Cover page

Dynamic Data Management and Modeling in a Lagoon Based Wastewater Treatment Plant

By

Gautam G. Narasimhan

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Civil and Environmental Engineering
June 2000

Abstract

The wastewater treatment facility at Riviera de Sao Laurenço, Brasil was upgraded in January 2000 by the inclusion of Chemically Enhanced Primary Treatment (CEPT). The lack of a comprehensive data management scheme at the plant made the effects of the upgrade on the facility difficult to quantify. Data management at two treatment plants in the U.S. is reviewed and a solution for Riviera is presented. This scheme is then implemented in the framework of a desktop database application after a review of software development practices. Analyses from the application appear to demonstrate the validity of CEPT as a low-cost method of improving the performance and flow capacity of the Riviera facility. Load analysis on the Riviera system suggest that the installation of the CEPT clarifiers has decreased the load that would have been placed on the first lagoon in the treatment train, perhaps resulting in an increase of aerobic activity. The importance of modeling as an analytical tool is discussed and previous work on lagoon modeling is also reviewed. Rate constants for the carbon cycle at the Riviera lagoons during the Carnival period of 2000 are evaluated through a procedure developed to isolate global minima. In general, the model predicted effluent quality well, and the presented rate constants bear close agreement with those of previous investigators. However, the monitoring regime at Riviera should be expanded, both for environmental soundness and to provide higher-quality data for subsequent models. A natural extension of this expanded regime would be a combined data management and modeling application for the wastewater treatment facility at Riviera de Sao Laurenço.

Thesis Supervisors: Professor Donald R.F. Harleman and Susan Murcott
Titles: Ford Professor of Civil and Environmental Engineering, Emeritus and Lecturer.

Acknowledgements

To my family, for their uncritical and often undeserved support in letting me be the person I want to be.

To my thesis supervisors Dr. Harleman and Susan Murcott, it has been a privilege to work with both of you.

To all the people in Brasil who opened their hearts to us; Dr. Tsukamoto, Christian, Carlos, Godoy, the lab ladies, and Adriano, who I hope to see again in Bertioga someday.

To my project partners, Ir-eeene, Mike "On-line ALL the time" Bourke, and Heidi-Ho.

To Frédéric Chagnon, for making all his previous work on Riviera freely available to me.

To Dr. Al Pincince at CDM and Maury Hall at the MWRA for the information and advice they supplied.

To the John Freeman fund of the BSCE and the Civil and Environmental Engineering Department at MIT for financial support in making the amazing trip to Brasil possible.

To the Triple-B Double-P, for bringing the beach to me in November, introducing me to Jay Leno, and for showing me that the wings indeed come off the eagle.

And to the M.Eng. Class of 2000, for showing me that MIT can be a fun place to go to school.

Table of Contents

List of Tables

List of Figures

1 INTRODUCTION

1.1 BACKGROUND ON RIVIERA DE SAO LAURENÇO

Riviera de Sao Laurenço is a beachfront resort community 140 km from Sao Paulo, the largest city in South America. Riviera, as it is commonly called, is privately owned and operated by the Sabloco Construction Company, which has installed a waste stabilization lagoon system for the treatment of the community's wastewater. The facility is recognized as one of the premier wastewater treatment facilities in the state of Sao Paulo, and consists of two clarifiers for Chemically Enhanced Primary Treatment (CEPT), four lagoons and chlorine disinfection prior to discharge into the Itapanhau River. The river is also the source for drinking water for the community for which the collection point is a few hundred meters upstream of the discharge point. Thus, most of the river water that is withdrawn for use is treated and returned, which contributes to Riviera's reputation in the region for environmental soundness.

As can be expected in a resort town, the population is largely seasonal, with a huge increase during the summer months of January through March, and especially during the weekend when the festival of Carnival is held. Plant records show that the system load increases drastically for this period for both Biochemical Oxygen Demand (BOD) and Chemical Oxygen Demand (COD), as seen in Figure 1-1.

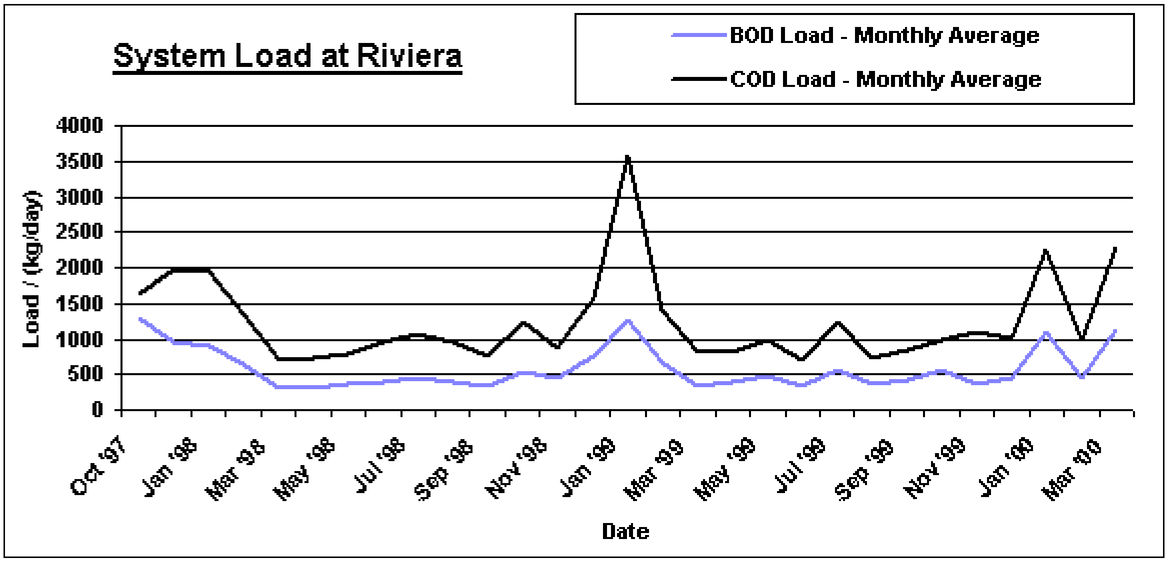

Figure 1-1 : System Load in kg/day for the Riviera WWTP

To address concerns of system efficiency and effluent quality during the summer months and to plan for projected increases in population, the CEPT clarifiers were added to the existing treatment train in January of 2000. CEPT involves the addition of chemical salts and polymer to augment the primary settling process. Experience with CEPT in other parts of the world has shown that it results in increased efficiency in addition to large savings of cost and space. Well operated plants equipped with CEPT have shown dramatic improvements from conventional primary treatment in terms of the removal of Total Suspended Solids (TSS), BOD, and phosphorus, which is often the bane of many a conventional treatment train.

	Conventional Primary Treatment	Chemically Enhanced Primary Treatment
Total Suspended Solids	60	85
Biochemical Oxygen Demand	30	65
Nitrogen	30	30
Phosphorus	30	85

Table 1-1 : Typical Removal Efficiencies in the Primary Treatment Stage (Morrisey and Harleman, 1992)

Chemical salts (typically ferric chloride or sulfate) are added at a dosing station about 3 kilometers upstream of the wastewater treatment plant. At the feed channel to the CEPT clarifiers, the mixture is dosed with anionic polymer to aid the flocculation process. The wastewater then proceeds into two clarifiers (to simulate a slower mixing time) and the resultant floc settles into a sludge blanket. The sludge blanket is removed through an automated sludge scraper. The effluent from the CEPT clarifiers is then fed into a lagoon system. The entire flow from the CEPT clarifiers passes into an anaerobic lagoon which has a depth of 3.2 m and a surface area of 6600 m^2, after which flow is divided and passed through three facultative lagoons in parallel. The facultative lagoons have an average depth of 1.5 m and a combined surface area of 45000 m^2. The effluent from the facultative lagoons is recombined and dosed with chlorine prior to river discharge.

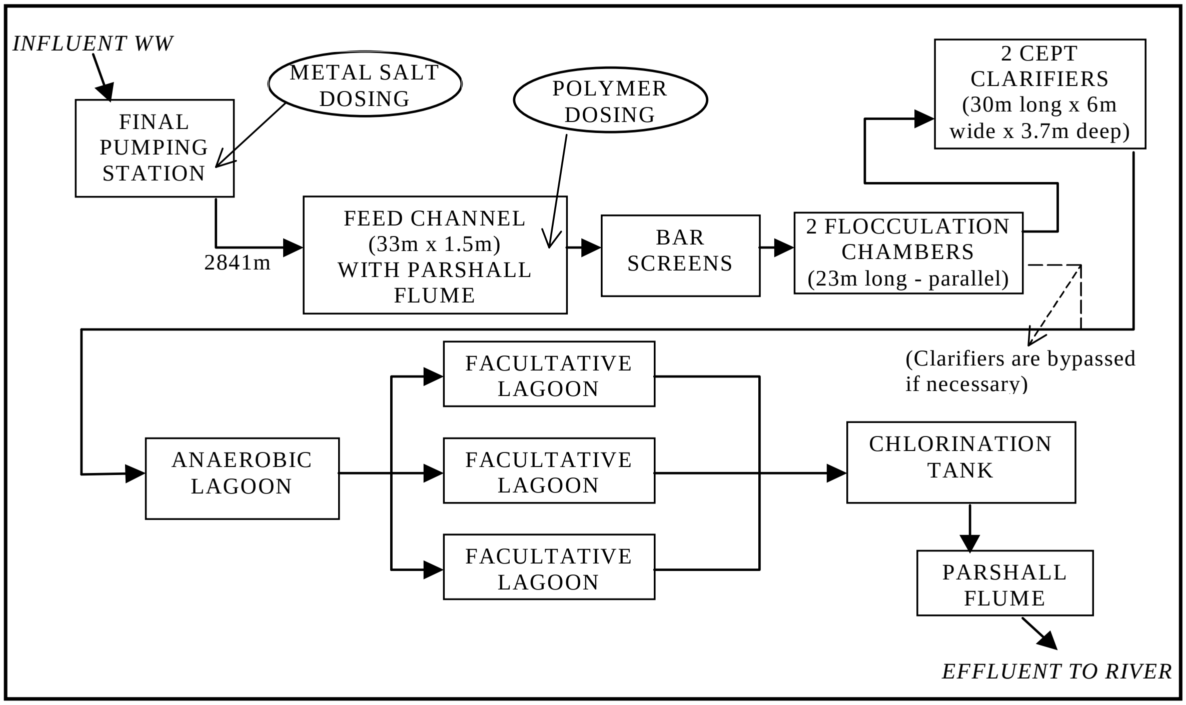

Figure 1-2 : Schematic of Upgraded Treatment Train at the Riviera WWTP

At present, only one of the clarifiers is being used for chemical treatment. The other is employed as a sludge storage tank where biosolids are stabilized with lime prior to disposal. CEPT was deemed to be the optimal choice for the system upgrade as the existing treatment system of the lagoons could be retrofitted with minimal effort. In addition, during the winter months, when the unaided lagoon system is sufficient, the CEPT addition can be taken off line with little effort; wastewater dosing can be ceased and the flow routed directly to the lagoon system.

Preliminary results indicate that the CEPT system is improving system performance admirably. Not only has the overall efficiency of the plant increased, but this increase in efficiency has occurred during periods of high influent load.

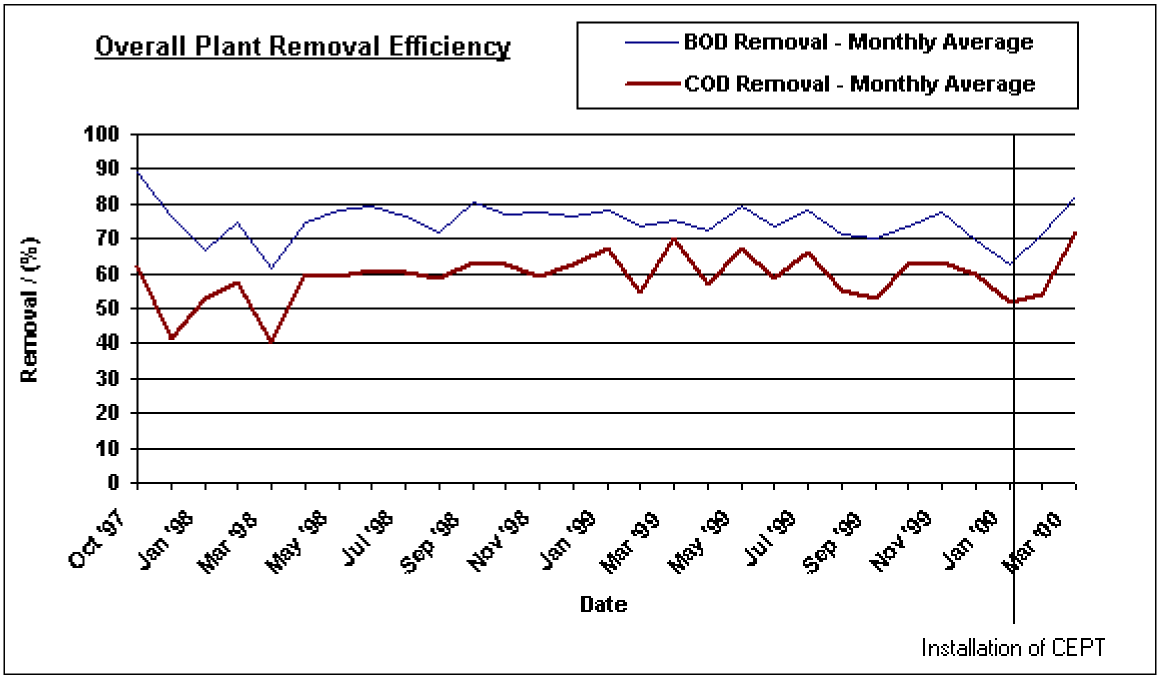

Figure 1-3 : Overall BOD and COD Removal for the Riviera Facility

Furthermore, the load to the biological lagoon system has been decreased. This is consistent with results obtained from well-operated chemical treatment at other locations.

1.2 PROJECT MOTIVATION

While the efficiency of the Riviera treatment facility and the benefits of CEPT are generally accepted claims, it is difficult to assess their veracity without a comprehensive data management scheme. Such a scheme must not only store and analyze data collected by the plant staff, but also present the results in a manner that can be clearly understood. Indeed, graphs demonstrating the increase in system load during the summer months or the effectiveness of CEPT would be extremely difficult to generate without such a process. In its absence, plant operators would have to go to extra lengths for decision support and environmental compliance reporting. With the advent of the information revolution and the ubiquity of desktop computers of significant computational power, implementation of a data management scheme has become easier than ever before. Indeed, implementation on such a platform is a vital step in assuring that the entire world has the maximum benefit of the information age.

Modeling the Riviera lagoon system is also extremely important, as it provides insight into the actual processes underlying wastewater treatment. This is especially true after the treatment train has been modified. A comparison of the parameters used to estimate effluent quality before and after the system upgrade can give valuable information as to the processes in a biological system that are most affected by chemical treatment.

