Science & Technology Facilities Council

Technical Report

Guidelines for the development of HSL software, 2008 Version

J. K. Reid and J. A. Scott

August 2008

RAL-TR-2008-027

Library and Information Services
SFTC Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Didcot
OX11 0QX
UK
Tel: +44 (0)1235 445384
Fax: +44(0)1235 446403
Email: library@rl.ac.uk

ISSN 1358-6254

RAL-TR-2008-027

Guidelines for the development of HSL software, 2008 Version[1]

by

J. K. Reid and J. A. Scott[2]

Abstract

HSL is a collection of portable, fully documented, and tested Fortran packages for large-scale scientific computation. It has been developed by the Numerical Analysis Group at the Rutherford Appleton Laboratory, with additional input from other experts and collaborators.

The aim of this report is to provide clear and comprehensive guidelines for those involved in the design, development and maintenance of software for HSL. It explains the organisation of HSL, including the use of version numbers and naming conventions, the aims and format of the user documentation, the programming language standards and style, and the verification and testing procedures.

This version supersedes RAL-TR-2006-031.

Keywords: HSL, software development, software testing.

---

[1] Current reports available from "http://www.numerical.rl.ac.uk/reports/reports.html".

[2] The work of the second author was supported by the EPSRC grant GR/S42170.

Computational Science and Engineering Department,
Atlas Centre, Rutherford Appleton Laboratory,
Oxon OX11 0QX, England.

August 11, 2008.

Contents

1 Introduction

HSL is a collection of portable, fully documented, and tested Fortran codes for large-scale scientific computation. HSL began as the Harwell Subroutine Library back in 1963, making it one of the oldest such libraries. Over the past four decades, HSL has continually evolved and been updated as new algorithms and codes have been developed and older codes have been superseded. It moved to Fortran 77 in 1990 and now consists of a mixture of Fortran 77 and Fortran 95 codes that are all threadsafe. The majority of HSL codes are written and developed by the Numerical Analysis Group at the Rutherford Appleton Laboratory (`http://www.cse.scitech.ac.uk/nag/`), with additional input from other experts and collaborators.

Many of the older HSL codes have gradually been superseded by newer versions, with increases in functionality, improved interfaces, or speed of execution. As a result, HSL is now arranged in two parts, the main library (currently HSL 2007) and an archive, HSL Archive. The HSL Archive comprises older codes that were part of previous releases of HSL, many of which have been superseded by more modern codes. We do not maintain the codes in HSL Archive and the rest of this report refers to the main library except where HSL Archive is explicitly mentioned.

HSL 2007 and the HSL Archive are arranged as collections of Fortran 77 and Fortran 90/95 **packages**, each of which consists of either a single program unit or a set of program units. Almost all the Fortran 77 program units are subroutines, but there are also some functions and some HSL Archive packages contain block data subprograms. The Fortran 90/95 program units are mostly modules, but there are some external procedures. Each package performs a basic numerical task and has been designed to be incorporated into programs. Each has its own specification document, which gives full details about how to use the package (see Section 3.1), and is available as a PDF file.

As well as being widely used by the academic community for both teaching and research purposes, HSL packages are regularly incorporated into advanced software applications. More than 2000 organisations world-wide use HSL, with packages incorporated, under licence, into more than 90 commercially available software products. Such users must contact the software distributors, Aspen Technology Inc, for a licence (see Section 7).

HSL offers users a high standard of reliability and, over the years, it has gained an international reputation as a source of robust and efficient numerical software. Furthermore, HSL software is fully supported and maintained.

The aim of this report is to provide clear and comprehensive guidelines for those involved in the design, development and maintenance of software for HSL. Section 2 explains the organisation of HSL, including the use of version numbers and naming conventions. In Section 3, we discuss the user documentation, which is an important part of the Library. The programming language and style that is used by HSL is discussed in Section 4. Then, in Section 5, we explain the verification and testing procedures that we employ before a new package is accepted. The role of the Librarian is discussed in Section 6 and commercial issues are the subject of Section 7.

2 Organisation of HSL

2.1 Versions

Prior to 1990, HSL evolved continuously. New packages were added whenever they were ready. Since then, additions have been collected into releases, with intervals of around two to three years between releases, although packages are made available as soon as they ready. The original motivation for having formal releases was for bound sets of documentation (both the catalogue and the specification documents for the individual packages) to be printed. Nowadays, the catalogue is made available online (`http://www.cse.scitech.ac.uk/nag/hsl/contents.shtml`), but it remains convenient to construct a new catalogue only occasionally and it is convenient for publicity, too, to have distinct releases.

Academic licence holders (see Section 7) can download PDF versions of individual specification documents while, for those purchasing an HSL licence, the documentation is distributed with the Fortran source code. At the time of writing, the current release is HSL 2007.

With HSL 2004, we started to mark each version of a package with a version label of the form $l.m.n$ where

l is a digit that labels major changes that affect the interface,

m is a digit that labels a technical change, usually a bug fix, that does not affect the interface, and

n is a digit that labels an editorial change such as the correction of a spelling error in an output format or a non-trivial alteration to the documentation (typos in the documentation do not lead to a new label) .

The first version of a package is labelled 1.0.0. We did not attempt to mark the changes prior to HSL 2004 with version labels.

The documentation, which is described in Section 3, includes the version label. In addition, the leading comments in the source code for each package provide a copyright statement and, for each version, a brief description of the changes that led to it.

2.2 Naming conventions

Each Fortran 77 package has a sequence of two letters and two digits as its name. The name of each program unit within the package starts with these four characters. The fifth character, which is always alphabetic, is used to distinguish between the program units. The sixth character of the name (or its absence) identifies the type of its principal argument. The possibilities are

Absent	Single precision.
`D`	Double precision.
`I`	Integer.
`L`	Long integer.
`C`	Single precision complex.
`Z`	Double precision complex.

If the principal argument is always complex, the sixth letter is absent or `D`. If there are more than 26 procedures in a package, another four-character name is used for these but they are still

regarded as a part of the package; for example, `EA16` has procedure names that start with `EA16`, `EA17`, and `EA18`.

