SCRIBE: AN INTERACTIVE SYSTEM FOR COMPOSITION OF METEOROLOGICAL FORECASTS

R. Verret, G. Babin, D. Vigneux, J. Marcoux,
R. Parent and J. Boulais

Development Branch
Canadian Meteorological Centre

1. INTRODUCTION

The Canadian Meteorological Centre (CMC) has developed RAPELS (Verret, 1990), a prototype plain language public forecast generation system for days 1 to 3, to produce automated forecasts at 264 Canadian stations. The initial reaction from the Regional Weather Centres evaluating it was positive, although the system needed refinements. The CMC designed the **SCRIBE** system (Boulais *et al*, 1992; Verret *et al*, 1993) based on previous expertise, to provide to the Regions an operational system tailored to their needs to help the composition of public forecasts. The prototype version of **SCRIBE** has been developed at CMC with the close collaboration of the Québec Weather Centre and the Maritimes Weather Centre.

The philosophy behind **SCRIBE** is as follows: a set of objective weather element guidance matrices at stations or sample points, prepared centrally at CMC, transmitted and decoded at the Regional Weather Centres are edited through a user-friendly graphical interface. Plain language bilingual (English and French) public forecasts are then generated from the modified guidances.

The weather element forecasts found in the matrices are prepared at 3-hour intervals giving **SCRIBE** a 3-hour time resolution. Weather element matrices for forecast areas (sub-regions) are generated locally from those available at the sample points within each area. These matrices are then combined in space and time by the Space and Time Combination system to create matrices for forecasts regions or zones. Then, they are combined to produce matrices for groups of regions (super-regions). The three levels of matrices (sub-regions, regions and super-regions) are all in the same format.

The weather element matrices are preprocessed by the Concept Generator before being displayed on the interface. The Concept Generator works through a domain space of rules to generate the concepts which are the results of a semantic numerical synthesis of the weather element matrices content. The main task of this module is thus to extract from the raw data the ideas or constructs that are hidden behind the digital weather element forecasts, or the numbers in the matrices. The concepts follow standards of codification based on the normalized Backus–Naur notation form. The concepts provide a simpler way to display the content of the weather element matrices rather than displaying the raw numbers.

The concepts are displayed on the graphical interface to be modified if needed. The concepts are represented by a series of bar graphs and/or icons. They can be modified through simple actions on the mouse or through simple quickly accessed menus and tool boxes. The interface has to be as user-friendly as possible so that the operational forecasters can perform this editing task efficiently. The interface is bilingual and file driven so that it can be adapted to the local needs. The inputs and outputs (concepts) of the interface are all in the same format.

The modified concepts are then quality checked, before being fed to the Text Generator. If inconsistencies are found, the system reverts back to the interface for corrections. Version 2.0 of **SCRIBE** has been released in June 1993.

2. WEATHER ELEMENT MATRICES

Figure 1 shows an example of a weather element matrix for the station YAW (Shearwater, Nova Scotia). FT is the projection time in hours and ZT is the valid time in UTC. CT2 is the climatological maximum/minimum temperature on a local time window, in degrees C, based on the data from 1949 to 1982. TST is the spot time temperature forecasts at every three hours in degrees C. TLO is the maximum/minimum temperature forecasts, calculated from the spot temperatures (TST) on a local time window in degrees C. Each maximum/minimum temperature forecast is repeated twice allowing for a range of values if needed (there is no range allowed for a single station forecast). CP1 is the climatological frequency of a trace (0.2 mm) or more of precipitation over 6-hour periods, CP2 is the climatological frequency of a trace or more of precipitation over 12-hour periods and CP3 is the climatological frequency of 10 mm or more of precipitation over 12-hour periods. The climatological frequencies of precipitation are in percent and are based on the data from 1963 to 1984. CLD is the total cloud opacity forecasts at every three hours in tenths of sky cover. P06 and P12 are the 6- and 12-hour forecasts of probability of precipitation (trace threshold) over 6- and 12-hour periods respectively. P10 is the probability forecast of getting 10 mm or more of precipitation over 12-hour periods. P06, P12 and P10 are in percent. QPS is the precipitation amount forecasts over 3-hour intervals in tenths of mm. CVI is the Showalter index. WW8 is the vertical velocity at 850 hPa in tenths of microbar/second (positive means upward vertical motion). TYP is the conditional precipitation type forecast. The numbers are a code (8 means snow and 2 means rain). AC is a companion to the precipitation type (the numbers are a code; 0 means no companion forecast). The AC column is used to forecast blowing snow and fog. DPD is the surface dew point depression in degrees C. TH5, TH7 and TH8 are the 1000-500 hPa, 850-700 hPa and 1000-850 hPa thicknesses respectively in decametres. DD is the wind direction in degrees and FF is the wind speed in km/h.

