

# Optimization of hybrid ethanol dehydration systems

Z. Szitkai, Z. Lelkes, E. Rev \*, Z. Fonyo

*Department of Chemical Engineering, Budapest University of Technology and Economics, H-1521 Budapest, Hungary*

Received 19 February 2001; received in revised form 17 May 2001; accepted 16 October 2001

## Abstract

Hybrid ethanol dehydration systems are modeled and optimized using MINLP. The systems consist of a distillation column for approaching the ethanol/water azeotrope and of a pervaporation unit for producing pure ethanol. The optimal design and operating parameters including number of trays, feed location, reflux ratio, number of membrane sections in series and the number of membrane modules in each section are determined. Regression equations fitted to solutions of differential equations are employed for modeling the membrane modules. Quadratic and exponential regression, as well as metric and linear interpolation are studied for approximating the integral membrane model; the exponential approximation is selected. A new mathematical representation of the superstructure of the membrane subsystem is suggested and applied. A successive refinement method with non-increasing number of binary variables is developed and successfully applied. Computational experiences with GAMS/DICOPT are presented. Using our new membrane superstructure representation, the hybrid system can be optimized effectively. The optimization method developed is also successfully applied for process intensification of an industrial scale dehydration plant. Compared to the existing plant, 12% savings in the total annual cost can be achieved by applying 32% additional membrane surface, in consequence of a radical decrease in the reflux ratio (3.3:1.4) in the column, and of producing less concentrated alcohol in the distillate. Sensitivity of the total annual cost to the specified ethanol yield, overall membrane surface and membrane replacement cost is studied. Total permeate recycling is found to be more economical, compared to partial recycling. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Ethanol dehydration; Pervaporation; Optimization; MINLP

## 1. Introduction

Though many techniques are known for ethanol dehydration, the following four process classes are the most widespread in practice: adsorption, distillation, pervaporation and hybrid processes. Distillation processes are the most applied in industry. Either a pressure-swing distillation is used, with the consequences of extra costs; or a third component as entrainer, with its unfavorable side effects, is applied in the extractive and azeotrope distillation processes [1,2]. Pervaporation is an emerging membrane separation technology with the merit of low operating costs. The disadvantages of the pervaporation are the low maximal capacity and the high capital cost [3]. The most promising technologies are the hybrid processes, espe-

cially the distillation–pervaporation system. Although these hybrid methods are most economical, the design and optimization of these systems are numerically difficult.

In this article the hybrid distillation–pervaporation processes are dealt with. In our MINLP formulation, both the distillation and the membrane modules are rigorously modeled. Optimization of the pervaporation system is already presented by Srinivas and El-Halwagi [4]. They used a state space model, optimized by MINLP, but did not address the hybrid system. Viswanathan and Grossmann optimized the distillation column with rigorous MINLP modeling [5]. On the other hand, they did not consider capital and operating costs but optimized for the number of theoretical trays at minimum reflux ratio. Sander and Soukup experimentally determined the permeate concentration and flux in function of the feed concentration at different temperatures in a pilot plant of ethanol dehydration [6].

\* Corresponding author. Fax: +36-1-463-3197.

E-mail address: ufo@sunny.vemt.bme.hu (E. Rev).

Table 1  
Utility costs

Item	Specific cost
Low pressure steam (160 °C, 5 barg)	13US\$/1000 kg, 6.225 US\$/GJ
Cooling water ( $\Delta T = 10$ °C)	0.16US\$/GJ
Electricity	0.06US\$/kWh
Permeate condenser cooling medium	2.64US\$/100 m <sup>3</sup>

Our work was initially motivated by an industrial site of a hybrid separation process, and started with the problem statement described in Section 2. The size and the complexity of the problem soon turned out to be a challenge in the sense of the model building and the representation of the superstructure. The model building was just an interesting but tedious search for the best appropriate approximation; but the representation of the superstructure generated a new type of problem about structural multiplicity and redundancy. A kind of solution for minimizing the necessary number of binary variables in the representation is given in this article, that made possible to solve problems of greater scale that would be impossible otherwise. This methodology has made possible to determine the optimal configuration and to perform a wider analysis of the industrial problem itself.

## 2. Problem statement

The system for producing absolute ethanol consists of a distillation column for purifying the crude ethanol inlet and a succeeding network of pervaporation membrane modules with liquid pumps, a vacuum pump and heat exchangers. An MINLP representation for minimizing the total annual costs (TAC) of the system is to be set up and solved with one of the available solvers. The optimization is to be performed over the design and operating parameters including number of trays, feed location, reflux ratio, number of membrane sections in series and the number of parallel membrane modules in each section of the membrane train. The utility costs are given in Table 1. The annual number of working hours is taken to be 8000. The GAMS/DICOPT package [7] is available as a tool for solving MINLP problems.

In order to perform the given task, appropriate superstructure, unit models and representation are to be given.