The name of each Fortran 90/95 package starts `HSL_` and is followed by a sequence of two letters and two digits. In most cases, this is its whole name, but occasionally this is followed by a qualifier such as `_ELEMENT`. The letters in the name are written in upper case. The package usually consists of one or more modules, but may also contain some external procedures. Each module name starts with the package name and is followed by a qualifier to identify the type of its principal argument. The possibilities are:

`single`	Single precision.
`double`	Double precision.
`integer`	Integer.
`long_integer`	Long integer.
`complex`	Single precision complex.
`double_complex`	Double precision complex.

If the principal argument is always complex, the qualifier is `single` or `double`. Each public name is the sequence of four characters (excluding the leading `HSL_`), followed by a qualifier (if any), followed by further characters.

With the appearance of hardware for which single precision working is significantly faster than double precision, we anticipate that the main use of the single precision versions will be in mixed precison computations where the single precision result is computed and then refined to double precision accuracy.

2.3 File naming conventions

The naming convention that is used for the files of a double-precision package that uses the free source form are summarized in Table 2.1. For files in the fixed source form, the suffix `.f` is use instead of `.f90`. All the Fortran 77 packages[1] use the old source form, of course. All of the Fortran 90/95 packages except `HSL_MP01` use the new source form. By 'source', we mean the source code of the package itself, excluding any modules accessed from elsewhere (e.g. other HSL packages) or procedures invoked from elsewhere. By 'test', we mean the comprehensive test (see Section 5.2). By 'spec', we mean the first or only simple example in the specification document (see Section 3.1.5). By 'other spec', we mean another simple example in the specification document.

Single-precision versions are as for double with the `d` replaced by `s`. Integer versions are as for double with the `d` replaced by `i`. Where a package contains both real and complex versions, the complex version is as for double with the `d` replaced by `c` and the double complex version is as for double with the `d` replaced by `z`.

3 Documentation

Each HSL package is accompanied by a specification document that the user will need to read to find out how to use the package. For many of the more complex packages, which cannot be fully described in this way, a separate technical report is written. In this section, we discuss the

[1]The default versions of `FD15` and `ZA12` call Fortran 95 intrinsic functions.

Table 2.1: The file naming convention for double-precision Fortran 95 packages

directory:	*package*
source:	*package*`d.f90`
test:	*package*`dt.f90`
test data:	*package*`dt.data`
test output:	*package*`dt.output`
spec:	*package*`ds.f90`
spec data:	*package*`ds.data`
spec output:	*package*`ds.output`
other spec:	*package*`ds1.f90`
other spec data:	*package*`ds1.data`
other spec output:	*package*`ds1.output`
other spec:	*package*`ds2.f90`
other spec data:	*package*`ds2.data`
other spec output:	*package*`ds2.output`

specification document in detail and then briefly comment on the aims and objectives of writing a report to provide users and other researchers with further information.

3.1 HSL specification documents

For most potential users, their first in-depth contact with the software is likely to be through the specification document. It is therefore essential that it is clear, concise, and well written. Writing good user documentation is an integral part of the design process and developers of HSL software are strongly encouraged to draft the user documentation in the early stages of code development. This allows others in the Group to comment on the design and make suggestions on the options offered and on the user interface.

The specification documents are written in Latex or TSSD. New specification documents will usually be written in Latex. A template Latex file together the necessary style files are provided to assist developers. As already noted, the typeset versions of the specification documents are made available to users in PDF format.

The complexity of a specification document will, in part, depend upon the complexity of the package and the range of options that it offers. But, for each package, the specification follows a similar format. In the rest of this subsection, the various sections of the specification document are explained.

3.1.1 SUMMARY

The summary begins with a single sentence that states the purpose of the package, with the most relevant part in bold. Often the reader should be able to decide whether or not the package is suitable by reading this one sentence.

It continues with a fuller description that introduces any mathematical ideas and notation that are needed in the rest of the document. This is usually less than about half a page in length, but sometimes it has to be longer than this. The summary ends with a list of attributes that include:

Version: The version label (see Section 2.1).

Types: The types offered (see Section 2.2).

Remark: (Optional) Any useful information not covered otherwise. Examples are that it uses MPI, that at least 8-byte arithmetic is recommended, that the package supersedes or is superseded by some other package, or that it is a front end for some other package. If the development of the package was supported by a grant (or other funding), this should be highlighted here.

Language: Fortran 77, Fortran 95, or Fortran 95 plus allocatable dummy arguments and allocatable components of derived types.

Parallelism: Either 'MPI' or 'OpenMP available'. This must be noted for packages that use parallelism directly or through the procedures that they call.

Calls: The names of the packages containing procedures that are called or the generic names of other procedures that are called or used. This will include other HSL packages, BLAS routines and any LAPACK routines that are called (see Section 4.9).

Original date: The date that the package entered HSL.

Origin: The names of the authors and their affiliations at the time of writing the package.

3.1.2 HOW TO USE THE PACKAGE

The main body of the documentation explains how to use the package. This should be straightforward to understand, avoiding as far as possible technical terms and details of the underlying algorithms because the reader may not be an expert in the area. The input required from the user and the output that will be generated should be clearly described, with full details of any changes that the code may make to the user's data. This section is divided into a number of subsections.

If the package contains modules, the first subsection starts with an explanation of the `use` statements that are needed. Two modules of different types may contain the same public name if it is expected that they will rarely be used together. In this case, the documentation explains what renaming will be needed should they be used together.

If the package uses OpenMP, there should be a warning that the user who wishes to take advantage of this needs to tell the compiler about OpenMP at compile time.

The procedures that are available to the user are then listed, with a very brief description of each (this should just be one or two sentences). If reverse communication is used, this is explained here. The aim is to provide readers with a quick overview of the procedures and introduce them to the calls that they will need to make to use the package. It should be made clear whether a procedure must be called or is optional.

If the package contains public derived types, these are explained in the next subsection. A full description of each component that the user is expected to access is either given here or in a later subsection. Since Fortran 95 does not allow individual components to be declared `public` or `private`, sometimes there are additional components that are not documented.

