1. INTRODUCTION

1.1 Background

The Natural Environment Research Council (NERC) was established in 1965 by Royal Charter, and is the British organisation with responsibility for research into the physical and biological sciences relating to our natural environment. NERC's component institutes possess a wide capability for planning, encouraging and carrying out research which leads to a better understanding of the nature and processes of our environment, and on whose resources we depend.

NERC currently has around 2,500 staff. It is funded in part by direct grant from the UK Government, and part by commissioned research (ie: performed under contract). Such contracts are won from a number of sources, notably other public departments and governments of foreign countries, especially in the developing world. In the financial year 1987/88, the total budget of NERC amounted to some $100 million, of which 30% came from commissioned research.

Much of the work of NERC therefore is not done under contract, and is simply directed towards obtaining a better understanding of the working of our environment. However, the Council is uniquely qualified in being able to provide governments, industry and commerce with all the facilities for contract research across a very broad spectrum of environmental science. Its institutes operate either individually or in collaboration, and can provide research in areas such as resource exploration and evaluation, environmental management and impact assessment.

1.2 Component Bodies

Although NERC operates as a single entity, it is composed of a number of separate institutes whose activities span a wide range of scientific disciplines. There are currently 15 institutes, of which the main ones include:

- British Geological Survey (BGS)
- British Antarctic Survey (BAS)
- Institute of Hydrology (IH)
- Institute of Oceanographic Sciences (IOS)
- Institute of Terrestrial Ecology (ITE)
- Institute for Marine Environmental Research (IMER)
- Freshwater Biological Association (FBA)
- NERC Scientific Services (NSS)

NERC also has strong links with the UK academic community, to the extent that some 13% of the total budget is committed to supporting academic research within the Universities and Polytechnics. Such support is effected through direct research grants, by fellowships and by commissioned "special topics".

The last of the component bodies listed above, NSS, acts as a service body to the other institutes and has several spheres of operation. One part of NSS is NERC Computer Services (NCS), which has responsibility for running the general-purpose computing facilities throughout NERC, and of which the authors are members.

The NERC institutes are based at sites (currently about 30) which are widely distributed throughout the UK, ranging from Plymouth on the south coast up to the Shetland Isles. Since computing equipment is not always available at a local site, much use is made of network links and remote computers. In fact, all of the NERC sites are connected into the Joint Academic Network (JANET), which is a wide-area network shared by the universities, polytechnics and the UK research councils. The major regional computer centres are also linked to this network, allowing remote access to the super-computers installed at them.

1.3 Computing Requirements

The different institutes between them have a very wide range of computing needs. The diversity occurs in terms of both the level of familiarity of the users, and the functional requirements. The users range in experience from the novice up to the extremely competent. The functional requirements
vary from, for example, a basic need to use electronic mail, up to major database users and three-dimensional modellers.

In more general terms, the functional requirements can be categorised as follows.

1. **Data management.** There is a major need to be able to manipulate data in meaningful ways. The data volumes are often large, with queries that often cannot be determined beforehand.

2. **Statistics.** Statistical analysis remains the most common form of data evaluation within NERC. The required procedures range from basic descriptive statistics up to modern multivariate methods.

3. **Graphics.** Two levels of graphics are required, namely interpretative and presentation quality. The types of graphs to be produced reflect the diverse nature of the underlying sciences.

4. **Mapping.** Many of NERC's data are spatial in nature, and an increasing requirement is to be able to handle and manipulate them topologically. Automated production of maps is also needed.

5. **Modelling.** Mathematical modelling in two dimensions or more is regularly done, particularly in the areas of fluid dynamics and geology.

6. **Reports and word processing.** This is a rather newer requirement for multi-user machines. Until recently this was always handled by typists, but users are now wishing to incorporate data and graphs into reports, and to use the high quality printing devices attached to the computers.

These requirements of course can no longer be viewed as separate, since any single application is likely to break down into two or more of the categories. The ability to provide good links between software items is thus of considerable importance.

