of the final term,

$$b_4 \frac{q_m}{p_m} \left(\frac{\alpha}{a_r} \right)^2 ,$$

which was the least significant remaining term but still had a F value in the analysis of variance almost as large as the second-least-significant term M. It therefore appeared best not to eliminate that term, particularly because the increase in standard deviation started to weaken the representation of the pressure correction near the limits of the flight envelope without this term.

For the alternate pair of measurements on the C-130, PSFRD and QCFR, the same five-coefficient fit was used rather than going through the full elimination of terms as for the primary measurement pair (PSFD and QCF). The standard deviation about the reference values was almost as low for the alternate sensor pair as for the primary pair, so the selected parameterization appeared to provide a satisfactory calibration for both.

At the time this study was conducted, the primary processing scheme for the LAMS measurements (Scott Spuler's PCA method as modified by Matt Hayman) was not yet available for ARISTO-2016 flights, so all identification of peaks in the LAMS histograms was based on the SG-polynomial method. It may be useful to repeat this one day with the reprocessed data using the PCA method, because that likely will add additional measurements to the dataset used for the calibration.

8 Discussion of Sect. 5, LAMS-based measurement of temperature

This section led to incomplete results with mixed success, so this section won't receive extensive discussion here. The approach was to examine many cloud passes from ARISTO-2015 and ARISTO-2016, calculate the temperature based on LAMS measurements of airspeed and the LAMS-corrected measurement of Mach number, and compare the resulting temperature to the conventional measurement ATX. Plots were generated using Ranadu::plotWAC(), after some smoothing of the LAMS-based measurement to reduce noise. Figures 11, 13, and 16–20 showing some of the cloud passes were constructed in two-panel plots using this function.

9 Discussion of Sect. 6 (Transient effects)

There is an earlier study related to this section but for the GV, discussed in this report and this presentation. Those documents provide additional details on the approach taken here, but that study did not have a separate workflow document so some of the details are missing. An important aspect of this study is that the normal acceleration of the aircraft in the upward direction, BNORMA, is used, and that is often not present in standard files. The earlier study showed that this measurement can be deduced with sufficient accuracy by differentiating the ground-speed components of the aircraft (GGVNS, GGVEW, GGVSPD) to find the accelerations and then transforming those accelerations to the aircraft reference frame. That was not done in the present study because BNORMA was available, but it might be useful in future work.

An additional special step was needed for this section of the study. The standard routine for access to the netCDF files, Ranadu::getNetCDF(), constructs a data.frame where all variables have the highest rate present in the input file. Because in this case the input file still has the beam-histogram data (BEAM[1-4]_LAMS) present at 50 Hz, it would produce 50-Hz data without a special step, so just before this section of the report a redefined version of that routine is defined that forces the sample rate to 25 Hz.

10 Comments on reproducibility

The publicly accessible archive for this document is the github file listed at the end of the report, in the section on reproducibility. However, the special data.file records (SaveRData.zip) are omitted from that archive because that file is relatively large. The file is present on this EOL location: /h/eol/cooperw/RStudio/LAMS3D/SaveRData.zip. That file contains almost all the data needed to reproduce the report. The exceptions are a few ancillary plots, like the pressure difference plot for FRAPPE (Fig. 9), which can be regenerated from the processed files for that project.

References

Cooper, W. A., Spuler, S. M., Spowart, M., Lenschow, D. H., and Friesen, R. B.: Calibrating airborne measurements of airspeed, pressure and temperature using a Doppler laser air-motion sensor, Atmospheric Measurement Techniques, 7, 3215–3231, doi:10.5194/amt-7-3215-2014, URL http://www.atmos-meas-tech.net/7/3215/2014/, 2014.

Cooper, W. A., Friesen, R. B., Hayman, M., Jensen, J. B., Lenschow, D. H., Romashkin, P. A., Schanot, A. J., Spuler, S. M., Stith, J. L., and Wolff, C.: Characterization of uncertainty in measurements of wind from the NSF/NCAR Gulfstream V research aircraft, NCAR technical note NCAR/TN-528+STR, Earth Observing Laboratory, NCAR, Boulder, CO, USA, doi:10.5065/D60G3HJ8, URL http://dx.doi.org/10.5065/D60G3HJ8, 2016.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org, 2013.

RStudio: RStudio: Integrated development environment for R (Version 0.98.879), URL http://www.rstudio.org, 2009.

Xie, Y.: Dynamic Documents with R and knitr, Chapman and Hall/CRC, Boca Raton, Florida, URL http://yihui.name/knitr/, iSBN 978-1482203530, 2013.

Xie, Y.: knitr: A general-purpose package for dynamic report generation in R, URL http://yihui.name/knitr/, r package version 1.6, 2014.

– End of Memo –