There follows a full description of all the argument lists of procedures that the user may call and their calling sequences. Usually, there is a separate subsection for each procedure. In each of these, the type, shape and intent of each argument is documented ahead of a description of its purpose. If a full description would be long, e.g. for all the possible values of an error flag, it is better to give a short description here and refer the reader to a full description in a later subsection. Any restrictions on the argument (e.g. $n > 0$) are listed in bold at the end of the description of the argument. In each call, `OPTIONAL` arguments are listed as the last arguments and square brackets `[ ]` are used to indicate these in the description. For example, in the call

```
call MA77_open(n,filename,keep,control,info[,nelt,path])
```

the arguments `nelt` and `path` are optional. Since we reserve the right to add additional optional arguments in future releases of the code, we advise users that all optional arguments be called by keyword, not by position.

After the subsections that describe the argument lists there may be a number of additional sections describing, for example, a derived type that has many components, information that is returned to the user, possible error and warning diagnostics, and other features of the package that have not been described elsewhere.

3.1.3 GENERAL INFORMATION

The general information section has the following headings:

Input/output: This includes diagnostic printing plus any I/O to direct-access and/or sequential access files.

Restrictions: A list of the restrictions mentioned for one or more of the arguments and/or control parameters.

Changes from Version $l.m.n$**:** ($l.m.n$ is replaced by the version label) If a major change has been made (that is, l has changed), this should be explained here.

3.1.4 METHOD

The method section provides a brief self-contained description of the method, which should give the reader some understanding of it without delving into fine details. It is normally less than a page long.

For details, the reader should be given a short list of references to papers and to the accompanying report, where available (see Section 3.2).

3.1.5 EXAMPLE OF USE

The calling sequence for the package is illustrated through the inclusion of a simple example, which is supplied complete with the input data and expected output. In our experience, if it is well written and fully commented, this kind of example provides a template that is invaluable, particularly for first-time users of a package.

Some packages have a second example to illustrate some other aspect of the way the package is used. Very occasionally, for a complex package with a number of different possible calling sequences, there is a third example.

The examples that are included in the specification documents generally use the default settings (see Section 4.3) and do not attempt to illustrate the use of all the facilities of the package.

3.2 Accompanying reports

For many users, a report that provides full details of the capabilities of the package and the options it offers is useful. This will describe in greater depth the algorithms used, highlight important or novel implementation details, present theoretical results, and show numerical results for practical applications. It will also provide references to other relevant research reports and papers. The report should enable advanced users to select suitable control parameters (see Section 4.3) and to understand the ways in which the current package differs from other packages. For the software developer, explaining a new code in a report is a useful exercise; a careful review frequently leads to modifications and improvements and so should be undertaken before the package is formally included within HSL.

The reports that accompany HSL software are made available via the Group's webpage (`www.numerical.rl.ac.uk/reports/reports.shtml`). They are often also submitted (possibly in a modified form) for publication in a leading international journal, such as ACM Transactions on Mathematical Software.

3.3 HSL catalogue

The catalogue contains a complete list of all the packages in the current release of HSL and gives for each one a brief outline of its purpose, method, origin, language, and other attributes. It was originally hoped that this data could be extracted automatically from the specification documents, but this has not proved possible and separate source files are maintained.

The catalogue also contains an extensive index which is designed to help a potential user identify the appropriate package for a particular task.

There is a separate catalogue with the same format for HSL Archive.

4 Programming language and style

4.1 Use of Fortran 95

All the packages in HSL are written in Fortran 77 or Fortran 90/95. The first Fortran 90 package to be included in HSL was `HSL_MA42` in 1995. Since 1995, developers have individually chosen to use Fortran 77 or Fortran 90/95. A major consideration when choosing the language was, in the past, portability. Fortran 77 compilers are almost always freely available on all platforms. Good quality free Fortran 90/95 compilers were slow to appear and was one reason why many HSL users were reluctant to switch from Fortran 77 to Fortran 90/95. However, in the last couple of years the free Fortran 95 compiler `g95` has become available. Our experiences of using it have been very positive and so we are happy to recommend it to HSL users. As access to `g95` is widespread (at least on serial machines), new HSL packages should normally be written in Fortran 95. The developer may choose whether or not to also supply a Fortran 77 version (note that a number of Fortran 77 packages within HSL, including `MA57` and `MA48`, have Fortran 95 wrappers).

Packages must adhere to the Fortran 95 standard except that, from September 2006, we allow the use of allocatable structure components and dummy arguments. These are part of the official extension that is defined by Technical Report TR 15581(E) and is included in Fortran 2003. It allows arrays to be of dynamic size without the computing overheads and memory-leakage dangers of pointers. Addressing is less efficient in code that implements pointer arrays since it has to allow for the possibility that the array is associated with an array section, such as `a(i,:)`, that is not a contiguous part of its parent. Furthermore, optimization of a loop that involves a pointer may be inhibited by the possibility that its target is also accessed in another way in the loop.

Since Fortran 77 has no dynamic memory, working memory required by a Fortran 77 package must be provided by the user and passed to the package as real and integer arrays. If the memory is insufficient, the code should terminate and give advice to the user on how far the computation has proceeded and how much additional memory is needed. The user will then need to provide larger work arrays and, very often will be required to restart the computation from the beginning (although sometimes it is possible to restart at some point before it was found that the memory was insufficient). This is much less convenient for the user than working with a Fortran 95 code that automatically allocates and deallocates memory as required (especially since once larger arrays have been chosen, there is still no guarantee that they will be large enough and the process may need to be repeated). Relying on the package to allocate and deallocate memory also simplifies the specification documentation since the number of arguments can often be reduced and this helps restrict the number of possible errors the user can make.

The use of dynamic memory allocation does come with a memory-management issue: the developer should provide the means for all temporary storage that has been allocated by the package to be released when the computation terminates. In particular, if the code fails before the computation is complete (perhaps because of an error in the user-supplied data), the developer should ensure that all temporary storage is released. This can be difficult, especially if there are many possible return paths back to the user's calling program. For storage that cannot be released until the user has finished with the computation, a straightforward method of deallocating the storage should be included within the package. A sparse solver is very commonly used in the inner loop of an iterative process, such as in solving Jacobian linear systems in the solution of differential-algebraic equations. If pointer memory is used in such a setting, even a small memory leak can lead to memory exhaustion over the course of the iterative process.

