SCOR WG145: Responses to web questionnaires from potential users

An important element of the Group's work is to specify the design for a user-friendly chemical speciation computer program that will meet user needs. The process of assessing these needs started with a Town Hall meeting at the 2016 Ocean Sciences meeting, and was followed up by two web surveys in 2016 – 2017 using the Survey Monkey package.

The first survey was aimed at academic users, using the Working Group's initial mailing list, supplemented with the Town Hall participants. This survey received 109 responses.

The second survey was aimed at non-academic users, and was distributed through organisations working at the academic / non-academic interface. This survey received an additional 57 responses, of which 39 were from academics and 18 from non-academics. The response from non-academics was therefore relatively small, which should be borne in mind when looking at the percentage responses from this group.

Apart from the questions on the respondent's backgrounds, the surveys allowed (and in many cases encouraged) multiple answers so that many of the responses add up to much more than 100%

1. Respondents' Backgrounds

The respondent's subject areas are available from the first survey. As one would expect (Fig. 1a), most academic respondents classify themselves as chemists, and there are small numbers of biologists and geologists. Other subject areas given were biogeochemistry, marine sciences and physics (4 respondents each); earth sciences and engineering (2 respondents each); economics, ecosystem modelling, environmental sciences and mathematics (1 respondent each). The second survey asked for information on the respondents' employment sector; the distribution of the 18 non-academic respondents is shown in Fig. 1b.

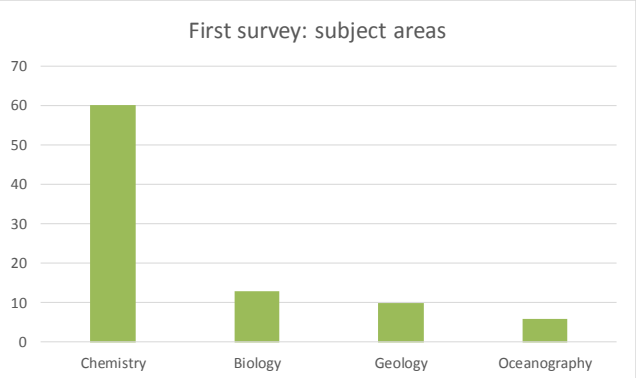

Fig. 1a. Subject areas (%) of the respondents to both surveys (166 responses).

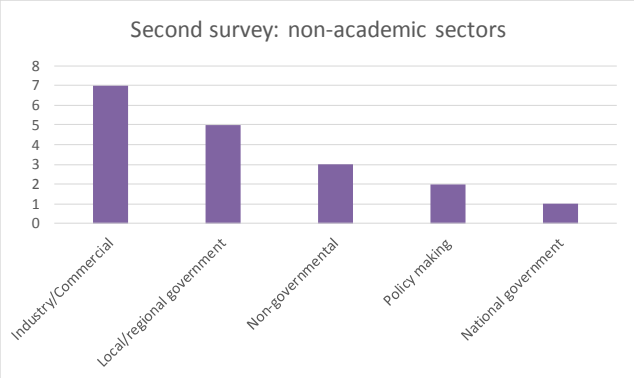

Fig. 1b. Employment of respondents from non-academic sectors (18 responses).

2. Chemical systems and species

Four areas of application dominated the responses: CO_2 and pH; trace metals; bioavailability; coastal and estuarine systems. Note, however, that the open ocean was not included as an option since it was assumed that many marine scientists have an open ocean focus. The other three categories (pore waters; hydrothermal systems; salt lakes and brines) did not attract a great deal of interest. The chemical species that were given priority followed a similar pattern, although redox processes and sulphides attracted more interest than might have been expected from the responses about target systems. Respondents also had the opportunity to name particular chemical species or groups of species of special interest. Trace metals dominated the replies, the most frequent being Fe (20 mentions), Cu (16), Zn (8), Mn (7), and Ni(4).

