Invited Review

Models for the optimization of regional wastewater treatment systems

João Joanaz de Melo and António S. Câmara

Environmental Systems Analysis Group, College of Sciences and Technology,
New University of Lisbon, 2825 Monte da Caparica, Portugal

Received August 1993

Abstract: The problem of the optimization of regional wastewater systems may be generally formulated as follows: to define the transport and treatment system, in a region or water basin, which assure compliance with given pollution control criteria, with minimum cost. In addition, one may try to satisfy other objectives, such as minimum environmental impact, better effluent reuse or adequate phasing. From the optimization point of view, the two main problems that render the solution difficult are the dimensionality and the concavity of cost functions. The matter has been dealt with by many authors, who have produced varied techniques to try to solve this problem. This paper begins with a brief review of the work of those authors who have produced models specifically designed to study the problem. Then, solution strategies are discussed concerning three major items: definition of the objective function and constraints, optimization method and practical applicability of the models. The paper concludes with the discussion of topics for future research.

Keywords: Optimization; Regional wastewater treatment systems; Facility location

1. Introduction

The planning of regional wastewater treatment systems is a classic optimization problem, which may be generally formulated as follows: to define the characteristics of the treatment and transport system, in a region or water basin, which assure compliance with given pollution control criteria, with minimum economic cost.

In addition, one may try to satisfy other goals, which render the problem multiobjective:
- To minimize the environmental impact;
- To maximize system reliability;
- To maximize system flexibility under uncertain conditions;
- To assure equity among users of the system;
- To maximize benefits from reuse of treated effluent.

The solution of the problem consists on the identification of a system composed by several treatment plants, each one treating effluents from one or more polluting sources. This solution should include the location, size and operating standards of treatment plants, as well as the layout of the necessary transport systems.

From the conceptual point of view, the optimization of regional wastewater treatment systems presents the following major difficulties:
- Nearly all objectives are difficult to quantify and even to define accurately;
- The number of potential solutions grows exponentially with problem size, creating the need to use computerized optimization techniques;
- Cost functions (which make up the nuclear objective function) are strongly non linear and concave, seriously limiting the application of most common optimization methods.

On the other hand, practical application of optimization models has two additional problems. First, such a model should be compatible with existing institutional water resources management procedures. Second, environmental engineering projects are usually designed, and decisions made, by engineers and politicians not familiar, and indeed suspicious, of mathematical modelling; which is the reason why such models are seldom used in common practice.

2. Review of proposed models

Many models have been devised to solve the problem of optimization of regional wastewater treatment systems (see Table 1). Critical descriptions of most of those approaches were produced by GUGENHEIM (1979), MANDL (1981), LEIGHTON (1982), CÂMARA (1985) and MELO (1992), among others. In order to provide a historical backgroung and basis for the following discussion, those models will be briefly reviewed.

DEININGER (1965) studied the optimal distribution of discharges to assure pre-determined water quality levels, using linear programming. LIEBMAN and LYNN (1966), and SHIH (1970), used dynamic programming to identify the optimal distribution of discharges along a river. LOUCKS et al. (1967) suggested the application of linear programming to the same problem.

AKFIRAT and DEININGER (1966) seem to have been the first to propose the optimization of regional wastewater treatment systems allowing for transport between sources. CONVERSE (1972) used dynamic programming to relate the number of treatment plants with the extension of transport systems, with given treatment levels. The problem is formulated as a transhipment model with a linear structure, a formulation that would be used by many authors thereafter. The author reports a case-study in the bay of the Merrimack River, USA.

GRAVES et al. (1972) proposed a non linear formulation which allows for local treatment, joint treatment and effluent transfer, respecting pre-determined water quality levels. The identification of global optimum is not guaranteed.

WANIELISTA and BAUER (1972) defined the problem as a transhipment model. They used cost functions approximated to linear segments. The solution is found with a mixed integer programming algorithm. The model is used to study the expansion of the sewer system of Little Econlockahatchee River, USA.

DEININGER and SU (1973) defined the regional treatment system as a network. They used concave cost functions and linear restrictions, applying a convex programming method to identify extreme points in the solution space. The extreme point hierachization method of Murphy is used to find an optimal solution. No application to a large real case is reported.

HAHN et al. (1973) proposed a branch and bound method which performs an implicit enumeration of the alternatives. At each step, the configuration found so far is examined to check how interesting it is, eliminating inferior alternatives and reducing the dimension of the tree. Practical experience is not reported in detail.

MCCONAGHA and CONVERSE (1973) presented a heuristic algorithm that proceeds by successive iterations, starting from one initial configuration and testing, for each node, if another treatment location is more interesting than current solution. The solution depends on the starting point.