2. **COMPUTING OVERVIEW**

2.1 **NERC Computer Services**

As mentioned in the previous chapter, NCS is responsible for running the general-purpose computing facilities within NERC. These include the multi-user machines with all their associated peripherals, various image analysis systems, and parts of the network links. The latter two items, although listed for completeness, are not directly relevant to the current paper. The former item involves supporting both systems and applications software, operating the machines, and the day-to-day support of the various user communities.

Until 1985, computing in NERC was based on a Honeywell mainframe and a small number of GEC minicomputers. At that time though, these machines were approaching the end of their useful lives, and a replacement plan had to be developed.

With a large body of computer users having an increasing requirement for computing facilities, NCS needed to formulate strategies for both software and hardware procurements. The development of these strategies was subject to two particular constraints, namely finance and manpower. The financial aspect is a common enough one, and directs that the Computer Services makes sure that all purchases represent the best possible value for money, with no funds available for "trial" procurements; the manpower aspect dictates that both development and support be minimised.

2.2 **Hardware Strategy**

The basis of the hardware strategy is that of a dual architecture policy. This is that two types of
hardware would be purchased, and these types should also be in common use in the academic community (to aid program and data transfer). With two architectures, NERC would not be tied in to a single manufacturer, and the differing styles of operating system would enable users to choose the most appropriate for their applications.

The architectures selected were Digital Equipment Corporation's VAX range running VMS, and IBM's range using VM/CMS. In order to minimise effort, the operating systems would be run as supplied, thereby saving manpower in both development and subsequent support. At the time of writing, NCS has two IBM 4381 machines and 10 VAX installations (ranging in size from MicroVAX-IIs to an 8600/8530 cluster). The first machines were purchased towards the end of 1985, and further procurements within each architecture are likely.

Two other architectures are worthy of mention, although they are supported at a much lower level. These are IBM PC (and clones), and UNIX machines, normally workstations. Both of these are becoming increasingly important for certain applications.

2.3 Software Strategy

The basis of the software strategy is that products would be bought in, and that NCS itself would not do any development beyond very minor tailoring. Products purchased should be few in number, and should provide most of the required functionality. There would thus be a number of generalised packages with some additional ones for specialised needs. While this approach would inevitably mean that a small proportion of the requirement would not be met, this could be minimised by buying the right range of software.

To help fill the remaining requirement, there would also be certain third generation languages provided, namely FORTRAN, PASCAL, BASIC and C. A large number of scientists within NERC program regularly in FORTRAN, and have been used to developing applications in this language.

The software products purchased should be:

(1) established. NERC cannot afford to base its scientific computing on new or unproved software.
(2) high quality. NCS should obtain the best products for the requirement within the financial limitations.
(3) well supported. With limitations on manpower, it is essential that support for products rests with suppliers and not with NCS.
(4) available on both machine architectures (though this is not always possible). With the increasing availability of personal computers within NERC, it is also highly desirable that software products run on an IBM/PC.

While this paper is concerned principally with the applications software, the same strategy applies for systems software.

2.4 Software Base

The current range of applications software includes the products given below. While some of them have a long history of usage within NERC, others of them have only been obtained since the new hardware was installed. The more expensive products were obtained through competitive tender exercises.

(a) Data handling needs are met by the relational system ORACLE, and the free-text system STATUS.
(b) A range of general-purpose statistics products are provided:

- MINITAB as a simple, easy to use system,
- SAS as a mid-range product that also offers other facilities,
- GENSTAT as a rigorous tool for "real" statisticians.

In addition, the more specialised products GLIM and MLP are available.

(c) The mathematical and statistical subroutine libraries installed are NAG and IMSL.

(d) The graphics subroutine libraries provided are UNIGKS and GINO at a low level, and an old high level library GRAFIX. This was developed some time ago in-house, and is currently interfaced to UNIGKS; its functionality though is not yet available from any commercial product.

(e) Until recently, the only graphics program available was SAS. However, the UNIRAS suite of products (UNIGRAPH, UNIEDIT and UNIMAP) has just been acquired.

