EARTH OBSERVING LABORATORY

Memorandum:

31 October 2016

TO:	ORCAS processing file
FROM:	William Cooper
SUBJECT:	Workflow to find sensitivity coefficients

1 Purpose

This memo describes the workflow for finding new sensitivity coefficients characterizing measurements of angle of attack using the radome gust system. It applies specifically to the ORCAS project, where the R script producing the document and performing the calculations is "AKRD-forORCAS.Rnw." That file generates the memo "AKRDforORCAS.pdf" and the archive file "AKRDforORCAS.zip" that contains the generating file, the document in PDF format, and this workflow memo. The specific subset of measurements used here is also saved as AKRDforORCAS.Rdata, but that is not included in the zip file because it is somewhat large for inclusion in the github archive.

The final authority regarding workflow is the ".Rnw" document itself, but this overview and diagram is intended to help explain the workflow at a general level and so might supplement reading the R and LaTeX code. The intent is to describe the workflow in sufficient detail to support replication of the analysis and figures presented in "AKRDforORCAS.pdf" and also to enable changes based on new data or new analysis approaches.

There are references and citations for the tools used (R, RStudio, knitr) in this technical note: Characterization of Uncertainty in Measurements of Wind from the NSF/NCAR GV Research Aircraft. The workflow document for that technical note also contains some additional explanations of the procedures that are used again in generating the present memo and the steps taken to ensure reproducibility of the results.

2 Acquisition of the primary data

The best measurements for the purpose of this analysis are those obtained during "speed runs," constant-altitude flight maneuvers where the flight speed is varied through the flight envelope of the aircraft,, preferably with modest rates of acceleration and deceleration. Those maneuvers cause the angle of attack to chance with airspeed through the normal range of measurements. Two aspects of the maneuver are crucial, that the vertical wind be near-zero and that the *geometric* altitude be constant. Unfortunately, there were almost no speed runs for ORCAS (only one partial one during a test flight), so fits will need to be based on data collected in variable conditions. This is not as satisfactory because real updrafts and downdrafts may bias the results.

The measurements used in this case were collected using the NSF/NCAR GV research aircraft during the set of flights in the ORCAS project. A description of the data-acquisition process and on-board data system, as well as post-processing software, is contained in this technical note:

Characterization of Uncertainty in Measurements of Wind from the NSF/NCAR GV Research Aircraft (called the "Wind Uncertainty Technical Note" in the remainder of this document) and in the workflow memo for that technical note. The resulting data files have netCDF format and contain measurements in scientific units along with variable attributes that give brief descriptions of each measurement. These are the data archives used for the processing described by this workflow document. At the time of this analysis, a preliminary processing run had been made and values of the standard vertical-wind variable WIC as well as standard AKRD were present in the files. Attributes stored with AKRD indicate that three different relationships were used to determine the values used, but the attributes do not indicate where the altitude breaks were that determined which formula was used. These files were released, but in case of future change the file AKRDforORCAS.Rdata contains the variables at the time the AKRDforORCAS.pdf document and this workflow description were generated.

3 Overview of the workflow

A flow chart describing the workflow is provided by Fig. 1. The workflow is organized into five tasks, each of which is discussed below. Everything in this flow chart, both generating the text document and performing the calculations, is embedded in AKRDforORCAS.Rnw. Where particular tasks require calculations using R, the R code is isolated in "chunks" that have header titles, so those titles are referenced below where appropriate. The first chunk is "initialization," which loads some required R packages. An important one is "Ranadu", a set of routines used for accessing, manipulating, and plotting variables in the archived data files. This package is available at this github link and must be installed. Also required is "knitr" so that the text document can be constructed. A list of required variable names is also contained in this chunk; later, when data.frames are constructed for analysis, this list is used to determine the variables to load from the netCDF files.

In the case of ORCAS, the rate-of-climb variable is GGVSPD, provided by the GPS. This has been available since the SPRITES-II project in July 2013, and is preferable to use when it is present, but prior to that time it is necessary to use VSPD_G or VSPD_A. They have the disadvantage that there are filters and associated delays imposed on these variables before they are provided from the avionics package on the GV.

3.1 Task 1: Background

1. This task normally would start by considering if there is a problem with the standard processing. Figure 3 in the primary document illustrates that there is a problem with the standard coefficients and also with the processing that has been applied in the first-released files. Therefore this section has a structure a little different from most other projects. It begins with two text sections. The first is a very brief statement explaining why this problem is being considered.