JOERES et al. (1974) formulated a mixed integer programming model, with cost functions approximated to linear segments. Treatment levels are pre-defined. The model is applied to a case-study in Dane County, Wisconsin, USA.
<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>YEAR</th>
<th>CRITERIA</th>
<th>SOLUTION TECHNIQUE</th>
<th>OTHER OBJECTIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEININGER</td>
<td>1965</td>
<td>distribution</td>
<td>linear programming</td>
<td></td>
</tr>
<tr>
<td>LIEBMAN and LYNN</td>
<td>1966</td>
<td>distribution</td>
<td>dynamic programming</td>
<td></td>
</tr>
<tr>
<td>LOUCKS et al.</td>
<td>1967</td>
<td>distribution</td>
<td>linear programming</td>
<td></td>
</tr>
<tr>
<td>SHIH</td>
<td>1970</td>
<td>distribution</td>
<td>dynamic programming</td>
<td></td>
</tr>
<tr>
<td>CONVERSE</td>
<td>1972</td>
<td>emissions</td>
<td>dynamic programming</td>
<td></td>
</tr>
<tr>
<td>GRAVES et al.</td>
<td>1972</td>
<td>water quality</td>
<td>non linear programming</td>
<td></td>
</tr>
<tr>
<td>WANIELISTA and BAUER</td>
<td>1972</td>
<td>emissions</td>
<td>mixed integer programming</td>
<td></td>
</tr>
<tr>
<td>DEININGER and SU</td>
<td>1973</td>
<td>emissions</td>
<td>convex programming</td>
<td></td>
</tr>
<tr>
<td>HAHN et al.</td>
<td>1973</td>
<td>emissions</td>
<td>int.progr.+branch &amp; bound</td>
<td></td>
</tr>
<tr>
<td>MCCONAGHA and CONVERSE</td>
<td>1973</td>
<td>emissions</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>WHITLATCH</td>
<td>1973</td>
<td>emissions</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>JOERES et al.</td>
<td>1974</td>
<td>emissions</td>
<td>mixed integer programming</td>
<td></td>
</tr>
<tr>
<td>ROSSMAN and LIEBMAN</td>
<td>1974</td>
<td>water quality</td>
<td>dynamic programming</td>
<td></td>
</tr>
<tr>
<td>ECKER</td>
<td>1975</td>
<td>distribution</td>
<td>geometric programming</td>
<td></td>
</tr>
<tr>
<td>KLEMETSON</td>
<td>1975</td>
<td>emissions</td>
<td>dynamic programming</td>
<td></td>
</tr>
<tr>
<td>LAURIA</td>
<td>1975</td>
<td>emissions</td>
<td>mixed integer programming</td>
<td></td>
</tr>
<tr>
<td>MCNAMARA</td>
<td>1976</td>
<td>emissions</td>
<td>geometric programming</td>
<td></td>
</tr>
<tr>
<td>WEETE and BELARDI</td>
<td>1976</td>
<td>emissions</td>
<td>heuristics</td>
<td>cost uncertainty</td>
</tr>
<tr>
<td>WHITLATCH and REVELLE</td>
<td>1976</td>
<td>emissions</td>
<td>dyn.prog. + lin.prog. + heur.</td>
<td>water quality</td>
</tr>
<tr>
<td>BAYER</td>
<td>1977</td>
<td>emissions</td>
<td>non linear programming</td>
<td></td>
</tr>
<tr>
<td>CHIANG and LAURIA</td>
<td>1977</td>
<td>emissions</td>
<td>heuristics</td>
<td>phasing</td>
</tr>
<tr>
<td>LASHKAR et al.</td>
<td>1977</td>
<td>emissions</td>
<td>gener. gradient reductin</td>
<td></td>
</tr>
<tr>
<td>BRILL and NAKAMURA</td>
<td>1978</td>
<td>emissions</td>
<td>int.prog. + branch&amp;bound</td>
<td></td>
</tr>
<tr>
<td>JARVIS et al.</td>
<td>1978</td>
<td>emissions</td>
<td>network algorithm + int.prog.</td>
<td></td>
</tr>
<tr>
<td>LOHANI and THANH</td>
<td>1978</td>
<td>distribution</td>
<td>lin.prog. with stochastic cons.</td>
<td></td>
</tr>
<tr>
<td>ROSSMAN</td>
<td>1978</td>
<td>emissions</td>
<td>heuristics + dynamic prog.</td>
<td>phasing</td>
</tr>
<tr>
<td>GUGENHEIM</td>
<td>1979</td>
<td>emissions</td>
<td>lin.prog. (adap. simplex)</td>
<td></td>
</tr>
<tr>
<td>PINGRY and SHAFTEL</td>
<td>1979</td>
<td>water quality</td>
<td>heuristics + gradient red.</td>
<td></td>
</tr>
<tr>
<td>NAKAMURA et al.</td>
<td>1981</td>
<td>emissions</td>
<td>mixed integer programming</td>
<td>groundwater recharge</td>
</tr>
<tr>
<td>LEIGHTON</td>
<td>1982</td>
<td>emissions</td>
<td>mixed integer programming</td>
<td></td>
</tr>
<tr>
<td>PHILLIPS et al.</td>
<td>1982</td>
<td>emissions</td>
<td>mixed integer programming</td>
<td></td>
</tr>
<tr>
<td>SMEERS and TYTECA</td>
<td>1982</td>
<td>water quality</td>
<td>shortest path+non linear prog.</td>
<td></td>
</tr>
<tr>
<td>KANSAKAR and POLPRASERT</td>
<td>1983</td>
<td>water quality</td>
<td>lin.prog.+objective prog.</td>
<td>land use impact</td>
</tr>
<tr>
<td>KITABATAKE and MIYAKI</td>
<td>1983</td>
<td>water quality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEIGHTON and SHOEMAKER</td>
<td>1984</td>
<td>emissions</td>
<td>mixed integer programming</td>
<td>groundwater recharge</td>
</tr>
<tr>
<td>CAMARA</td>
<td>1985</td>
<td>emissions</td>
<td>heuristics + shortest path</td>
<td>multiple solutions</td>
</tr>
<tr>
<td>KLEMETSON and GRENNEY</td>
<td>1985</td>
<td>emissions</td>
<td>dynamic programming</td>
<td>phasing</td>
</tr>
<tr>
<td>ONG and ADAMS</td>
<td>1987</td>
<td>water quality</td>
<td>poliedron random search</td>
<td></td>
</tr>
<tr>
<td>ZHU and REVELLE</td>
<td>1988</td>
<td>emissions</td>
<td>linear programming</td>
<td></td>
</tr>
<tr>
<td>JOSHI and MODAK</td>
<td>1989</td>
<td>distribution</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>MELO</td>
<td>1992</td>
<td>emissions</td>
<td>cluster analysis heuristics</td>
<td>multiple solutions</td>
</tr>
</tbody>
</table>
ROSSMAN and LIEBMAN (1974) developed a dynamic programming model with water quality constraints. The model is designed for essentially linear systems. It is applied to the estuary of Delaware River, USA, and is able to identify better solutions than current approach.

ECKER (1975) proposed a geometric programming model to define optimum allocation of polluting charge in a river, to minimize the cost while respecting preset levels of dissolved oxygen.

