

A real-time database/models base/expert system in predictive microbiology

Julie E. Jones

Department of Mathematics and Computing Sciences, Campden Food and Drink Research Association, Chipping Campden, Gloucestershire, UK

(Received 10 February 1993)

Key words: Database; Models base; Predictive microbiology

SUMMARY

This paper describes the development and operation of a database/models base/expert system funded by the Ministry of Agriculture, Fisheries and Food in the UK. As part of an on-going coordinated program on predictive microbiology, the system being established involves storage of data and models relevant to changes in populations of food-borne pathogens under given conditions. The system is due to be completed by March 1994.

INTRODUCTION

An on-going coordinated program on predictive microbiology is being funded by the Ministry of Agriculture, Fisheries and Food (MAFF) in the UK. In total, the UK Government are investing £7 million in predictive microbiology to extend the current understanding of food microbiology. Within the program, various sites throughout the UK are generating both data and models. As a direct result, a large amount of data have been generated, and the MAFF felt maximum use could be made of these data if they were held in a central computerized database accessible to all authorized workers. Furthermore MAFF felt that a models base holding 'state of the art' predictive models – accessible through either personnel at various centres, able to use and interpret the models, or directly through an intelligent interface – would also be of considerable value. MAFF recognized that the technology was available and consequently commissioned the Campden Food and Drink Research Association to design and construct such a database/models base/expert system.

The work itself, a 5 year project was divided into three overlapping phases with staggered parts. The first phase of the work started in April 1989 with the design and construction of a database to hold data generated through MAFF funded research throughout the UK. It was considered that although this would not be of direct use to the food industry it would facilitate modeling research. Access to this data would be available to authorized users. The second phase started in April 1990 with the design and construction

of a models base – a library of computer programs and data – containing the results of modeling research. This was thought to be of direct benefit to the food industry but, because of the expertise needed for safe use and interpretation, it would be mediated by experts from service centres. These service centres would be the laboratories being funded by MAFF for data generation. A list of these laboratories is given in Table 1 and their activities within the program can be seen in Table 2. The final phase, the intelligent interface, started in April 1991. Eventually, this will allow non-expert users to gain direct access to the models base, ensuring that the models and associated information used are appropriate to the applications sought.

The database

The database has been constructed using a relational database management system called Rdb and is housed in a DEC VAX computer operating under VMS. Data are submitted to the database in a computer readable format. These data conform to a standard protocol which has been established in a set of documents called 'the data submission protocols documents'. These documents specify criteria such as appropriate flags for data and a standard format of a spreadsheet for the submission of data. For general convenience, data may be submitted in any spreadsheet package providing that the data within the spreadsheet are laid out in the format specified in the documents. In addition to this spreadsheet, an associated text file, referenced by name within the spreadsheet, is also submitted. This text file provides various general details about the experiment in which the data were generated – information that it is essential to know but which would be impractical to store in separate fields within a database. This text file is stored alongside the data and is despatched with the spreadsheet when data are requested by one of the authorized users.

TABLE 1
Organizations involved in the UK predictive microbiology initiative

CFDRA	Campden Food and Drink Research Association, Chipping Campden, Glos., GL55 6LD
DANI	Department of Agriculture (Northern Ireland), Food Microbiology Research Division, New Forge Lane, Belfast, BT9 5PX
FMBRA	Flour Milling and Baking Research Association, Chorleywood, Herts., WD3 5SH
IFR-NL	Institute of Food Research, Norwich Laboratory, Colney Lane, Norwich, NR4 7UA
IFR-RL	Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading, RG6 2EF
LFRA	Leatherhead Food Research Association, Randalls Road, Leatherhead, Surrey, KT22 7RY
TORRY	Ministry of Agriculture Fisheries and Food, Torry Research Station, P.O. Box 31, 135 Abbey Road, Aberdeen, AB9 8DG
UNILEVER	Unilever Research Laboratory, Colworth House, Sharnbrook, Bedford, MK44 1LQ
BATH	University of Bath, School of Biological Sciences, Claverton Down, Bath, BA2 7AY
CARDIFF	University of Wales – College of Cardiff, School of Pure and Applied Biology, P.O. Box 915, Cardiff, CF1 3TL
SURREY	University of Surrey, School of Biological Sciences, Guildford, Surrey, GU2 5XH