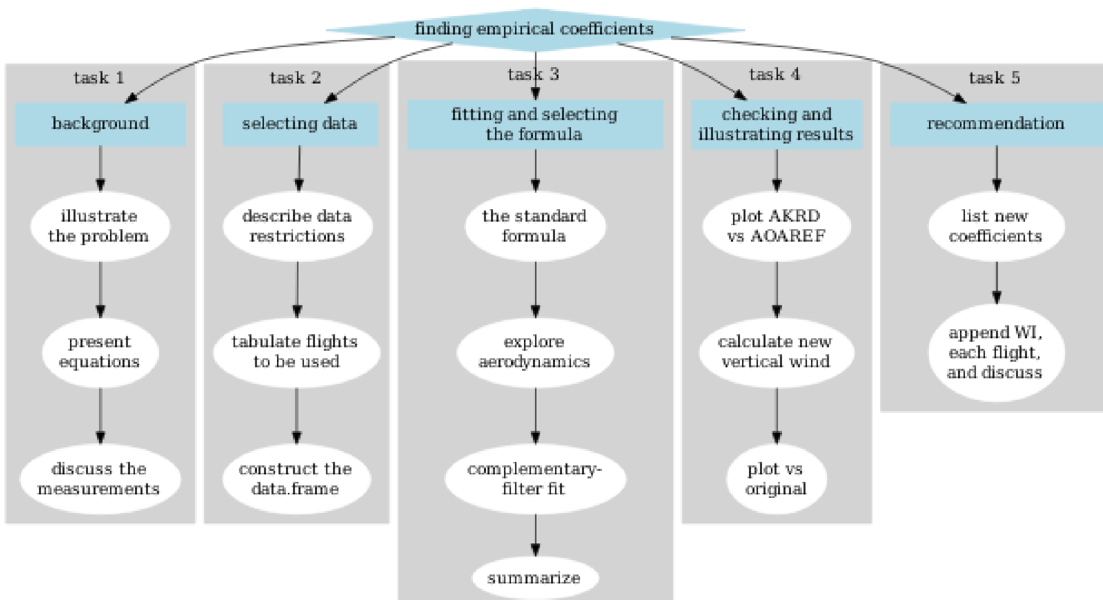

Figure 1: Flow chart describing the workflow.

2. The second section contains, in Eqs. (1) and (2), the basis for the conventional fit. The standard document on processing algorithms, available at this URL, documents the procedure in more detail and provides the "standard" coefficients previously found for the GV.

3. The symbols used here are the following:

   α the angle of attack, normally represented in units of degrees by the variable AKRD

   Δp_α The pressure difference between upward and downward ports on the radome, represented (in hPa) by the variable ADIFR

   q the dynamic pressure, used before correction and represented (in hPa) by QCF. The uncorrected variable is used because the correction to dynamic pressure involves angle of attack, so use of QCFC would lead to circularity in processing.

   θ the pitch angle, represented in degrees by PITCH

   w_p the rate of climb of the aircraft, represented in m/s by GGVSPD

   V the airspeed, represented in m/s by TASF

The code chunk associated with this task is that labeled "initialization." That is where the required library "Ranadu" is loaded, as well as the "knitr" library, where netCDF variables to be used in the analysis are declared, and where the location of the data files is defined. The latter location is only used for a first run or when the data-frame (constructed in the next task) needs to be changed. Once read, the required data.frame is saved, so on subsequent runs the compilation of data can be skipped and the data.frame re-read from the much shorter file where it is saved. The location used for the netCDF files is selected by the Ranadu::DataDirectory() function, which on EOL linux computers returns /scr/raf_data. In some versions of similar processing files, additional functions are defined in this chunk, but none are needed or used in the ORCAS case.

Sometimes there are erroneous values for some of these variables, Most often this arises from blockage of lines used for the measurement of ADIFR. Where this has occurred, the measurement will be set to the missing-value flag in the netCDF file, and when data are read into an R data.frame such missing values will be translated to the NA flag. When fitting to equations like (1), it is desirable to eliminate very slow values (because flaps may be in use, distorting the airflow and the radome sensitivity) and periods with significant bank (because the calibration procedure represented by Eq. 2 is only valid when the wings are level).

3.2 Task 2: Selecting data

1. *Data restrictions:* Some study of the measurements is needed to select data to which an equation like (1) will be fitted. In the case of ORCAS, some of the flights appear different from normal research flights because they are of short duration or appear to have been loaded especially heavily. These flights were apparently diversions for weather rather than normal research flight, and it seems appropriate to exclude them from the data to be used in fits.