KLEMETSON (1975) used dynamic programming to select, among pre-determined potential configurations of the treatment system, optimal solutions along time and construction phasing.

LAURIA (1975) used mixed integer programming to solve the problem of optimal effluent transfer in the regional system. The programming model is used as part of an iterative process. Linear cost functions were adopted, and the author discussed several ways to proceed to the linearization of the functions.

MCNAMARA (1976) proposed geometric programming to identify optimal solutions of a regional treatment system, given the treatment levels.

WEETER and BELARDI (1976) applied a revised version of the algorithm of McCONAGHA and CONVERSE (1973) to two regions of Pennsylvania, USA. Costs are modeled with a probabilistic function, using Monte Carlo technique to evaluate the probability of a given alternative being the most advantageous.

WHITLATCH and REVELLE (1976) proposed a model based in the work by WHITLATCH (1973). The model combines heuristic concepts with optimization algorithms, allowing the integration of user expertise in the solution procedure. The model studies systems with a linear structure. In a first stage, the model deals with the problem of configuration of the regional system given the level of pollution abatement. In a second stage, water quality constraints are incorporated. The models provides tools to help the optimization of each sub-problem. No global optimum is guaranteed.

BAYER (1977) proposed a non linear programming method to identify the optimal solution for the regional treatment. However, it is too demanding in computer time for problems of large dimension.

CHIANG and LAURIA (1977) proposed a heuristic algorithm with the peculiarity of performing intertemporal comparisons, in order to assure pre-defined treatment levels. A regional basic plan, over a number of time horizons, is used as starting point. First, the model checks, for each period, the advantage of substituting one treatment plant by a pipeline to another plant, producing the optimal configuration for that period. Then, each plant with expansions over time is analysed, to verify if the building of a larger plant, sooner, would be less expensive.

LASHKARI et al. (1977) presented a model to study the optimization of effluent treatment in the basin of Lake Utah, USA. The goal is to minimize treatment cost with treatment of combined sewage (from different sources) respecting the standard for effluent concentration. The solution is found with the generalized reduced gradient method. The model does not take into account possible scale economies in joint treatment, nor is it formulated as a general regional treatment system optimization.

BRILL and NAKAMURA (1978) proposed an integer programming method based on a branch and bound technique. Possible links between source and regional treatment plants, and treatment levels, are given data of the problem. Cost functions are approximated by linear segments. During the partition of the tree, feasible solutions are generated by inspection. The method is more directed to generate alternatives than to assure solution optimality.

JARVIS et al. (1978) defined the problem as a flow network with fixed charges. Concave cost functions are approximated by linear segments. A penalty system based on group theory is used to facilitate the solution of the problem by mixed integer programming. The model was applied to a small basin in Jefferson County, Kentucky, USA.
LOHANI and THANH (1978) proposed a model to allocate polluting charges, using linear programming as solution method. Uncertainty factors are incorporated by using stochastic constraints.

ROSSMAN (1978) proposed a model to study phased planning of regional treatment systems, with pre-defined treatment levels. The model is formulated as a network of "sites" (polluting sources, treatment plants and/or passage nodes). The algorithm is divided in three stages: first, with heuristic criteria and dynamic programming, a starting configuration is defined; second, the model determines the construction phasing that optimizes the passage from one period to the next; third, the solution is examined to check if parts of the configuration can be locally improved, going back iteratively to the second stage. The model was tested for examples of dimension up to 20 sites and 10 periods.

PINGRY and SHAFTEL (1979) proposed a heuristic model to optimize wastewater management, considering its reuse. Realistic concave cost functions are used. The solution is found with the reduced gradient method. The model only finds local optima, and is overburdened with water quality constraints.

GUGENHEIM (1979) proposed a variation of the Simplex method. Constraints are linear, but the objective function is, realistically, non linear and concave. The algorithm performs a systematic search between extreme points of the objective function, from one point no neighbouring points, in order to assure that any solution found is not significantly worse than any other. However, the model formulation for practical application is little elaborated, and no global optimum is guaranteed.

LEIGHTON and SHOEMAKER (1984) developed a model, first proposed by ATKINSON (1979) and reformulated by LEIGHTON (1982), in order to use commercial mixed integer programming packages. The basic formulation is a transhipment model. Additional constraints are used to reduce problem dimensionality: minimal flow in each pipeline section, number of pipes linked to a node, political and administrative requirements. Besides minimizing cost, the aim is to provide groundwater recharge, and the model is designed to study the trade-off between the two objectives. Linear cost functions are used. The model is applied to Western Suffolk County, New York, USA.

NAKAMURA et al. (1981) proposed a method to generate and assess alternatives for regional treatment systems, taking time variability into account. The solution depends on starting point and optimality is not guaranteed.

PHILLIPS et al. (1982) used mixed integer programming to study the location of treatment and transport systems, given the required treatment level.

SMEERS and TYTECA (1982) proposed a model to determine simultaneously the transport system, number, location, capacity and treatment level of several treatment plants, to assure pre-fixed water quality levels at several points of the water basin. A water quality model is used to compute water quality, using dissolved oxygen standards as a quality parameter. The model is restricted to dendritic structures. A shortest path algorithm is used to determine transport paths, and the reduced gradient method to define treatment efficiency allocation. No global optimality is guaranteed. The model is applied to Sambre River, Belgium.

KANSAKAR and POLPRASERT (1983) proposed a goal programming method. Objectives considered are cost, water quality and land use impact. The solution is found with a linear programming algorithm, using as objective function the weighed sum of deviations in relation to pre-determined objectives. Cost functions are linear and not generalized. Environmental criteria are not based on standards, but only on deviation in relation to an arbitrary objective. The model is conceptually interesting, but the poor accuracy of cost functions and subjective goal hierarchization seriously limit its application.
KITABATAKE and MIYZAKI (1983) analyzed the problem of treatment plant location in a water basin, considering spatial continuity of population distribution and receiving water body. However, the principles of this approach are very seldom realistic.

