Spreadsheet Documentation for Students and Engineers*

KEN R. MORISON and PATRICK J. JORDAN
Department of Chemical and Process Engineering, University of Canterbury, PB 4800, Christchurch, New Zealand. E-mail: k.morison@cape.canterbury.ac.nz

Spreadsheets are used widely by engineering professionals and students but a disturbing proportion of spreadsheets has been found to contain errors. An appropriate level of documentation is required for engineering spreadsheets so that others can use, understand, modify, verify and validate them. Methods of documentation and how they can be taught are described. Students, while reluctant at first, learn to document spreadsheets quickly and when working as engineers realise the value of good documentation.

INTRODUCTION

SPREADSHEETS are in widespread use through-out the world for many business activities such as accounting, asset recording, production scheduling and engineering design. There are now a number of specialised texts on spreadsheets written for engineers [1–4] and specialised applications of spreadsheets are being developed and reported for a range of engineering calculations [5–7]. There is, however, considerable evidence that many spreadsheets contain mistakes [8] and that many are difficult for others to understand [9].

Because of the ease of producing and modifying spreadsheets, they often develop in an unstructured and untidy manner. Spreadsheets seem to have escaped the disciplines that have traditionally been applied to engineering design and to computer programming, although there has been some interest in the accuracy of business and financial spreadsheets [8]. The concepts and rigours of software engineering that were developed for programming are too often not applied to spreadsheets. The documentation of traditional computer programs is now so widely accepted that many modern texts do not make explicit mention of the need for documentation, but rather computer program documentation is now an implicit part of practically all textbook programming examples. In contrast, textbooks on spreadsheeting seem to be lacking in examples of good practice of documentation, apart from a spreadsheet title, and in some cases, an author and date. An exception to this is Julian [10] who includes a page on documentation and style.

In courses we teach we aim to set a standard for layout, documentation and verification of spreadsheets that is adequate for a practising professional engineer.

The aim of this paper is to fill the gap left by many texts about spreadsheeting and to explain the practice of documentation and how it is taught to engineering students during courses on engineering computation. There are different styles of documentation that are appropriate in different contexts and the programmer can exercise personal choice. Our objective is to show the variety of techniques that is available rather than to be prescriptive.

This paper is based on Microsoft Excel but applies, sometimes with different terminology, to any spreadsheet program. The use of Visual Basic for Applications offers further scope for documenting Microsoft Excel spreadsheets, but is beyond the scope of this paper.

SPREADSHEET DOCUMENTATION

The required amount of documentation is related to:

- the purpose of the spreadsheet;
- the users of the spreadsheet;
- whether it will be used on-line or must be understood from hard copy;
- the consequence of errors and the need for validation.

There are a number of different levels at which spreadsheets can be used:

1. Quick calculations as a 'scratch sheet' and probably not saved.
2. Short-term personal use, perhaps during the lifetime of a project.
3. Long-term personal use for a standard calculation or operation, where use may be frequent or intermittent over months or years.
4. Modification and development by others or with other related spreadsheets.
5. Use by others with a similar background and knowledge.

Table 1. Some basic elements of documentation.

Item	Cell in Figure 1
Title	A1
Statement of purpose	A2
Scope and limitations	E11
Author's name	A3
Date written	C3
User instructions	A4
Variables used	
full name	A6 : A16
name used in formulae	B6 : B16
units	D6 : D16
Formulae	
using cell references	E13
using cell names	E15, E16
mathematical form	F13
Text box	E6
Borders to group sections	A1 : F4
Bold or italic font	C16

6. Use as a 'black box' by others where limited understanding is required.

Documentation requirements increase from levels 1 to 6. There is a tendency for spreadsheets to progress over time from one level to a higher level, and thus more documentation will be required than initially thought. Most spreadsheet users have experienced the problem of referring back to an old spreadsheet only to find it unintelligible.

Furthermore, spreadsheets could be classified according to the importance of the conclusions obtained from them. Results may have major, minor or no financial or social consequences. An incorrect design may lead to loss of profit, or equipment failure. Clearly the more the spreadsheet is relied upon, the more the requirement for documentation.

Unfortunately, good documentation takes time but becomes easier with a knowledge of the techniques, with practice and from habit. Engineering students must be introduced to techniques that are effective and they should be given practice through examples and assignments.

METHODS OF DOCUMENTATION

The material that is taught to engineering students is presented in this section.