TABLE 2
The activities of organizations involved in the UK predictive microbiology initiative

Organizations	Microbiological kinetic studies	Modeling	Database construction
CFDRA	◆	◆	◆
FMBRA	◆	◆	
IFR-NL	◆		
IFR-RL	◆	◆	
LFRA	◆		
TORRY	◆		
BATH	◆		
CARDIFF	◆	◆ ^a	
SURREY	◆	◆ ^a	
UNILEVER	◆	◆	

^aNot part of models base.

Once submitted, the data go through an input gate, an experienced microbiologist based at Campden. Careful checking is necessary to ensure that data going into the database are of the highest quality and that all related information is available. Any missing information is requested

from the site submitting the data. This completed, the data are then automatically loaded straight from the spreadsheet into the database.

Data within the database are made available to those developing and/or producing predictive models within the MAFF initiative via a set of query screens. These screens have been written using DECforms, a form management system based on the proposed FIMS (Forms Interface Management System) standard, FORTRAN77 and SQL (an ANSI/ISO-compliant structured query language). The user is given the opportunity to search the database for data in terms of any combination of: various characteristics of an organism; factor measurements – both planned (the design measurements) and actual (the response measurements); the substrate used in the experiment, which may be a bacteriological medium or a food product.

The user requests his data by a method called 'drilling down'. When he first enters the database, all data are available. Then, as the user specifies his query, experiments become eliminated, perhaps because they involve an organism or substrate not of interest. Having selected the experiments of interest, the user may then 'post' a request for the data. When this request is 'posted', the title from each of the experiments involved is displayed on the screen, and the user is informed if he has previously selected the data and, if so, the date it was requested and the date it was sent from Campden. At this point, the user has the option to accept or reject this experiment.

Once a request for data is stored in the database, the system triggers an electronic mail message that is sent to a computer operator. The computer operator runs a program to retrieve both the data and the associated text file from the database, loads these onto the appropriate computer media for that registered user and then dispatches them. Any data issued will be subject to a confidentiality agreement, so as a direct result the system has been designed specifically so that no actual data are ever visible on the screen.

Models base

Whereas the database is a repository for large quantities of data, the models base is an analogous version but for models. The models base consists of two parts. Firstly, a library of subroutines for functions such as the Gompertz [2,4] equation and a general polynomial (not order specific), and secondly, a database for storing data related to a model. The subroutines take input from a specified source and direct output again to a specified source. For example, if the models base contains a predictive model constructed using a Gompertz-Polynomial [2,4] technique, then producing a prediction for this type of model requires at least two of the subroutines: the first for the polynomial function and the second for the Gompertz equation. This is illustrated in Fig. 1. Input to the subroutine to calculate the polynomial function requires: the factors involved (e.g. temperature, salt and pH); a table of coefficients used in the polynomial to calculate each parameter; the factor levels (e.g. temperature 12.2 °C, pH 6.7 and salt 0.9%) at which it is required to calculate the polynomial value. Directing these as input into

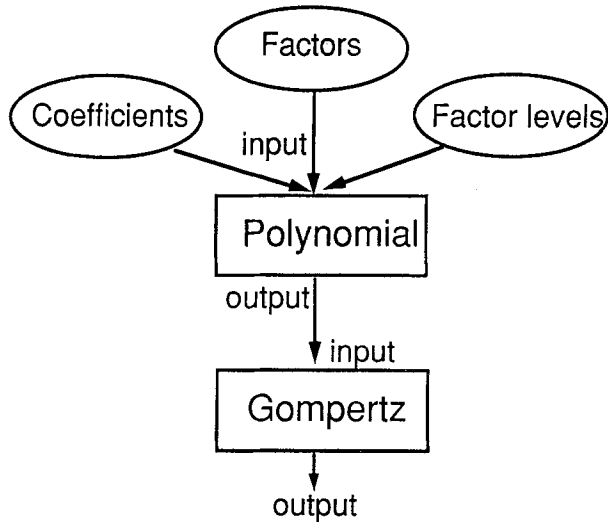


Fig. 1. Flow chart for calculating predictions.

the polynomial subroutine, the output from this is then directed into the Gompertz subroutine and the prediction is immediate (Fig. 1). This approach allows for modularity in programming, is easy to maintain and conforms to guidelines suggested for accreditation standards such as BS5750.