CÂMARA (1985) proposed a heuristic method using network theory to screen inferior solutions. Cost functions are point estimated, like in discrete dynamic programming, producing a network of non-inferior alternatives; solutions are found with a k-shortest path algorithm. The method aims at the identification of several optimal and sub-optimal solutions. It was further developed by CÂMARA et al. (1987).

KLEMETSON and GRENNNEY (1985) presented a dynamic programming model that optimizes the construction or phased expansion of treatment and transport systems, given pre-defined treatment levels. Each stage of the dynamic programming algorithm represents a period of time. Each node corresponds to a pre-defined treatment configuration. The state variable is the discounted annual cost. The model identifies the global inter-temporal optimum, as long as it is implicit in the set of pre-defined configurations. The authors tested the model for the case of the Jordan River, Utah, USA.

ONG and ADAMS (1987) proposed a solution method with the "poliedron random search" algorithm. It does not guarantee global optima and is overburdened by water quality constraints. ZHU and REVELLE (1988) proposed a solution method for the chain configuration case. The problem is formulated as a fixed-charge facility location model and is solved with linear programming. The treatment levels are pre-defined. For each potential treatment plant site, the additional cost of linking another polluting source is defined, considering that no source may be treated in a plant if there is another plant in the way. Cost functions are linear, although the authors show, for the cases tested, that the error due to linearization is minor.

PINEAU et al. (1985) and LÖWGREN et al. (1989) performed a comparative analysis between local and central solutions, confronting different technologies. However, their aim is more to gain insights into the structure of the problem, than to actually optimize a regional system. JOSHI and MODAK (1989) presented a series of heuristic models to determine the optimum allocation of polluting charge in a water basin, as a function of different treatment criteria and in order to respect water quality standards. They considered known discharge points and variable treatment levels, related by pre-determined criteria. The authors suggest that heuristics are quite interesting in this problem when compared to classical mathematical programming methods.

MELO (1992) developed further the work by CÂMARA et al. (1987). The revised model includes a heuristic cluster analysis submodel that reduces the solution space by establishing links between polluting sources and treatment plants. Several optimal and sub-optimal solutions are identified with a modified dynamic programming algorithm. A scenario analysis submodel integrates results of optimization for different scenarios. The equity principle and user-defined source-plant relations are used as screening criteria to improve computational efficiency. Case-studies show the model to be quite efficient up to 30 polluting sources and 3 regional treatment plants without user-defined screening criteria; by applying such user-defined criteria, much larger problems can be handled.

3. Definition of objective function and constraints

Economic cost

From the decision-maker point of view, the minimization of economic cost is doubtlessly a major objective to pursue, in a wastewater treatment system as in any other project. It is also the easiest-to-define goal, because it can be readily measured in currency units (if not always easy to quantify
accurately). For this reason, it is the chief, and often the only goal that is considered in the objective function of regional wastewater treatment optimization problems.

For the sake of economic analysis, it is important to differentiate investment cost and operation cost. In any case, one may distinguish direct cost, that is, directly dependent on a given project (civil works, equipments, financial interest, services, energy, labour, maintenance, pay off), and indirect cost, related to the institutional system (administration, professional training), as referred by PEREIRA (1988).

Direct cost can be computed by adding up component costs and/or by statistical analysis of a set of case-studies. However, some sources of uncertainty may remain, such as: insufficient detail in the component cost analysis; unpredictable variations in relative cost of different system components, by differential inflation or appearance of new technologies; distortion of market mechanisms that lead to unforeseeable costs, such as subsidies, under-rated job offers as marketing technique, or contract renegotiation. Indirect cost, given the difficulty to compute it accurately, is usually either disregarded or assumed as a function of direct investment.

Indirect costs to the surrounding community, such as impacts over land use, inconvenience during facility construction or unpleasant smells during facility operation, are not usually included in cost estimates and are better dealt with through careful site selection and planning.

**Environmental impact and pollution control criteria**

Although the ultimate goal of wastewater treatment is to assure the suitability of water quality for human needs and nature conservation, in practice pollution control criteria are most often reduced to emission standards and/or water quality standards. Water resources management strategy may thus be classified in three types:

- Uniform treatment (based on universal emission standards);
- Treatment at strict minimum cost (based on water quality standards only);
- Segmented uniform treatment (based on emission standards by section of the receiving water body, taking into account ecologic sensitivity and required needs).

In a model for the optimization of regional treatment systems, such criteria are normally considered as constraints of the model, rather than objectives to optimize. The definition of the constraints is linked to the institutional system and cost allocation criteria, and has a deep influence in the general formulation and results of an optimization model.

The uniform treatment strategy, by sector of activity or any other criteria, has the main goal of assuring identical market conditions. It has the significant disadvantage of not taking into account ecologic sensitivity or water uses. It may lead either to the subutilization of the self-cleaning capacity of the water body (rendering useless part of the treatment cost), or to its rapid degradation (implying serious disruption of the ecosystem and impeaching water uses).

The strategy of strict minimum cost, with constraints formulated exclusively in terms of water quality, leads to inequitable, highly variable solutions, difficult to compatibilize with any viable institutional structure. In fact, in order to guarantee predetermined quality objectives in the receiving water body, global optimum cost will theoretically be reached with different treatment levels in different sources or plants; but a decision demanding more stringent treatment in some sources would leave their owners in an unfavourable position, unless compensation procedures were set up. On the other hand, environmental quality criteria, complex by themselves, are very difficult to apply directly to pollution control at source.

The strategy of compartmented uniform treatment tries to combine the advantages, and reduce the disadvantages, of both other strategies. Environmental quality criteria are used along with policy-administrative criteria, as an element for the definition of emission standards. This approach is more practicable and effective, and in fact it is nowadays much more often used than the others. In the
European Community, it is typical that general emission and environmental quality standards are defined to begin with. Then, each country, region or water authority may establish more restrictive regulations, accounting for peculiar characteristics such as local ecosystem sensibility and polluting sources location.