2 DATA MANAGEMENT

2.1 IMPORTANCE OF AN EFFECTIVE DATA MANAGEMENT SCHEME

An effective data management and presentation scheme is integral to the ability of plant personnel to make informed decisions as to the operation of a wastewater treatment plant. Effective data collection is only half the battle since over any reasonable length of time the amount of data quickly becomes overwhelming unless it can be presented in an effective manner. Perhaps the most important application of such a scheme is the ability of plant operators to view influent and effluent water quality with respect to time. This is especially important in a community like Riviera where both influent and effluent vary considerably over the course of a year, with the summer conditions being drastically different from the winter.

Unless data on influent conditions are collected and effectively managed, it is difficult for plant operators to ascertain if the installed treatment process is suitable. Operators must be able to compare influent quality to other dates to help determine if the wastewater is typical for that time of year and the population. Accurate measurements of flowrate are also important. Significantly lower inflow compared to other periods of similar population might be indicative of a leak in the sewer system. In a coastal community such as Riviera, there exists the possibility that such a leak could remain undetected due to seepage into the ocean.

If accurate measurements of influent and effluent quality at various stages in the treatment process are also measured, it allows for the calculation of removal efficiencies of these stages. These efficiencies can then be compared to values in the literature to gain an estimate of how the plant performance rates with others of its type. Presenting efficiencies of the various plant stages over time is integral to determining if each of the stages is performing correctly. Unless component efficiencies are continuously plotted over time, it is difficult to gauge whether one of the stages is suffering from a loss in efficiency before it becomes painfully obvious, and after it has perhaps exacerbated the problem. If a problem in one of the stages can be detected and rectified at an early stage, it could save the plant huge expenditures.

At Riviera, one of the facultative lagoons is sometimes taken off-line during the winter months. Traditionally, the lagoon that has the lowest performance is removed from the treatment train and the flow routed to the other two lagoons. Continuous determination of the efficiency of each lagoon stage can help the plant staff quickly decide which of the lagoons need to be taken off-line, a process that currently relies on extensive repetitive testing.

Before a more effective data collection and management scheme for Riviera can be suggested, it is useful to examine the way in which other wastewater treatment plants perform these tasks. The facility at Attleboro, Massachusetts, is a privately run facility about which information was obtained. In addition, the data collection and management process at Deer Island, a government-operated facility that treats wastewater from metropolitan Boston, is also presented.

2.2 DATA COLLECTION & MANAGEMENT AT THE ATTLEBORO MUNICIPAL FACILITY

It is valuable to compare the sampling regime at Riviera with that at a private wastewater treatment facility in the United States, as it reflects the degree to which environmental regulation drives measurement practice. Two data logging processes at the Attleboro Water Pollution Control Facility are implemented to reflect climatic conditions, one during the warmer months and the other during the cooler seasons when organism growth is generally slower. In the period from May to October 1999, the plant flowrate is continuously monitored, and the free residual chlorine in the effluent is measured thrice daily from the results of grab samples. pH is measured daily from a grab sample while Carbonaceous Oxygen Demand (CBOD), TSS, fecal coliform, ammonia nitrogen and total phosphorus are measured three times each week from a 24-hour composite sample. In addition, total nitrogen is measured once a month. Toxicity tests such as the LC-50 and the C-NOEC are conducted six times a year from composite sampling. During the cooler seasons, when organism growth is not such a pressing factor, nutrient measurements are not conducted as frequently. Ammonia nitrogen is only measured twice a week, while total nitrogen and phosphorus are measured once a month. Fecal coliform is not measured.

Effluent standards are also specified for heavy metals content, and these are typically measured once a month from 24-hour composite sampling. The metals for which standards are specified are copper, zinc, chromium, silver, nickel, lead, aluminum, cadmium and cyanide. Monitoring results have to be summarized and presented to regulatory agencies once each month.

The Attleboro plant was using a DOS (non-graphical interface) system developed by Cochrane Associates for the collection of plant data and correlation with laboratory measurements. However, data management was recently upgraded to a Microsoft Windows based system with a graphical user interface. At present, the new system only contains plant data that was collected after the implementation of the Windows application. After plant instrumentation has been upgraded, the new system will have the ability to import data directly from sensors located around the plant. Currently, all data has to be entered manually. The new system allows much greater functionality and has the ability to perform trend and probability analysis and to automatically generate reports for plant personnel and environmental authorities. Security controls are also in place; the plant manager has complete control, laboratory technicians have data entry privileges and plant operators have data review permissions. Operators use generated reports to determine process needs while the managers use reports to track the overall process. (Wessel, 2000)

2.3 DATA COLLECTION & MANAGEMENT AT DEER ISLAND

For the sake of comparison, the measurement practices of the Massachusetts Water Resources Authority (MWRA), a governmental institution were studied. Information was obtained about the data collection and management scheme at Deer Island, which treats the wastewater for Metropolitan Boston. According to plant engineers, BOD, CBOD and TSS are measured daily. Correlations are not developed between BOD and CBOD; they are both measured directly. The values for water quality are measured for 24-hour aggregate samples. In addition, weekly measurements are performed on nitrogen and phosphorus components. Total Nitrogen, NO_3, and NO_2 are measured weekly, as are total phosphorus and PO_4. In addition, the water is measured for the presence of metals around six times a month. This sampling regime is very similar to that at Attleboro, suggesting that measurement practices are largely standardized over the private and public sectors, since the same environmental regulations must be met.

The flow data measurement is automated at Deer Island and fed directly into a computer. Together with the water quality data, the information is entered into an ORACLE database, which is a high-performance package designed for multi-user input. A detailed description of ORACLE capabilities and applications can be found at http://www.ORACLE.com. Since the administrative offices of the MWRA are not at the same location as the plant, a database application that can be accessed remotely is crucial to decision support and operation. This database is queried in different ways for two different functions; process control as well as compliance and monitoring. Monthly compliance reports are generated from automated queries followed by human inspection to check for outliers. Automated queries also generate daily reports for in-plant use. Plant operators have access to the entire database and use a proprietary package to view short-term data. For longer-term data retrieval, Process Book, a professional software package, is used. More information about the MWRA data logging and collection process can be found at http://www.mwra.state.ma.us/sewer/html/reg10.htm#a10_008.

2.4 DATA MANAGEMENT AT RIVIERA

The Riviera plant staff currently measure water quality four times a week. At two of these times, water quality at only the plant influent and effluent are measured. BOD and COD are the only water quality parameters that are measured, in contrast to the MWRA, where suspended solids, phosphorus and nitrogen are also measured. The other two measurement regimes are more stringent – BOD and COD are measured at the plant influent, the effluent of the anaerobic lagoon, the effluent for each of the facultative lagoons, and prior to discharge. However, all the measurements were done as grab samples and cannot be taken as a representative aggregate. In addition, flowrates at the plant influent, plant effluent and at the effluent of the drinking water plant are also measured on a daily basis, as is the rainfall for the day. A major drawback of the Riviera sampling procedure is that composite sampling is not performed. Composite sampling, usually taken from a mixture of samples collected over 24 hours, is essential to ensure that the results reported are actually representative of the Riviera wastewater and do not represent local fluctuations.

Data has been logged for the last two years, since 1997. The data set is more or less complete, though there are some gaps. In addition, it has been indicated by plant staff that some of the measurements may be suspect. The data has been stored in a single table in Microsoft Word, a document that spans 20 pages.

CONTROLE ANALÍTICO DO SISTEMA DE TRATAMENTO DE ESGOTO

DATA	ENTR. DA E.T.E.		SAIDA ANAERB.		SAÍDA LF-1		SAÍDA LF-2		SAÍDA LF-3		SAÍDA E.T.E.		REDUÇ	VAZÕES ADUZIDAS			CHUVA	OBSER-VAÇÕES
D/M/A	DBO	DQO	DBO	DQO	DBO	DQO	DBO	DQO	DBO	DQO	DBO	DQO	EFIC. %	ESGOTO	EFLUENTE	ÁGUA	IND. PLUV.	OBS.
24/12/97	358	1450	75	210	45	150	52	160			39	170	89,1	3607	1596	6431	4	
25/12/97	217	460	87	290	55	210	52	210			43	210	80,18	4845	3348	5225	4	
26/12/97														4760	5362	8241	4	
27/12/97	148	278									39	210	73,65	4984	7038	9676	3	
28/12/97														5450	7236	9677	22	
29/12/97	130	278									42	198	68,46	5358	8965	9253	11	
30/12/97	235	476	143	288	60	208	60	108			37	198	84,26	5907	6930	10842	10	
31/12/97														6774	8748	12929	4	
01/01/98	392	862	169	343	59	235	77	225			54	235	86,22	7189	6912	12063	3	
02/01/98														6741	9108	13083	0	
03/01/98	181	369									75	243	58,56	6625	8658	12339	0	
04/01/98														5513	7074	10094	5	
05/01/98	215	486									58	282	73,02	4344	5238	9073	4	
06/01/98	209	614	204	374	84	269	79	259			71	278	72,94	4471	5400	7929	24	
07/01/98														6241	6966	7383	64	
08/01/98	147	365	159	355	59	269	75	269			76	269	51,7	5226	1037	5511	27	
09/01/98														5433	1177	6705	0	
10/01/98	140	346									57	259	60	4920	8640	7247	0	
11/01/98														4472	6264	7036	0	
12/01/98	252	548									58	259	77	3968	4482	7290	2	
13/01/98	252	548	164	263	74	217	81	227			67	236	73,41	3802	3240	6744	0	
14/01/98														3975	4428	7895	0	
15/01/98	246	628	162	296	78	222	68	222			69	231	72	4659	5292	6565	25	
16/01/98														5107	6858	8093	0	
17/01/98	157	351									90	231	47,68	5284	6264	8269	0	
18/01/98														5266	6534	7779	2	
19/01/98	164	330									77	229	53,05	4125	4793	7058	0	
20/01/98	253	522	175	324	85	225	84	225			97	252	61,66	4146	4590	7163	0	
21/01/98														4645	4158	7531	0	
22/01/98	180	137	137	282	85	243	71	233			76	233	57,68	5252	4374	8539	0	
23/01/98														5173	4806	8361	0	
24/01/98	162	408									82	243	49,38	5093	5508	8756	0	
25/01/98														4670	7992	8804	0	
26/01/98	184	525									92	262	50	5060	5220	6584	0	

Figure 2-1 : Sample Page from Data Management Procedure At Riviera – Early 2000 (Riviera, 2000).

There was no way of graphically representing the data. In a document that size, it was virtually impossible to view trends over time. Only the overall efficiencies for the plant were calculated and tabulated, in some cases erroneously. There was no record of individual lagoon efficiencies. While it was a commonly accepted notion that the plant had to be upgraded due to overloading, there was at the time no obvious way of confirming this, since it was difficult to ascertain the spread of influent quality and quantity over time.

Because it was in a single Microsoft Word file, the data was extremely susceptible to corruption. For example, while entering data, all it took was a misplaced keystroke to change the data for another date. It is extremely likely that this has occurred on several occasions, since the original file had instances of letters being interspersed in the numbers. There were also duplicate values for a date in some instances. This made analysis of data almost impossible, since it was difficult to ascertain which of the two values (or what type of aggregate function) should be used in determining efficiencies. Furthermore, the date format was not consistent throughout the data set. The usual pattern was the dd/mm/yy format, but in some cases, dates were reported as mm/dd/yy.

As described, the plant at Deer Island avoids these and other complications through the use of data management applications. Since the Riviera data set was relatively small and data management packages are widely available for desktop computers, an application within this framework was devised.

3 SOFTWARE DEVELOPMENT

The laboratory at Riviera has a desktop computer on which the plant measurements are stored, but whose computational ability is under-utilized in evaluating plant performance. To streamline this process, a software application was written which manages plant data and performs analyses upon it. Although software development is relatively new, there is a considerable body of work devoted to it. An investigation of the basic principles behind this science is essential to delivering an effective software package within the desired time frame that has the potential to serve the needs of the Riviera plant staff. Only after an appropriate model is selected can implementation on a software platform compatible with Riviera hardware be implemented effectively.

3.1 PRINCIPLES OF SOFTWARE DEVELOPMENT

The first stage of almost any effective software development project is effective planning. This planning process takes into account the requirements of the target customer (the Riviera plant staff), the nature of the available data and the available time frame. Due to the information revolution, a large body of work exits on the software development process. Various models have been proposed which cater to different combinations of client, time, and complexity of data. The three most common classes of these models are the code-and-fix approach, the waterfall model, and the spiral model.

3.1.1 The Code-and-Fix Approach

This software development model relies on no planning. Instead, the programmer simply jumps headlong into the project with a vague idea as to what the final product should look like. Code and functionality are added according to the programmers whim, modified when possible, and discarded when not. In short, there is no systematic plan, code is simply added until the project is done or time runs out. The hope is that the former is true.

The advantage of this type of development model is that it allows the maximum time for actual program construction; there is no time spent on project visualization or systematic construction. It thus has the potential of allowing the programmer to demonstrate progress at an early stage. However, it relies largely on luck – the hope is that a major problem does not materialize late in the project schedule which then requires the developer to think of a way around the problem, or, more often than not, simply start over. This type of programming approach is most suited to small demonstration programs or other packages whose envisioned life is short. It is not recommended for projects of any size, which have to envision and serve the needs of a customer or which has a responsibility to support environmental compliance, or for software teams where the integration of different phases (often done by different programmers) becomes a crucial issue.

3.1.2 The Spiral Model

This type of software development scheme is at the other end of complexity, and is thus ideally suited towards large projects and software teams. The basic idea of this model of development is that it breaks the software project down into a series of mini-tasks that are combined to target areas of potential risk. Areas of potential risk include poor understanding of the requirements, problems with project integration and shortcomings of the hardware and software upon which the project will be implemented.