Two sets of matrices are prepared at each production cycle (00 and 12 UTC): one based on the Canadian Global model (operational version in November 1993: spectral, 119 waves with a triangular truncation, 21 levels in the vertical) (Béland and Beaudoin, 1985) and one based on the Regional model (operational version in August 1993: finite element, 50 km horizontal resolution, 25 levels in the vertical) (Staniforth and Daley, 1979). The Global model matrices cover a 72-hour period while those generated from the Regional model cover a 48-hour period. The matrices based on the Regional model are extended an extra 12 hours with the Global model data of the previous production cycle. This is done to allow **SCRIBE** to generate forecasts out to day 3 with the Regional model matrices.

TLO, CLD, P06, P12 and P10 are statistical forecasts based on Perfect Prog (PP) (Klein *et al*, 1959) linear regression equations. These forecasts are all operational at CMC (Verret, 1992). The conditional precipitation type forecasts are prepared using an algorithm to study the vertical temperature profile (Koclas, 1987) as produced by the driving model. The thicknesses are used to do a consistency check between the precipitation type forecasts and the spot time surface temperature forecasts. All the remaining weather element forecasts in the matrices are direct model outputs. There is no time interpolation in any of the forecasts included in the matrices, which means that **SCRIBE** has a true 3-hour time resolution.

The weather element matrices are prepared at 264 Canadian stations or sample points in **SCRIBE** 2.0. The spatial resolution of the system needs improvement and a doubling of the number of stations available in **SCRIBE** is foreseen in future versions of the system.

The matrices are compressed and transmitted to each Regional Weather Centre on the Wide Area Network. The total amount of data transmitted on the network for the purpose of **SCRIBE** is about one megabyte per day (including both production cycles and both numerical models).

3. COMBINATION SYSTEM

Upon reception of the weather element matrices at the Weather Centre, the Space and Time Combination system begins its work. The first task of the system is to go from point forecasts to area forecasts. The area forecasts will be in the same matrix format as the original ones. In **SCRIBE** 2.0 this is done by forcing the combination of the matrices available at the stations (sample points) within the same forecast zone. If there is only one sample point within the forecast zone, the matrix at that sample point becomes the matrix for the area forecast. The forecasts of TST, TYP, DPD, TH5, TH7 and TH8 valid between 06 and 15 UTC are taken from the matrix where the lowest temperature is found. In the valid period between 18 to 03 UTC, these parameters are taken from the matrix where the highest temperature is found. DD and FF are an average of the forecasts available in the different matrices. The remaining elements are taken from the matrix which shows the highest 12-hour probability of precipitation, for each 12-hour period. **SCRIBE** 2.0 does not consider sub-divisions of a forecast region.

The system then tries to combine the area forecast matrices for neighbouring regions. The approach taken follows Miller and Glahn (1985). The system scans in time each weather element in the different matrices potentially to be combined, and calculate the number of occurrences of no differences, small, medium and large differences. The thresholds used to define a small, medium or large difference are adjustable. The system will decide whether there is combination or not depending on adjustable preset thresholds for the number of differences. Directive lists are built in the system to control the way neighbouring regions can be combined. The lists allow the users to specify that a particular region never combines and is always by itself, or that some regions are normally combined together but not with others. If the system decides that there is combination, it will produce a weather element matrix valid for the group of regions, in the same format as the original ones. The resulting weather element matrix is produced using the same combination rules as those used to go from point forecasts to area forecasts, described above.

4. CONCEPTS GENERATOR

The concepts are the meteorological events coded according to a specified standard called METEOCODE. This codification standard is a context-free representation of the content of the weather element matrices based upon the Backus-Naur notation form (BNF) (Centre météorologique du Québec, 1992). The concepts always give the conditions at the beginning and at the end of the valid period of a particular event applicable to each weather element that can be displayed on the graphical interface. They also include the beginning and ending times of the validity period.