Emission standards may be formulated based on many criteria: maximum effluent concentration, percent reduction in pollution, maximum polluting charge per product unit (for each activity), maximum polluting charge per source or treatment plant (based for instance on tradeable emission permits), or combination of those. Pollution may be measured by many parameters, the most common being biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids, nutrients and several ecotoxic chemicals; or with indexes composed with those parameters. The most common index, for domestic and compatible effluents, is the number of inhabitants-equivalent.

One must note that emission standard formulation is not innocuous. Different types of regulations — with the common goal of guaranteeing equity and environmental quality levels — have different consequences, in terms of global cost, distribution of cost by the polluters, global polluting charge, and water quality effectively resulting. The problem was studied namely by CHADDETON and KROPP (1985). Doubtlessly this is an important matter to study on defining emission standards, although it is out of the scope of the optimization of a regional treatment system.

Models for the optimization of regional wastewater systems can be classified in three main groups, in face of the pollution control criterion (CÂMARA et al. 1987):

1) Models that determine the distribution of discharges in a water basin, which allows for given water quality objectives;
2) Models that specify the treatment and transport system, so that every effluent is subject to treatment according to given emission objectives;
3) Models that define the treatment and transport system, as well as treatment level, in order to comply with given water quality objectives.

In models of the first group the problem is formulated as follows: given a receiving water body, the goal is to determine the level of treatment (hence the distribution) of known discharges which, at minimum cost, comply with given quality objectives. This is the oldest formulation, which does not allow for wastewater transport from one location to another (thus producing sub-optimal solutions) and ignores the equity problem. Usually, the formulation of the optimization model simplifies significantly the behaviour of the water body: only a couple of water quality parameters are considered, dissolved oxygen being the most common.

This constraints seriously limit the practical applicability of these models. A less rigid approach of the concept, applied to the study of management strategies for a water basin, seems to have much more potential (JOSHI and MODAK 1989).

Models of the second group use emission standards as pollution control criteria. The aim is now to establish the transport network and wastewater treatment plants which assure, at minimum cost, the required emission levels. This approach is followed by most authors, with varying degrees of success. Although it does not consider explicitly the ultimate goal — water quality — it is a realistic formulation, since, in practice, water resources planning systems usually apply emission standards as control criteria.

Models of the third group try to combine the goals of the former: to identify transport and treatment systems, as well as treatment levels, which minimize the cost while complying with explicit water quality standards. This formulation is twice as complex, and does not seem to have been satisfactorily solved. GRAVES et al. (1972) and KANSARKAR and POLPRASERT (1983) used unrealistic convex cost functions, making the solution procedure easier. PINGRY and SHAFTEL (1979), SMEERS and TYTECA (1982) and ONG and ADAMS (1987) get only local optima. Not
satisfying equity, all this formulations are also burdened by oversimplified water quality models, which would not be needed if emission objectives were to be adopted.

Given the advantages and disadvantages of those three formulations, it seems more fruitful the separate aproach of the problem of discharge location and the problem of system configuration. On one hand, each problem is complex enough by itself, and its coupling is not quite feasible. On the other hand, definition of emission criteria, as an intermediate step between both problems — although not being the most elegant formulation — is quite compatible with prevailing institutional arrangements.

Reliability

Reliability may be defined as a measure of how well a goal is reached — in this case, how well are emission or environmental quality standards respected by a given treatment system.

Some authors have argued that, in a wastewater treatment system, reliability depends on the global configuration of the system. The issue was studied by, among others: FITZPATRICK (1977), in the perspective of effluent charge variability; ADAMS and GEMMELL (1980), comparing centralized and decentralized solutions; NIKU and SCHROEDER (1981), working from the knowledge of distribution functions of system effluent concentration; and SYKES (1984), taking into account the variability of river flow and rejected polluting charges.

It has been argued that ecologic risk is lower in a decentralized solution, because pollution concentration is less, and failure of several small plants will not cause as much damage as failure of a large centralized unit. On the other hand, the larger the plants, the more sophisticated and complex they become, implying higher personnel training and administrative support requirements. ADAMS and GEMMELL (1980), among others, presented examples showing that local treatment solutions are more reliable then a highly centralized solution.

Many other authors, since WESTON (1971), have argued exactly the opposite. Some of the reasons put forward are the following: risk of failure in small plants is higher than in large ones, because in the last there are more sophisticated equipment and control procedures; it is normal to operate a large plant by modules, which reduces risk of total failure; it is easier to have well trained staff; flows and polluting charges are better regulated; and it is possible to combine wastewater from several origins, making joint treatment easier if effluent characteristics are complementary. However, theoretical contradictions between those authors, pro and against centralized solutions, prove to have no great real significance. On one hand, there are limits to super-centralization:

- Ecologic reasons — it is self-defeating to build a giant treatment plant if its effluent implies the transgression of water quality standards after the discharge point;
- Technical-economic reasons — above certain size, there are no standard or largely tested equipments, and so technology development cost erodes scale economies;
- Institutional reasons — it is complex to involve many institutions in the management of a wastewater treatment system, in particular if there is an overlay of competences;
- Space availability — the larger the facility, the more space is required, and the harder it is to find an appropriate site (unless compact technologies are adopted, which in turn are costlier).

On the other hand, practice shows that the reliability of a wastewater transport and treatment system depends mostly on the supporting institutional structure. That is, construction and installation quality, adequate staff training, responsible operation and maintenance, and demanding fiscalization, are much more important than the fact of a system being more or less centralized. Portugal is a paradigmatic example. Common inefficiency of wastewater treatment plants and ecologic accidents because of polluting discharges do not happen, as a rule, because of any random causes (either natural or technology related). They happen because of shortcomings of the
supporting institutional structure: incorrect design due to wrong forecasts, poor conception or lack of timely decisions; poor staff training; careless or insufficient fiscalization. Therefore, reliability does not fit well in the problem of regional wastewater treatment systems, either as a goal or as a constraint. It is nevertheless fundamental from the point of view of construction and operation standards.