4.2 HSL rules and conventions

While it is recognised that each software developer will have his or her own programming style, there are a small number of rules and conventions that all contributors to HSL should adhere to. These are currently as follows:

- Each package starts with comments that specify the Copyright, original date, and give a brief summary of each change that has been made since the original version.

- Since 2002, every package in the Library must be threadsafe. This allows a user to safely run multiple instances of the package simultaneously in different threads or on different processors. A consequence of this requirement is that no common blocks or Fortran `save` statements may be used within HSL packages.

- The source code should conform to the Fortran standard; in particular, it should contain no tabs. `forchk` (see Section 9.4) checks for these.
- Neither the code itself nor any of its tests should have any variables that are not referenced or are referenced without being defined.
- For Fortran 90/95 packages, each line of the source must be at most 80 characters in length; for Fortran 77 packages, the limit is 72 characters (see Section 9.3).
- No implicit typing should be employed. Fortran 90/95 packages should use `implicit none`.
- There should be no `go to` statements that branch backwards (that is, the branch target statement should not be earlier in the code than the `go to` statment). They should be used in Fortran 90/95 only when they make the code clearer.
- Packages that allocate memory internally must either deallocate all such memory or must explicitly explain in the specification document if there are any arrays that have been allocated but not deallocated. An example might be arrays that the package has allocated to hold matrix factors and which the user may wish to retain for later solves. If there remains allocated memory it must be possible for the user to successfully deallocate it.
- Packages that hold data in direct-access and/or sequential-access files must provide a means of releasing file storage once it is no longer required.
- Optional arguments are allowed and are always at the end of the argument list.
- The source code must contain no `stop` statements (it should always be possible for the calling code to recover from error situations).
- There must be a way to suppress the printing of warning and error messages and, if printing is offered, the user must be able to specify the unit number(s) for this.
- List-directed I/O (that is, `print` or `write` in which the format specification is an asterisk) should be avoided in the source code and in the simple and comprehensive tests (see Sections 5.1 and 5.2). This is because the output is not portable, and can vary with both compiler and/or computer platform.
- From September 2006, all new Fortran 90/95 packages must specify the intent of each argument of each user-callable procedure. All information parameters (see Section 4.5) must be of `intent(out)`; except in a procedure that sets default values for control parameters, each control parameter (see Section 4.3) must have `intent(in)`.
- Allocatable arrays should be used in preference to pointer arrays.
- Where a format is used only once, it is preferable to place it as a character constant within the output statement, e.g.

```
write(*,'(a,es12.4)') " 2-norm of residual =", resid
```

- Pointers should be used only where they offer a significant advantage.

4.3 Use of control parameters

Since the early 1990s, many HSL packages have made use of control parameters, which have default values but may be set by the user to control the action; they are not altered by an HSL procedure unless it is initializing them with default values. Some HSL packages are designed to offer the user a large degree of control, while others leave few decisions open to the user. Clearly, a balance has to be achieved between flexibility and simplicity and this will, in part, depend on the target user groups/application areas and the complexity of the algorithms being implemented.

Apart from simple controls (such as those that control diagnostic printing and whether or not the user's data is to be checked for errors), default values for control parameters should normally be selected on the basis of numerical experimentation as being in some way the "best all round" values. Getting these choices right is really important because, in our experience, many users (in particular, those who would regard themselves as non-experts) frequently rely on the default settings and are reluctant to try other values even if their hardware is very different from the original test platform (possibly because they do not feel confident about making other choices). Thus when developing a new package experiments should be performed on as wide a range of computing platforms and practical problems as possible; this is discussed further in Section 5.3. There will inevitably be situations when the defaults may give far from optimal results and the user will then need to try out other values. Note that the ability to try different values can also be invaluable to those doing research. The developer may want to change the default settings at a later date as a result of hardware developments; within HSL this would lead to a new leading digit in the version label.

HSL routines that are written in Fortran 77 use an array for integer controls and another for real controls. In general, the packages include an initialisation subroutine that the user may call to assign default values to these control arrays. If other values are wanted, the relevant individual entries of the arrays can be reset by the user after the initialisation and prior to calling other subroutines in the package. We recommend using arrays for the controls and allowing some extra space for controls that may be wanted in the future.

Most of the Fortran 90/95 packages in the Library currently adopt a similar approach for initialising the controls but, in place of the arrays, use a derived datatype whose components are the control parameters. Extra components can be added, if necessary, in a new version of the package. If the default value gives the behaviour of the unmodified code, only the second digit of the version label need change; otherwise, the first digit should change. Using a derived type can be more user-friendly as it allows meaningful names to be used for the controls. Since we started to use Fortran 95, the components of the control derived type can be given default values when a variable of this type is declared. We do this in some packages, including `HSL_MA77`.

We note that it is important that the control parameters are fully documented within the specification document. However, as many of them are primarily intended for tuning and experimentation by experts, they should not over complicate the documentation for the novice user. The specification document needs to explain what, in broad terms, the effect of resetting a control parameter is likely to be.

If a parameter is important enough for the HSL developer to want the user to really consider what value to use (which may be the case if, for instance, the best value is too platform or problem dependent for a default to be confidently selected), then that parameter should not be a control but should be passed to the subroutines as an input argument. An example might be the use of

scaling for a linear solver. Our experience has been that the benefits of different scaling strategies are highly problem dependent and so we feel a user should really be aware of what scaling, if any, is being performed and that it should not be "hidden" away in a control parameter.

4.4 Checking of the user's data

To make a package robust, it needs to incorporate checking of the user's data. This is done (sometimes optionally) by most of the packages in the Library. The main exceptions are routines that are primarily intended for use by other HSL packages so that the data will have already been checked before they are called (and rechecking would not only be unnecessary and would complicate the code but could also add an unacceptable overhead). An example is `MC34`, which generates the expanded structure for a symmetric sparse matrix given the structure for the lower triangular part.

As already noted, in the description of the argument lists, the specification document highlights what restrictions there are, if any, on the parameters that must be set by the user. The code should normally check that the user-supplied controls are feasible; if one is not, we suggest that either a warning is raised and the default value is used or a return is made after setting an appropriate error flag. The user's control parameters **should not** be overwritten; if one or more is unsuitable and is replaced by the default, this should be done using an internal copy.