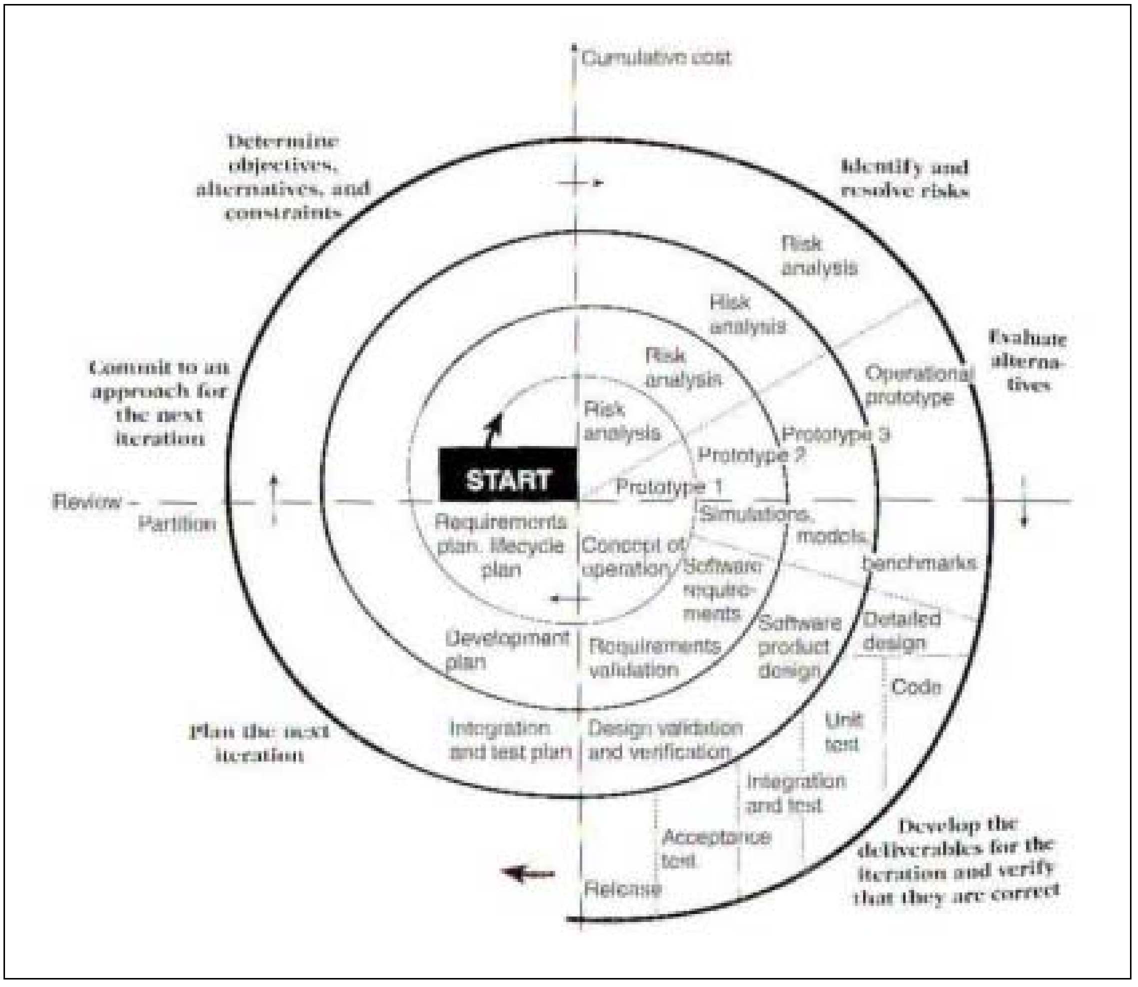

Figure 3-1 : Spiral Model of Development (McConnell 1996)

The goal of this model of development is a series of iterations around a spiral. Each iteration ends in a prototype, which can then be analyzed to determine if it is on track with customer requirements. Each iteration also accounts for time spent on risk analysis when the project team brings to light potential trouble areas and proposes solutions. A much smaller fraction of the time is actually spent on writing code than in the earlier model, and a much greater fraction ensuring that effort is not wasted and that a deliverable can be produced within the slated time.

The advantages of this development scheme are that it addresses areas of risk before they become obstacles to the completion of the project. Since the prototype is compared to project requirements at every iteration, it ensures that application development is proceeding in the intended direction. By ensuring that development proceeds in iterations, it allows for implementation of milestones that can aid in project development. This development approach is also suited to large software teams, as it ensures that integration of various stages of the project is continually considered at the end of each iteration. Thus, issues related to project integration are unlikely to manifest themselves as major problems as the final project deadline approaches.

However, this model of software development is the most intensive in terms of time spent in management. In instances where the envisioned project is small in terms of staff and objective, and where the risks are easily handled, it might provide an excessive degree of management that detracts from actual application construction. It is also complicated and relies heavily on competent management, the lack of which can negate the advantage of an individual programmer's skill.

3.1.3 The Waterfall Model

The waterfall model is in between the code-and-fix and spiral models in terms of complexity. Like the spiral model, it breaks up the development process into a series of small tasks, beginning with concept development and an understanding of the project requirements, and proceeding towards system testing en route to delivering the final product. Reviews are conducted to ensure that one stage of the project is completed before the next stage is started.

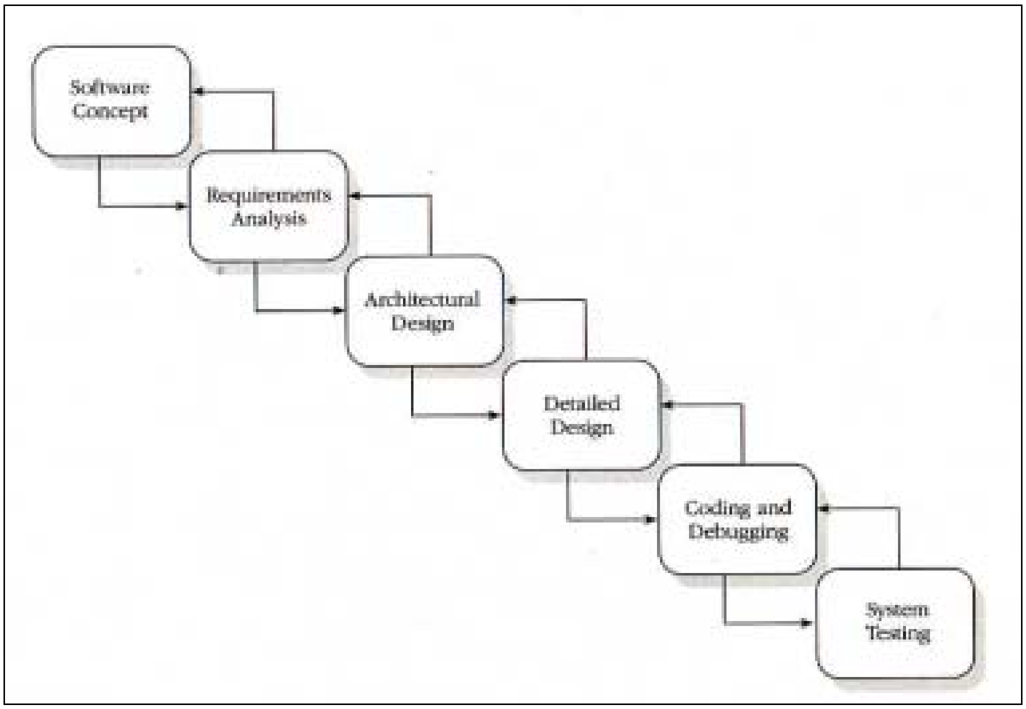

Figure 3-2 : The Waterfall Model of Development (McConnell 1996)

The waterfall model can be understood to be the spiral model condensed into a single iteration. It thus relies on a clear understanding of requirements at the very start of the project cycle, before any of the code has been written. If stages of the product cycle are handled by different development staff, problems may arise with stage integration. In addition, prototypes are not produced along the development process, and in the event that the customer wishes to see regular evidence of progress, the waterfall model can fail to satisfy. However, this type of software development model provides a level of discipline lacking in the code-and-fix approach without requiring the extensive management discipline and time-commitment of the spiral model.

3.1.4 Selected Method – The Sashimi Method

A modified version of the waterfall model, sometimes called the ‘sashimi’ model of development was chosen for the implementation of the Riviera application. This model is close to the ideal waterfall model, but allows for a greater degree of overlap among the various product development stages. Insights gained during the development project can be incorporated into the design strategy, an option that is not easily implemented in the pure waterfall model.

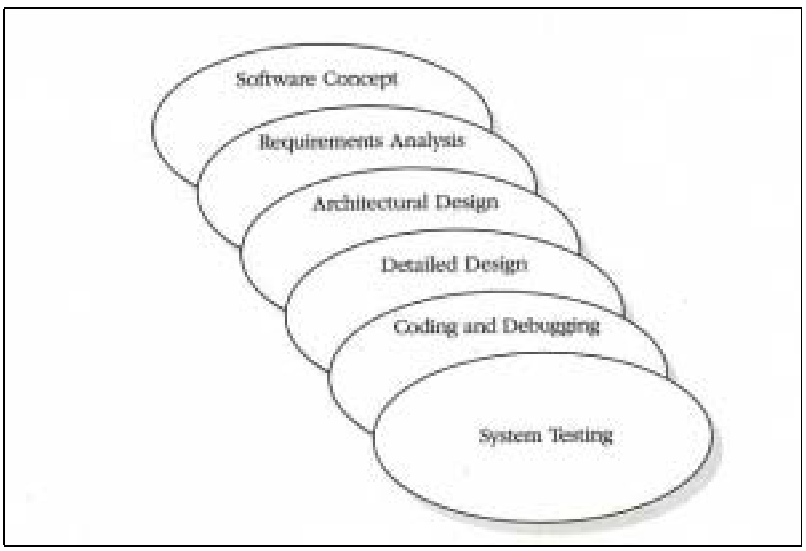

Figure 3-3 : The Sashimi Model of Software Development (McConnell 1996)

The sashimi model may suffer from problems of coordination if different teams are working on different aspects of the projects. Since there was only a single programmer for this project, the risk of coordination problems between the various stages was minimal. Although the project requirements were fairly well understood, room for maneuver was still desired, since it was conceivable that the requirements would change upon feedback from the Riviera staff. It was decided that the extensive time in management specified by the spiral model was not required since the project team was small enough to avoid the miscommunication mistakes that occur in software development projects. In addition, the size of the project was modest enough to avoid complications with system requirements and data integration. There was no pressure from the customer to produce evidence of progress and thus intermediate prototypes were not required. The code-and-fix approach was deemed inappropriate since it was hoped that the resultant application directly addressed the needs of the customers and was not a throwaway demonstration model. Furthermore, there was a fixed time frame within which the project had to be completed, and at least some degree of development discipline was deemed necessary to make sure that it had a good chance of success.

As shown in the diagram for the Sashimi model, the development process begins with a software concept. Once this has been conceived, the goals of the application are decided. From an understanding of the goals, a rough blueprint of the steps required to implement this vision is outlined. This is then refined to provide as good an understanding of what the project involves as possible. Only after there is an understanding of the entire project does coding commence. The final stage is system testing, closely related with the coding stage in the Sashimi model.

3.2 REQUIREMENTS ANALYSIS

The goal is to design and construct a data collection and logging schema that minimizes the possibility of data corruption and maximizes the potential for analysis. To ensure that the data painstakingly collected over the past two years is not discarded, any schema must be compatible with this data. The application must be able to effectively present the data collected by the Riviera staff over time, as this is currently the greatest limitation. This is usually best done in the form of a graph. It would also be valuable if this presentation was in a layout, which allows the plant staff to immediately see which aspect of the lagoon system the presented data is for. It should allow the plant staff to immediately select which aspect of the data they wish to view without a lengthy procedure.

The type of analyses that should be performed include calculating efficiencies for the various lagoons and presenting this information in such a way that trends over time can be easily recognized. Data that is potentially erroneous should be brought to the attention of the plant staff so that they can determine if their measurement procedures are flawed. Furthermore, the existing data logging process has no way of determining average values, a drawback since it does not allow the plant staff to determine if their lagoon performance is drastically below acceptable values. In addition, there is no way to determine if the values are typical for the time of year or system load.

The Riviera staff currently makes measurements of both BOD and COD. However, the procedure done in many other plants is to develop correlations between the level of BOD and COD in the plant at the different unit operations. If effective correlations can be developed, it might save the Riviera staff from having to perform five-day BOD tests on every sample. Instead, BOD would be derived from the measured COD (a test that takes only 3 hours) and the correlation. Direct measurement of BOD to assess the continuing validity of the correlation could suffice.

Although it is generally accepted that the flowrates increase during the summer months when the area population more than doubles, there has been no way of viewing this data or determining the extent of difference between the input to the plant or the plant effluent. There is also no way of viewing the difference between the influent to the wastewater treatment plant and the effluent from the drinking water plant, which is an estimate of the water consumed by the community at Riviera de Sao Laurenço.

3.3 Architectural Design

Rather than store all the data in a word processing application, which his poorly suited to data analysis and is susceptible to data corruption, a spreadsheet or database application is more viable. Spreadsheets are interfaces that are conducive to computation but whose data storage is not systematized. The priority of a database is efficient storage to facilitate retrieval. Of these two alternatives, a database package is more sensible since it has more options for data protection and recovery. While a spreadsheet might have greater facilities for numerical computation, this complexity is not required for the Riviera data and any potential benefit is more than offset by the possibility of data corruption inherent in a spreadsheet application. A database application is also more suited to selection and analysis of particular records. While a spreadsheet can perform the same calculations, doing so on the entire data set can be time consuming when only a subset of the data needs to be analyzed. Not only can analyzing the entire data set be time-consuming, but it can also be confusing and irrelevant, hiding true trends within a morass of data. Database applications that support SQL (Structured Query Language) are extremely well suited to the extraction and processing of data, especially if the analysis required is computationally simple. SQL is a language that forces the user to break down the retrieval and storage process into logically simple steps that enhance system efficiency.

Because the database application must be compatible with the Riviera system, Microsoft Access appeared to be the most sensible alternative, since it was already loaded onto the Riviera system and is compatible with the host of other Microsoft applications that are used by the staff. Whenever possible, there must be also a graphical user interface, so as to increase the accessibility of the model to all the plant employees. This must be done since a key disadvantage of databases is that they can simply overwhelm the viewer with data to the extent that any idea of trends is lost. The application must immediately make apparent to the user what its purpose is.

The application must make the viewer consciously recognize and choose to change existing data. This would involve at least a two-step procedure before existing data is changed, and make data corruption much harder than in the current word processor based data collection scheme. The application also must make evident to the viewer the steps that went into any analysis or calculations so that an informed decision can be made as to the veracity of the results. This includes the presentation of missing values, what values were excluded from the analyses and why. Formulas used in calculation must also be presented along with the results.

Any data that is evaluated should also be plotted. This allows the operator to decide for him or herself how much faith to place in aggregate values. This is especially important in areas such as mean values of influent BOD and COD, and the correlations conducted between them to develop the BOD/COD curves. Furthermore, any efficiencies and plots must be calculated on the fly from the most recent copy of the data set. If this is not the case, then the resulting analysis is not updated and thus not accurate. In addition, performing calculations and generating graphs rather than calling upon stored values ensures that valuable disk space is not utilized storing information that can be synthesized in a fraction of a second by a standard desktop processor. The only disk space that is taken up besides the disk space for the raw data set is that required to store the functions that the queries perform, not the actual results of the queries themselves.

3.4 DETAILED DESIGN

A four-tiered design approach was implemented in Microsoft Access. The first stage was Table Design, which involves an inspection of the available data and its organization in such a way that it is conducive to analysis. The second stage is Query Design, where the required analysis framework is devised and implemented upon the tables. The third stage is Form Design, where the results of the queries are presented in an effective manner. The last stage is the design and construction of an introductory screen where access to the diverse forms is integrated.

3.4.1 Table Design

Effective Table Design is of paramount importance in a database application. The performance of the most elegant queries would be compromised if the underlying tables in which the data are stored were not designed to facilitate the accessibility of records. All records in a table should be tagged by a Primary Key, which is a unique identifier for each record. A record is thus made up of the Primary Key and other fields which are called Non-key Fields. Ideally, the best tables are those that are normalized – as these have been demonstrated to be the most robust against corruption and to maximize the ease of data retrieval. Normalization can be defined as adherence to five successively stringent standards, called the Normal Forms (Kocur, 1999).

- 1st Normal Form: All records must be of the same length. All records in the same table must have the same number of columns.
- 2nd Normal Form: All Non-key fields must be a function of the Primary Key
- 3rd Normal Form: No Non-key field should be a function of another Non-key field. Only the Primary Key uniquely identifies all fields.
- 4th Normal Form: A row should not have more than one independent fact about each object.
- 5th Normal Form: A row in the table cannot be effectively reconstructed from several smaller tables.