The inputs to the Concepts Generator come from the weather element matrices, either at the sample points (stations) or after being processed by the Space and Time Combination system. It queries approximately 540 decision rules (**SCRIBE** 2.0) each one having one or more conditions and actions. The concepts are generated for the complete period of interest, from 0 to 72 hours after initialization time. The concepts have a 3-hour time resolution, dictated by the weather element forecasts found in the matrices. All the processes to generate the concepts are based on a scanning in time of the weather element forecasts with cross-referencing between them. The Concepts Generator uses the **SCRIBE** knowledge base system (Boulais, 1992) (see section 7).

Version 2.0 of the Concepts Generator can produce more than 40 precipitation concepts (rain, rain heavy at times...) including 3 types of concepts applicable to thunderstorms (risk, possibility, a few) at up to 3 levels at the same time (ex.: rain and snow mixed with risk of freezing rain). It can also produce 2 types of concepts applicable to precipitation accumulation (liquid and frozen), 6 classes of probability of precipitation concepts, 13 sky cover concepts (11 stationary states and 2 evolving states), 14 classes of wind speed with 8 directions, 2 types of visibility concepts (blowing snow and fog) and 10 types of maximum/minimum temperature concepts. More concepts will be produced in future versions as new weather elements are added into the system.

5. INTERFACE

The interface used in **SCRIBE** 2.0 is called METEOPUB and has been designed and developed at the CMC with the close collaboration of the Québec Weather Centre. The interface is used to display and edit the concepts produced either by the Concepts Generator or inputted into the system by the forecasters. It is also used to modify the proposed groupings of forecast regions or create new groupings of regions. The inputs and outputs of METEOPUB are all in METEOCODE format. The forecasters use icons, simple menus and tool boxes to do the work at the interface, and everything can be done with the mouse. The interface is also used to trigger the Quality Control module and the Text Generator.

Figure 2 gives an idea of what the screen of the interface looks like (the actual screen interface may differ slightly from Figure 2). The bar at the very top of the screen is used mainly to select the bulletins, to load the guidances, to display the forecast regions and to trigger the Text Generator. Tool boxes appropriate to each weather element are activated by clicking on buttons on the left hand side of the screen.

The actual working area occupies most of the screen, with a horizontal scrolling bar at the bottom. It has a one-hour resolution and is divided into sub-areas for each weather element from top to bottom: sky cover, precipitation, probability of precipitation, temperatures, winds, visibility, precipitation accumulation and warnings. All the concepts displayed in the working area are completely editable. However a concept cannot last less than three hours. Figure 2 gives an example of the concepts that can be displayed on the screen (the concepts displayed are those generated from the matrix of Figure 1). Although the concepts generated by **SCRIBE** are at a 3-hour time resolution, the interface allows the user to work at a one-hour time resolution. Versions of this interface for marine and aviation forecasts are under development.

6. TEXT GENERATOR

The concepts, once modified at the interface, are fed into the Quality Control module. This system does a horizontal consistency check of each weather element over time, and a vertical consistency check of the different weather elements against each other. Messages are displayed at the interface to signal the main inconsistencies that may be found in the modified concepts file. The users have the choice to do the necessary corrections on the concepts or simply to by-pass the Quality Control system, and trigger the Text Generator.

The quality controlled concept file is preprocessed before going into the Text Generator as such, to make sure that there are no more than 3 concepts over a particular forecast period (normally a 12-hour period) for the clouds and precipitation. The concepts will be generalized in the event that they are too numerous.

The Text Generator uses the **SCRIBE** Knowledge Base system (see section 7). The rules (approximately 700) in the knowledge base are queried by forward chaining, although backward chaining is also possible depending on the problem to solve. The rules can thus be considered as the branches in a decisional tree. The design of the rules is largely based on Marcoux (1992) for precipitation and sky cover. Most of the rules are used for logical decisions, and only the terminal nodes of the branches in the tree do the selection of the words in the dictionary to construct the sentences. The Text Generator will create a basic sentence structure that can be matched into different structures representing different semantics expressing the same content, following a case base reasoning approach. Grammar rules are also available to make sure that the text comply to the syntax both in English and French. A graphical editor is in the process of being developed to provide and easy way to interactively modify and maintain the dictionary of words or lexicon.