A number of packages in HSL, including most of the sparse direct solvers, allow the matrix data that is input by the user to include out-of-range indices and duplicated indices. In this case, the specification document should clearly explain what action is taken if such indices are encountered. We recommend that out-of-range indices are ignored and that the real (or complex) values corresponding to duplicated entries are summed (although sometimes the developer may offer the user, via a control parameter, an alternative course of action, such as taking the first occurrence and ignoring subsequent duplicates). In each case, a warning should be issued and the number of such indices returned to the user. Ignoring out-of-range indices and summing those that are duplicated may result in changes to the user-supplied data, in which case this should be clearly explained. Note there are instances when the software developer may not wish to allow out-of-range indices and/or duplicated indices. An example is the frontal solver `MA42`. The user has to specify the largest integer used to index a variable (the variables need not be numbered contiguously) and if any user-supplied variable indices exceed this it is considered to be an error and an immediate return is made after setting the relevant error flag.

4.5 The role of information parameters

Most HSL packages provide the user with information, both on successful completion and in the event of an error. The information that is provided does, of course, depend very much on the package but, as a minimum, this information will include an error flag. In the event of an error, the user needs to be given sufficient details to understand what has gone wrong together with advice on how to avoid the error in a future run. All error and warning returns should be fully explained in the specification document.

HSL routines that are written in Fortran 77 use "information arrays" to return information to the user; in general two arrays are used (one for real and the other for integer information) and should be slightly larger than needed to allow for later additional requirements. The arrays

need not be set by the user, although some codes that have a reverse communication interface currently rely on the information in these arrays not being altered by the user between calls; from January 2007, new HSL packages will not do this. In Fortran 90/95 packages, derived types (with `intent(out)`) can be used in place of arrays. As with the control parameters, this can be more user-friendly as it allows meaningful names to be used. However, printing a simple list of the information generated is then less straightforward; one possibility to assist users is to provide a separate printing routine to do this or, alternatively, to use a control parameter to allow the user to select an option that prints the information parameters once the computation is complete.

4.6 Communication between procedures

It often happens that data that is constructed during an invocation of a procedure in an HSL package is needed on a later invocation of the procedure or on an invocation of another procedure. The requirement that the code be threadsafe (see Section 4.2) means that this data must be passed through the argument lists.

This can often be done naturally by changing one or more of the principal arguments; for example, an array might be overwritten by its LU factorization and a permutation array might be set. If this is not realistic or further data is need, variables with the name `keep` (or a name including these four characters) are normally used (note that some existing HSL packages use the name `save` in place of `keep`). In Fortran 77, they are usually arrays; in Fortran 90/95, they are usually scalars of a derived type that is part of the package.

In Fortran 77, there is no way to stop users altering a keep array, which would probably have a disastrous effect, so the documentation tells users not to do it. In Fortran 90/95, the derived type may be defined in a module with its components declared `private`, which will mean that the user cannot alter them. However, it may be desirable that some of the data is visible to the user or to other HSL packages. In this case, only those components that are intended to be visible to the user should be mentioned in the specification document.

4.7 Use of MPI

When writing a package that uses MPI, it is important to avoid interfering with communications in other libraries or on other processes.

We recommend creating a new communicator and using it exclusively within the package. The easiest way to do this is to duplicate the user's communicator by calling `MPI_COMM_DUP`. If a smaller communicator is needed, `MPI_COMM_SPLIT` is available. When the package has finished, it should destroy its communicators by calling `MPI_COMM_FREE`, in order to avoid a memory leak.

The use of a separate communicator means that the package's communication is not confused with the user's, which aids debugging for both.

4.8 Use of OpenMP

An HSL package that supports OpenMP parallelism may need to call some of the procedures that are part of OpenMP. In particular, `omp_get_num_threads()` may be used to find out whether the code is executing in a uniprocessor environment, which allows key parts to be written differently for this case.

For users that are not running OpenMP, HSL has a replacement module `omp_lib` that provides suitable constants and procedures for a uniprocessor. The package code must include a `use` statement for this module.

4.9 Use of other library routines

Where appropriate, we recommend that developers of new HSL packages make use of existing HSL packages. Many of the packages whose names begin with the characters `MC` are sparse-matrix manipulation routines that have been included as separate packages so that those designing, for example, sparse direct solvers can use them to simplify both the development and maintenance of what is inevitably very complex code. Another advantage of using existing packages is that they will have been tested separately and so can be used with confidence, without retesting. Sometimes when developing a new package it will be appropriate to put part of the code into a separate HSL package to improve the modularity of the design and to allow others access in the future. For example, the out-of-core solver `HSL_MA77` has at its heart a call to a routine that efficiently performs a partial factorization (and corresponding partial solve) of a dense matrix. It was felt that this would be useful to have as a separate package (`HSL_MA54`). Similarly, all the handling of the out-of-core working is done by a separate HSL package (`HSL_OF01`), which we anticipate reusing for future out-of-core codes.

We also encourage the use of BLAS (Basic Linear Algebra Subroutines). BLAS are an aid to clarity, portability, modularity, and maintenance of software, and they have become a *de facto* standard for elementary matrix and vector operations. Performance (speed) will typically be critically influenced by the use of appropriately tuned BLAS so, although the BLAS source code should be used when testing (and is supplied with HSL), users are advised to replace this with BLAS that are tuned for the target machine, for example, vendor-supplied BLAS or ones generated by ATLAS (see `http://math-atlas.sourceforge.net/`).

We also allow the use of LAPACK routines. The main problem associated with this is that many of the LAPACK routines involve a large number of dependencies, which must be identified and the source compiled when testing. As with the BLAS, source code is provided with HSL but performance will be dependent on using a tuned version.

The packages in the EP and MP sections of HSL rely on the MPI (Message Passing Interface) library. If this is not already available on the user's computer, it may be downloaded from

`http://www-unix.mcs.anl.gov/mpi/`

The package `HSL_MP01` consists solely of a module with an `include` statement for the MPI constants. All Fortran 95 packages in HSL should use this module since this will permit them to be written in the free source form.