**Flexibility and phasing**

Polluting sources are not static predictable sources, but have rather a dynamic and random behaviour. One of the main problems of optimizing treatment systems is therefore their evolution in time.

A configuration considered optimal today may well be far away from optimum on the day construction ends. Systems evolve (and data upon which decisions were made will probably be out of date anyway): hence, the problem of optimization of system configuration must be considered dynamic in nature.

ROSSMAN (1978) argued that both things — configuration definition and phasing — are indissociable. But that would be true only if the evolution of the system could be known accurately. With the existing uncertainty of socio-economic scenarios evolution, this relation, though important, is not mandatory.

The issue is more complex than it appears at first sight. For instance, an error in the definition of capacity of a water supply or sewer system in an urban area has the only effect of limiting the useful life of the facility. In the case of regional treatment systems, a different conclusion may imply choosing a completely different system configuration, and even another institutional arrangement.

The problem is further complicated because expansion or reequipment of a treatment plant is significantly cheaper than a new plant, especially if that had been previewd to begin with. ONG and ADAMS (1989) showed that substantial differences result fron design criteria directed towards long term, or towards short term combined with phased expansion. If expansion periods are not too short, the last approach tends to result less expensive.

In short, a model for the optimization of regional treatment systems should look at probable or possible evolutions of the system, either in an explicit or an implicit way.

Most authors study optimization of wastewater treatment as a static problem. This aproach has two advantages: it is methodologicaly much simpler and it does not require the definition of evolution scenarios. However, it is adequate only with predictable and relatively invariable systems. Many authors treat the problem of optimal solution variability (as a function of different scenarios) merely as sensitivity analysis.

BHALLA (1970), CHIANG and LAURIA (1977), ROSSMAN (1978) and KLEMETSON and GRENNEY (1985) proposed models that specifically consider phasing. These authors integrated the concepts of phasing wastewater treatment plant building by, among others, BERTHOUEX and POLKOWSKI (1970), LAURIA et al. (1977) and BAZIW and SCHERER (1979), with the problem of optimizing of regional systems. Those models come after potencial solutions are identified, that is, they presume a known evolution of the system. MELO (1992) proposed a scenario analysis model to integrate results of optimization with different scenarios.

Given the uncertainty associated with system evolution, it is important to incorporate somehow this uncertainty in the decisions, which means to establish a measure of resilience/flexibility of a proposed system configuration before different scenarios. The problem of phasing can be dealt with separately, once the system configuration has been established, using specific methods (for example the one proposed by ONG and ADAMS, 1989). This perspective seems more effective than a deterministic consideration of phasing.
On the other hand, it is imperative to integrate existing facilities in the optimization of the system. Although many authors allow for this, many others do not, rendering their models useless for the study of system expansion.

**Equity**

In a set of polluting sources and respective treatment system, the costs are seldom paid by a unique entity. Normally, there are different polluters (industries, municipalities), with some autonomy and who supposedly should pay for their own pollution. This is called the polluter-pays principle. This principle has seldom been applied to the last consequences — either in the perspective "pollution abatement cost", "tax per unit of residual pollution" or "incentive for reduction of pollution generated". However, the trend in real institutional systems is to apply it more and more. Therefore, any cost allocation method should respect equity among participants. This is not a minor problem. Any attempt to apply an apparently not equitative system will inevitably find severe opposition from the interested parties, which will make unfeasible theoretically optimal solutions (GIGLIO and WRIGHTINGTON 1972).

Because of scale economies in transport and treatment, it is often cheaper to treat together effluents from several sources, than to treat them separately. However, because treatment scale economies imply increasing transport costs, it may happen that an autonomous source is not interested in a coalition because it would be cheaper to treat its effluent locally — depending on cost allocation criteria. Such a decision to protect the interests of one participant may increase the total cost of the system.

The problem is complex, because equity can be understood in different ways. Treatment requirements and cost allocation criteria can be set up with multiple formulae, with a common objective of assuring equity and acceptable water quality levels. However, some basic principles should be respected:

- If other circumstances are similar, a source with higher polluting charge should pay more than another with lower charge;
- Scale economies from joint transport and treatment should benefit all the participants;
- Addition of a new polluting source to a coalition of sources contributing to a treatment plant should not imply added cost for any of the sources that were there before.

It should be noted that these principles leave a wide margin to define detailed cost allocation criteria. There are question like: should cost be proportional to pollution, or should pollution removal efficiency be equal, whatever the cost? Should the distribution of scale economies in a given solution depend on the cost of that solution, or be a function of alternative solutions? This is a classical game theory problem, which has been studied by several authors, in the perspective of the definition of cost allocation criteria in water resources management (LOUGHLIN 1977, HEANEY and DICKINSON 1982). However, it has seldom been considered in the optimization of regional wastewater treatment systems.

Most authors study optimization of regional wastewater systems using only global cost as an objective function. This approach implies that all costs are paid by a unique entity, or there is a satisfactory mechanism to distribute costs among interested parties. Equity is usually one overlooked issue, explicitly or implicitly (this failure is indeed surprising, since the polluter pays principle tends to be universally accepted). The excellent work of CHADDERTON e KROPP (1985) must be referred here, because it provides important insights to the problem at hand; however, the subject of these authors was not system optimization, but the comparison among types of emission standards.

Unfortunately, the "unique entity" situation is very seldom realistic. Although several polluting sources may belong to the same entity (for instance, villages in the same municipality), the typical
situation will be one of different sources, with responsibilities of their own and autonomy of decision (for instance, different towns or industrial plants), not willing to accept a more expensive solution if they can choose a cheaper one. Therefore, theoretically optimal solutions may well become impracticable due to lack of cooperation of interested parties.

It is thus fundamental that a model for optimization of regional wastewater systems considers equity. As equity may be defined in different ways, realistic formulations should be preferred, both institutionally and from the point of view of autonomous decision-makers.

**Effluent reuse**

Treated effluent has several possible uses, like irrigation, groundwater recharge, industrial use and fire fighting.