Elements of documentation

There are a number of obvious and simple elements of documentation that can be used in any spreadsheet. Table 1 and figure 1 show some basic elements of documentation on a simple spreadsheet.

The choice of which type of formula to use may depend on the purpose of the documentation. A mathematical equation (e.g. in cell F13 on Fig. 1) allows a technical person to see the engineering form but does not confirm that the equation has been entered into the spreadsheet correctly. The spreadsheet formula with cell references (E13) enables checking but requires that row and column headings are shown on the printed sheet. However, cell references in documentation can become incorrect if new rows or columns are inserted after the documentation was written, so spreadsheet authors can be reluctant to document sheets during development. The third option using cell names (cells E15, E16) is easier to read and will be correct after row and column insertions, but it requires that cell names are clearly identified on the sheet (see below). Another option, not shown, is to print the spreadsheet formulae (with Tools, Options, View in Excel). This can provide a complete record for archive or quality assurance purposes but the result is difficult to read because
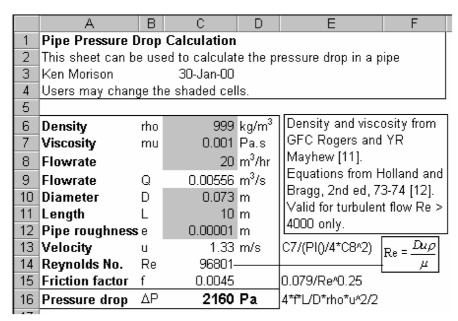

Fig. 1. A simply documented sheet.

	A	B	C	D	E
6			equation		
7	**Density**	rho		999	kg/m^3
8	**Viscosity**	mu		0.001	Pa.s
9	**Flowrate**			20	m^3/hr
10	Flowrate	Q		0.0055556	m^3/s
11	**Diameter**	D		0.073	m
12	**Length**	L		10	m
13	**Pipe roughness**	e		0.00001	m
14	Velocity	u	1	1.327	m/s
15	Reynolds No.	Re	2	96801	
16	Friction factor	f	3	0.0045	
17	**Pressure drop**	**ΔP**	**4**	**2160**	**Pa**

Spreadsheet Equations

1. $$u = \frac{Q}{\pi D^2 / 4}$$

2. $$Re = \frac{Du\rho}{\mu}$$

3. $$f = \frac{0.079}{Re^{0.25}}$$

4. $$\Delta P = \frac{4fL}{D} \frac{\rho u^2}{2}$$

Fig. 2. Spreadsheet and documentation using equation numbers in column C.

numerical values are not shown on the same sheet as the formulae.

Equations can also be referenced by an equation number in a separate column as shown in Fig. 2. The equations can then be written separately in another part of the sheet, on another sheet or on a completely separate document. This can be very useful if a sample calculation is also produced (see verification below).

The use of named cells and ranges can be useful but presents some problems. Cell names must be clearly identified as shown in column B on Fig. 1. Names make the writing and reading of cell formulae clearer and so can aid verification of the spreadsheet. Range names are also useful, e.g., the Vlookup function used in the formula:

'=VLOOKUP(B4,$D$4:$G$20,3)'

may be easier to understand expressed as:

'=VLOOKUP(B4,Properties, 3)'

where 'Properties' is the name for the range $D$4:$G$20.

The names used on an Excel workbook can be documented by using the Paste List facility provided within the Insert, Name, Paste menu. Comments explaining the use of each variable can be added in the next column [10]. Unfortunately, once pasted, the name list does not get updated when the locations of the cells change and this discourages this style of documentation during spreadsheet development.

The principle disadvantage for named cells is that they cannot be copied as relative addresses. Names are an absolute address. This becomes a problem when a user wishes to copy a calculation for another set of data. In Fig. 1, for example, if the cells C6:C16 were copied to column G named references would refer to the original location, whereas relative cell references would refer to column G. The resulting sheet could easily contain errors. Similar confusion can occur if named cells are referenced from other sheets, or if an entire sheet is copied. We therefore recommend that names be used only for values that will be constant everywhere in a sheet (e.g. gravity).

Text made be added to a sheet for many purposes:

- literature references to equations or correlations used;
- literature references to data sources, e.g. physical properties;
- instructions for data entry;
- instructions for use of the spreadsheet, e.g. for Solver;
- statement of decisions made, e.g. turbulent or laminar flow;
- file and path names;
- references to associated reports.