The only other external library that may currently be called from an HSL package is the ordering library METIS. This is optionally used by `MA57`. If used, it must be downloaded separately by the user from the METIS website

`http://glaros.dtc.umn.edu/gkhome/metis/metis/overview`

5 Verification and testing

5.1 Simple examples

One or more simple examples are included in the specification documents, see Section 3.1.5, and are intended to provide a simple illustration of the use of the package. While they do provide a check for gross blunders, they rarely exercise enough of the code to give confidence in its correctness.

5.2 Comprehensive test

The comprehensive test is intended to provide a fuller check on the correctness of the code. It (and the simple examples) are rerun whenever a change is made to a package so it is important both that its execution time be not too long and that it explores as much of the code as possible. Furthermore, it should be designed so that the output is (as far as possible) machine and compiler independent. For example, for a linear equation solver, the output should not be the computed solution or the residual; instead, the test code should check that the solution and residual are as expected and then only perform printing in the event that it is not.

These objectives mean that it has to be designed specifically for testing the package. It would not be satisfactory, for example, to employ a code that was used to tune its performance or compare it with other codes. The aim is to check that it really is performing the algorithm that was finally chosen and this can usually be achieved with small test cases, perhaps constructed with a generator for pseudo-random numbers.

It should exercise all the modes that the user is permitted to employ (including testing of different settings for the control parameters, errors and warnings, and the different levels of printing). Ideally, it would exercise all possible execution sequences, but we recognize that this would usually not be practicable. Instead, we aim to exercise every statement at least once. Even this is often not practicable; for example, it is desirable to test the `stat=` variable of a `deallocate` statement, but it is probably impossible to provoke the statement to fail.

A tool should be used by the developer to discover which statements are not being exercised. Each such statement should be examined to make sure that it is not practicable to exercise it and to check by eye that the statement is correct and that the statements that would be executed following its execution are correct. In Section 9.1, we explain how we currently use the g95 compiler for this purpose.

5.3 Independent testing

In addition to the simple examples that are included in the specification document and the comprehensive test, testing should be performed on a set of problems representing those at which the package is primarily aimed. These tests do not form part of the Library but they will typically be reported on in any accompanying report (see Section 3.2).

Testing on practical applications needs to be performed continually while the package is under development. Such tests often result in changes to the user interface, in particular, it may become apparent that further options are needed (perhaps via optional arguments or additional control parameters). The tests may be used to select defaults for the control parameters. The test set should be as large and varied as possible. If a package is

intended to be general-purpose, the test set not only needs to include problems of different sizes but also problems from a range of application areas. For testing sparse linear solvers and sparse eigensolvers, matrices from the University of Florida Sparse Matrix Collection (`www.cise.ufl.edu/research/sparse/matrices/`) may be used. This is an extensive set of matrices arising from real problems. It includes the original Harwell-Boeing Collection (which members of the Group were responsible for creating back in the 1980s) and is constantly being updated to include ever larger problems. For optimisation, there is the CUTEr suite of Fortran subroutines (see `http://www.cse.scitech.ac.uk/nag/cuter/cuter.shtml` for details), which is developed by the Group and collaborators. A key advantage of using these collections is that they are available via the web, which facilitates comparisons with other packages (both HSL and non-HSL packages). Where appropriate, comparisons with other packages needs to be part of the validation and testing process.

As well as testing on a range of problems, a package should be tested on as many different platforms as possible. This is to help ensure portability and robustness and also to check that the default settings for the control parameters are appropriate.

We highly recommend that the developer of an HSL package involves someone who has not themselves been closely involved in the design and development of the code in the testing. This may be another Group member who has an interest in the package or, ideally, a potential user from outside the Group. In our experience, feedback from an independent user is invaluable, both for the specification document and the package.

5.4 Use of compilers

We are fortunate in having access to a range of up-to-date Fortran compilers (currently, Nag, g95, gfortran, lf95, ifort). The developer should test his or her software using each of these compilers. We often find that different compilers will unearth errors that others failed to detect and can provide assistance in tracking down and correcting an error. When compiling the simple test(s) and the comprehensive test, the highest level of error checking offered by the compiler needs to be used. Any warnings returned at compile time should be checked. Ideally, no compiler warnings should be issued but, as we now allow allocatable structure components and dummy arguments, these will currently lead to a warning message. In our opinion, warnings can be off-putting for users, who may be concerned that they have made an error in the compilation process, and it is normally very easy to make simple changes to the code to eliminate most of these warnings.

6 Role of the Librarian

The Librarian is responsible for maintaining an up-to-date copy of HSL and associated tools. For HSL 2007, this is currently in the directory

```
/numerical/num/hsl2007
```

on the Group's Linux computer `nag`. This directory is maintained under version control with the Linux tool `svn`. Everything associated with a particular package is held in a subdirectory named after the package in the subdirectory `packages`. For example, everything associated with `HSL_MA48` is held in the directory

`/numerical/num/hsl2007/packages/hsl_ma48`

For the HSL package *package*, the names of the files for source, data, and output are summarized in Table 2.1. The output files are those produced by the Nag compiler on the Linux computer `nag`.

In addition, there are the following files:

- `makefile`, whose purpose is explained in the next paragraph
- `blas` holds the names of the files for the source for the BLAS procedures called directly or indirectly in the single-length case. `dblas`, `cblas`, ... hold the names for other cases.
- `lapack` holds the names of the files for the source for the LAPACK procedures called directly or indirectly in the single-length case. `dlapack`, `clapack`, ... hold the names for other cases.
- `deps` holds the names of the files for the source for the HSL codes called or used directly or indirectly in the single-length case. `ddeps`, `cdeps`, ... hold the names for other cases.
- *package*`.tssd` or *package*`.tex` holds source for the specification document.
- *package*`.ral.pdf` and *package*`.aspen.pdf` hold the RAL and Aspen Technology versions of the specification document in PDF format (they differ only in that the footnotes refer to different web addresses for the licence conditions).