## 3. Superstructure

The superstructure applied for the hybrid system is presented in Fig. 1. The top product of the column is pumped in the first section of the membrane train. In

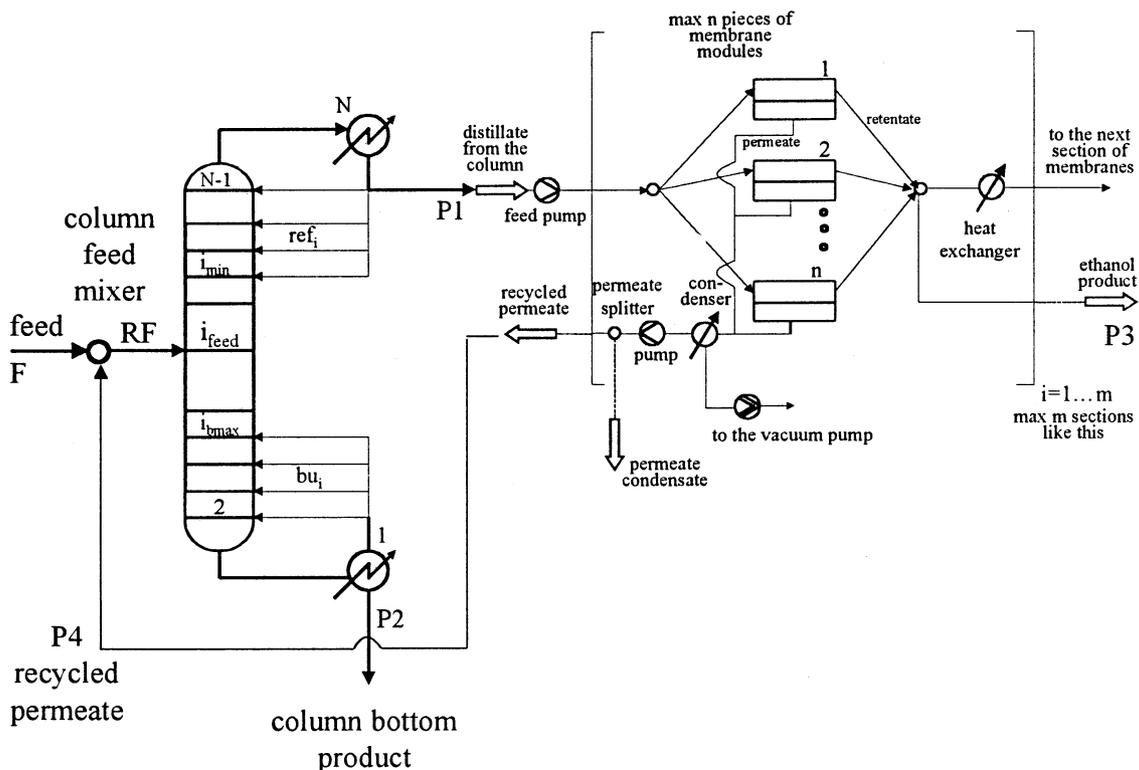


Fig. 1. Superstructure of the system.

each membrane section the retentate is collected and fed to a heat exchanger for re-heating. There is a vacuum pump at the permeate side of the membrane system. The permeate is withdrawn as a product stream and/or recycled to the column feed.

#### 4. Distillation column model

The MINLP model and superstructure of Viswanathan and Grossmann [5] has been adopted for the distillation column except the cost function that was modified. Margules equation, with parameters [8] taken from DECHEMA data bank [9], is used for calculating activity coefficients. Column cost functions are taken from Novak et al. [10], although the column shell and valve tray costing equations are used in the following form:

$$C_{\text{shell}} = \text{M\&S} \frac{937.61}{280} D^{1.066} H^{0.802} \quad (1)$$

$$C_{\text{trays}} = \text{M\&S} \frac{136.14}{280} D^{1.55} H \quad (2)$$

$$A = 0.621 \sum_c (1 + R) F_c^D \sqrt{\frac{M_D}{\rho_m^D}} \quad (3)$$

The number of trays, feed location and reflux ratio are the quantities varied for optimizing the objective function.

#### 5. Membrane network model

The representation of the pervaporation membrane network is more complicated than that of the distillation column because, depending on the mathematical model of the network, structural multiplicity may occur that hampers the optimization [11]. Modeling the pervaporation mass transfer is also not a trivial task; and it is based on experimental data.

##### 5.1. Integral model of the membrane modules

To be able to model the behavior of a membrane module, it is necessary to know its characteristics. The pervaporation membrane (in case of a binary liquid mixture) is characterized by two experimental parameters, namely the permeate flux  $J'$  and the selectivity  $\beta = c'/c$ . Here  $c$  and  $c'$  are the concentrations of the faster permeating component in the feed ( $c$ ) and in the permeate ( $c'$ ), respectively [6]. The basic model equations of Neel [3] are the differential equations for the elementary volume of flowing liquid, within  $z + dz$  cross-sections (Fig. 3):

$$dJ = J'p \, dz \quad (4)$$

$$dc = \frac{(c' - c)J'p}{J} dz \quad (5)$$

$$dH = J'\Delta H'p \, dz \quad (6)$$

$$dT = \frac{dH}{Jh} \quad (7)$$

where

$$H' = H'(c') \quad (8)$$

$$J' = J'(c, T) \quad (9)$$

$$c' = c'(c, T) \quad (10)$$

Using the experimentally determined characteristic pervaporation functions of Sander and Soukup [6], the differential equations can be numerically integrated for a desired membrane surface [3]. However, the equation oriented optimizer system we use (GAMS, version 2.25) cannot deal with differential equations. Instead, the differential equations are in advance integrated numerically for the potential values of feed concentration ( $c_0$ ) and feed flow rate ( $J_0$ ). In our case, Eqs. (4)–(7) were numerically integrated for the membrane area of 1/3 m<sup>2</sup>. The result surfaces [ $J(c_0, J_0)$ ;  $c(c_0, J_0)$ ] (see Figs. 4 and 5) are represented later by regression and interpolation. Here  $J$  and  $c$  are the retentate flow rate and concentration, respectively. The differential equations are linear and metric interpolations as well as regression with polynomial and exponential functions are examined. The retentate flow rate  $J_T(c_0, J_0)$  is approximated by linear surface all over this study.