7. THE KNOWLEDGE BASE SYSTEM

The Knowledge Base system (SCRIBE/KBS) is used by the Concepts Generator and the Text Generator. The three major functional constituents of the SCRIBE/KBS are the compiler to compile the rules, the inference engine to query the rules and the fact database management system. They work together to extract and manipulate the concepts imbedded in the raw data, based on the knowledge represented by an ensemble of rules and to solve the truth system particular to each meteorological situation. The compiler is used only once when the rules are modified or updated.

A rule is divided into three parts: pre-conditions, conditions and actions. The first two parts are used to determine whether the actions prescribed by the rules have to be executed. More specifically the pre-conditions allow to regroup the rules under different themes, for instance all the rules applicable to the sky conditions, or to the precipitation regime. The conditions determine the constraints governing the truth system.

The content of the three parts of a rule is expressed using a predicate language. Each part of a rule involves logical expressions which in turn may involve mathematical expressions. The compiler validates the syntax of the rule file with respect to a finite set of entities expressed as functions, predicates, variables and constants and of relationships expressed as logical, relational or mathematical operators. At the same time, the compiler codes the rules in formalized tree structures, easily manageable by the other functional components of the SCRIBE/KBS.

Once compiled and coded, the rules are selectively queried by the inference engine. The inference engine uses a set of functions and/or the fact database management system to solve each rule that is being looked at. A rule will see its actions executed only if the pre-conditions and conditions are tested as true. The actions are used to create or modify the facts, and to execute functions which modify the facts, generate products and/or redirect the inference engine. The fact database management system updates, adds or withdraws facts or informations in the database. The facts are used to solve the rules or to realize the actions prescribed by the rules.

Forward chaining is normally used to query the rules, although backward chaining is possible depending on the problem to be solved.

The main advantage of the SCRIBE/KBS is that it uses a predicate language. It is thus possible to generate and modify rules easily and rapidly. It is also possible to develop a knowledge base editor that will allow the users of **SCRIBE** to tailor the knowledge base system to their local needs.

SCRIBE/KBS has been coded in standard C. It is thus completely portable and can be run on any platform with a C compiler eliminating the need of a dedicated machine.

8. CONCLUSIONS

SCRIBE 2.0 runs on an HP 9000-735 machine at CMC. It needs about 10 megabytes of memory, including a fifteen day archiving of the data. Once the weather element matrices are available, it can generate public forecasts, in English and French, for 244 zones (forecast regions) in 26 bulletins, for days 1 to 3 in approximately 15 minutes real time, when there is no intervention at the interface level. The weather element matrices are produced on the main frame frontal computer at CMC.

Considerable efforts will be dedicated to include the recommendations of all Weather Centres in future updates of the system. In particular, the Text Generator should be tailored to the local needs. Additional weather elements, such as moisture, should also be added in the Concept Generator and at the interface level. A follow-up system on the warnings/watches should be developed to keep track of such events. This system will have to be tied with the Space and Time Combination system. Expansions of **SCRIBE** will include the production of marine and agricultural forecasts and, in the longer terms, aviation terminal forecasts.

An archiving system will have to be developed to archive the data at different steps in the system, for verification purposes and to make the data available to be used again. A prototype verification system in **SCRIBE** is under development.

It is planned at a later stage to develop a module to locally update the weather element matrices with actual observations. This will allow the system to compare the forecast weather with past or present weather situations. Finally a knowledge base editor could be developed to allow on site interactive editing of the rules used in the Text Generator, so that the latter can be easily and quickly tailored to the local needs.

9. ACKNOWLEDGEMENTS

The authors wish to acknowledge the participation in the project of the following people: D. Layton, J. Murtha, A. MacAfee and F. Petrucci at the Maritimes Weather Centre; M. Larivière, A. Méthot and D. Bachand at the Québec Weather Centre; M. Mondou, L. Pelland and S. Payer at CMC; and G. Viau at Training Branch.

10. REFERENCES

Béland, M. and C. Beaudoin, 1985: A global spectral model with a finite element formulation for the vertical discretization: adiabatic formulation. Monthly Weather Review, 113, 1910-1919.

Boulais, J., 1992: A knowledge base system (SCRIBE/KBS) analysing and synthesizing large amount of meteorological data. Preprints 4th AES/CMOS Workshop on Operational Meteorology, Whistler, B.C., Sept. 15-18, 1992, 283-286.