There is a variety of methods for adding text comments to a sheet. Text can be entered directly into a cell (Fig. 1, A1) but the text box (Fig. 1, E6) can be tidier and more flexible. If a significant amount of text is required another worksheet, perhaps titled 'Read me', provides space without cluttering up a worksheet. Notes can be added as 'cell comments' and these can be printed on the sheet or at the end. Cell comments, however, are less flexible than text boxes and we find that they are most useful as reminders to the spreadsheet author only.

In many cases hand-written text can provide adequate documentation but this is not kept with the electronic version of the spreadsheet and thus is only appropriate for one-off use.

Diagrams are used to clarify the symbols used. For example, the spreadsheet in Fig. 1 is easily enhanced by a diagram as shown in Fig. 3.

Where a series of data is presented a graph of the data may help a user understand the spreadsheet and may help verification of it.

History of changes

One feature of many computer applications is that a record is kept of changes to the programs as they are developed. This can also be desirable with spreadsheets, particularly those shared between users who may all make changes. The Track Changes command on Excel's Tools menu provides powerful features to identify and log changes made by any user and can conveniently provide a printout of full details to changes made to the spreadsheet. The drawback to using this is that the workbook must be specified as a shared workbook and in this condition certain changes
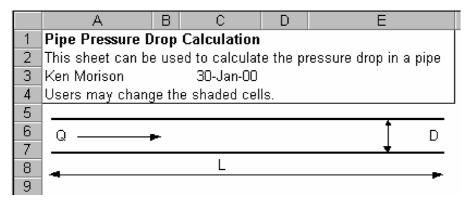
Fig. 3. A simple diagram can aid documentation.

cannot be made. Using the Track Changes feature is more useful when the structure of the spreadsheet is settled, rather than at the early stage of development.

Where a spreadsheet is modified periodically, it can be desirable to show the date it was last modified (manually entered in a cell) as well as, or instead of, the date the hard copy was printed (e.g. using the cell formula = today()).

Hard copy or electronic documentation

Hard copies of spreadsheets are frequently appended to reports. However when the work is revisited some time later, it is often impossible to reproduce the spreadsheets from the hard copy. In other cases all electronic versions may have been lost and a spreadsheet needs to be reproduced from the hard copy. This leads to the simple principle that documentation should be complete in the hard copy and should not rely on an electronic version.

However the electronic version of a spreadsheet allows the user to select a cell to see the cell formula, cell comments, cell name, and changes made to the cell contents if Track Changes has been activated. All of these types of information can be readily printed to form part of the hard copy documentation of the spreadsheet.

With the electronic form of a spreadsheet, the user can see how Solver was last used. This information can also be incorporated in hard copy by printing the screen with the Solver window open. However a text box of instructions as shown in Fig. 4 will usually suffice.

The choice between printed or electronic documentation will depend on the particular situation, though for teaching and marking the authors generally insist on the former.

Spreadsheet writing styles for minimising errors

There are a number of items of style in writing spreadsheets that can be used to minimise the chance of errors occurring:

* numerical values of constants and parameters should not be used in formulae, but they should be put in cells that are referenced by the formulae;
* the input, constants, calculations and outputs of a spreadsheet should be grouped with their own sections if possible;
* cells should not be hidden as their contents cannot then be checked;
* data is usually clearer when written in columns;
* only as many significant digits as required should be displayed.

VALIDATION AND VERIFICATION

The experience of the authors is that well over half the spreadsheets produced by students contain spreadsheeting errors. Panko [8] reviews numerous studies of spreadsheet error rates finding that typically at least 20% of spreadsheets contained serious errors. He stated concern that spreadsheet users tend to be overconfident in the accuracy of their own spreadsheets. In contrast, users of conventional programs may be more sceptical of

Calculation instructions

1. Add in Solver (Tools, Add-Ins, Solver) if it is not present on the Tools menu.
2. Select Tools, Solver to Set Target Cell C10 to a minimum by Changing Cells C11:C14, Subject to the Constraints C12 >= 1E-6.
3. Select Solver, Options to ensure that the Convergence is set to 0.00001.
4. Click on Solve
5. Solve again using the solution found as the starting values.

Fig. 4. Textbox with instructions for a manual operation.

the accuracy of programs written by others so may test them more rigorously. There is a clear need for systematic methods of verification and validation.

Edwards and Finlay [13] distinguish between validation and verification. The verification process ensures that the spreadsheet is a correct implementation, but the validation process ensures that it is a valid representation of the original problem.