The second part of the models base involves the storage of information associated with a model. This information is stored in its own database. Within this database, information is stored, for example, on the origin of the model, the goodness of fit to the data, the boundary conditions for the model, the type of equations involved, the parameter values, and the foods that the model has been validated against. This information is important to enable the models base to function, and vital for the expert system to reach a decision.

As with the data, when a predictive equation is submitted to the models base it has to go through an input gate. However, unlike the data, and due to the importance of having reliable models in the models base, the input gate in this instance is a team of people called the Models Selection Group. This team of experts, drawn from various expert centres, has the task of deciding the fate of specific models – when a model is submitted to the models base it has to undergo two stages of acceptance.

The first stage involves considerations of the model itself: Has an accepted modeling technique been used? How reliable were the data used? Does the models base already contain a similar model? How well does the model fit the data? Does the model make biological sense? If the Models Selection Group decides that a model has satisfied part 1 of the selection procedure, it can then go onto part 2. This involves validation – validation against other data obtained in both food and media studies, and perhaps even comparisons with other models. This validation is on-going. As long as a model is accepted for use in predictions for at least one food commodity it can then be officially accepted onto the models base, but obviously the more food data the model is validated against the better.

The Models Selection Group also directs work on other

UK Predictive Food Microbiology Database			
Prediction Model for <i>Listeria monocytogenes</i> main			
Version 1.02			
Condition	VALUE	Minimum	Maximum
NaCl /% w/v	7.5	0.5	8.0
pH	5.5	4.8	7.4
NaNO ₂ conc. /mg l ⁻¹	.8	0.0	200.0
Temperature /C	5.0	5.0	30.0
Initial count:		100	
Maximum Population Size:		100,000,000	
Enter values for each of the conditions within the given limits			
<input type="button" value="Predict time/counts"/>	<input type="button" value="Draw graph of time/counts"/>	<input type="button" value="Temperature fluctuations"/>	<input type="button" value="Exit model"/>

Fig. 2. Input screen for model factor values.

aspects of modeling, such as the merging of data sets to produce one model, thus preventing a fragmentation of the models base with similar models with narrow ranges. Also it may advise on matters such as areas where additional data may be required.

Direct access to the models base, available for selected microbiologists from the expert centres involved in the program, is via a set of front-end query screens – again written in FORTRAN77 using SQL to query the database and DECforms as the screen display package – as described below.

The first screen shows a list of all models available. On-line help is available by simply moving the cursor to the appropriate organism and pressing the help key. Once a model has been selected, a screen appears (Fig. 2) listing the conditions or factors available for the model and their associated maximum and minimum ranges. The user has the opportunity to enter his own condition simply by altering the numbers in the column marked 'value'. Also a default initial count of 100 colony forming units is assumed, although the user can change this. Along the bottom there is a series of four option boxes. Selecting the first box 'predict time/counts' takes the user into another screen showing the maximum growth rate, lag and generation time (Fig. 3). Entering a time under the time/hours heading will result in the count for that time appearing under the 'number of organisms' heading and vice versa.

Returning to the first screen (Fig. 2) a user can select an option which will draw a graph of predicted log counts against time. Additionally, there is an option to create a

UK Predictive Food Microbiology Database	
Prediction Model for <i>Listeria monocytogenes</i> main	
Version 1.02	
Growth rate	.1206
Lag time/hours	44.8
Generation time/hours	2.5
Use the arrow keys to move to the time or number of organisms then enter the value and press RETURN to find the corresponding predicted value.	
Time /hours	Number of organisms
60.0	6,770
<input type="button" value="Previous screen"/>	<input type="button" value="Model parameters"/>

Fig. 3. Display screen for predicted values.