Boulais, J., G. Babin, D. Vigneux, M. Mondou, R. Parent and R. Verret, 1992: SCRIBE: An interactive system for composition of meteorological forecasts. Preprints 4th AES/CMOS Workshop on Operational Meteorology, Whistler, B.C., Sept. 15-18, 1992, 273-282.

Centre météorologique du Québec, 1992: MÉTÉOCODE: Norme de codification. Version 2.2, révision 92-04-01, 42 pp.

Klein, W.H., B.M. Lewis and I. Enger, 1959: Objective prediction of five-day mean temperature during winter. J. Meteor., 16, 672-682.

Koclas, P., 1987: Study of a winter storm using numerical output of the EFR model. CMC Monthly Review, 4, 2-12.

Marcoux, J., 1992: Liste des critères objectifs pour un générateur de prévisions météorologiques. Centre météorologique du Québec, 35 pp.

Miller, R.L. and H.R. Glahn, 1985: Techniques used in the computer worded forecast program for zones: Interpolation and combination. TDL Office Note 85-5, NOAA, NWS, 22 pp.

Staniforth, A. and R. Daley, 1979: A baroclinic finite element model for regional forecasting with the primitive equations. Monthly Weather Review, 107, 107-121.

Verret, R., 1990: Automated plain language composition of weather forecast: RAPELS. CMC Technical Document, 34, 49 pp.

Verret, R., 1992: CMC operational statistical products. Preprints 4th AES/CMOS Workshop on Operational Meteorology, Whistler, B.C., Sept. 15-18, 1992, 119-128.

Verret, R., G. Babin, D. Vigneux, R. Parent and J. Marcoux, 1993: SCRIBE: An interactive system for composition of meteorological forecasts. Preprints 13th Conference on Weather Analysis and Forecasting, Vienna, Virginia, August 2-6, 1993, Amer. Met. Soc. 213-216.

SHEARWATER YAW

920122 00Z

FT	ZT	CT2	TST	TLO	CP1	CP2	CP3	CLD	PO6	P12	P10	QPS	CVI	WWB	TYP	AC	DPD	TH5	TH7	TH8	DD	FF
00	00	-7	-14		25	34	8	1	1	11	0	0	16	0	8	0	4	503	143	122	300	14
03	03		-14					0				0	14	0	8	0	5	503	143	122	290	19
06	06		-14		26			0	10			0	14	0	8	0	5	504	143	123	280	22
09	09		-13	-9				0				0	18	0	8	0	2	504	142	122	280	25
12	12	-1	-14	-9	25	37	6	1	0	0	0	2	21	0	8	0	0	504	142	121	300	27
15	15		-13					0				3	22	-4	8	0	0	502	142	120	310	33
18	18		-11		25			0	0			3	23	-7	8	0	0	504	142	120	320	39
21	21		-10	-13				1				1	21	-11	8	0	0	509	144	121	330	36
24	00	-7	-9	-13	25	34	8	0	0	0	0	0	20	-11	8	0	1	515	147	122	330	31
27	03		-10					0				1	20	-11	8	0	2	521	148	123	330	22
30	06		-12		26			0	0			0	20	-4	8	0	6	526	150	124	340	12
33	09		-10	1				0				0	21	0	8	0	6	528	151	125	290	8
36	12	-1	-11	1	25	37	6	5	16	61	26	1	22	0	8	0	7	530	151	125	260	9
39	15		-5					9				1	20	0	8	0	6	532	152	126	180	9
42	18		-3		25			10	53			4	19	0	8	0	3	532	152	127	170	20
45	21		0	0				10				10	13	-4	8	0	1	535	153	128	170	29
48	00	-7	3	0	25	34	8	10	100	100	34	26	7	4	2	0	1	539	154	130	160	37
51	03		4					10				31	3	11	2	0	0	543	155	132	170	42
54	06		5		26			10	90			42	2	11	2	0	0	547	156	134	180	42
57	09		5	7				10				48	1	14	2	0	0	549	157	134	180	44
60	12	-1	4	7	25	37	6	10	100	100	29	81	2	14	2	0	0	551	157	135	180	53
63	15		6					10				80	4	32	2	0	0	555	158	135	170	65
66	18		6		25			10	56			8	4	22	2	0	0	553	158	134	180	59
69	21		4					10				0	7	-4	2	0	1	545	155	132	210	40
72	00	-7	1		25	34	8	10					11	0	2		2	535	152	130	250	36

Figure 1: Weather element matrix for Shearwater Nova Scotia, prepared at 00 UTC January 22, 1992. See the text for more details.