The `makefile` is used by the Librarian for the following tasks:

- Run the comprehensive test. This relies on a fresh compilation of all the code invoked (including any BLAS and/or LAPACK routines called by the package), which allows full tests to be made for errors such as out-of-range subscripts or argument list mismatches wherever they occur.
- Run the simple tests. Again, all the invoked code is compiled.
- Check the 80-character line length for Fortran 90/95 code or the 72-character line length for Fortran 77 code.
- Run a coverage check of the comprehensive test.
- Construct all the files needed by Aspen Technology.
- Clean the directory of all temporary files.
- Merge the versions of the source code, the comprehensive test and the specification tests into master files with suffix `.coco`.
- Expand the master files (usually after editing them) back to their orginal form, overwriting existing versions.

For each release of a package, the Librarian's responsibilities are:

- to check the leading comments in the code;

- to check conformance with the relevant language standard;
- to run the comprehensive and examples codes under the Nag compiler and check that the output remains unchanged, or that the changes are acceptable;
- to check that the tests do not have any unused variables or references to undefined variables;
- to check that the line-length limits have not been exceeded;
- to construct the PDF files of the specification document; and
- to construct the files needed by Aspen Technology.

The Librarian is also responsible choosing new HSL names (in consultation with Group members) and for maintaining lists of HSL names. All names, including those under development and those that have been deleted from HSL, are listed in

```
/numerical/num/hsl2007/hslnames.txt
```

A slightly shortened version, containing only those currently in HSL and HSL Archive, is contained in

```
/numerical/num/hsl2007/schedulec.txt
```

This is an Annex of the contract with Aspen Technology.

In addition, the Librarian has responsibility for all aspects of the catalogue, which is normally not changed except at a new release of the whole Library.

The Librarian is not responsible for constructing the code and its tests, eye-ball checking of the code and its comments (other than the leading comments), performing performance checks, running the tests on compilers other than Nag, constructing and checking the specification document, or correcting bugs. These are all the responsibility of the developer. Of course, the Librarian is a member of the Group, so he may perform these tasks as a developer or as a colleague of the developer.

6.1 Check lists

Starting in January 2007, before passing a new package to the Librarian for inclusion in HSL, a pre-release check list needs to be completed. Similarly, after a level one revision is accepted (the first digit in the version label is incremented), a revision check list must be completed. Sample pre-release and revision check lists are in Appendices 9.9 and 9.10. These check lists are intended to remind the developer of the steps that need to be gone through when preparing the code and the accompanying documentation and should prevent packages from being passed to the Librarian prematurely. In particular, the specification document and the main source code must be signed off by a senior member of the Group (currently Duff, Reid and Scott may approve a package).

7 Licence issues

All use of HSL codes requires the user to have a valid licence. Academics may apply for free access to HSL for non-commercial purposes. Access is via a short-lived individual password-controlled account, details of which are sent on completion of a short online form (see `http://hsl.rl.ac.uk/hsl2007/hsl20074researchers.html`). Other potential users must contact the software distributors, Aspen Technology Inc, for a licence (details are available at `http://www.cse.scitech.ac.uk/nag/hsl/howto.shtml`). Each package is distributed as Fortran source code, starting with comments that detail the Copyright and conditions of use, followed by comments on the version history (see Section 2.1). There follows the code itself.

The Copyright lines specify all the institutions and/or individuals that have any Intellectual Property Rights in the package. Where this extends beyond Science & Technology Facilities Council (or CCLRC, RAL, etc) and Aspen Technology Inc. (or Hyprotech, AEA, etc), there should be an agreement in place that specifies the rights and how any revenues will be shared. In the case of a collaboration in which the outside party does not claim any Intellectual Property Rights, there should be a memorandum signed by the outside party that makes this clear.

8 Acknowledgements

We would like to express our thanks to Iain Duff, Chris Greenough, Jonathan Hogg, and David Worth for their careful reading of drafts and many helpful suggestions. We are grateful to Coralia Cartis and Sue Dollar for commenting on a draft and for putting together the check lists in Appendices 9.9 and 9.10. We also appreciate the comments that were made by the late Lawrence Daniels of Aspen Technology Inc.

9 Appendices

9.1 Coverage with g95

The tool that we recommend comes with the Fortran compiler g95 and may be run on one of the Group's computers `nag` and `softeng` as is illustrated by the example

```
/home/mathsoft/bin/g95 -fprofile-arcs -ftest-coverage \
          fa14d.f90 of01d.f90 of01dt.f90 ddeps.f
./a.out > temp
/home/mathsoft/bin/gcov4 of01d
gedit of01d.f90.gcov
```

The steps in this example are as follows:

1. Compile the program, adding counting instructions to the executable. This also creates files with suffix `gcno`.

2. Run the program. This also creates files with suffix `gcna`.

3. Construct an annotated listing of `of01d`. This needs the files `of01d.gcno` and `of01d.gcna`.

4. Examine the annotated listing with an editor. Executable lines that were not executed are marked `#####`.

9.2 Coverage with nag_coverage95

An alternative is the Nag tool `nag_coverage95`. The way the Librarian uses it is illustrated by the example

```
nag_coverage95 -source -load all_together.f90 -first ddeps.o
all_together.inst.exe
```

The code to be profiled must include the main program. Here it has been collected into `all_together.f90` and the rest has been compiled under `f95` into `ddeps.o`. The first line creates instrumented code in `all_together.inst.f90`, compiles it, and links it with `ddeps.o` to `all_together.inst.exe`. The second line runs this executable and produces counts in `all_together.out`.

For other ways to use `nag_coverage95`, see its documentation

```
http://www.nag.co.uk/nagware/nq/man/f95_tools/nag_coverage95.html
```

9.3 Checking line lengths

The commands

```
g95 /numerical/num/hsl2007/checks/check72.f90
a.out <ma48ad.f
```

check `ma48ad.f` for lines longer than 72 characters, copying any such lines to unit 6. The code `check80.f90` in the same directory checks for lines longer than 80 characters.

9.4 Debugging and checking conformance with standards

For debugging and checking conformance with standards, it is important to run several compilers. Our experience is that different compilers can find different problems. The following command lines may be used on the Group's computer `nag`:

Nag compiler:

```
f95  -C=all -C=undefined -nan -u -gline
```

For Fortran 77 packages with double complex versions, the option `-dcfuns` may be needed.