Organic effluents, like domestic and agro-industrial ones, have high nutrient contents, namely nitrogen and phosphorus. It is thus interesting to use them in irrigation, particularly in dry regions where no other organic fertilizers are available. However, the use of effluent from treatment plants in irrigation presents some problems. Firstly, organic and microbiologic polluting charges must be reduced to levels compatible with the culture to irrigate (in forestry and other non-consumption cultures this restriction is less important, sometimes even raw wastewater is used). Secondly, many pollutants common in urban effluents — like heavy metals — have to be monitored and eventually removed. Thirdly, there is the problem of transport: most convenient treatment plant sites are commonly away from areas to irrigate, or located in depressions, rendering effluent transport for irrigation economically uninteresting.

Optimum reuse of effluents to irrigation has been studied by many authors, although not in the perspective of global optimization of a regional treatment system. Typically, a given effluent is assumed and the economic viability of using it is tested against the basic hypothesis of sending it into a water body. An interesting example of this perspective was presented by DINAR and YARON (1986). These authors developed a model to optimize regional income through agricultural production, with domestic effluent reuse, in a rural dry area in Israel.

As for aquifer recharge, constraints are similar to those for agricultural use, regarding namely microbiological and heavy metal contamination. Pollution limits are determined by hydrogeologic features and proximity of springs or waterheads. The objective of maximum recharge may be conflicting with global minimization of cost, as shown by LEIGHTON and SHOEMAKER (1984).

In the industrial use of effluents, two cases can be clearly distinguished: internal treatment and recycling, or creation of a regional structure to distribute effluent from other sources. Internal recycling is becoming quite common, but not redistribution, because it implies building costly reservoirs and pipelines. Again, the work performed in this area has not been related to regional wastewater system optimization.

4. **Solution strategy regarding optimization methods**

As in other combinatorial problem, the optimization of regional treatment systems (even with the single objective of minimizing cost) is always a challenge (EISELT and LAPORTE 1987). The number of alternative configurations for the system grows exponentially with the number of polluting sources. In a case with \( n \) sources and \( m \) alternative locations for its treatment, without any other constraints, the total number of possible combinations goes up to \( m^n \).

Problems of such size (even with relatively few sources and plants) are not solved with empirical solutions. An experienced engineer may undoubtedly find reasonable solutions, even close to the optimum. But he can never guarantee that his solution cannot be significantly improved. Hence the usefulness of computer applications which allow the identification of optimal solutions with
minimal work. In the current state of computer science, no explicit enumeration method is viable to this kind of problem. It is then necessary to apply optimization techniques. Unfortunately, the concavity of the objective function (which is a sum of cost functions for treatment plants and transport systems), due to scale economies in the main cost factors, renders the issue more complex; this is so because the optimality of solutions found with mathematical programming algorithms is not guaranteed. The concavity of the objective function implies the existence of multiple local optima, whose identification does not give any clue to the distance to global optimum.

The joint problem of explosive combinations and objective function concavity, typical of optimization of regional wastewater treatment systems, has been handled by different authors with three types of methods — none of which assures the finding of global optima in reasonable computation time:

- Use of search methods which increase the probability of local optima, identified with non linear programming algorithms, being close to global optimum;
- Elimination of objective function and/or solution space concavity, namely through linearization of cost function and constraints;
- Heuristics which improve the objective function by successive changes over a given basic configuration of the system.

**Non linear programming strategy**

Generally, non linear programming methods tend to be computationaly inefficient and do not guarantee optimality of solutions. This characteristic is due to the search method which, from one initial point, tries to improve the objective function in the neighbourhood of the solution space. GUGENHEIM (1979) showed, based on the model of JOERES et al. (1974), that mixed integer programming is computationaly inefficient in problems of great dimension. As for dynamic programming, it implies the establishment a priori of possible solutions (thus not ensuring optimality) or the enumeration of solutions (implying excessive computation time). Geometric programming and other non linear mathematical programming methods are inadequate to solving problems with concave cost functions (CÂMARA 1985).

Authors like GUGENHEIM (1979) tried to go around the problem, with methods that increase the probability of local optima, identified with programming algorithms, being close to global optima (as far as the value of the objective function is concerned). However, it is never possible to guarantee the identification of the global optimum.

**Linear programming strategy**

Linearization of cost functions, which eliminates the concavity of the objective function, is one approach followed by many authors. With linear restrictions, it is possible to use more efficient algorithms to reach a solution. This approach bears nevertheless significant problems. First, we have integer decision variables: to build or not to build a wastewater treatment plant, to use or not to use a transport link. But integer programming, even with a linear objective function, is computationaly much less efficient than linear programming. Authors like LEIGHTON (1982) went around the problem by increasing the number of constraints. Others, like ZHU and REVELLE (1988), applied linear programming directly, benefiting of the frequence of integer solutions in the linear problem, complementing it with branch and bound techniques.

However, the strict linearization of cost functions leads to distortions. Although it may be possible to find, in an efficient manner, the optimal solution of the linearized problem, there is no warranty that it is the optimal solution of the original non linear problem. In fact, there is a high probability that these solutions do not coincide, given the influence of scale economies in the definition of
costs. Few authors analyse their solutions to check on the deviations provoked by this approach. In general, one can state that a reasonable approximation is possible only within narrow limits, that is, when the range of dimension of all transport and treatment systems are in the same order of magnitude, or if scale economies are weak. Distortions will be greater with increasing scale economies and larger range of polluting charges and system dimension in the real problem.

As for the segmented linearization of cost functions, the result is a composed multiplication of the linear programming problem, the greater the higher precision of the linearization. Results of this type of methodology do not seem to be significantly better, because the gains in precision are rapidly eroded by larger computation time.

**Heuristic method strategy**

Heuristic methods as proposed by Whitlach e REVeLLe (1976), CHIANG and LAURIA (1977) and JARVIS et al. (1978) are especially attractive because they try to benefit on the problem structure in order to light the solution procedure. However, they do not explore those techniques very far. The approach of all these authors is based on heuristics as a mean to reach successive approximations of optimal solution in iterative procedures.