##### 5.1.1. Interpolation

Interpolation techniques have the advantage that the approximation function always takes exact values at the basis points, i.e. in the points where the function values are originally given. In the case of many given basis points [in our case many  $c(c_0, J_0)$  points calculated by the differential equations] and of surfaces not following the form of any well-known function, interpolation is expected to be superior to regression. That is why we first tried to apply Shepard's 'metric bivariate interpolation' [12]. This method applies just a single parametric equation for the whole range, and works in the following way: Let  $P_i$  denote a  $(J_i, c_i)$  point ( $i = 1, 2, \dots, N$ ); and  $r_i(P) = |P - P_i|$  be the Euclidean distance of this point  $P_i$  from the point  $P = (J, c)$  where the surface is to be approximated. Then the surface is approximated by

$$\hat{c}_T(P) = \frac{\sum_i (c_T)_i \prod_{i \neq j} r_j^\alpha(P)}{\sum_i \prod_{i \neq j} r_j^\alpha(P)} \quad (11)$$

where the  $\alpha$  exponent is a free parameter. However, this equation cannot be practically used with arbitrary large number of base points. This is because the products appearing in both the numerator and the denominator

of the equation become too high or too low at a given high number of basis points. In our case using a Sun SparcStation, maximum 20 basis points can be used. In case of more basis points numeric overflow occurs. The equation, therefore, is fitted to 20 basis points. The approximation can be improved by tuning the equation through the parameter  $\alpha$ . The optimum value of  $\alpha$  can be determined by non-linear programming. Among the more than 3000 potential basis points calculated by a numerical solver for the differential equations, 20 are selected to represent the whole range of the  $(c_0, J_0)$  domain. Then the values of the interpolation function at 60 other representative checkpoints are calculated using  $\alpha$  as a parameter to be optimized, so as to minimize the sum of the square error between the approximated and the known values. Using this method, the optimum value of  $\alpha$  is found around 2 for the particular selections of the checkpoint sets. The mean estimated deviance is 1.33 mass%; the maximum is 4.08 mass%. In this case, however, the shape of the approximating function (Fig. 6) includes local minima; and, as a consequence, the optimization routines find different local minima with different initial values. This unfavorable property is found even in a small optimization problem involving just one membrane module. At lower  $\alpha$  values, the local minima become more explicit (Fig. 7). At higher  $\alpha$  values the local minima disappear; but the slope of the function at about halfway between the basis points becomes very high, causing numerical problems for the NLP solver (Fig. 8). Therefore, to our greatest disappointment, Shepard's metric interpolation had to be discarded.

Though *linear interpolation* seems to be much simpler than metric interpolation, its application is not a trivial task in some MINLP solver environments. We, for example, use the GAMS package, as a widespread tool for translating equations into the input requirements of the MINLP solvers. The parameters of an equation in GAMS cannot depend on the actual value of a variable. Several different linear functions, section by section, are applied in linear interpolation; and which one is to be used depends on the actual value of the argument. In case of equidistant basis points, however, this interpolation can be represented in GAMS by introducing new variables and equations, as follows.

Let the set of the equidistant basis points  $\mathbf{x}$  with distance  $\Delta$ , and the set of their known function values  $\mathbf{f}$  be  $\mathbf{x} = \{x_0, \dots, x_i = x_0 + \Delta i, \dots, x_N\}$  and  $\mathbf{f} = \{f_0 = f(x_0), \dots, f_i = f(x_i), \dots, f_N\}$ .

A continuous variable  $k = (x - x_0)/\Delta$  measures the distance of  $x$  from  $x_0$  in the units of  $\Delta$ . Define the following sets of binary variables that are applied to point to the location of the actual  $x$  point:  $\mathbf{A} = \{y_0^{\mathbf{A}}, \dots, y_i^{\mathbf{A}}, \dots, y_N^{\mathbf{A}}\}$  and  $\mathbf{B} = \{y_0^{\mathbf{B}}, \dots, y_i^{\mathbf{B}}, \dots, y_N^{\mathbf{B}}\}$ . Only one of the elements of  $\mathbf{A}$ , and respectively  $\mathbf{B}$ , may be equal to 1; this is enforced by the following two equations:

$$\sum_i y_i^{\mathbf{A}} = 1 \quad (12)$$

$$\sum_i y_i^{\mathbf{B}} = 1 \quad (13)$$

The only binary variable equal to 1 in set  $\mathbf{A}$  defines the index of basis point  $x_i$  that directly precedes  $x$ :

$$k - 1 \leq \sum_i i y_i^{\mathbf{A}} \leq k \quad (14)$$

The only binary variable equal to 1 in set  $\mathbf{B}$  defines index  $i + 1$ :

$$\sum_i i y_i^{\mathbf{A}} + 0.5 \leq \sum_i i y_i^{\mathbf{B}} \leq \sum_i i y_i^{\mathbf{A}} + 1.5 \quad (15)$$