Visual checking of cells is seldom effective for the elimination of errors. Verification will often require that another person check all the steps from engineering fundamentals to the solution. This requires complete documentation of the original methods used.

We aim to teach methods that will reduce the likelihood of errors. There are numerous reasons for incorrect calculations which include:

- wrong equations chosen—engineering mistake;
- incorrect manipulation of equations;
- correct equation wrongly coded;
- incorrect specification of constants, properties or parameters;
- formula copied from wrong cell;
- incorrect use of relative addresses and absolute addresses leading to errors on copying;
- incomplete iteration when using iterations or Solver;
- automatic calculation turned off;
- wrong specification of target or adjustable cells for Solver.

Verification cannot be done only by reading the spreadsheet, but there are a number of techniques which reduce errors and aid verification:

1. Systematically verify each cell formula perhaps using built-in auditing tools.
2. Solve the same problem independently perhaps by another person.
3. Solve by an alternative method or use an alternative equation.
4. Use redundant equations as a check (e.g. overall mass balance as a check on the sum of the component mass balances).
5. Avoid complex cell formulae by using simpler formulae that calculate intermediate values;
6. Show intermediate results on the sheet and make them easy to locate.
7. Check for engineering sense, e.g. are fluid velocities in a sensible range?
8. Graph results.
9. If cells are copied into a table, check a random row or column rather than the first cell.
10. Write out a full sample calculation using a piece of paper and a calculator without referring to the spreadsheet.

The authors have seen all of these methods fail; none seems to be foolproof. The most reliable, and most educationally useful, is to require a complete worked example showing all the calculations done by calculator for a typical set of data.

INSTRUCTIONS FOR THE USE OF THE SPREADSHEET

There is a need to provide sufficient documentation to suit the user. The examples given in Fig. 1 may suit a novice user. Expert users should not rely on their own memory but could consider that they themselves might need to be reminded about the use of their own spreadsheets and thus they should write documentation accordingly.

The on-line user will need to know how to enter the data, ensure that any necessary manual operations are performed and identify or display the required results. The clarity of a sheet may be enhanced if gridlines are replaced by borders that are specifically chosen to aid readability.

Data entry
Data entry cells can be grouped together and given a background colour, or shading, as in Fig. 1. It is sometimes desirable to include derived values between the input cells if that makes a logical progression. The value of the flow rate converted to $m^3 s^{-1}$ in Fig. 1 is an example. Once development of the spreadsheet is complete, it is worthwhile to lock the worksheet and unlock only those cells used for data entry. It is an advantage when viewing hard copy of a spreadsheet to be able to see which values were entered by the user for a specific case and which cells contain fixed or calculated values.

Many other techniques, such as input forms and on-line help produced by Visual Basic macros, can make spreadsheets even easier to use, but these are not covered here.

Manual operations
Instructions for manual operations can be detailed in a text box or on a separate 'Read me' worksheet. An example is provided in Fig. 4 for the use of Solver.

Other areas where instructions for manual operations may be required include specifying a procedure for restarting an iterative calculation following an error condition, and in the creation of graphs of the results.

Spreadsheet results
The important results are likely to be calculated in diverse parts of the spreadsheet. These can be copied to cells grouped and formatted with border and colours so that the key results can be identified and read easily. In the same manner as written reports, the reader should be informed (perhaps by text box) of the purpose and location of each figure or chart.

Undisciplined students, and engineers, have a tendency to believe that the use of a spreadsheet removes the need for appropriate presentation of results. In any work, results and workings must be clearly shown and summarised, and graphs must have titles, axis labels and legends. The same disciplines also apply to spreadsheets.

TEACHING METHODS

Documentation is shown by example in all lectures and one lecture out of the five available for Excel is specifically about documentation.

Initially students find documentation time consuming and will make such comments as 'it took two hours to solve the problem but four hours to tidy the spreadsheet'. As they progress, students find it increasingly easy to add documentation and learn to set out spreadsheets from the start so that less tidying is required. Some instruction makes the task much easier. In particular we encourage the effective presentation of equations by providing students with a one-page sheet of Microsoft equation editor short cut keys that includes a simple exercise. This is well received and well used.

Assessment

Normally assignments are marked on hard copy only and students are informed that spreadsheets will not be accepted in electronic form. This has the advantage of placing greater importance on documentation.