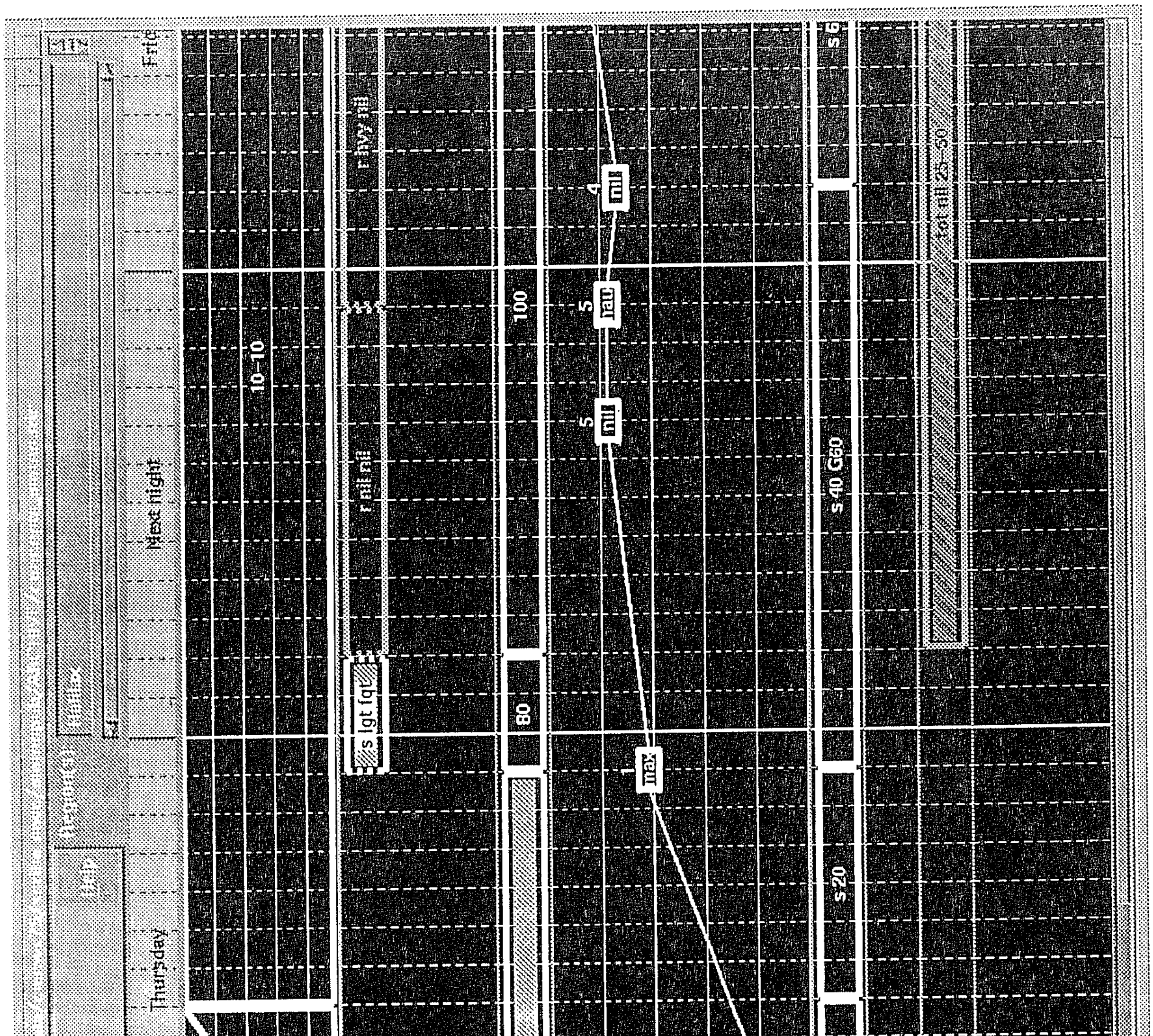

Figure 2: Representation of the graphical interface screen, with concepts displayed. The actual interface may differ slightly from this figure.