Within heuristics, one should refer to genetic algorithms. Although it does not guarantee global optima, genetic algorithms seem to allow for a quick search of the universe of solutions and the identification of near optimal configurations. It is a recent method, which was successfully applied by GOLDBERG and Kuo (1987) to pipeline optimization and by PEREIRA (1988) to drainage network optimization. No applications to the optimization of regional wastewater systems seem to have been done so far.

When bulk investments are at stake, the knowledge of the distance to theoretical optimum is an important piece of data, even if in practice this solution may not be the best. So, the principle proposed by CAMARA (1985) and developed by MELO (1992) becomes very interesting: use of heuristics to reduce the solution space to a dimension that can be tackled by optimization algorithms. Although other authors have used heuristics to reduce the space of alternatives (LEIGHTON 1982, among others), it had not been done in such a systematic way.

**The multiple solution problem**

In a problem so impaired by uncertainties and conflicting objectives, only by mathematical abstraction can someone declare any system configuration as the ultimate optimal alternative (HARRINGTON and GIDLEY 1985). It is therefore very important that a model provides, not only one, but several near-optimal alternatives. In other words, this is not only a multi-objective, but also a multi-solution problem.

Most models ignore the problem, considering it merely as a sensitivity analysis issue. In fact, it is not. Widely different configurations of a wastewater system may have similar costs but different consequences at other levels. This should be explicitly accounted for in an optimization methodology. Heuristic models are the ones that approach this issue better.

Furthermore, one should bear in mind that decisions over wastewater systems are taken by politicians, who may consider political opportunity criteria over and above technical criteria. Of course, political opportunity criteria are not suitable to enter formally in the optimization problem, but it is another reason why one should produce more than one interesting solution for any given problem.
5. Model applicability

COCKLIN (1989a,b) demonstrates the need to understand optimization models, not as identifying ideal solutions, but to gain insights at the problem and as a decision support tool. CHANG et al. (1982, 1985) showed the utility to apply mathematical models to the generation of solutions for water resources management problems, similar in reaching modelled objectives, but dissimilar in configuration and in the prossecution of non modelled objectives. The case of regional wastewater systems is paradigmatic of the need to put those concepts to practice. Any engineer, politician or official, involved in water resources management, will want to study several alternatives. They will not be interested in a theoretical optimum because they know it rarely has correspondence in practice. They will want to comprehend clearly the problem, to examine a relatively small number of good alternatives, and to be sure that no potencially interesting solutions escape to their attention.

Computerized mathematical models have always been looked upon by laymen has misterious entities, to be handled only by scientists and with doubtful similarity to day to day realities. However, the causes of this mistrust are not price or difficulty of access to computers. It has to do with computer science language, odd to common experience, and with the distant position of researchers towards practical applications (FEDRA and LOUCKS 1985).

Modelling of regional wastewater treatment systems has been a typical case of divorce between researchers and users. Potential interested people, including engineers, officials and decision-makers, are suspicious of cloudy mathematical models. The only way to convince those people to use models, is to present them in a transparent, ready to use way.

Transparency has to do with clear comprehension of principles and limitations of the model. A model has a set of basic principles, an array of questions it is able to answer, and limitations — and it is essential that users understand all those points. Basic principles should be fairly simple, even if they are treated in a complex manner at computation level.

When the user of the model is an engineer or planner, the model should be able to integrate the experience of those people, and at the same time contain some expert-based information. A good example of this type of aproach is given by PEREIRA (1988). On the other hand, if the model is to be used by people without technical training, for decision support, data and results should be plain enough. A local government member or official should be able to work with the model without excessive technical background.

Those apparently minor conditions are really critical, because model users must be confident with what they are doing; they must be aware that the model is producing realistic results, useful for them (or not).

The ease of using a computer model has yet to do with the concept of "user-friendly interface". This philosophy puts to work techniques like option choice by menus, open action sequence, graphics applications, multiple interfaces (such as mouse and keyboard), and geographic information systems.

For years now computers have been cheap enough for anyone to purchase them. But their recent popularity is undoubtedly associated with the development of graphical user-friendly interfaces. Several authors have shown the usefulness of applying such philosophy to water resources and other environmental problems (for example LOUCKS et al. 1985).

Many models in the literature are applied to case-studies, and part of them will have contributed to decions over those cases. However, the perspective of those applications is generally, either a theoretical demonstration, or a consultancy job: that is, the model is applied by its authors, and the results are transmitted to the interested party, with little or no contact between the interested people and the model.
Among the reviewed models, the most significant exceptions to this perspective may be those by Whitlatch and ReVelle (1976) and Melo (1992), that include the integration of end-user expertise as a major component of the decision process. However, even in those cases no report of end-user views is presented.

6. Conclusion

Many advances have been made in the field of optimization of wastewater treatment systems. However, there is still a realm of possibilities that are open to research.

We should aim at decision support systems that allow the mathematical models to be readily used by decision-makers. This is not a simple task, because such systems must be supported by optimization models which are flexible, reliable, work on real time, and, above all, are designed as an interactive decision support tool.

New interfaces should thus be based on new models that integrate a number of developments. The most efficient models seem to be those that combine a heuristic approach with optimization techniques, and that integrate the user experience and the real world institutional arrangements into the mathematical problem.

Four main areas of research may then be considered:
- First, further development of heuristics and optimization techniques that profit as much as possible on the peculiar structure and known data of these category of problems;
- Second, further development of robust interactive models that allow for input of user-provided information;
- Third, automation of input of detailed information (on cost functions, on facility location, on geographic data, on source-plant relationship, on legal requirements) in such a way that it improves the accuracy of the results, without impairing the computational efficiency;
- Last, but not the least, the design of interfaces, which is a critical condition for effective application of these models by end-users.

References


CONVERSE, A.O. (1972). Optimum number and location of treatment plants. Journal of the Water Pollution Control Federation, 44 (8), 1629.