Then the interpolation function is defined as

$$\hat{f}(x) = \sum_i y_i^{\mathbf{A}} f_i + \frac{z - x}{\Delta} \left( \sum_i y_i^{\mathbf{B}} f_i - \sum_i y_i^{\mathbf{A}} f_i \right) \quad (16)$$

where

$$z = \sum_i y_i^{\mathbf{B}} x_i \quad (17)$$

This method of linear interpolation, which can also be used for any other problem, works well for bivariate functions, and for a single membrane module, too. The estimated mean deviance is 0.27 mass%; the maximum deviance is 2.46 mass% for 30 basis points.

The main drawback of this method is that, in our case, a set of such binary variables should be defined to each membrane section; therefore, the number of defined binary variables multiplies to a large extent.

### 5.1.2. Regression

The main advantage of regression techniques is their simplicity, which makes them easy to use in the environment of GAMS. The main drawback of regression is its inaccuracy when the points to be approximated do not follow the form of any well-known functions. Another problem in our case is that the surfaces should be regressed in a way that the approximating functions satisfy the material balance of the membrane module. Another specific constraint is that both surfaces should be fitted exactly to the points of zero feed flow rate and/or zero water inflow in order to avoid numeric error at zero or very small feed flow rates. Quadratic or higher order polynomial regressing functions satisfying these requirements are characterized by rather poor approximation properties. Application of exponential regression in the form of

$$J_{\text{T}} = 0.999 J_0 - 0.031 C_0 \quad (18)$$

$$C_{\text{T}} = 0.55 C_0 J_0^{0.145} \quad (19)$$

gives, on the other hand, satisfying results. The parameters of Eqs. (18) and (19) considering all constraints were fitted to more than 3000 calculated basis points

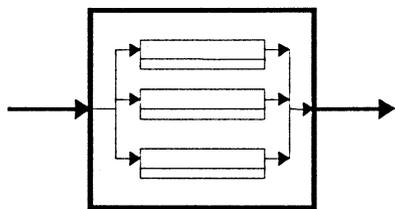


Fig. 2. A membrane module, the basis of our membrane structure calculation. This 1 m<sup>2</sup> module consists of three pieces of 1/3 m<sup>2</sup> cells.

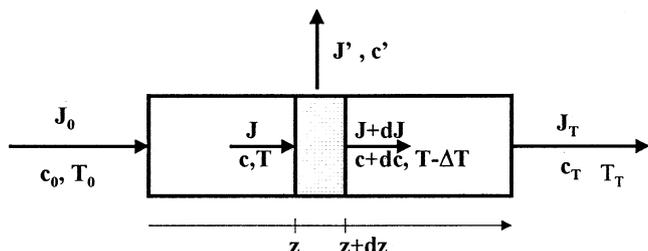


Fig. 3. Analysis of continuous-flow pervaporation.

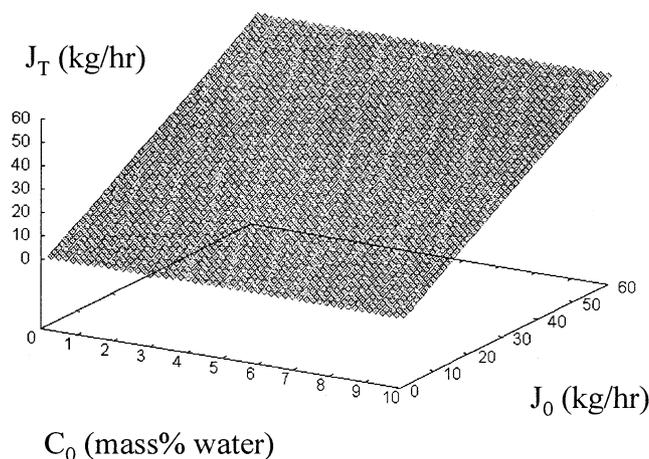


Fig. 4. Retentate flow rate in function of the feed concentration ( $c_0$ ) and the feed flow rate ( $J_0$ ). The surface is almost linear. The points are calculated by numerical solution of the differential equations (Eqs. (4)–(7)).

(Figs. 4 and 5), using the method of least squares. This NLP optimization was carried out using a separate GAMS input code.

Comparing all the interpolation and regression techniques outlined above, exponential regression in the form of Eqs. (18) and (19) is found to be the most appropriate in our case. (Still the selection amongst the interpolation and regression techniques is not generally solved since the best method is necessarily problem specific.)

### 5.2. Structure of the membrane subsystem, membrane superstructure and mathematical representation of the superstructure

Having the integral model of the membrane modules

constructed, we are in position to deal with the problem of how to represent the structure of the whole membrane subsystem. When using MINLP for design, the existence of a structural unit is usually represented by a binary model variable. This binary variable takes the value of unity when the unit is needed, and equals zero when the unit is not needed. We have to decide what the basic elements of our membrane system should be, and how to use binary variables to represent the structure of the whole membrane system consisting of these basic elements.