Normal forms are sometimes violated for good reason, but these standards serve as a guide to efficient Table Design. In addition, there are various rules of thumb that generally serve the developer well. However, in this case, since all fields are essentially functions of the date and are not strict functions of each other, all the data can remain in one table, which makes Table Design a trivial task. The date serves as the Primary Key, since only one measurement is usually made each day and the date is the characteristic that intuitively identifies a particular record. Breaking up the data any further would simply cause a series of tables that have an explicit one to one relationship with the same Primary Key, the date. Tables with such a relationship should be merged into a single table.

Problems arise when multiple observations are reported for a single date. Not only does this cause problems with using the date as a Primary Key (which must be unique) but the issue arises as to which record to use for analysis. This problem could be avoided by including separate ‘observations’ columns where additional information could be stored. Thus, no measurements are lost, and it forces the operator to recognize that multiple values exist for a date and to decide which of the values are more suitable for use in analysis.

Field Name	Type
Date_Field	Date/Time – Primary Key
BOD – Plant Entrance	Number
COD – Plant Entrance	Number
BOD – CEPT Exit	Number
COD – CEPT Exit	Number
BOD – Anaerobic Exit	Number
COD – Anaerobic Exit	Number
BOD – Facultative Lagoon I Exit	Number
COD – Facultative Lagoon I Exit	Number
BOD – Facultative Lagoon II Exit	Number
COD – Facultative Lagoon II Exit	Number
BOD – Facultative Lagoon III Exit	Number
COD – Facultative Lagoon III Exit	Number
BOD – Plant Exit	Number
COD – Plant Exit	Number
Efficiency calculated by hand	Number
Flowrate – Entrance	Number
Flowrate – Exit	Number
Flowrate – Drinking Water Station	Number
Rainfall	Number
Observations	Text
Observations – II	Text

Table 3-1 : Structure of Riviera Plant Data Table

3.4.2 Query Design

Queries, which extract and analyze selected portions of the dataset have to be designed so they perform the required function without extracting more data than is needed, as this is detrimental to speed of operation. This is especially true when all forms are generated on the fly, and the queries are run each time the form is invoked. While this fully utilizes the processing power of the standard desktop and ensures that no extra disk space is used to store data, it places the burden of efficiency on the programmer.

In performing the calculations for efficiency, it makes sense to disregard those removal efficiencies that are greater than one or less than zero. Both of these usually imply erroneous measurements. However, the latter might indicate a serious flaw in the lagoon performance, since BOD or COD is actually being produced in the lagoon rather than being removed. Thus, while it makes sense not to include these in aggregate estimates such as the average, they should be presented in such a way that the operator sees when they occurred. If errors seem to be sustained over a period of time, it is indicative that the measurement regime is flawed and needs to be investigated.

In developing correlations between BOD and COD, it makes sense to disregard those measurements where the measured BOD is greater than the measured COD, since BOD is traditionally a fraction of COD. However, it is important to present those instances where BOD is greater than COD to the plant operators so that decisions can be made as to the efficacy of their measurement regime.

The first query that was written was done to ensure that there were not multiple sets of data for the same day. This would interfere with the normalization rules described above. Where these were detected, one set of the data was copied to the ‘observations’ section, as described in Table 3-1 instead of having a Primary Key indexed value. After duplicates were deleted, the date field could be set as the Primary Key.

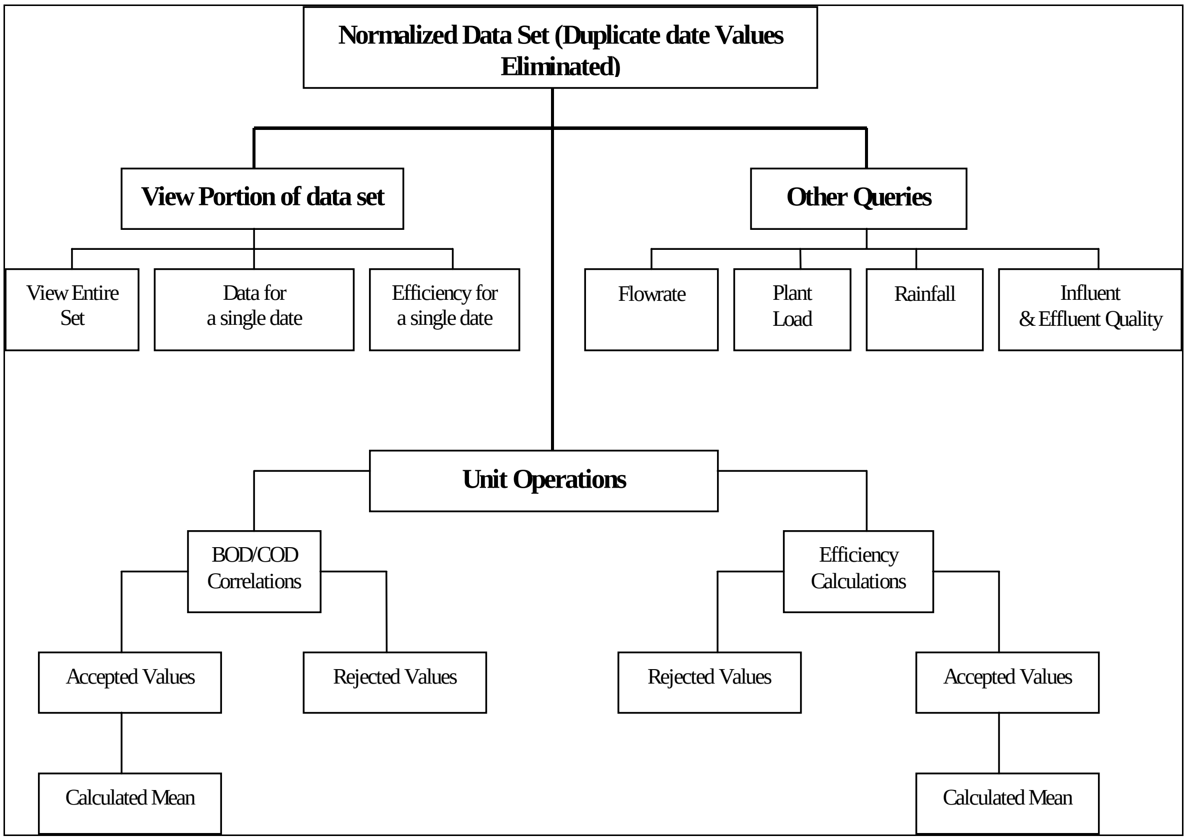

Figure 3-4 : Query Organization Structure

The other queries were written to evaluate a parameter for a particular unit operation. That is to say, queries that calculated the removal efficiency of the CEPT clarifiers, the anaerobic lagoon, each of the three facultative lagoons, and the total efficiency were written. Queries were written to extract the effluent quality of each of the unit operations, the BOD and COD removal efficiencies of each section, and the BOD/COD ratio at the effluent to each lagoon. Points were deemed to be valid if nonzero values were present for the influent and effluent concentrations of BOD and COD, and if the calculated efficiency fell between zero and one.

An efficiency less than zero indicates that there is more BOD or COD in the effluent of that particular unit than in the influent, and is usually indicative of an error in measurement. However, the possibility that lagoon performance is substandard to such a degree cannot just be ignored. This is especially true if no attempt is made to represent these instances to plant operators in the application, as in the case where data depicting a potentially serious situation (where the lagoon is adding rather than removing oxygen demand) is simply swept under the carpet. In addition, if these errors are sustained over time, it could also represent a flaw in the measurement regime for the plant, which should be corrected as soon as possible.

The implemented solution was to run separate queries for each of the unit operations extracting the values that were ignored in the analysis. These values were the cases where the calculated efficiency was either less than zero or greater than one. A similar query was written for BOD/COD correlation values – dates that yielded a value greater than one (which indicated that the BOD was greater than the COD) were pulled into a separate query so plant operators could see when they occurred.

Queries were also written to calculate the means of these values for each section of the plant. Mean efficiencies and BOD/COD ratios were calculated from the results of the query that calculated 'valid' efficiencies and BOD/COD ratios. Invalid entries (that resulted in a BOD/COD greater than one or efficiencies below zero or greater than one) were not used in calculating means. To help assess the impact of the new CEPT clarifiers on the performance of each of the unit operations, queries that calculated means were separated by date. January 21, 2000 was used as the date when CEPT became operational at the plant. This date was the first for which valid efficiencies were provided for the clarifiers. Thus, four sets of queries that extracted mean values were constructed for each unit operation; mean values for BOD and COD removal efficiency before CEPT, after CEPT, BOD/COD ratios at the effluent before and after CEPT.

Queries were also written for general data that is collected at the plant. This included extracting the flowrate entering the plant, leaving the plant, and measured at the drinking water station. The difference between the flow at the drinking station and at the entrance to the wastewater treatment plant can be taken as an approximation of water lost – either water that is collected but not discharged into the treatment system or that is lost in the sewerage system through leakage. Queries were also written to extract the rainfall measured at Riviera and to calculate the load to the treatment plant, the product of average influent concentration and flowrate.

In the case where only the values for a particular date had to be extracted – such as determining the efficiency through the lagoon system for a particular day, a macro was written that prompts the user for the date in question. A macro that serves as the date select clause was deemed to be more efficient than running the entire query and then selecting the relevant record from the result. In addition, the same macro – “select which date to use” - could be used for the function that presents the data for a particular day. However, two versions of the macro had to be constructed – one that allows for edits (used for the function that adds and edits data for a particular day) and one that was read-only, used for virtually everything else.

3.4.3 Form Design

<u>Data Entry/Edit/View Form</u>

The same template was used for both the data entry/edit form and the data view form. Once an efficient method was found, it was maintained since the same data were being represented. At attempt was made to organize the form in such a way that the fields pertaining to data most often collected were placed at the top of the form. Thus, the first set of fields prompt the user for the influent and effluent levels of BOD and COD for the plant, the measured flowrates, and the measured rainfall. These are the parameters for which measurements are conducted most often; flowrates and rainfall are recorded for each day and influent and effluent BOD and COD are collected four times a week.

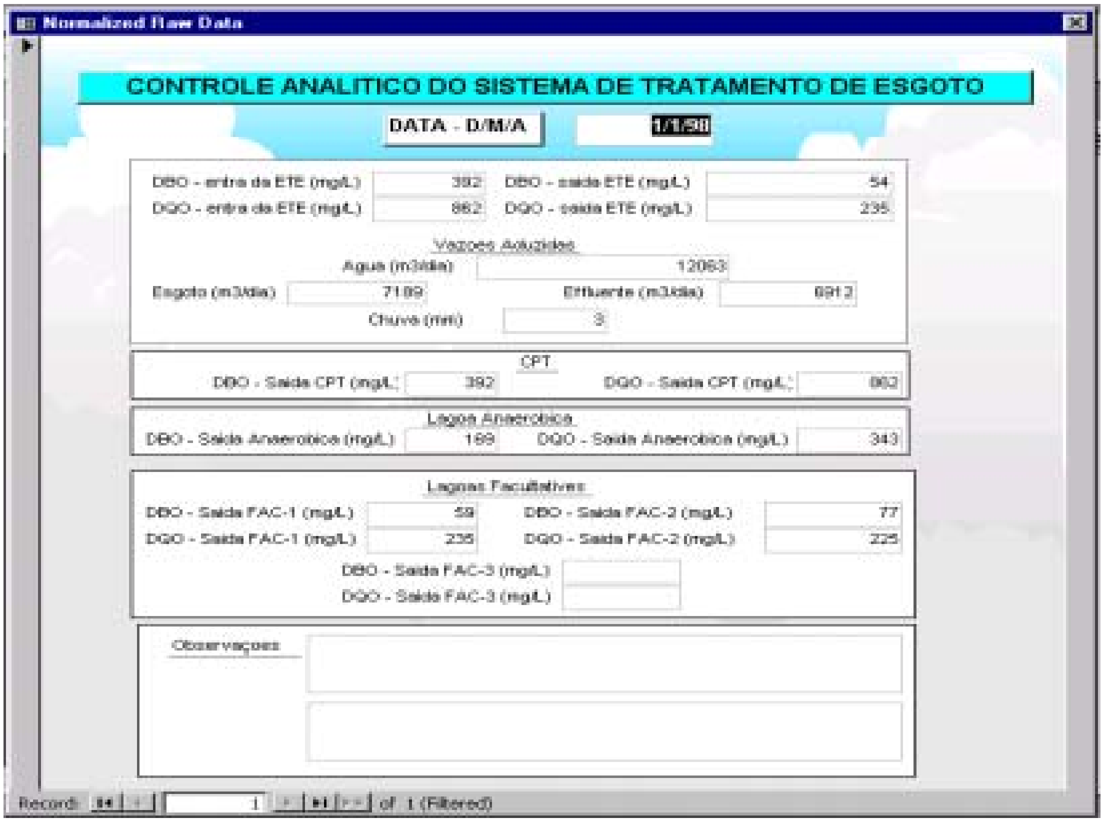

Figure 3-5 : Data Entry/Data View Form

Additional sections of the form largely followed the unit operations. Thus, the next section contains fields for entering the BOD and COD levels in the effluent of the CEPT clarifiers, followed by a section for the anaerobic lagoon, and then by one for the three facultative lagoons. A large section for observations brings up the rear. The observations section was made purposely large so that it could accommodate notations as to errors in observations or contain duplicate sets of measurement. Therefore, while only one set of data is deemed valid for a particular day, additional sets can be stored in the observations sections so as to ensure that no data collected is ever lost.

Both the data entry/edit form and the data-view form are invoked by a date macro. This macro prompts the user for the date under consideration, and then uses that as the clause to extract or insert records into the database. This allows the invocation of a select clause as quickly as possible, thus reducing the number of additional records that must be extracted. As described, two versions of the macro were written, one which allows editing of the database that it calls (for the data entry/edit form), and one which invokes the underlying table in a read-only mode (for the data view form).

Removal Efficiency for a particular day form

To assist plant operators in visualizing point efficiency of the lagoon system, a form was constructed that presents all the water quality measurements taken during the day along the lines of a plant schematic. It is important to note that the efficiencies are not lagged by the hydraulic retention time, and thus provide only rough estimates of how each unit operation is performing. Although this is an engineering shortcoming, it represents a rough value that plant operators can use to judge if a particular unit operation is performing far below specifications. Values were not lagged by the hydraulic retention time as this constantly fluctuates and the measurement regime does not lend itself to this procedure. Since measurements are not taken at each point every day, determining lagged efficiencies would involve interpolating the flow data between known values. Thus, the choice is either between admitting that the efficiency represented is not lagged or interpolating the data to a potentially erroneous value. The former approach was chosen.