(f) Gridding and contouring requirements are satisfied by ISM:

(g) LEX is installed for word processing applications.

Between them, ORACLE, SAS and ISM account for some 95% of the total package usage.

2.5 Position of SAS

SAS was first purchased by NCS to coincide with the introduction of the new hardware, ie in 1985. The intention in purchasing SAS was to provide a general purpose tool for data handling and manipulation, statistics, reporting and graphics. In particular, a product was required that would satisfy a significant proportion of the user requirements in these areas. An integrated system and a user interface that was readily understood were of primary importance.

It was never expected that SAS would be able to satisfy totally the requirements in the above areas, and more specialised products are available for the applications that go beyond its limits. However, the aims of NCS in providing SAS were to make available a sufficiently wide-ranging and powerful product that would improve the throughput of scientists by:

(a) encouraging those not using machines into the benefits of computer analysis of their data,

(b) attracting FORTRAN and BASIC users onto a higher level of programming, and

(c) showing users of several different packages the improvements attainable from an integrated system.

3. SAS IN NERC

SAS was first installed (on a VAX 8600) in early March 1986. There was no existing experience of SAS within the Computer Services or within the user community of the NERC, although a number of people had been involved in the evaluation of suitable statistical packages, including SAS, in the previous year.

NCS now has SAS/Base and SAS/GRAPH installed on both of its IBM systems, and 8 of the VAX systems. SAS/ETS is available for special applications at two of these installations.
3.1 Changes and Additions

Very few changes have been made to the basic SAS product and modules provided. To date no attempt has been made to develop additional procedures, functions, format or informats. A number of additional logicals, under VMS, or filedefs, under CMS, have been added to ease user access to particular areas of SAS, for example the samples library. No macros have been developed, mainly because they are not available under VAX/VMS version of SAS. A number of SAS command files have been developed to carry out more specialist requirements of NERC users under VAX/VMS. On the MicroVAX-II implementations, the limitation of disk space lead to a reduced version of SAS. The solution adopted was to exclude PROC GMAP together with its related procedures and datasets. A universal graphics interface has been developed using the low level graphics package GKS. It has been designed to be graphics device independent. This allows the support of non-SAS graphics devices, such as off-line plotters, which are available widely within NERC.

NERC Computer Service does not provide an in-house data processing department or a statistical service as might be found in some commercial organisation. The majority of scientist carry out their own statistical analysis, usually interactively, with statistical advice available from their own statisticians within their institute. For this reason it was necessary to embark on a programme of introductory training for all users. The varied levels of computer literacy and specialist requirements of scientists within NERC, as well as the potentially high cost, made the use of SAS Institute run courses unsuitable. An in-house training course was developed which was designed to give scientists an introduction to SAS, a basic coverage of the datastep, basic statistical procedures and simple graphics. The Computer Service started to promote SAS actively towards the end of the summer of 1986, although a number of users had started to use SAS prior to this date. The introductory courses were started at the end of the summer of 1986 and lasted one day, and included "hands on" experience of SAS. Despite the shortness of the course, this approach appears to have been very successful with an increasing number of people using SAS. In the near future the course is likely to be supplemented by an in-house introductory guide or handbook.

Information additional to that provided by SAS is disseminated by two online methods. Permanent information is stored in help files, and information of a more temporary nature and "hints and tips" are stored on a bulletin board. Users have free access to both areas which are controlled by NCS.

As part of the NERC's policy of greater integration of packages, work has started on the development of a link between the relational database management system Oracle and SAS.

3.2 SAS Applications

SAS has been used widely over a number of applications and scientific disciplines. A few graphical examples of which are given here.

3.2.1 NERC Scientists

SAS has proved to be very popular with scientists throughout NERC. The ease of use of the SAS/Graph module has, in particular, resulted in heavy and varied use (Fig. 1 to Fig. 4). As well as a graphics tool though, SAS has found a niche in the areas of simple data management and manipulation, and statistics (although it does not suit the more ill-conditioned applications). Worthy of special mention is the way in which some users have been able to abandon writing BASIC programs to manipulate data, and have achieved significant savings in their time by using the SAS DATA step instead.