A pervaporation plant usually consists of many membrane cells connected both in parallel and in series. To use just one large membrane unit is not possible in practice for several reasons; one of them is the high temperature drop of the retentate, which makes the permeate flux to decrease drastically. The retentate has to be re-heated repeatedly. This requires a serial arrangement of the membrane cells. Since one membrane cell is not able to let through great rate of feed with acceptable purification, parallel connection of membranes is also required. In our industrial case, 1/3 m<sup>2</sup> membrane cells are chosen as the basic elements of the membrane system.

Since large membrane surfaces are needed in case of pilot-plant or industrial scale pervaporation, it is enough to determine the membrane surface in a section (Fig. 1) with the accuracy of 1 m<sup>2</sup>. Supposing this, imaginary 1 m<sup>2</sup> modules consisting of three 1/3 m<sup>2</sup> membrane cells in parallel (Fig. 2) can be used to construct the membrane subsystem. Eq. (11–12) are, therefore modified to describe the transport properties of a 1 m<sup>2</sup> module like this, without recalculating the differential equations.

According to the previous considerations, our membrane network superstructure is defined as an arbitrary large  $n$  times  $m$  rectangle of 1 m<sup>2</sup> modules. The design task is to determine the total number of modules and to find how to fill out this ‘rectangle’ with 1 m<sup>2</sup> modules.

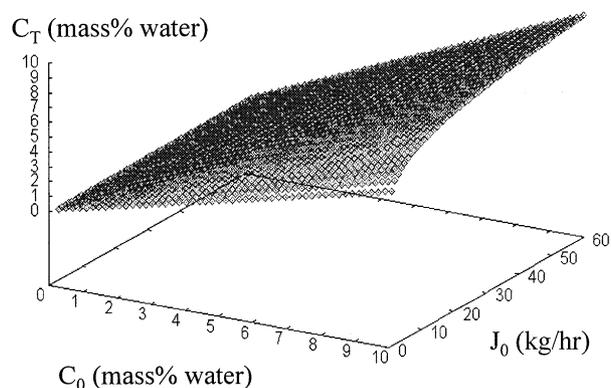


Fig. 5. Retentate concentration in function of the feed concentration ( $c_0$ ) and the feed flow rate ( $J_0$ ). The points are calculated by numerical solution of the differential equations (Eqs. (4)–(7)).

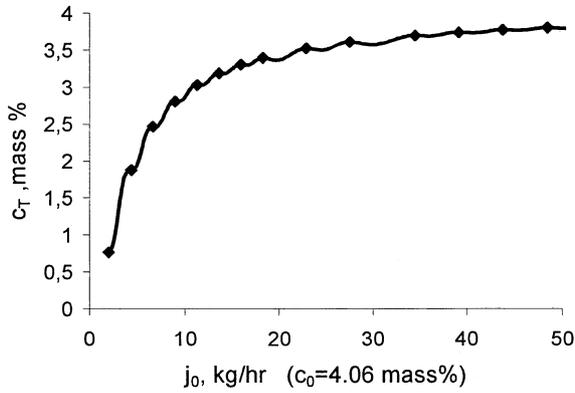


Fig. 6. Shepard interpolation with optimized  $\alpha$  parameter ( $\alpha = 2$ ). Wavy function with local minima.

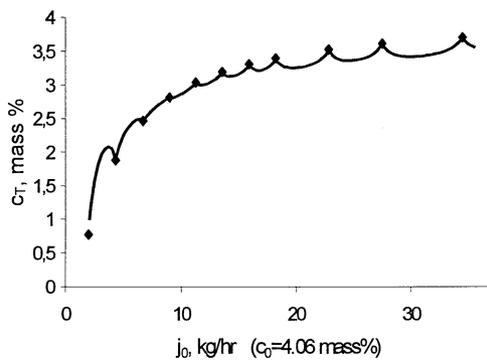


Fig. 7. Shepard interpolation with  $\alpha = 1$ . The local minima become more explicit.

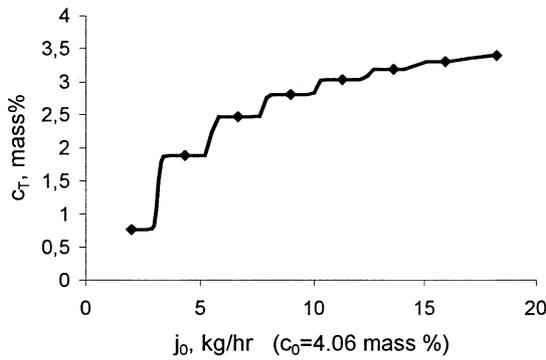


Fig. 8. Shepard interpolation with  $\alpha = 10$ . With increasing  $\alpha$ , the interpolating function takes a step function form.

If we denote the existence of each  $1 \text{ m}^2$  module by a separate binary variable, two problems arise. For large structures, the total number of binary variables will exceed a practical upper limit, which is set by the computer we use. Using more than 80 binary variables in a model results in enormous computation time on our Sun Ultrasparc-1 workstation. On the other hand, using one to one correspondence between the binary variables and the  $1 \text{ m}^2$  modules (this is one possible mathematical representation of the superstructure) the

same practical membrane structure can be described by different values of the set of binary variables in most of the cases.