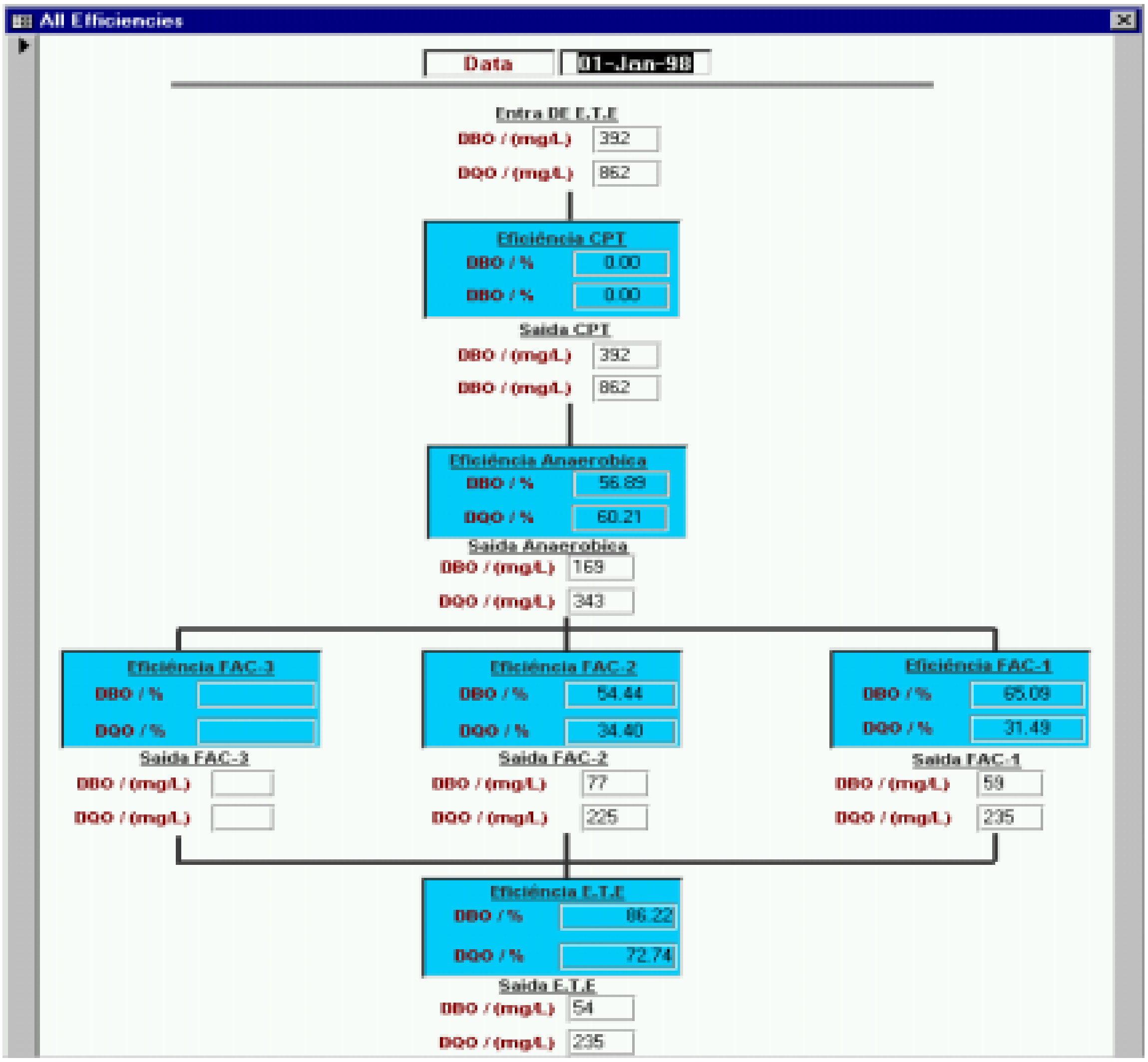

Figure 3-6 : Removal Efficiency for a Date Form

Entire Dataset Form

A read-only form was created that allows the user to view the entire underlying database. While this is not particularly useful in terms of analysis, it can be used to quickly scroll through many records and allows the user to determine the nature of the underlying database.

System Load Form

The primary motivation behind the installation of the CEPT clarifiers was concern for the efficiency of the wastewater treatment plant during the summer, when the loading is much higher than in the winter. However, in the previous measurement scheme, no method existed of quickly assessing this load. To address this issue, a form was created that graphically presented the results of the query that calculated system load, as shown in Figure 1-1. The graph generates shows monthly averages of system load of BOD and COD and clearly demonstrates the issue of concern – much higher values of load during the summer months.

Flowrate and Rainfall Forms

These forms graphically represent monthly averages of measured flows in the drinking water station, the entrance to the treatment plant, and at its exit and the measured rainfall at Riviera de Sao Laurenço. These forms do not perform any complex analysis, but are included for the sake of completeness. The flowrate form can also be used to alert plant operators to possible inaccuracies in their measurement apparatus; a sustained discrepancy between inflow and outflow becomes immediately evident, for example. In addition, if the discrepancy between the measured flowrate at the drinking water system and at the influent to the wastewater treatment plant grows alarmingly high, it can alert authorities to a possible leak in their sewerage collection system.

BOD/COD Correlation Forms

Since direct flow measurements were available for the flowrate into and out of the lagoons, the BOD/COD correlation forms at the entrance and the exit graphed the BOD/COD ratio as a function of both flow and time. This was to assist plant operators in deciding the degree of faith that they wished to place on the calculated mean value, and to better understand the nature of the system. However, preliminary results show that the BOD/COD mean ratio of 45% is largely independent of both flow and time, represented as a monthly average. At the bottom of the screen, the results of the query that selected for invalid values are displayed chronologically along with the mean value.

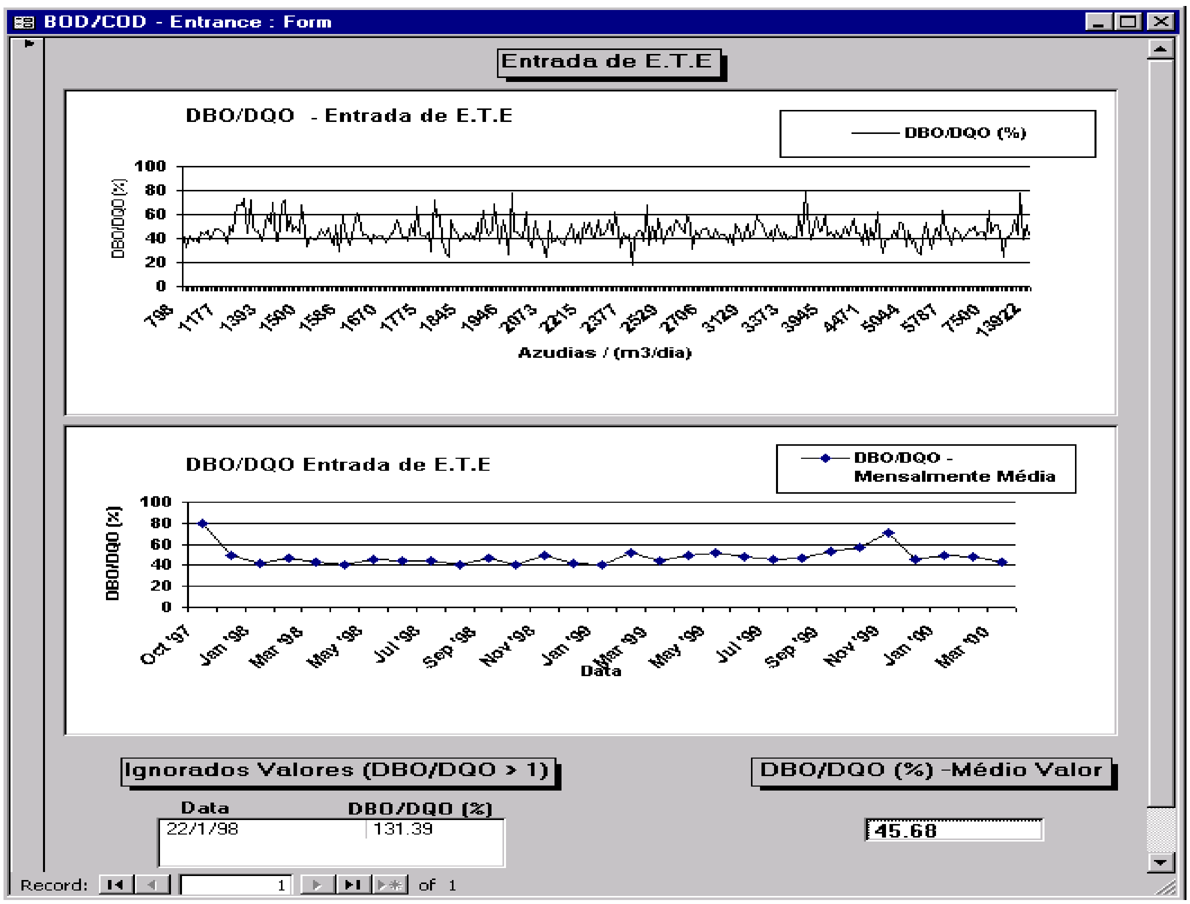

Figure 3-7 : BOD/COD for Influent Wastewater Form

In the case of the other unit operations, such as the CEPT clarifiers, and the lagoons, flow data was not available. Rather than interpolating flow data and potentially providing plant operators with erroneous information, BOD/COD ratios were simply plotted as a function of time. As before, monthly averages were calculated and graphed. In addition, the ignored values are shown at the bottom of the screen. The results of two of the mean queries are also shown, the average before and after the CEPT clarifiers were installed. Although CEPT has not been running long enough to tell with a reasonable degree of certainty, preliminary indications are that the BOD/COD ratio has risen after the installation of the clarifiers. Forms of the following type were written for the BOD/COD ratio at the exit of the CEPT clarifiers, at the exit of the anaerobic lagoon, and for the exit of each of the three facultative lagoons.

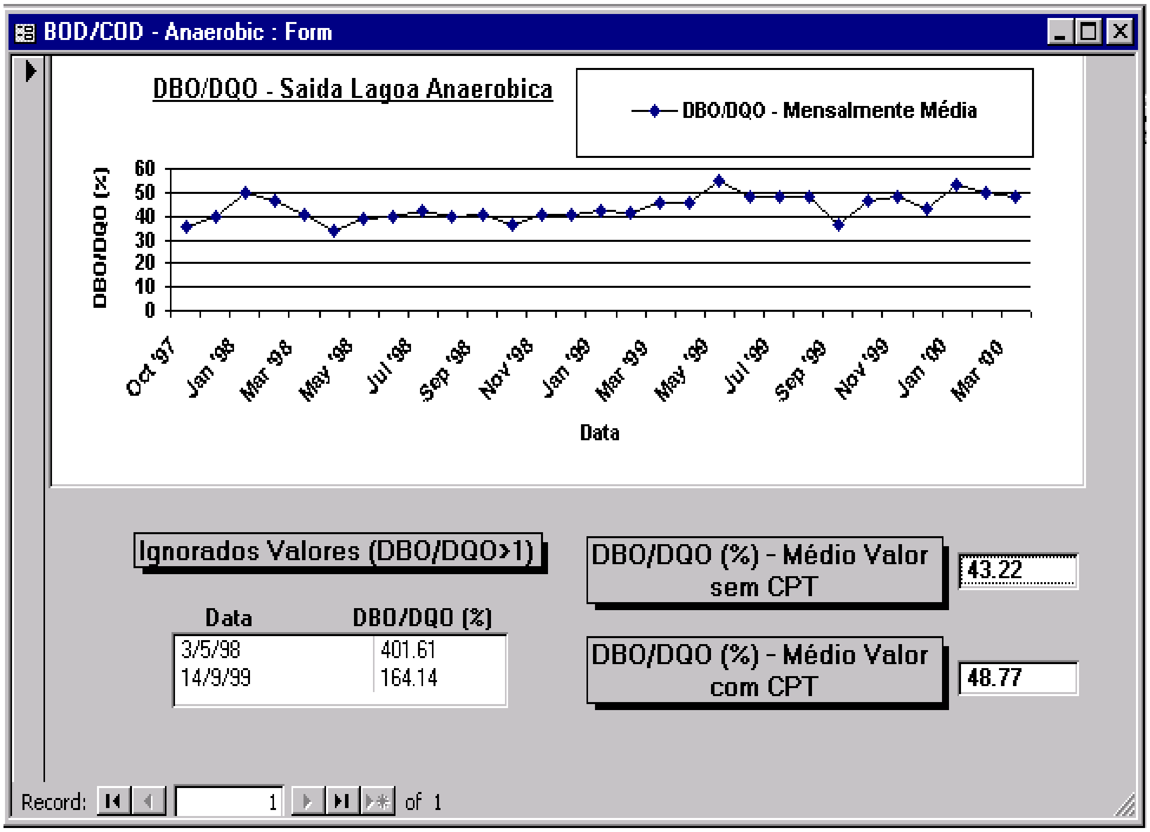

Figure 3-8 : BOD/COD Correlation for Anaerobic Effluent Form

Unit Operation Removal Efficiency Forms

A second set of forms that graphed the values for lagoon efficiency was also constructed. Forms were made for each of the unit operations. In addition, a form was constructed to illustrate the removal efficiency of the entire system. For each of the unit operations, the form contained two graphs, one which simply graphed the monthly averages of the effluent (BOD and COD values) and one that graphed the monthly averages of removal efficiency. As in the case of BOD/COD correlation forms, the values that were ignored in the analysis (in this case when the efficiency of removal was less than zero or greater than one) were placed in a scrollbox at the bottom of the screen. Also shown are mean values for the removal of BOD and COD before and after CEPT was installed.