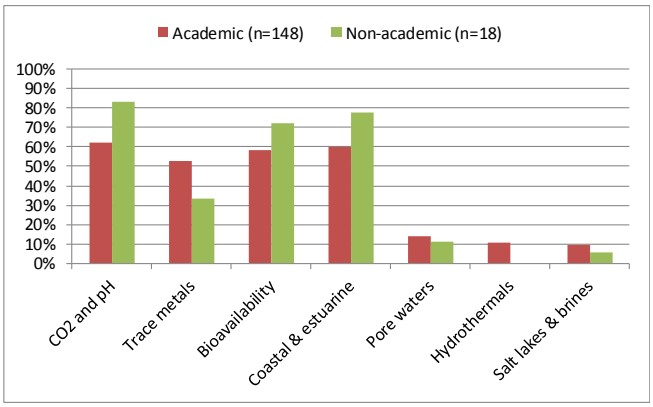

Fig. 2a. Chemical and natural water systems of interest to the respondents (%), for both academic and non-academic sectors (166 responses).

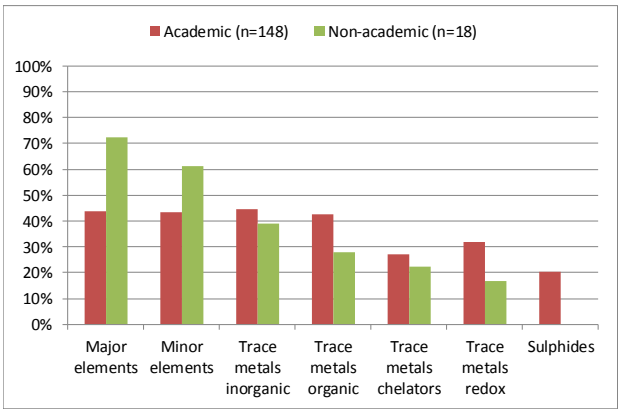

Fig. 2b. Chemical species of interest to the respondents (%), for both academic and non-academic sectors (166 responses).

3. The Chemical Speciation Program

(a) Program operation and access

Survey results are shown in Fig. 3 on the next page. The questions concerning access and mode of operation showed a clear preference for a standalone program package rather than on-line access. There was significant interest in a capability to call the speciation package from programs: the free text responses showed preferences for R and MATLAB (3 responses each) and Python (2 responses). The free text responses included a call for free access and for open source code (see Appendix 1). While many respondents are happy with a “black box” package, a substantial proportion would like to be able to edit the database allowing, for example, new ligands to be added. Importing large files was also of interest, and ODV-compatibility was mentioned in the free text responses.

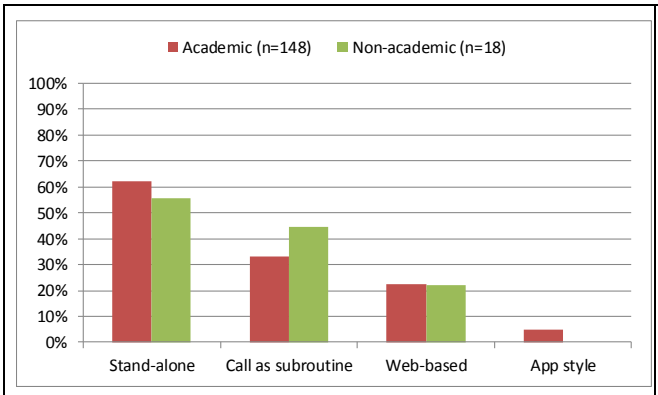

Fig. 3a. Preferred type of program (%), for both academic and non-academic sectors (166 responses).

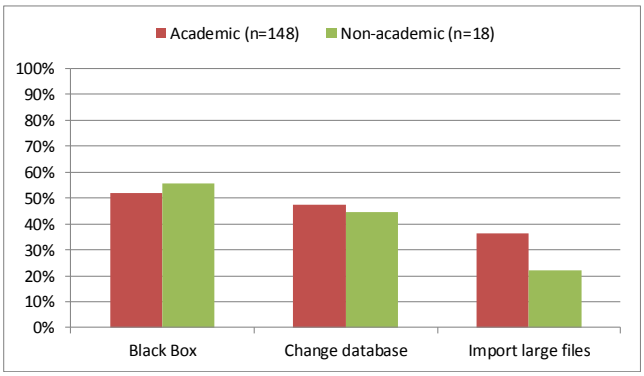

Fig. 3b. Preferred type usage of the program: simple ("black box"), the ability to change the chemical database, and the ability to process large datasets (166 responses).