The selection of flights was made by performing preliminary fits to equations like (1) or (6) and looking for consistency with other flights. This was a rather subjective judgment, but on that basis I selected flights 1, 2. 3. 6. 8, 9, 11, 14, and 18 as providing relatively consistent results and therefore to be candidates for a composite fit. Inclusion of other flights appeared to distort the fit coefficients unacceptably, probably because of extreme errors in the representation of angle of attack arising from gaps and other problems in the data. Some cases had gaps already flagged as missing in the data files, so those regions did not need to be excluded in any special way. However, the some additional restrictions were applied before performing the fits. These are discussed below in connection with the fit procedures.

2. *Accepted flight periods:* In some projects, a table was constructed to indicate the flights and periods of those flights that were used for the fits. However, here the full duration of all flights was included in the composite data.frame, except for ORCAS flight 12, for which the netCDF file on the EOL workspace had no valid data and so could not be used.

3. *Construct the data.frame.* The R chunk "construct-dataframe" was then used to read the selected flights and construct a composite data.frame containing measurements from all flights except flight 12. To construct the data.frame, the function "Ranadu::getNetCDF()" was used to read the desired variables for each individual flight, and the results were concatenated using "rbind()" to get a single data.frame. The variables included are specified by the array of names in 'VarList' in the 'initialization' chunk. Within that data.frame, each flight was tagged with a variable "RF" set to the (integer) flight number, so that subsets of the data.frame could be selected to examine individual flights. In addition, within the "construct-data-frame" chunk a number of new variables were added to the individual data.frames before concatenation into the composite data.frame:

   (a) A variable DENS, equal to the air density (kg/m^3); this was used in preliminary study of aerodynamic effects and as a candidate in exploration of empirical terms to include in the representation of angle of attack.

   (b) Some exclusion of individual measurements was needed to avoid problems:

      i. If the Time variable was missing, the data point was skipped. (Later inteerpolation will fill in short periods.)

      ii. If QR=ADIFR/QCF (below) was infinite it was set missing (NA). Such points are skipped in fit routines and plots.

   (c) A time-of-flight variable (TF, units seconds from the start of the flight), for some later studies of aerodynamic effects where it was necessary to consider the varying weight of the aircraft.

   (d) W, representing the weight of the aircraft. There are some flight-dependent branches used to estimate this, but these have not been verified against pilot records (and should be if they are used other than as rough exploratory variables, as in this study).

   (e) K, calculated as described later, and related to the lift of the wing; see the discussion later re aerodynamic effects.

(f) Some variables used later for fitting, including QR=ADIFR/QCF, M=Mach number based on uncorrected pressure and dynamic pressure without humidity correction, and AOAREF, as described by Eqn. (2) in the main document.

(g) A smoothed version of WICS, used for some plots in the main document. This is calculated flight-by-flight to avoid possible errors introduced by smoothing across flight boundaries.

(h) Low-pass and high-pass versions of AOAREF, QR, and M, for use in the study where the fit is performed for these components of angle of attack independently.

(i) The filter function used (signal:filtfilt()) combines forward and backward filtering to avoid phase shifts. The result is that there are some possible edge effects and the start and end of results, so results from the filtering were set missing for 1.5 times the cutoff period at the start and 2 times the cutoff period at the end of each flight. This had the effect of removing most initial climbs and final descents from the data used in fitting to the filtered values.

(j) The first time the data.frame (named Data, containing all the flights concatenated together via the R command 'rbind' with identifying variable RF containing the flight number) was constructed it was saved in a local file named "AKRDforORCAS.Rdata". When the logical variable ReloadData is FALSE, this saved file is used to restore ('load()' the saved data.frame to avoid the longer time required to re-read all the netCDF files. If any change in the contents of this variable are needed, 'ReloadData' should be set to TRUE so that the branch of the code constructing this data.frame is repeated. In some versions of the AKRDforProject files this data.frame is saved with the archive file, but in this case the size of the data.frame is about 60 MB so it is not included in AKRDforORCAS.zip.

3.3 Task 3: Fitting and selecting the formula to use

1. *Fit to the standard formula.*

   (a) This task begins by finding fit coefficients using the standard formula (1). The standard coefficients in the document on processing algorithms did not work well for ORCAS, so the first step was to see if new coefficients might work better. For this task, the data used in the fit were restricted to ORCAS flight 6 and measurements with TASF > 110 and abs(ROLL)<2. The R chunk "fits" performs this fit. The linear-model R function "lm()" is used to find the coefficients, and the resulting coefficients are incorporated into the text via "\Sexpr{}" expressions included in the LaTeX text so that, if there is any change, the coefficients quoted will be updated automatically. The fit standard deviation was reasonably good in comparison to other projects, but there were regions where the results appeared to be suspicious, esp. during a descent, low-level flight segment, and climb back to altitude in the middle of the flight.