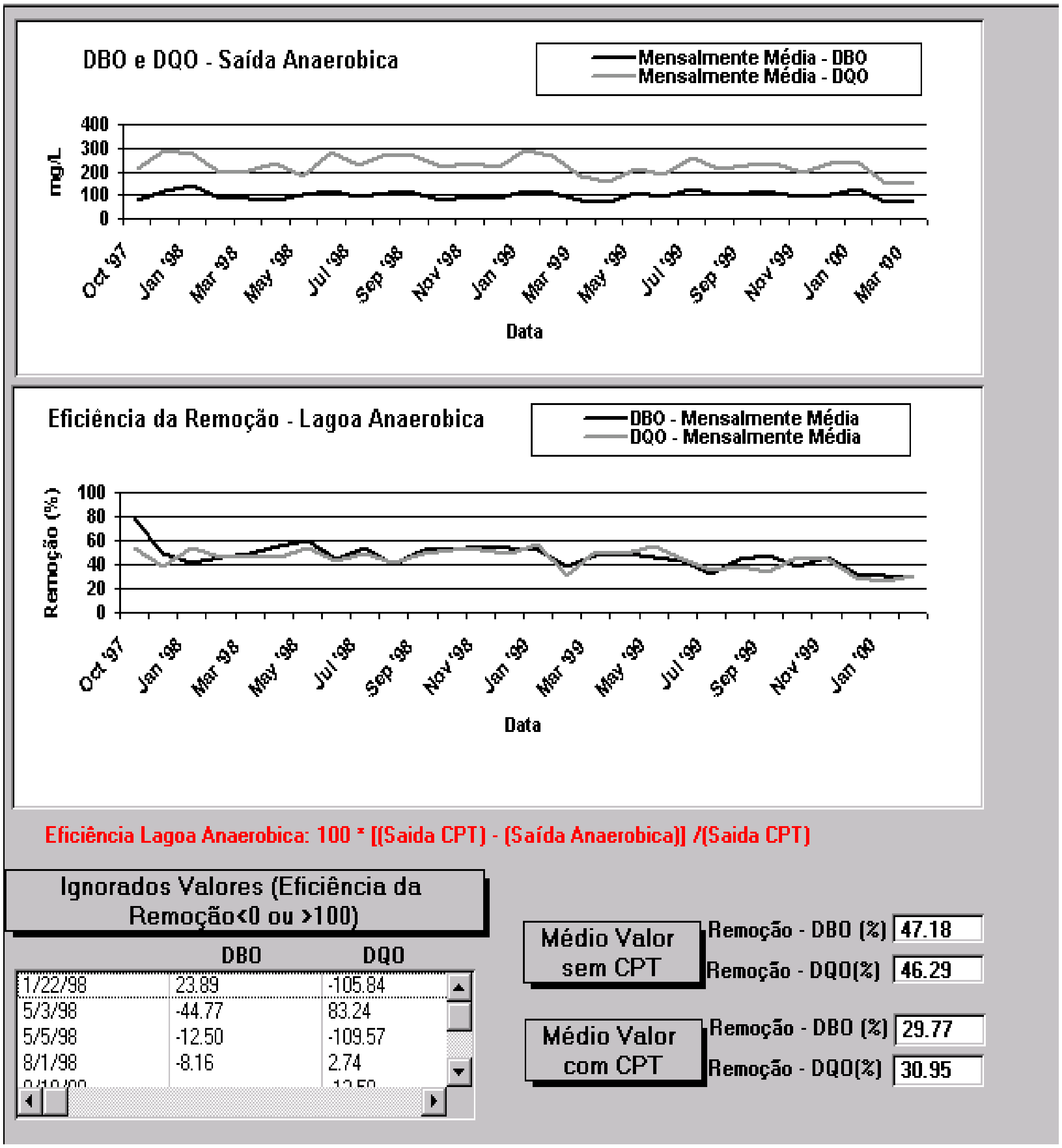

Figure 3-9 : Effluent Quality/Removal Efficiency for Anaerobic Lagoon Form

In the case of the total efficiency for the treatment plant, simply the monthly averages of efficiency were graphed, as shown in the introduction. Although CEPT has not been operation for a long period of time, the graphical indication is that the removal efficiency of both COD and BOD has increased. Although the mean values are not significantly higher than before CEPT was installed, it must be kept in mind that this higher removal efficiency has been achieved during a peak loading period, while the average value prior to CEPT installation also incorporates times when there was much lower loading.

Rather than place a graph of effluent quality along with overall removal efficiency, it was decided to be more effective if this information was put on the same form as a graph that showed influent quality. This would allow plant operators to compare at a glance the influent quality that the system had to handle, and the effluent characteristics that were met. In addition, this graph is probably the one that plant operators will find most useful in terms of submission to environmental authorities.

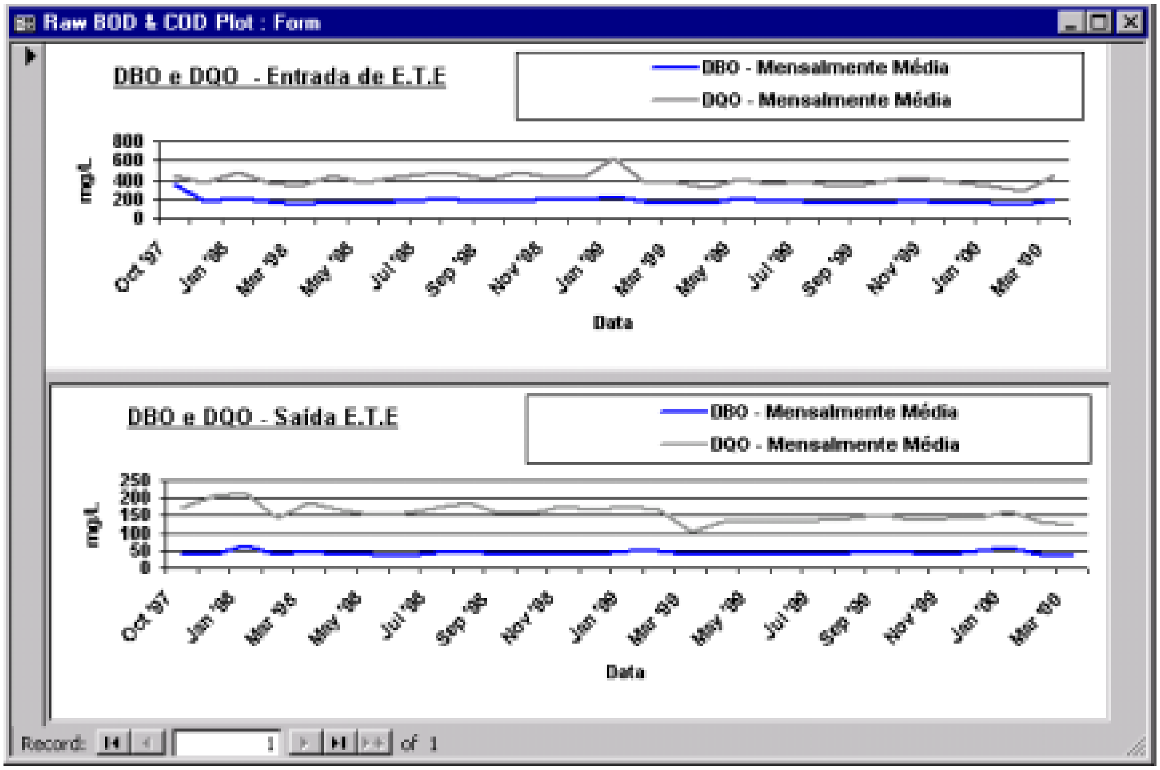

Figure 3-10 : Influent and Effluent Quality at Riviera Form

3.4.4 The Main Form

The main form is the introductory form that loads up automatically when the application is invoked and provides the link to all the other functions described above. The layout of the form was chosen to combine aesthetics and functionality. All the forms that correspond to specific unit operations are laid out in the form of a plant schematic so that operators can, at a glance, decide which aspect of the plant is being measured. Command buttons that do not correspond to any specific unit operations are placed on the left, with the exception of the system load and influent/effluent forms which were placed in the center of the form for prominence and symmetry. In addition, the buttons that invoke forms for data entry/edit and data view are separated so as to minimize the risk of invoking the edit form when only read-only functionality is desired. Two exit buttons in red are placed at the bottom of the form, one which exits from Microsoft Access and one which exits from the graphical user interface and allows manipulation of the underlying queries, table, and forms. Thus, to directly change the dataset, this function must be invoked and the database opened, a two step procedure that is designed to thwart accidental data corruption.

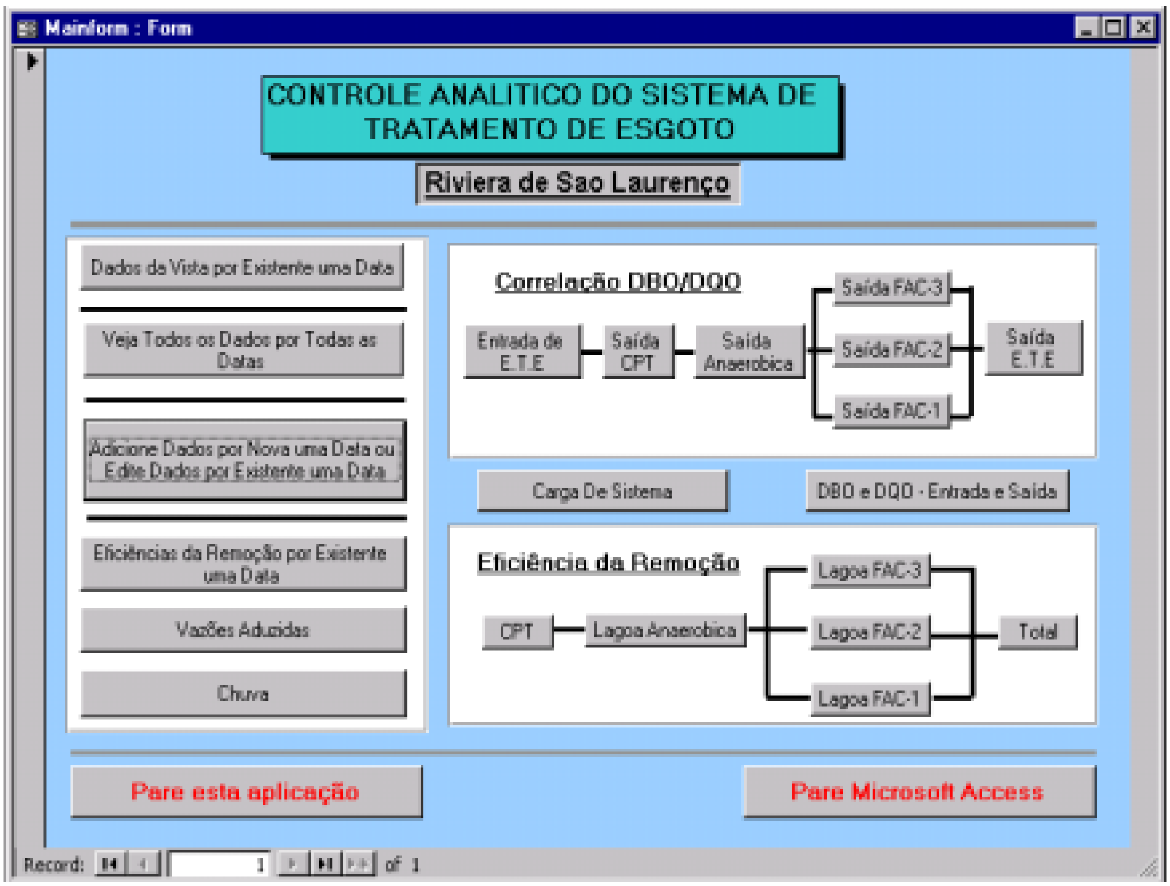

Figure 3-11 : Introductory Form

3.4.5 Coding and Debugging

A modular approach to coding and debugging was adopted, as it appeared to generate the greatest functionality in the shortest time. In other words, a form was created which would perform the intended function. Once the goal of the form had been decided upon, the underlying queries were constructed. This top-down approach was felt to avoid spending time writing extraneous queries that would not be utilized in the final application. Extraneous queries and forms are also a waste of system disk space.

Once functional forms had been designed and constructed, a main page was designed to contain and present the major functions of the application. The layout of the main form had to be chosen so as to group together functions of a similar nature. In addition, there had to be a respectable space on the screen between the “Edit Data” and “View Data” functions to decrease the possibility of data corruption. Thus the coding approach was top-down with respect to the forms and underlying queries, and bottom-up from the forms to the main front page.

Once functional forms had been allocated on the main page, macros were instituted to accelerate the speed of use and to decrease the need for user familiarity with the application. Macros that were written included those that prompted the user for a date to serve as input to the “view data” and “edit data” functions, as well as those that connected form-opening functions to command buttons on the mainform. This allows the user to specify which part of the lagoon was to be considered for analysis simply by selecting the relevant location on a schematic of a lagoon.

The debugging procedure mainly involved trying to think of procedures around certain annoying features of Access. One of them was not being able to change the axis value labels on a graph if the ordinate was an aggregate function – a monthly average in this case. Access' functions can usually be instituted by menus, but this one cannot. Instead, the raw SQL had to be edited in a separate window to make the change.

Another issue that had to be resolved was setting read/write permissions on some of the functions. A function was required that allowed the user to view the entire database as it was entered, but without allowing for data editing. However, a direct query on the database produces a result that can be changed. To resolve this issue, a dummy macro had to be inserted that simply opened the form, but allowed for a read-only toggle that would prevent the form being opened from inadvertent editing.

A special consideration of null values also became necessary. In calculating efficiencies, which require division, a null value in the divisor resulted in an undefined value, in many cases causing Access to crash. Special functions had to be inserted that would specifically exclude null columns and only use those data in analysis that resulted in arithmetically feasible values.

Finally, the inclusion of data from the CEPT clarifiers presented a unique challenge, since they were installed after the plant had been running for several years. This meant that there were additional columns of data (BOD and COD from the effluent of the clarifiers) that would have null values for the majority of the data set. In addition, this complicated the design of queries that calculated efficiencies for the anaerobic lagoon. Before the installation of CEPT, these are calculated from the raw influent and the effluent from the anaerobic lagoon. After the clarifier installation, efficiencies are calculated from the effluent from the CEPT system and the anaerobic lagoon. Thus, the same query cannot be used for the entire dataset. The problem was circumvented by setting the values for the effluent from the CEPT system prior to installation to the raw influent; a dummy value that allowed for smoother computation. This is also correct from an engineering standpoint, as in the absence of chemicals, the effluent from the CEPT system would be the same as the influent.

3.5 SYSTEM TESTING

System testing was integrating with coding and debugging, as befitting the Sashimi model. Since a modular approach was adopted to coding, once a module – consisting of a form with underlying queries was developed, it was tested. Testing involved both making sure that it integrated with the underlying data set and that it was sufficiently robust to prevent data corruption. Computational speed was also a factor – when two possible queries would have performed the same task, one that would rely upon selection of fewer records or performs fewer calculations to achieve the same result was selected.

System testing also involved determining the application response to erroneous data entry. This involved attempting to insert multiple records for the same date, which correctly resulted in an error message. Error messages were also received when trying to enter letters instead of numbers in the fields for water quality parameters or incorrect date values. These were tested to reduce the risk of typographical errors such as misplaced keystrokes where a letter was entered instead of a number.

All the forms that were intended to be read-only were tested to make sure that they did not allow for the possibility of data entry. As hoped, the forms that displayed the data did not register any of the user keystrokes. The only form that was writeenabled (the form for entering or editing values) was tested to make sure that changes were reflected in the underlying database.

Another test involved temporarily changing one of the values (such as influent BOD) that would be used in efficiency calculations and determining if generated values and graphs reflected this change. This was used to test whether graphs and averages were generated on the fly as designed, or used previously stored values.

Finally, all the generated graphs were inspected to make sure that the results they represented made sense. For example, the efficiency of the anaerobic lagoon or each of the facultative lagoons for a day could not be greater than the total efficiency. Monthly averages of rainfall could not be greater during the dry season than during the wet. Monthly averages of load applied to the wastewater treatment system could not reasonably be expected to be lower during the summer months when the area population more than doubles than during the winter.

3.6 NOTES ON TRANSLATION

The best data management application for the Riviera system would be rather useless unless plant operators can understand the system functions. It was thus necessary to translate all the command buttons and forms into Portuguese. An Internet translator provided by the Altavista portal was used for the translation. A summary of system development and a suggested method for changing functions was also provided to the Riviera staff. It is anticipated that the most important of these will be instructions for providing the application with knowledge as to when CEPT is operational and when it is off-line, so that mean values for efficiency and BOD/COD correlations with and without chemical treatment can be updated.

4 LAGOON LOAD AND MODELING ANALYSIS

4.1 LAGOON LOAD ANALYSIS

Analysis of the load placed on the lagoon system is an important step in determining if the system is performing according to specifications. Since lagoons are often designed upon the basis of the amount of load they are expected to receive, deviations from the design load are important to recognize. Anaerobic lagoons, which are characterized by a thin aerobic layer near the surface and the lack of dissolved oxygen at depth, are designed to operate under high organic loading conditions. Anaerobic systems generate less sludge than aerobic systems, but require a higher operating temperature to be effective. Lagoon load is usually specified on a per-area basis, which is the norm for clarifier-based systems since fractional removal based on settling velocity is a function of the surface area. A wide range of values is reported in the literature for optimal loading criteria of anaerobic lagoons.