Lahey-Fujitsu compiler:

```
lf95 --chk aesux
```

Gnu g95 compiler, allowing allocatable components and dummy arguments:

```
g95 -std=f95 -ftr15581 -Wall -Wimplicit-none -fbounds-check -ftrace=full
```

Intel compiler:

```
ifort -C -u
```

Forcheck syntax analysis for Fortran 77:

```
forchk -f77 -nwarn -ninf -nff
```

The following command lines may be used on the Group's computer `fox`:

Gnu compiler:

```
gfortran-4.3 -std=f95 -fbounds-check -Wall
```

Since each of these find different problems, they should all be used.

9.5 Polishing Fortran 95 code

Fortran 95 code may be polished with the Nag tool `nag_polish95`, as in the example

```
nag_polish95 temp.f90
```

which copies `temp.f90` to `temp.f90.original` and overwrites `temp.f90` by a polished version. It uses a default set of options unless it finds a polish options file with name `.polish_options`. It looks first for the environment variable `NAG_POLOPT95` or `NAG_POLOPT90` containing the full name of the file, then in the current directory, then in the home directory. We recommend that you copy the file `/numerical/num/hsl2007/tools/.polish_options` to your home directory. You can edit it there with `nag_polopt95`. There are man pages for both `nag_polish95` and `nag_polopt95`.

9.6 Polishing Fortran 77 code

Fortran 77 code may be polished with the Nag tool `nag_polish`, as in the example

```
nag_polish -po ~/polish_options temp.f
```

which copies `temp.f` to `temp.f.orig` and overwrites `temp.f` by a polished version. It uses the polish options file `polish_options` in the home directory. We recommend that you copy the file `/numerical/num/hsl2007/tools/polish_options` to your home directory. You can edit it there with `nag_polopt`. There are man pages for both `nag_polish` and `nag_polopt`.

9.7 Time Profiling

Time profiling reveals the amount of CPU time that each subroutine uses and can indicate which statements take the longest to execute. This in contrast to the coverage profiling of Subsections 9.1 and 9.2, which give the number of times each statement is executed.

There are two approaches – code instrumentation and hardware performance counters. We describe the latter, since we have found it to be the more useful. Both provide only approximate timings because of compiler optimizations and because they sample the execution.

Hardware performance counters work by counting specific events on each CPU core. When a count reaches a configurable value, it triggers an interrupt which is recorded against the currently executing instruction. The interrupts slow the execution slightly, but the approach has the benefit that code does not need to be instrumented and therefore runs at near-production speed.

To obtain output that is easy to read, the user must compile the source with the `-g` option.

We now explain how to use `oprofile` on the Group's computer `fox`. We need a separate window in which `root` is logged in and two windows in which the user is logged in normally. On one of the user windows, execute

```
top
```

This will dynamically show how the machine is being used. If anyone else is making significant use of the machine when you are profiling, you will be monitoring (and slightly slowing) their use as well as your own. **Therefore, all members of the Group should be warned beforehand.**

The profiling may be controlled by a GUI. To start this, execute

```
oprof_start &
```

in the `root` window. The GUI may be used to change settings. The most useful event is the unhalted clock tick, `CPU_CLK_UNHALTED`, which essentially measures time.

To profile a program, first use the start button in the GUI to start the profiler. Next, in a user window, run the program as normal. Use the GUI stop button to stop the profiler when the run has completed. The results of the profile run can be viewed by executing commands in the user window. To see a list of counts in procedures, execute

```
opreport -l a.out > temp
```

and view `temp` with an editor. To see a limited callgraph report, execute

```
opreport -l -c a.out > temp
```

To see annotated source code, execute

```
opannotate --source a.out > temp
```

For further profiling, the counts will be accumulated unless the 'Reset sample files' button in the GUI is used to reset the counts to zero.

`oprofile` has many more advanced options. These are detailed in its own documentation, available on the web.

9.8 Other tools

For advice on other tools see the web site of the Software Enginneering Group:

`http://www.softeng.cse.clrc.ac.uk/SESP/`

The Group has developed a tool

`http://www.sesp.cse.clrc.ac.uk/html/SoftwareTools/pyqa_userdoc/pyqa_forcheck.html`

that acts as a front end to the static analysis tool Forcheck. It

- automates most of the steps needed to run Forcheck, including searching for relevant source files, sorting them based on module dependencies, selecting program options, and building the Forcheck execution command; and
- parses the text-based output from Forcheck, and presents the results in a more manageable and comprehensible format.

9.9 Check list for new packages

Check List for a New Package

Package name:	.....
Start date:	.....
Initial documentation:	**Read by** .....
Check HSL rules (section 4.2):	YES/NO
Simple tests:	YES/NO
Comprehensive tests:	
Profile:	**Coverage:** YES/NO **Different modes:** YES/NO
Compilers:	**g95:** YES/NO **Nag:** YES/NO **ifort:** YES/NO **lf95:** YES/NO **gfortran:** YES/NO
Use source BLAS/LAPACK:	YES/NO
Output portability:	YES/NO
Independent testing:	**Tested by** .....
Generate other versions (e.g. single):	YES/NO
Simple tests:	YES/NO
Comprehensive testing:	YES/NO
Finalize documentation:	YES/NO
Signed off date:	.....

9.10 Check list for revised packages

Check List for a Revision

Package name:	.....
Start date:	.....
Version:	.....
Start with the latest HSL version of codes and documentation:	YES/NO
Update documentation:	
Version number:	YES/NO
Comments on changes:	YES/NO
Changes:	YES/NO
Add comments to start of code:	YES/NO
Simple tests:	YES/NO
Comprehensive tests:	
Profile:	**Coverage:** YES/NO **Different modes:** YES/NO
Compilers:	**g95:** YES/NO **Nag:** YES/NO **ifort:** YES/NO **lf95:** YES/NO **gfortran:** YES/NO
Use source BLAS/LAPACK:	YES/NO
Output portability:	YES/NO
Independent testing (optional):	**Tested by** .....
Generate other versions (e.g. single):	YES/NO
Simple tests:	YES/NO
Comprehensive testing:	YES/NO