In order to illustrate the difficulty arising from the opportunity of selecting any number (0 to the maximum) of a given maximum number of modules in a membrane train, we consider an example. Suppose that the maximum number of sections in the membrane train is 7, and the maximum number of modules in a section is also 7. Then there are  $2^{49}$  possible structures covered by the superstructure. However, many of these structures are equivalent, because the same type of modules is used throughout the seven sections of membrane train. This phenomenon is called here ‘structural multiplicity’. The multiplicity of some potential structures can go up to  $35^7$  ( $\sim 6.4 \times 10^{10}$ ). Only the trivial cases (no membrane is used or all the 49 modules are

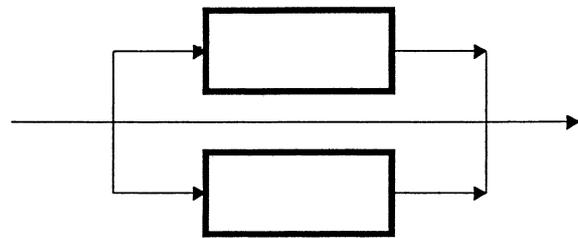


Fig. 9. Illustration of multiplicity arising from parallel arrangements. In case of using binary variables for denoting the existence of a module, the four combinations of the binary variables correspond to three real structures only.

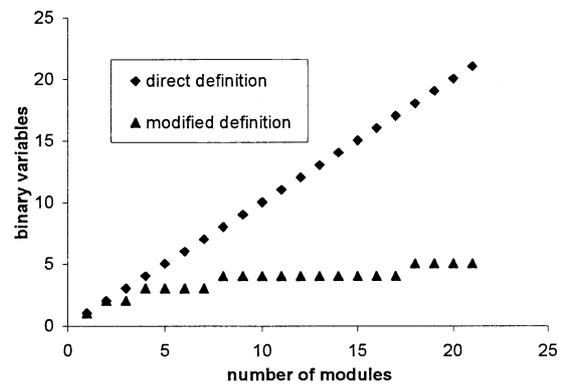


Fig. 10. Number of binary variables in the cases of direct definition and the modified one.

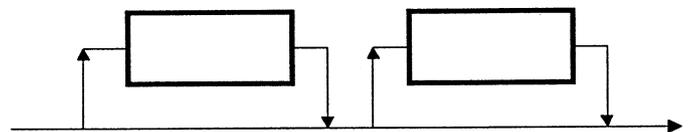


Fig. 11. Illustration of multiplicity arising from sequential arrangements. In case of using binary variables for denoting the existence of a module, the four combinations of the binary variables correspond to three real structures only.

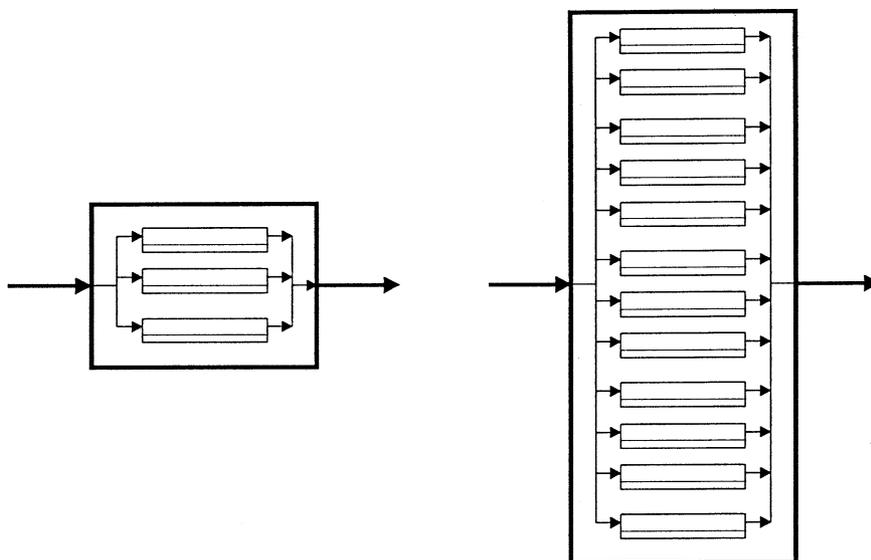


Fig. 12. Theoretical membrane modules consisting of different number of  $1/3 \text{ m}^2$  PVA membrane cells. The transport properties of a module can be calculated based on the equations derived for the  $1/3 \text{ m}^2$  cells (Eqs. (18) and (19)).

used) have no multiplicity. If only one module is used, or just 48 (maximum – 1) modules are used, then their multiplicity is  $7^2 = 49$ . A series of seven single modules in each section has the multiplicity of  $7^7 = 823543$ . This extreme multiplicity may involve incapability of finding the global optimum. Therefore, we developed a method to radically decrease this structural multiplicity.