(b) Output of Results

There is a clear preference for Excel-compatible formats (see Fig. 4, below), and an equally strong preference for information on the uncertainty associated with the calculated results, something that is lacking in current packages such as CO2SYS for the CO_2 system (although this was included in the original DOS programme). The free text replies took up again the need for compatibility with a variety of languages and programming environments including Fortran, MATLAB, C, and Python (see Appendix 1). Coupled to the interest in being able to process large datasets was a request to provide gridded data in NetCDF format.

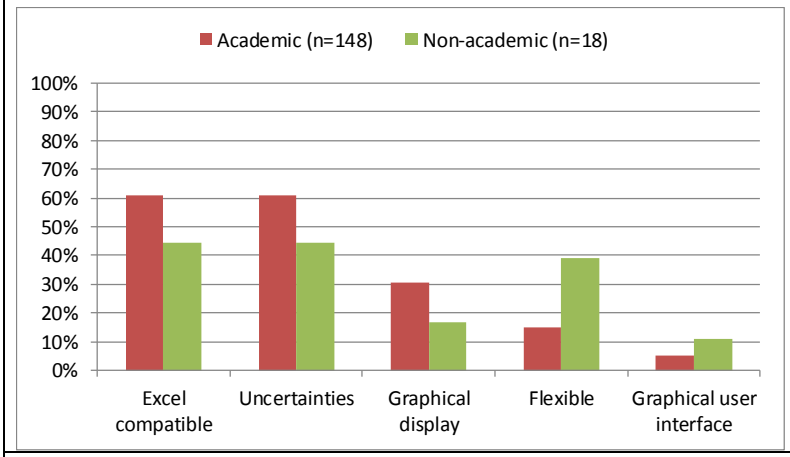

Fig. 4. Preferred types of output of the results of the chemical speciation program (166 responses).

(c) Help and Support

The final question concerned information and support: here there was a clear preference for written manuals and training materials rather than videos (Fig. 5, below). There was also a call for a discussion forum. Respondents were asked to select two options: an analysis of the combinations selected can be found in Appendix 2.

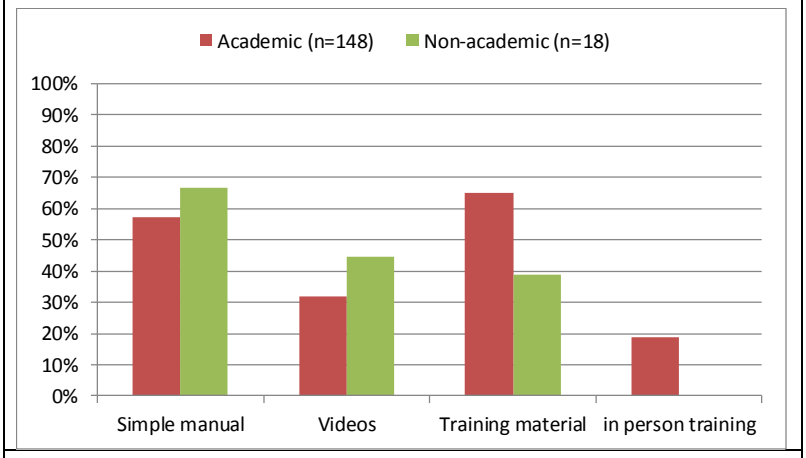

Fig. 5. Preferences for help and support for the chemical speciation program (166 responses).

4. Conclusion

The results of the surveys can be summarised as follows: Most of the respondents to the survey are academics (89%), of which the majority are chemists. The small number of non-academics are mostly from industry, and from local or regional governments. The main scientific interests of respondents, with regard to chemical speciation, are: CO_2 and natural water pH, and trace metals speciation and bioavailability. Their focus is on coastal and estuarine environments. Interest in trace metals includes interactions with organic components of seawater, chelation, and redox reactions.