(b) With the coefficients obtained from the fit, a new angle of attack was calculated (in R chunk 'S08summarize-fit') and plotted vs. the reference AOAREF, and a histogram was constructed of the difference between the resulting angle of attack and the reference angle. These appear in the main document as Figs. 1 and 2. The latter chunk also includes code to print a summary of the results from the fit into the document.

(c) An additional code chunk labeled 'S08new-vw' then calculated the vertical wind that resulted from the new angle of attack. In preliminary exploration this calculation was checked using the function Ranadu::WindProcessor(), but once checked a simpler formula was used that just corrected the vertical wind from the netCDF file by using the product of the airspeed and the change in angle of attack to corrected the original value WIC. The vertical wind that would be obtained using the standard coefficients was also calculated, and both the new values were smoothed using the function Ranadu::SmoothInterp() to obtain values smoothed with a 61-point Savitzgy-Golay filter (imposing smoothing with something like 20-s effective smoothing period) for convenience in plots. Using these results, a plot was constructed of the three values of vertical wind for ORCAS flight 6. The routine Ranadu::plotWAC() was used to generate this plot, which appears in the document as Fig. 3. That routine is just a convenience routine that sets some options favored by the author; this could as well be constructed using the standard R routine 'plot()', which plotWAC() calls after setting some options.

2. *Explore aerodynamics.*

(a) As discussed at the beginning of the subsection titled 'Consideration of aerodynamic effects", the results in Fig. 3 and other similar plots of ORCAS flights did not look satisfactory, so some exploration of alternatives seemed appropriate. This began with consideration of aerodynamic effects, because it appeared the problems appeared during climbs, descents, and low-level flight while results appeared satisfactory for level flight at upper levels. A prime candidate that could influence the sensitivity coefficients for angle of attack would be use of flaps, which could cause effects on the airflow around the airflow including near the radome. For this reason, a representation of the lift generated by the wings was sought that might indicate departure from normal performance. The theory underlying that study was included at the beginning of that subsection, and then a variable named K was defined to be proportional to the weight of the aircraft $W(t)$ divided by the dynamic pressure q.

(b) When divided by the angle of attack α, K/α was found to be approximately constant for the level-flight portions of most flights. However, as shown in Fig. 4, there are significant departures from the constant value during initial climb and final descent and during the descent-level-climb maneuver in the middle of flight 6. (Figure 4 is generated in R code chunk 'plot-K', where the plotted data are restricted to TASX>85 to avoid periods before takeoff or after landing.) There are high values of K/α near the start and end, likely caused by deployment of flaps, but it is not as clear why there are reduced values during climbs, descents, and low-level flight. The referenced code chunk also generates Figs. 5 and 6, respectively a histogram of K/α^* and a scatterplot of K vs α^*. These plots suggested that departures from the normal value of K/α^* might be used to identify periods of abnormal aerodynamic characteristics of the aircraft, and so might point to some way of representing those effects.

(c) R code chunk 'flight-mode' then generated Fig. 7 to show that using K/α^* to characterize normal vs abnormal flight periods provided good correspondence to periods when the results for vertical wind appeared questionable. 'layout()' and 'par()' commands are used to combine the two panels of this plot, with an offset used to cause only one abscissa label to appear. The data.frame D6 used in this code chunk was defined in the previous chunk. The syntax '(with (D6, ...)' is a useful way to avoid having to repeat D6 with each variable like WICS (normally D6$WICS) and to avoid the need to use quote marks around those variable names. This syntax is used often in the code of AKRDforORCAS.Rnw.

3. *Consider a new fit equation.*

   (a) Initial studies at this point explored incorporation of the term K in the fit. This appeared to work reasonably for individual flights, with appropriate tuning of the flight weights, but it was difficult to find a composite representation of all flights in this way. That might have been possible using the pilot records to find actual weights, but the dependence of K or $1/q$ suggested using 1/QCF as a term in the fit. This provided substantial improvement, and QCF instead provided still more improvement. A small additional improvement was found by adding log(GGALT) to the fit. This appeared to work much better when multiple flights were combined than did the use of the variable K, so a change to this fit approach was explored next.