If the binary variables denote the existence of the same type of modules (direct definition), and the modules are set in parallel, the multiplicity increases exponentially with the number of units. In the case of two possible modules in parallel (see Fig. 9), for example, there are four different combinations of the binary values, with only three different structures as solution. This is because the structures assigned by the existence arrays (1,0) and (0,1) are equivalent; i.e. these arrays denote the same structure. Generally, in case of  $n$  possible modules in parallel, the number of different structures is only  $n + 1$  of the  $2^n$  variations. This type of structural multiplicity can be eliminated by always selecting the first  $n$  modules of the possible maximum number. This could be achieved by additional logical constraints, but that method would lead to increasing the number of equations. A simpler, and more efficient, way is defining the actual number  $n$  in a binary number system, by explicitly using binary variables for the binary digits. This definition also involves a decrease in the number of binary variables. Dependence of the number of binary variables on the number of modules is shown in Fig. 10 for both the direct definition and the modified one.

If the binary variables denote the existence of the same type of modules (direct definition), and if the sections (with the same number of module) are in series,

then the multiplicity increases exponentially with the number of sections. Fig. 11 shows an example where a section contains one module only, and there are two sections following each other. There are four different combinations of the binary variables, in this case, with only three different structures, as solutions. Using a monotonic constraint for the number of modules in the sections along the membrane train can eliminate this type of structural multiplicity. The monotonic constraint means that the actual number of modules in the sections of the series may not increase from the feed to the end. The advantage of this monotonic constraint is that we do not need to change the definition of series, but only to use that extra constraint, instead.

In one of our examples of membrane superstructure, there are 16 modules in parallel, and 16 sections in series. This means 256 binary variables with direct definition; but the number of binary variables is only 80, using our method. Using the direct definition, we could not get solutions in reasonable time. Even our new, modified superstructure representation yields in high computation time, especially when structural multiplicity can occur. According to these results, the number of binary variables had to be further decreased.

### 5.3. Techniques for decreasing the number of binary model variables in case of the membrane subsystem

#### 5.3.1. Successive refinement

As we noted earlier [11], the solution time sharply increases with the number of binary variables. In case of a  $7 \times 7$  membrane module system, the search is always effective after eliminating the structural multiplicity with our, above explained, methodology that

involves a strong decrease in the number of binary variables. However, the search for the optimum is rather slow; and, in some cases, it is divergent with a superstructure containing  $16 \times 16$  membrane modules, even if the structural multiplicity is eliminated. The search, on the other hand, can be enhanced by successively refining the resolution grid on the membrane network in a way that the number of binary variables does not change. For this aim, we utilize the fact that the model does not contain directly the number of cells applied parallel in a membrane module. The integral model is valid for a membrane module with a given number of cells. If the number of the cells is changed, then the flow rates change proportionally, but the forms of the equations do not change. This enables us to optimize the hybrid system in more steps using different number of cells in the modules at the succeeding steps. Our new method, called ‘successive refinement’, is illustrated in Figs. 12 and 13, and works in the following way.

In order to use successive refinement, we have to apply membrane module models of flexible number of cells. This idea is illustrated in Fig. 12, where a module of three cells and another module of 12 cells are shown. Since each cell contains  $1/3 \text{ m}^2$  membrane surface, the membrane surface of the first module is  $3 \times 1/3 = 1 \text{ m}^2$ , while that of the second module is  $12 \times 1/3 = 4 \text{ m}^2$ .

Should the number of binary variables applied in a section be limited to three, the cells of a questioned part of a membrane section are to be distributed into maximum seven modules ( $7 = 2^0 + 2^1 + 2^2$ ). The sequence numbers 1–7 in Fig. 13 correspond to the maximum seven possible modules in question. Let the number of cells in a module be 12, i.e. let each module contain  $12 \times 1/3 = 4 \text{ m}^2$  membrane surface, then the total mem-

brane surface in the questioned part of a section can vary between  $0 \times 4 = 0 \text{ m}^2$  and  $7 \times 4 = 28 \text{ m}^2$ . At the first level of the successive refinement a suboptimal solution is sought, subject to the constraint of each module having 12 cells.

With the suboptimal result, the membrane surface in a section is determined up to  $4 \text{ m}^2$  accuracy. In our example, we got four modules as suboptimum i.e.  $4(12 \times 1/3) = 4 \times 4 = 16 \text{ m}^2$  membrane surface, denoted by shading in the leftmost column in Fig. 13. The optimal area is expected around this suboptimal value; therefore the existence of the first three shaded modules and the non-existence of the upper two modules are already determined, and just module 4 (the last shaded one) and module 5 (the first unshaded one) are to be considered in more detail. This is done at the second level of successive refinement, indicated by the middle and the rightmost columns in Fig. 13. Fixing the use of at least  $3(12 \times 1/3) = 3 \times 4 = 12 \text{ m}^2$  membrane surface is indicated by shading at the bottom of the middle and the rightmost columns. This part of the membrane section is already not questioned. Neither are in question the upper two modules of the leftmost and the middle columns. If the minimum number of cells in a module is limited to be at least three (i.e. 3 cells/module  $\times 1/3 \text{ m}^2/\text{cell} = 1 \text{ m}^2/\text{module}$ ) then the actual optimum may be  $12\text{--}19 \text{ m}^2$  (integral numbers only). The questioned part of the section is therefore partitioned into seven times  $1 \text{ m}^2$  possible modules, as indicated in the rightmost column of Fig. 13. A  $12 \text{ m}^2$  of membrane surface in this particular section is now added as a constant. The surfaces of  $12$  and  $19 \text{ m}^2$  can be calculated in the following way:  $12 \text{ m}^2 = 12 \text{ m}^2$  fixed surface + 0 modules,  $19 \text{ m}^2 = 12 \text{ m}^2$  fixed surface + 7 modules  $\times 1 \text{ m}^2/\text{module}$ . In order to optimize over the