3.2.2 NERC COMPUTER SERVICES

The traditional piecharts and similar graphical representation of machine usage has been produced by SAS. In addition slides, overheads and posters have been produced for conferences and meetings (Figure 5).

A more serious application has been "REPORTER" on the VAX systems. This allows the monitoring of all applications and other packages. The VMS accounting facility is used to generate daily summaries on
all VAXes. On a monthly basis these are transferred to the central machine. Quarterly graphical and statistical summaries are produced for senior management (Figure 6). This has permitted better monitoring and targetting of support. Further development of this system is likely.

4. CONCLUSIONS

4.1 Advantages of SAS

The addition of SAS to the packages supported has led to an improvement in the facilities available to users. Notable improvements have been seen particularly in the area of graphics. Here users expectations have risen considerably with high quality and quicker production of graphs.

With simpler data management, the usage of the machines has improved. Manipulation or transformation of data is possible without reverting to either Basic or Fortran; the benefits of this to a novice user are considerable.

The consistent interface and the good defaults of SAS make it possible for user to obtain a result quickly. The integrated nature of SAS allows access to a wide range of features.

4.2 Disadvantages of SAS

The most serious problem has been the documentation. While it is highly detailed, its usability is low. Here, usability is defined as the frequency with which users report a "problem" when the solution is in the manual. Regularly, there is the feeling that the information is there if only you could find it. Improved indexing, possibly with the use of bold page numbers for key areas, and by subject area may help.

Technical information is spread widely and often difficult to find or obtain. While Technical Support do a good job, more detailed technical information would be welcome.

Many of our users are "novice" statisticians. The large quantities of output and the frequently "unusual" statistical test is confusing to them. There is a need for SAS to be much more truly interactive than at present. The inclusion of more brief options to control output may help.

The failure of SAS to flag or warn users when SAS encounters a statistical problem is seen as a particularly serious omission. Such facilities are seen in other serious statistical packages such as Minitab or Genstat.

The learning curve for SAS is particularly steep. In the past the cost of the SAS/AF module excluded the widespread use of a menu driven system. The news of the inclusion of the PROC DISPLAY into the SAS/BASE module was welcome. The one prototype system that we have evaluated looked promising. The lack, at present, of a menu driven version of SAS/GRAPH is disappointing, since this is one of the most heavily used areas of SAS.

Finally, departing from SAS's own defaults can be both time consuming and frustrating. Despite the voluminous documentation some options are either poorly documented or, in some cases, not at all.

4.3 Wish List

A number of enhancements would considerably enhance the packages from NERC's point of view. The principal "wishes" would include the following.

(1) The lack of macros within the VAX/VMS version is a serious omission. It restricts the development and extendability of the package. At the present time, it is the only statistical package provided by NERC Computer Services that does not have this facility.
(2) The presence of non-parametric statistics would be welcome. While these tests are simple they are used heavily by our users in other packages.

(3) The availability of separate summary cards would be useful.

(4) An index to the sample library under VAX/VMS like that found for the IBM's. This would increase substantially their usefulness.

(5) A good introduction or "getting started" manual would be welcome. The present guide is not an introduction and contains serious omission like how to get out of SAS!

(6) The availability of accelerated training course would be useful for users or supports with a limited knowledge.

4.4 Summary

On the whole, SAS has fitted its role very well. Its range of options is impressive, and the additional modules available (which NERC will probably start to explore before too long) extend its capabilities into a number of other useful directions. While NCS would like to see better linkages to other software products and more in-depth support within the UK, we are, on balance very pleased with the choice, and find SAS a very good basic tool for data manipulation and analysis.

SAS then has been a useful addition to the applications software base. The users' expectations have been raised considerably in the last 2 years. At the same time the standard of the service has improved. SAS has made a significant contribution to this improvement, notably in the graphics area. However, the move to a truly interactive and user friendly operating system has contributed as well. Finally, but hopefully not least, NERC Computer Services themselves might have contributed in some not insignificant way.