   (b) Results from individual flights were examined, and it was found that there was reasonable consistency among these flights: 1, 2, 3, 6, 8, 9, 11, 14, 18. Other flights produced results somewhat inconsistent with each other and with the preceding set of flights, so a composite data.frame to use for fitting (named DF) was constructed by selecting only the listed flights and by restricting the measurements to those with TASX>110 and abs(ROLL)<2. In code chunk 'revised-fit', a new five-coefficient formula involving the added variables QCF and log(GGALT) was fitted to the data in that data.frame, again using the R routine 'lm()'.

   (c) The coefficients from that fit were used to calculate a new angle of attack (named AK) and a new vertical wind (named WIK), as before. The code chunk includes a function to print a summary of the fit into the LaTeX document and to construct a scatterplot of the fitted result vs. the reference value (Fig. 8), a histogram of the difference (Fig. 9), and an illustrative result showing the vertical wind for flight 6 (Fig. 10). These results were significantly improved over similar results for the standard formula, when applied to the same composite data set (not shown in the document). The LaTeX code at the end of this section references the fit result to print the resulting coefficients.

4. *Using a complementary filter.*

(a) The initial part of Sect. 4.3 in the main document explains the rationale for this approach and gives the equations used.

(b) The code that implements this complementary filter is included in the 'construct-dataframe' chunk discussed earlier, so the variables are ready for use in this section. The code chunk 'filterfit' does the fitting. It starts by computing again the same data.frame DF used in the preceding chunk. This is not necessary and is just a remnant of how the code was constructed; the recalculation of DF could be omitted. The first fit is to the high-frequency component and results in coefficients cff[]. In this case, only two coefficients were needed, representing an offset (turning out to be very close to zero) and a sensitivity to the high-frequency component of ADIFR/QCF. Additional terms in this fit, including the usual dependence on Mach number, did not lead to significant improvement in the residual standard deviation and so were not needed. Then a similar fit for the slowly varying component is generated, in this case using Eq. (8) from the first part of this subsection.

(c) Plots are generated (Figs. 10 and 11) to illustrate that reasonable matches are obtained for both components separately. The scatterplot in Fig. 10, for the high-frequency component, includes scatter caused by real vertical wind and is reasonable in comparison to expected variance in that wind. For the low-frequency component, Fig. 12 shows that a reasonable match is obtained although the fit is based on the full set of flights used for the fit and not only flight 6. Finally, Fig. 13 shows the resulting vertical wind, as the red line, and for reference other calculated vertical wind variables also.

(d) At the end of this subsection, the resulting coefficients for the fits are included by reference to the values obtained in the R code, vis \Sexpr{} statements in the LaTeX code.

3.4 Task 4: Checking and illustrating results

Much of the content often included here has been incorporated into earlier plots and discussion in the preceding sections. The key addition under this task is the generation of plots for each of the ORCAS research flights (except 12, for which no valid original file was present in the EOL directory containing ORCAS processing). Each plot shows four of the vertical-wind variables and includes mean values for the new variables in the title to the plots. These plots are appended to the end of the AKRDforORCAS.pdf report.

3.5 Task 5: Recommendation

The final task is to provide the recommended coefficients to use for processing ORCAS flights. In this case, the coefficients are not repeated but appropriate references to their values listed earlier in the document are given.

Appended information

The end of the memo includes a “Reproducibility” table that documents what is archived, where it is located, and how anyone could retrieve the program, text, and data and repeat the analysis. This would need to be changed if this file is applied to a new project, and care should be taken to avoid overwriting the github archive by failing to change the names of the archived files. The program itself generates the file “AKRDforORCAS.zip” but this and the other archived files must be uploaded to github manually.

Here is additional information on some of the files mentioned in the “Reproducibility” section, in addition to the main document (AKRDforORCAS.pdf). All the referenced files can be found on EOL data directories, many under ~cooperw/RStudio/Reprocessing, in addition to those included on the github archive. The files are:

- The AKRDforORCAS.zip file, containing the files listed in the attachment list. This is present in the github archive.
- The workflow document. This is this document, a more detailed description of the workflow leading to the final document. It does not document all steps (e.g., a discussion of preliminary results that led to a different focue for the work), but tries to illustrate the steps that would be needed to repeat the work.
- Original data: The netCDF files used were those on the listed directory as of Oct 2016. The contents of that directory may change if or when the project is reprocessed. For that reason, the subset of data actually used, consisting of the data.frame used, is saved as a 60 MB file that R can read to recover the data.frame, via a ’load(“AKRDforORCAS.Rdata”)’ statement. This Rdata file is not included on github because of its size, but it is present on EOL workspace as ~cooperw/RStudio/Reprocessing/AKRDforORCAS.Rdata.
- The github archive contains the file AKRDforORCAS.zip. It is available for download by anyone.

– End of Memo –