Source	Optimal Depth /m	Surface Loading / (kg/ha.day)	Detention Time / day	BOD Removal / %
Metcalf & Eddy (1993)	2.5-5	225-560	20-50	50-85
WHO EMRO Technical Report No. 10 (1987)	2.5-5	> 1000	5	25-30
Lagoon Technology International (1992)	2-5	> 3000	1-2	25
World Bank Technical Paper No. 7 (1983)	4	4,000-16,000	2	27-30
Riviera 1st Pond- Off Season	3.2	670	9	47
Riviera 1st Pond-Peak Season (1999)	3.2	1150	6	47
Riviera 1st Pond- Peak Season (2000)	3.2	1380	3	21

Table 4-1 : Optimal Surface Loading Criteria for Anaerobic Ponds (Chagnon 1999) and Riviera Values

Upon inspection of Table 4-1, it becomes apparent that the Riviera first lagoon is at the lower range of acceptable loading. Indeed, with the exception of optimal loading reported by Metcalf & Eddy that is far lower than other literature values, the off-season loading at Riviera is insufficient for the pond to be considered anaerobic. Even during the peak period of 1999 when the loading doubles compared to the value from the off-season, the loading is insufficient to safely characterize the first lagoon as anaerobic. In the peak period of 2000, the surface loading rate is slightly higher than the peak value for 1999.

However, it is widely accepted that a design criteria based on surface area is not applicable to anaerobic lagoons, which should be designed for load on a per-volume basis (Chagnon, 1999). 100-400 g of BOD/ m^3.day is the optimal range that is reported in the literature (Metcalf and Eddy, 1993). However, by this criterion, the disparity between optimal loading and that observed for the first pond at Riviera is even greater. For the off-season, the load applied is only about 20 g of BOD/ m^3.day. During the peak period of 1999, the loading is about 36 g/ m^3.day. During the peak season of 2000, after the clarifiers had been installed, the load was 42 g /m^3.day. Both of these loading rates are still insufficient to characterize the pond as anaerobic. The installation of the CEPT clarifiers in the early part of 2000 has decreased the load to the anaerobic lagoon by a substantial amount. Data from the Peak period of 2000 show that if the clarifiers had not been present, the load to the anaerobic lagoon would have been about 80 g/ m^3.day, which would make the pond approach that of a true anaerobic system. Thus, by decreasing the load to the anaerobic pond, the CEPT clarifiers might cause an increase in aerobic activity in the system. The BOD removal efficiency of the first pond has markedly dropped after the installation of CEPT compared to 1999. An explanation might lie in the quality of the wastewater entering the first pond. During the peak season of 1999, the BOD level was about 200 mg/L, while for the peak period of 2000, the level was only about 120 mg/L, a direct consequence of BOD removal in the CEPT clarifiers. Higher loading was observed for 2000 since the flowrate was also much higher, 7500 m^3/day as compared to 3800 m^3/day during the peak of 1999. The nature of the wastewater with respect to BOD might explain why the removal efficiencies are much lower; dilute mixtures are by nature harder to treat. However, the possibility that the chemicals themselves modify microbial degradation cannot be ignored, and deserves further study. As mentioned, an inspection of figure 3-7 indicates that the BOD/COD ratio at the effluent of the anaerobic lagoon has increased after the installation of the clarifiers. Regardless, as demonstrated in the introduction, the installation of CEPT has resulted in a higher overall plant efficiency than that which is usually observed during peak periods.

To determine the nature of the first pond at Riviera, it is useful to compare the loading to the literature design criteria of facultative ponds. Facultative ponds are characterized by both aerobic and anaerobic activity; a significant aerobic layer exists at the top of the pond followed by an anaerobic layer at depth. At intermediate depths, both aerobic and anaerobic processes occur. An inspection of the Riviera values from table 4-1 shows that the surface loading and detention time is higher than the optimal design values. The first pond at Riviera could thus be called a facultative pond with a high degree of anaerobic activity rather than a purely anaerobic pond. However, during the off-season, the loading of 670 kg/ha.day closely approaches that of an ideal facultative pond as reported by the World Bank in 1983.

Source	Optimal Depth /m	Surface Loading / (kg/ha.day)	Detention Time / d	BOD Removal / %
Metcalf & Eddy (1993)	1.2-2.5	60-200	5-30	80-95
WHO EMRO Technical Report No. 10 (1987)	1.5-2	200-400	-	80
Lagoon Technology International (1992)	1-2	100-400	-	70-80
World Bank Technical Paper No. 7 (1983)	1-1.8	200-600	-	-
Riviera Facultative Ponds - Off Season	1.5	60	26	50
Riviera Facultative Ponds - Peak Season (1999)	1.5	90	17	50
Riviera Facultative Ponds - Peak Season (2000)	1.5	160	9	57

Table 4-2 : Optimal Surface Loading Criteria for Facultative Ponds (Chagnon 1999) and Riviera Values

However, loading analysis is insufficient to truly understand the underlying processes in a lagoon system. Modeling the lagoons can be a valuable analytical tool for this purpose, as it forces an explicit understanding of major underlying processes. In addition, loading analysis can only be done after the fact and is incapable of predicting lagoon efficiency from knowledge of the influent. In the following section, the principles of modeling are outlined, previous work on lagoon modeling reviewed and a simple model for the Riviera peak season of 2000 presented.

4.2 INTRODUCTION TO MODELING

Modeling is a powerful analytical tool as it forces an understanding of the major forces at play in a natural system. Since the infinite processes at work in a natural system cannot all be included in a human simulation, it challenges the modeler to identify those processes that are truly major and those that can be neglected. Therefore, an effective model must be computationally simple enough to predict the action of a natural system within a reasonable period of time but complex enough to include all the major processes. The value of an effective model is not restricted to an analytical tool, however. Reliable models can serve as generators of estimated values in the absence of measured values in the natural system. This can result in large savings when actual measurement of the system is expensive and cumbersome. This is especially true in those systems where it is difficult to conduct measurements without altering the system of interest.

In wastewater treatment engineering, a model that can predict effluent quality given information about the influent is invaluable. Rather than having to deal with the ramifications of unacceptable effluent quality after the situation manifests itself, an effective model can alert plant operators to implement corrective action. Modifying plant-operating parameters can then avoid a situation that will be costly to correct later and potentially detrimental to public health. In addition, modeling can be an extremely cost-effective method to investigate the effects of changing plant-operating parameters. Rather than the construction of expensive pilot plants or the implementation of full-scale tests for the entire range of possible alterations, a model can be used as a cheap filter to limit these procedures to those that appear to show the most promise.

4.3 *PREVIOUS WORK IN LAGOON MODELING*

4.3.1 The Ferrara Model

A large body of excellent work has already been done on modeling of wastewater treatment lagoons, much of it from the Civil and Environmental Engineering Department at MIT. In 1978, Raymond Ferrara wrote his doctoral thesis on the processes underlying waste stabilization lagoons. In addition to an investigation into the design criteria behind waste stabilization ponds, he also presented a comprehensive model to describe their underlying processes.

The Ferrara model considered both hydraulic and biogeochemical influences. He considered the hydraulic effects of baffling and presented conclusions on how such influences could be applied to systems at steady state. In addition, he modeled the element cycles of carbon, nitrogen and phosphorus in a lagoon system as well as a method for predicting the concentrations of fecal coliform. The parameters under consideration in the model were inorganic carbon, organic carbon, phosphorus, nitrogen and fecal coliform. The factors that were assumed to determine the interplay between these various parameters included mineralization, organism growth, net loss by settling, atmospheric re-aeration, and death of fecal coliform. All of these factors were assumed to proceed along first-order kinetics. For example, the mineralization of organic compounds was assumed to be directly proportional to organic matter concentration, as was the organism growth. The loss of fecal coliform was assumed to be proportional to its concentration, and the atmospheric re-aeration of carbon dioxide was assumed to be proportional to the difference between saturation and actual concentration.

Raymond Ferrara calibrated and tested his model in stabilization pond systems in Corrine, Utah and Kilmichael, Mississippi. He presented rate constants along with corrections for temperature (Ferrara, 1978).

4.3.2 The Chagnon Model

In 1999, Frederic Chagnon at MIT's Department of Civil and Environmental Engineering modified the Ferrara model and applied it to the lagoon system at Riviera prior to the installation of the CEPT system and at an aerated lagoon system in As-Samra, Jordan. Confronted with lack of data about phosphorus and nitrogen at Riviera (since levels of these nutrients are not monitored), Chagnon reduced the system to three governing equations, which concerned the interplay between the organic and inorganic forms of carbon and the rate of change of fecal coliform levels. He also used meteorological data from Santos, a city about 50 miles distant from Riviera to estimate lagoon influent temperatures. Furthermore, he conducted sensitivity analyses to determine how effluent predictions were affected by estimates of pond temperature and influent levels of inorganic carbon.

The data set that Chagnon used is the same that was collected by the Riviera plant staff and which is described in Chapter 2. Levels of COD can be used to calculate the concentration of organic carbon, but no measurements are made of either inorganic carbon or fecal coliform. Chagnon conducted a sensitivity analysis that demonstrated that the influent level of inorganic carbon did not particularly influence the rate constants. Riviera does not measure levels of fecal coliform, and this lack of data made it impossible to ascertain the success of a model that predicted effluent levels. In addition, a constant influent temperature of 25°C was assumed. In general, the fit between the observed and modeled series was fairly good. A 4th order Runge-Kutta numerical approximation was applied to the governing differential equations in Visual Basic to derive rate constants, which generally compared favorably to those derived by Raymond Ferrara in 1978 (Chagnon, 1999).

4.4 MODIFICATIONS TO THE FERRARA & CHAGNON MODELS

4.4.1 Governing Equations

Since Brasil does not specify effluent requirements on fecal coliform, levels are not measured at Riviera. The lack of data on fecal coliform levels renders any attempt at modeling its level in the effluent useless, as there is no way to check a predicted series against an observed series.   An attempt was made to measure phosphorus levels in the effluent and effluent during January 2000 but the testing apparatus proved inappropriate to this task. For this reason, since no data was available, the model was further reduced to two governing equations concerning the carbon cycle.

$$\frac{d(OC)}{dt} = \frac{Q_i}{V}(OC)_i - \frac{Q}{V}(OC) - R_{12}(OC) - R_{1S}(OC) + R_{21}\left[\frac{IC}{K_{SC} + IC}\right](OC) \qquad \textbf{(4-1)}$$

$$\frac{d(IC)}{dt} = \frac{Q_i}{V}(IC)_i - \frac{Q}{V}(IC) + R_{12}(OC) - R_{20}(CO_{2,sat} - CO_2) - R_{21}\left[\frac{IC}{K_{SC} + IC}\right](OC) \qquad \textbf{(4-2)}$$

Symbol Used	Definition
OC	Organic Carbon
IC	Inorganic Carbon
Q	Flowrate
V	Lagoon Volume
K_{SC}	Half-Saturation Constant for Carbon
R_{12}	Rate Constant for Transformation from OC to IC
R_{21}	Rate Constant for Transformation from IC to OC
R_{1S}	Net Loss Rate of OC
R_{20}	Rate Constant for Atmospheric Re-aeration
$CO_{2,sat}$	Saturation Concentration of Carbon Dioxide
$_I$	Denotes influent

Table 4-3 : Legend for Governing Equations

Since the rate constants are highly sensitive to temperature, their value is reported at a reference, usually 20 degrees C. However, when used for predicting effluent quality, their value must be corrected to the temperature of the lagoon system. The following corrections were used.

$$R_{8S} = R_{8S,20}(1.05^{(T-20)}) \tag{4-3}$$

$$R_{12} = R_{12,20}(1.036^{(T-20)}) \tag{4-4}$$

$$R_{21} = R_{21,20}(1.066^{(T-20)}) \tag{4-5}$$

$$R_{1S} = R_{1S,20}(1.1^{(T-20)}) \tag{4-6}$$

Symbol Used	Definition
R_{xx}	Rate Constant at Lagoon Temperature
$R_{xx,20}$	Rate Constant at 20^{o} C.
T	Lagoon Temperature in Celsius

Table 4-4 : Legend For Temperature Correction Equation

4.4.2 Inputs to Model

Rather than simply extend the pre-CEPT data set that Chagnon used, only the post-CEPT data set for Carnival, 2000 was modeled. There were two reasons for this. The first was that the data collected during the five-day Carnival period (March 3 – March 7) was of much higher quality than the general data set. The second was that a comparison of the rate constants during Carnival, after the installation of CEPT, might grant insight into the wider effects of chemical treatment upon a lagoon system.

An inspection of the presented governing equations shows that carbon content (calculated from COD), temperature and flowrates are the inputs to the model. COD measurements in the general data are grab samples, and thus carry no guarantee that they are actually representative of the average concentration in the wastewater. The Carnival data, on the other hand are the results of 24-hour composite sampling. The general data set did not contain any measurements of water temperature, which is an important input into the model. During the period of Carnival, six measurements were taken each day at three different depths (top, middle and bottom) for each of the lagoons. The average temperature for each day served as the model input. Flowrate measurements for the general data set are made once every day by a one-time visual inspection at the Parshall flume. Therefore, there is little guarantee that the flowrate reported is the average for the day. On the other hand, flowrates were measured hourly during the Carnival season. A daily average of these measurements served as the model input for flowrate. A smaller data set of higher quality was thus used in place of a data set which contained many more points but which was less reliable.

Day	OC_i / (mg/L)	OC/ mg/L	T / (^oC)	Inflow (L/s)	Outflow (L/s)
3 March	92.88	62.92	31.05	39.73	38.32
4 March	96.63	70.41	28.23	76.69	73.41
5 March	81.65	67.04	28.03	94.74	92.17
6 March	88.01	73.41	27.65	100.85	99.07
7 March	80.52	76.40	28.39	120.86	125.94

Table 4-5 : Inputs to Anaerobic Lagoon Model

Day	OC_i / (mg/L)	OC / mg/L	T / (^oC)	Inflow (L/s)	Outflow (L/s)
3 March	62.92	49.81	26.99	38.32	29.66
4 March	70.41	48.31	27.21	73.41	53.33
5 March	67.04	44.57	27.2	92.17	76.37
6 March	73.41	51.31	27.6	99.07	88.14
7 March	76.40	49.44	26.61	125.94	157.16

Table 4-6 : Inputs to Facultative Lagoon Model

Values for inorganic carbon concentration and carbon dioxide levels in the water were not available and were assumed to be 0.5 mg/L and 0.2 mg/L, respectively.

4.4.3 Application of the Model

As in the Chagnon model, a Runge-Kutta 4th order numerical approximation to the given differential governing equations was applied. However, the Chagnon method presented rate constants derived from the application of local minima analysis. That is, one of the rate constants was varied until the closest agreement between the observed and modeled series was found. Then the value that gave this closest agreement was preserved and another constant varied until the error was minimized. However, the danger exists that these are local minima and not global minima.

The advantage of using fewer data points was that an attempt could be made to isolate global minima. Rate constants are varied along a four-dimensional axis (one dimension for each rate constant under consideration) and the set that has the lowest error presented. The model was written in MATLAB and used four successive nested loops to simulate the four-dimensional axis.