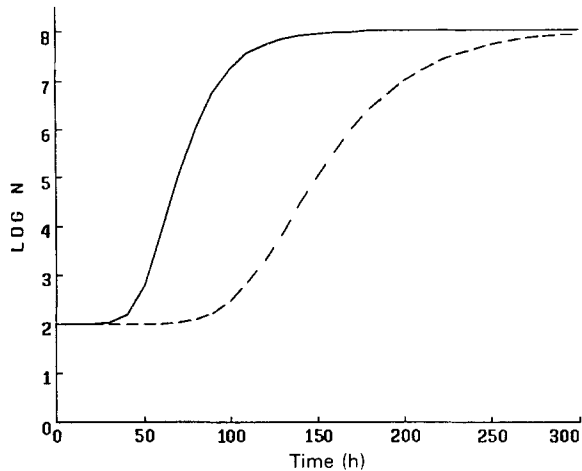


Fig. 4. Gompertz curve for *Listeria monocytogenes*.

new graph or overlay a new graph onto an existing one. The user may also modify the range of the plot. Figure 4 shows a graph where predictions at two different sets of conditions have been overlaid on the same axes.

The other type of graph that the system is capable of dealing with is predicted log counts based on time/temperature profiles. From the main menu, the 'temperature profiles' option is selected and a screen appears that allows the user either to call up an existing profile or to enter a new one. Again, once a profile is loaded, the user then has the option to plot a new graph, or overlay the prediction onto an existing graph (Fig. 5). On this graph the dashed line is the profile and the solid line indicates log counts. There are two vertical axes, the one on the far left is for temperature and the other is for log counts. On the screen the graphs actually appear in different colours.

The above summarizes the total capabilities of the models base to date. Following recent meetings with various representatives of the UK food industry, modifications have been suggested and it is anticipated that the majority of

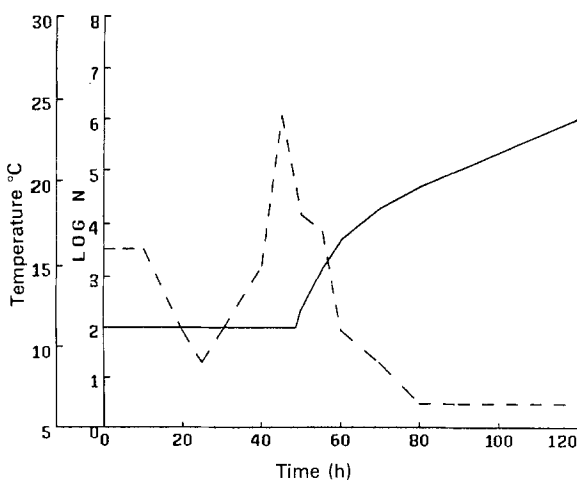


Fig. 5. Effect of temperature profile on growth.

these suggestions will be incorporated into the expert system to which they relate, and will not necessarily be implemented directly by the models base.

Expert system

In order for non-expert users to gain access to the models base, work has recently started on an intelligent interface. This will ultimately embody the expertise of experienced modelers and microbiologists. It will be in place to interrogate the user to determine his requirements in an interactive session, select the appropriate model and produce the required prediction.

The expert system is intended for use by experienced food microbiologists – it will not be designed for mathematicians or microbiologists already experienced in the use of predictive models in microbiology, who would probably find access via the expert system a little frustrating. In order to build the intelligent interface, the work has been divided into the following sections: determine the requirements; select the appropriate model or models; select the required data for input to the model; run the model; obtain the results; interpret the results.

The first section determines the requirements of the user. Does the user wish to modify a product or perhaps diagnose a cause of failure? Perhaps the issue is the safety of a product, requiring the identification of critical parameters in processing. How does the user want the results presented?

Having determined this, the next task is to select an appropriate model. This model will need information, including, for example, values for pH, temperature and water activity. This information may come from the user directly or, if the user does not know, the expert system will call on a reference database that contains the appropriate information. This reference database is comprised of data from various different sources such as Jowitt's Classification of Foods and Physical Properties [3] and the Royal Society of Chemistry's Composition of Foods [5]. The expert system may call on a thesaurus [1] containing alternative names covering products – i.e., Beans in tomato sauce is the same as Baked beans. Also, it has access to a hierarchical classification so that if, say, the product selected is strawberries but there are no models validated for strawberries it will search to see if there is a product similar to strawberries for which a model has been validated.