Students are told that enough documentation must be presented for the marker to understand the spreadsheet and to be able to reproduce it from the hard copy. A substantial portion of the marks (e.g. 30%) is awarded for the documentation and explanation with the spreadsheet. This is entirely consistent with the practice with hand-written assignments in which the working must be clearly shown. The student must document the spreadsheet sufficiently well so that the marker can check through the working to confirm the answer. Students are sometimes told that the easier it is to find any mistakes, the more marks they will be given.

Once documentation skills are well established students are asked in one of our courses to submit their spreadsheets electronically. The spreadsheets are then assessed for accuracy and for ease of use. Most of the skills learnt for hard copy documentation can be applied, but other techniques such as colouring cells can improve clarity.

STUDENT RESPONSE

A number of our graduates have commented after working for several years that they have been setting a higher standard in the companies in which they are employed because of their expertise both in spreadsheeting and in the documentation of sheets. They have been very clear in their acknowledgement of the value of high standards of documentation. The following comment received from a graduate sums up the usefulness of good documentation.

> I am definitely glad that you pushed the spreadsheet documentation as hard as you did. I have found it most valuable to be able to return to any of my spreadsheets and be able to recall exactly how I have calculated things. I regularly use spreadsheets and still document them in detail. I find it reduces the number of errors made as you check your calculations carefully as you note your method.

> I only wish that all engineers had the same habits. Numerous times in my company, when people have left their jobs and moved on, I have had to carry on work using other people's spreadsheets. The time involved in working out somebody else's spreadsheet can be even greater than redoing the calculations yourself.

> Spreadsheet documentation has been especially valuable to me as I work in a team where spreadsheets are often shared. Much time is saved in explanation if the spreadsheet documentation is done well.

CONCLUSIONS

Spreadsheets can and usually should be documented as rigorously as other forms of calculations or reports. Students can be taught a range of useful documentation skills which will make their assignment work more intelligible and will be invaluable in their later professional careers.

REFERENCES

1. W. J. Orvis, *Excel for Scientists and Engineers*, 2nd ed., Sybex, San Francisco (1996).
2. B. V. Liengme, *A Guide to Microsoft Excel for Scientists and Engineers*, Edward Arnold, London (1997).
3. B. S. Gottfried, *Spreadsheet Tools for Engineers: Excel 97*, McGraw-Hill (1998).
4. G. Filby (ed), *Spreadsheets in Science and Engineering*, Springer-Verlag, Berlin (1998) pp. 171–202.
5. C. Y. Lam, A non-traditional numerical solution to heat conduction in a rectangular prism, *Int. J. Eng. Educ.*, **13**, 1 (1997) pp. 52–53 .
6. P. K. Yin, Numerical methods on spreadsheet for machinery design projects, *Int. J. Eng. Educ.,* 13, 6 (1997) pp. 412–416.
7. E. G. John, Simplified curve fitting using spreadsheet add-ins, *Int. J. Eng. Educ.*, **14**, 5 (1998) pp. 375–380.
8. R. R. Panko, What we know about spreadsheet errors, *J. End User Computing*, **10**, 2 (1998) pp. 15–21.
9. P. B. Cragg and M. King, Spreadsheet modelling abuse: an opportunity for OR? *J. Opl. Res. Soc.*, **44**, 8 (1993) pp. 743–752.

10. F. M. Julian, Using spreadsheets in chemical engineering problems, in Filby G. (ed) *Spreadsheets in Science and Engineering*, Springer-Verlag, Berlin (1998) pp. 171–202.
11. G. F. C. Rogers and Y. R. Mayhew, *Thermodynamic and Transport Properties of Fluids: SI Units*, 5th ed., Basil Blackwell, Oxford (1995).
12. F. A. Holland and R. Bragg, *Fluid Flow for Chemical Engineers*, 2nd ed., Edward Arnold, London (1995).
13. J. S. Edwards and P. N. Finlay, *Decision Making with Computers, the Spreadsheet and Beyond*, Pitman Publishing, London (1997).

Ken Morison is a senior lecturer in the Department of Chemical and Process Engineering. He joined the department six years ago after a number of years in the dairy industry where he used spreadsheets extensively for engineering calculations and routine monitoring. He teaches spreadsheeting, programming, modelling, simulation techniques and fluid mechanics to engineering students.

Patrick Jordan is also a senior lecturer in the Department of Chemical and Process Engineering. He has been in the department for more than 25 years and uses spreadsheets extensively in research and teaching, particularly in the teaching of heat transfer. He also teaches spreadsheeting, optimisation and process modelling and dynamics.