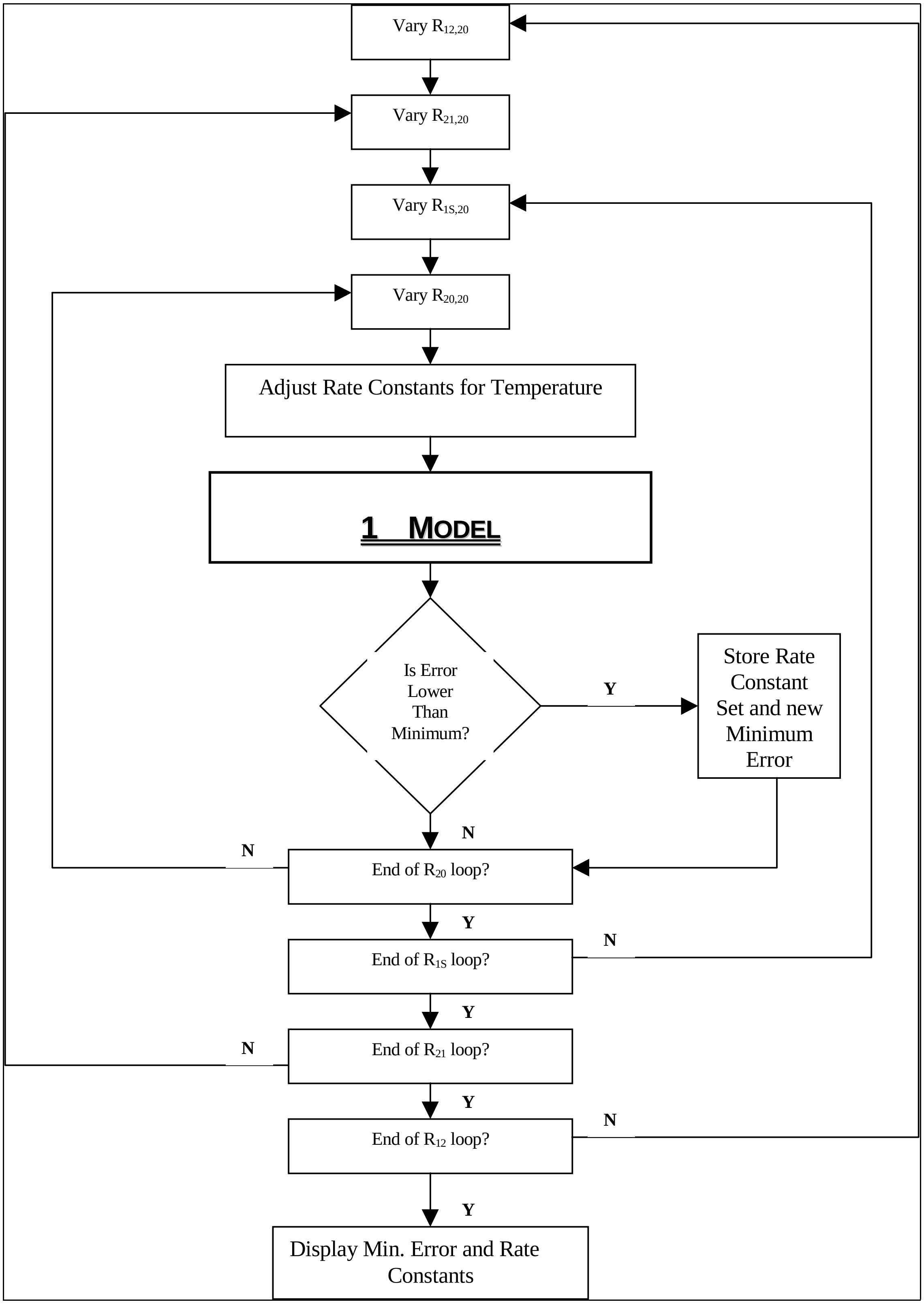

Figure 4-1 : Procedure for Evaluating Global Minima

The criterion of error that was chosen was the square root of the square of the difference between the observed and modeled series.

$$Error = \sqrt{(OC_{model} - OC_{observed})^2} \qquad \textbf{(4-7)}$$

4.4.4 Lagoon Modeling Results

Two sets of data were modeled for the anaerobic lagoon during Carnival, 2000. In the first set, the model was lagged by the hydraulic retention time of the anaerobic lagoon, averaged at two days. In the second, the effluent was not lagged. This was done to gain a greater understanding both of the robustness of the model and whether water quality at Riviera varied to such an extent that an extremely accurate value for hydraulic retention time was needed. If this were the case, it would have resulted in a dirth of data for the facultative lagoon, as the retention time would ensure that no matching influent-effluent data pairs would remain. Fortunately, the agreement between the rate constants with and without a lag for hydraulic retention time were surprisingly close, suggesting that the quality of wastewater for the time period under analysis was fairly consistent.

	Ferrara (1981)		**Chagnon (1999)**		**2000**		
	Kilmichael (1st Fac. Pond)	Corrine (1st Fac. Pond)	Anaerobic (all year)	Anaerobic (Carnival)	Anaerobic (lagged)	Anaerobic (unlagged)	Facultative
R12	0.05	0.05	0.05	0.11	0.01	0.015	0.11
R21	0.04	0.085	0.02	0.02	0.02	0.02	0.085
R1S	0.02	0.02	0.04	0.04	0.04	0.04	0.02
R20	8.6	8.6	8.6	8.6	10.2	10.2	6.4

Table 4-7 : Evaluated Rate Constants at 20^{o}C

The measured and modeled series were also plotted to gain a visual indication of how well the model performed. In general, the fit was rather good for both the anaerobic and facultative ponds.

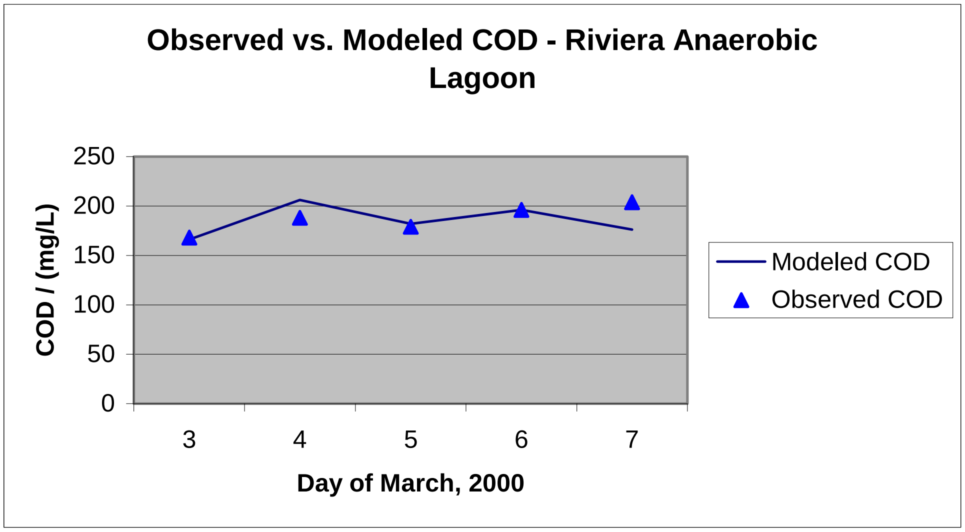

Figure 4-2 : Observed vs. Modeled COD - Anaerobic Pond (Unlagged)

As mentioned, the facultative pond model was not lagged by the hydraulic retention time. However, a comparison of lagged and unlagged models for the anaerobic lagoon indicated that the quality of the wastewater during Carnival was sufficiently constant to apply an unlagged model for the facultative pond.

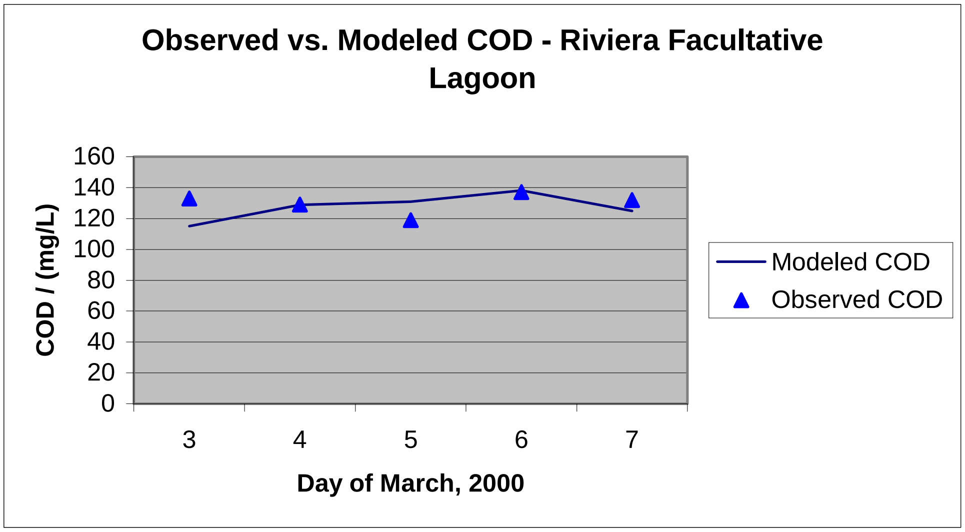

Figure 4-3 : Observed vs. Modeled COD - Facultative Pond

Unfortunately, the paucity of data points used for modeling makes it difficult to decide how much faith to place on the presented rate constants. It is comforting that they are generally in agreement with those that were derived with a much larger data set by Ferrara and Chagnon. However, the one parameter which varies considerably from previously reported values is that for R_{12}, the rate constant for the conversion of organic carbon to inorganic forms. The rate constants for the anaerobic pond are far lower than those reported previously by Chagnon, even for periods of comparable wastewater type and loading. However, before the modified rate constants can be accepted as those typical of the anaerobic pond following CEPT, the modeling analysis must be repeated with many more data points.

Perhaps the altered rate coefficients are a consequence of the CEPT clarifiers. The increased removal of organic matter before the wastewater reaches the anaerobic lagoon might lower the rate of its conversion to inorganic carbon. However, this implies that chemical precipitation has a greater affinity for organic carbon since the opposing rate constant R_{21} remains virtually unchanged. However, it could also be argued that R_{21} is dependent to a higher degree on organism concentration (which are required to transform carbon to its organic form), which might not be significantly greater than Carnival, 1999.

The rest of the rate constants compare extremely favorably with those reported by Chagnon and Ferrara. With the exception of R_{12}, which has been discussed, the other rate constants are in close agreement with those derived by Chagnon for the Riviera anaerobic pond in 1999. Furthermore, the calculated values of R_{21} and R_{1S} for the facultative pond at Riviera are virtually identical to those reported by Ferrara for the Corrine facultative pond in 1978 and are an indication of the robustness of his model.

5 Conclusions and Recommendations

A data management scheme was designed for the wastewater treatment plant at Riviera de Sao Laurenço, Brasil. This scheme was implemented in the form of an application for Microsoft Access, a widespread database package that is already installed on the Riviera laboratory computer. Although the system upgrade of Chemically Enhanced Primary Treatment was installed too recently to produce a wealth of data, all indications are that it is dramatically improving plant performance. Overall efficiency of the treatment system has increased, all the more remarkable since this was accomplished during peak-season loading. In addition, the load to the biological lagoon system has decreased. A possible consequence of this might be an increase of aerobic activity in the first pond. A model for the lagoon system during the Carnival period of 2000 was applied and first-order rate constants for the carbon cycle presented. In general, the presented constants are in excellent agreement with previous work. The one factor that drastically differed was the rate constant for the conversion of organic carbon to inorganic carbon, which was far lower than previous investigators had reported. It is possible that this is a consequence of the chemical precipitation process anterior to the biological system. However, since the modeling results were obtained with very few data points, the model must be re-applied with more data before the lower rate constant can be convincingly accepted. The real benefit of the presented model for 2000 might be the procedure for evaluating global minima, which can be equally applied to data from any period. Perhaps, in future, computational software packages that explicitly contain optimization algorithms could be employed to further accelerate the process of evaluating these rate constants.

The collection of composite water quality data along with accurate measurements of the lagoon temperature would also go a long way in advancing knowledge of the effect of chemical treatment on subsequent biological processes. The model for the Riviera lagoons should also be extended to include the nitrogen and phosphorus cycles, for which there is currently no data. The inclusion of nitrogen and phosphorus monitoring in the Riviera regime would not only be a boon to modeling efforts but would expand the facility's reputation for ecological goodness. Although there are currently no governmental pressures to maintain levels of phosphorus and nitrogen in the effluent, these nutrients can cause extensive ecological problems, and concern for the environment rather than regulatory compliance should drive these efforts. Furthermore, the logical next step after the expansion of the measurement regime would be a combined data management system and model that would categorically demonstrate that effective water treatment and the benefits of information technology do not have to be restricted to the First World.

6 REFERENCES

1. Altavista Babelfish Translator : Internet. http://www.altavista.babelfish.com/

2. Attleboro Water Pollution Control Facility, NPDES Permit No. MA0100595, 1 November, 2000

3. Chagnon, F., "Design and Dynamic Modeling of Waste Stabilization Ponds." Department of Civil and Environmental Engineering, Massachusetts Institute of Technology. June 2000.

4. Ferrara, R.A., Harleman, D.R.F., "A Dynamic Nutrient Cycle Model for Waste Stabilization Ponds." Department of Civil Engineering, Massachusetts Institute of Technology. December 1978

5. Ferrara, R. A., Harleman, D.R.F., "Hydraulic Modeling for Waste Stabilization Ponds." Journal of Environmental Engineering Division, ASCE, Vol. 106, No. EE1, Proc. Paper 1470, August 1981, pp. 817-829.

6. Hall, M. Personal Communication, 2 May, 2000.

7. Harleman, D.R.F. and Murcott, S, "The Role of Physical-Chemical Wastewater Treatment in the Mega-Cities of the Developing World", Water Science Technology, Vol. 40, No 4-5, pp.75-80.

8. Kocur, G. "Entity-Relationship Diagrams and Data Normalization: Lecture 5 – Information Technology and Computer Models in Logistics," Department of Civil and Environmental Engineering, Massachusetts Institute of Technology. November 1999.

9. McConnell, S., Rapid Development : Taming Wild Software Schedules, Microsoft Press, Redmond 1996.

10. Morrissey, S., and Harleman, D.R.F., "Retrofitting Conventional Primary Treatment Plants for Chemically Enhanced Primary Treatment in the USA", Chemical Water and Wastewater Treatment II. Proceedings of the Fifth Gothenburg Symposium, R. Klute and H. Hahn (eds). Springer-Verlag, Berlin. Pp.401-416.

11. MWRA Monitoring and Effluent Guidelines. Internet http://www.mwra.state.ma.us/sewer/html/reg10.htm#a10_008

12. Oracle Corporation: Internet

http://www.oracle.com

13. Pincince, A. Personal Communication, 27 April 2000.

14. Prague, C.N., and Irwin, M.R., Access 97 Bible, IDG Books, New York 1997.

15. Riviera de Sao Laurenço Plant Records, 2000.

16. Wessel, Mark. Personal Communication. 8 May, 2000.