Having obtained the appropriate input, the model can then be run with the results being passed back to the expert system for the interpretation required by the user. The actual interpretation of the results is probably the most difficult section, in that work to date has shown that users want the results in similar but not quite identical forms. Hopefully, the scope of this interpretation will suit most users.

At all stages the user will be able to see the reasoning of the expert system, for example in the choice of a pH value. Also, if as in the previous example the expert system is trying to find an alternative but similar product it will inform the user that it is doing so and that the results should be interpreted with caution. Furthermore, the expert system

will also be aware of its own limitations, so that if it can not carry out a specific request it will refer the user to a human expert.

The intelligent interface is being constructed using Nexpert object, an expert system shell, FORTRAN and SQL, and the forms package being used is Nexpert's own. Nexpert allows a direct integration with other environments: it uses a bridge to connect directly to the database to look up information such as the reference data, it runs the appropriate FORTRAN routines, it will receive the results back from the FORTRAN routines, and it will interpret them in the required format.

Access

In all cases access is (or in the case of the expert system will be) available by modem and phone lines using remote computers or terminals. Potential users will have to be authorized in order to access the system. Once authorized they can be issued with an account and passwords on the system. To access the system remotely, a user will require a modem and either a microcomputer with an appropriate terminal emulation package or a DEC compatible terminal. At present access is only through the standard public switched telephone network of British Telecom. In the near future it is anticipated that access will be possible via a packet switching system such as Dialplus.

Currently an account is set up so that when a user dials in he is taken straight into a menu system. If a user is authorized to use both the database and models base then both options appear on the menu. All systems – the database, the models base and the expert system – have been or are being developed with built-in accounting procedures. At present this accounting involves the logging of information such as the user name, the user's organization, the option selected, the start and finish times, and the associated costs.

Administration, maintenance and development

At the end of the 5 year project (March 1994), it is intended that the whole system will become self-financing. This will cover both the running of the system and administration structure. This structure will facilitate an administration for users (e.g. help desk, processing of request for data) and an administration of users (e.g. addition and deletion of accounts, accounting procedures, charging rates, invoicing procedures). It is also anticipated that there will be a need for maintenance of both software and hardware including an implementation of upgrades to the proprietary systems.

Perhaps the most important long-term feature is the development of the system to maintain its relevance. Funding will be required for, for example, modifications to user interfaces, changes to search facilities in accordance with user requirements and general technological developments. Both the data and models bases will need to be maintained through addition or deletion of data and models, and through modification of criteria for using models.

CONCLUSION

Through UK government funding, a system is being established which involves the collation of data and mathematical models relevant to the changes of populations of food-borne pathogens under given conditions. This system will be accessible to non-expert microbiologists initially via service centres (from October 1992) and subsequently via a 'user-friendly' expert system. It is envisaged that the development of the system will continue and that it will become self-financing by March 1994.

ACKNOWLEDGEMENT

The Research Association gratefully acknowledges the financial support of the Ministry of Agriculture, Fisheries and Food for the work described in this report and for its permission to make the results available. In giving such permission, the Ministry does not necessarily associate itself with the views expressed in the report. The results of the research, are the property of the Ministry of Agriculture, Fisheries and Food and are Crown Copyright.

REFERENCES

- 1 Centre for Food Safety and Applied Nutrition. 1990. LanguaL – an Automated method for Describing, Capturing and Retrieving Data About Food. United States Food and Drug Administration.
- 2 Gibson, A. M., N. Bratchell and T. A. Roberts. 1988. Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride and storage temperature. *Int. J. Food Microbiol.* 6: 155–178.
- 3 Jowitt, R. 1989. A Classification of Foods and Physical Properties. Food Science Publishers Ltd. ISBN 094832385X.
- 4 Palumbo, S.A., A.C. Williams, R.L. Buchanan and J.G. Phillips. 1991. Model for the aerobic growth of *Aeromonas hydrophila* K14. *J. Food Protect.* 54: 429–435.
- 5 Royal Society of Chemistry. 1988. 'The Composition of Foods' Database. Cambridge, UK.