The majority of respondents would like the chemical speciation program to be "freestanding", with a large fraction also wanting it to be callable as a subroutine from other applications. In addition to expensive applications such as MATLAB there is a clear need for interfaces to open source applications such as R. A small, but still significant fraction (20%) would like the application to be web based. About 50% of respondents would be satisfied with a program that functioned as "black box" (with few options or chemical properties changeable by the user), but a very similar percentage would like to be able to change the chemical database associated with the program (this contains equilibrium constants and activity coefficient parameters for the interaction of the ions). A significant number of respondents would like the capability to process large oceanographic datasets such as those now being generated by the GEOTRACES programme.

The output of the chemical speciation program should be Excel compatible, and include uncertainty estimates. A graphical display of results was also requested (but is less popular). Help and support in the form of a manual, more detailed training materials explaining the science, and videos are all desirable.

5. Appendix 1: free text replies

Free text replies to the questions concerning the chemical speciation program (section 3) are reproduced here.

(a) Program operation and access

- Third important option would be importing large data files
- Should be free software (licensed e.g. under GPL or BSD-2clause)
- Implement directly into ODV 7
- A short learning curve design, e.g. an Excel spreadsheet or GUI like CO2sys
- Please make source code open.
- If not freestanding, then written in language freely available (e.g., not Matlab)
- API via the web i.e. programmable online interface
- Incorporate into numerical models
- linked to R would be ideal to me
- Through MATLAB
- In case of "Stand-alone", should also be free software
- Open access code
- Python and R compatible
- Need for use in R, python, and Fortran
- Incorporated into your own computer programs
- Python and MATLAB
- Matlab based
- To allow for specific options
- Would incorporate routines into own programs
- In my own computer
- Linked into my own code(s)

(b) Output of Results

- Excel file compatible
- Flexible (e.g., Excel, text, Web-based)
- Graphical user interface (GUI) with fields to fill in
- Provides information on the uncertainties of calculated results
- Program displays results in graphs
- Input AND output in plain text format
- NetCDF should be used for gridded data. / I don't need graphs or uncertainties, but questionaire demands two ticked answers.
- Would be good if it showed which constants are used
- Python and MATLAB compatible inputs and outputs
- Matlab features will be most important; excel file and graphs are not important, but I had to check one or the other to move on in the survey!
- I probably wouldn't use the web feature.

- Possibility to work on many titrations on the same time if using the same ligand
- Callable from a fortran or C interface

(c) Help and Support

- Please no YouTube videos, they are a terrible way to learn
- Documentation needs to be correct and precise (like the OpenBSD man pages). / I checked "videos" only because I needed to check two answers.
- Manual and well-documented source code
- And a forum where it will be possible to ask questions if necessary

5. Appendix 2: Combinations of support options selected

The question posed was:

"What kind of user documentation and/or training options would you prefer (select the two most important options)?"

The four available options were:

Abbreviation	Full text
Simple Manual	Simple printable and/or online manual
Video	Tutorial videos (e.g., on YouTube)
Training materials	Training materials with tutorial examples, suitable for non-expert and teaching courses
In person support	In-person training/tutorials in conjunction with relevant international science meetings

The tables below show the number and percentage of academic and non-academic respondents selecting each combination.

Combinations with choice of simple manual

	academic	non-academic
Video	17 (10%)	5 (28%)
Training materials	60 (36%)	5 (28%)
In person support	6 (4%)	0

Combinations with choice of video

	academic	non-academic
Simple manual	17 (10%)	5 (28%)
Training materials	20 (12%)	2 (11%)
In person support	9 (5%)	0

Combinations with choice of training materials

	academic	non-academic
Simple manual	60 (36%)	5 (28%)
Video	20 (12%)	2 (11%)
In person support	13 (8%)	0

Combinations with choice of in person support

	academic	non-academic
Simple manual	6 (4%)	0
Video	9 (5%)	0
Training materials	13 (8%)	0