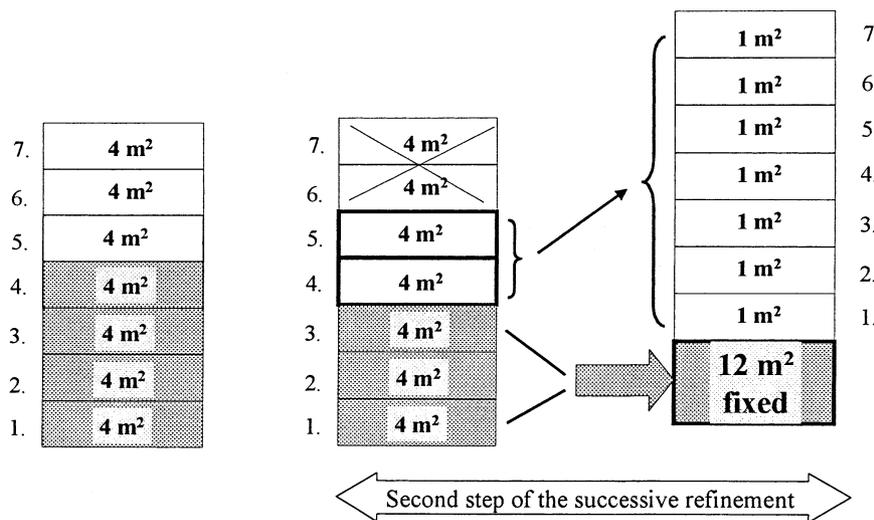


Fig. 13. The method of successive refinement. After obtaining the suboptimal solution of  $16 \text{ m}^2$  membrane surface in a section using  $4 \text{ m}^2$  modules ( $4 \text{ m}^2$  accuracy), the solution is refined in the neighborhood of this suboptimal solution using  $1 \text{ m}^2$  modules. In the second step, the membrane surface in the same section is expected to be greater than or equal to  $12 \text{ m}^2$  and less than  $20 \text{ m}^2$ . The number of binary variables remains the same.

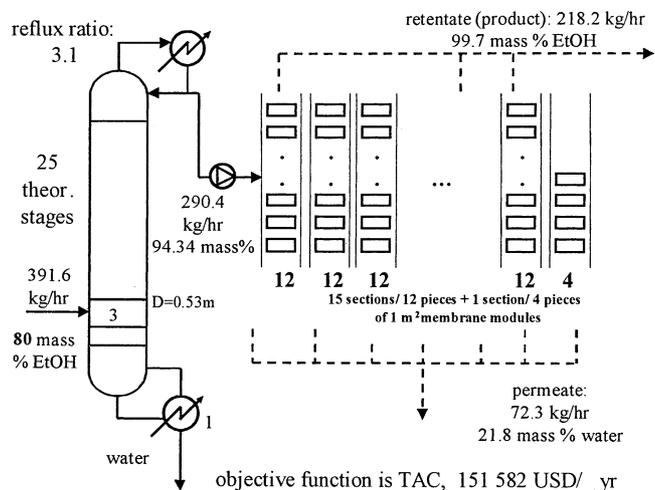


Fig. 14. Optimized system of Example 1

number of 1 m<sup>2</sup> modules, no more than the same three binary variables are needed that we used for discriminating between the 4 m<sup>2</sup> modules, as indicated in the leftmost column.

Using this method of successive refinement, arbitrarily large networks can be optimized with limited number of binary variables. The actual number of steps and the module sizes have to be chosen according to the problem to be solved. The succeeding steps are repeated until the minimum area applicable to a module is reached.

### 5.3.2. Continuous number of modules

The simplest way to reduce the number of binary model variables used for the membrane subsystem is regarding the number of modules in a section as a continuous variable. After obtaining the solution in this way, the optimization must be repeated in the neighborhood of the results, now using integer module numbers only. This solution type may be advantageous when too many binary variables have to be used for the distillation column i.e. many trays are needed. The drawback of this method is that after obtaining the continuous solution, the model has to be changed to binary representation and run again.

### 5.4. Membrane network costing

To determine the membrane network costs, we assumed that the total capital investment of the membrane network is linearly proportional to the overall membrane surface, and the proportionality constant is 1616.1 US\$/m<sup>2</sup>. This includes all the pumps, vacuum vessels, control equipment, etc. Equations for the variable cost calculations are taken from Srinivas and El-Halwagi [4] with the modification that the PVA membrane replacement cost was taken 775 US\$/m<sup>2</sup>, based on industrial practice. We assumed 3 years life-

time for the PVA membranes, thus one-third of the total membrane area has to be replaced each year, on average.

## 6. Computational results

### 6.1. Example 1

The flow rate of an ethanol–water mixture from fermentation is 391.6 kg/h; its composition is 80 mass% (this corresponds to approximately 6000 l/day absolute ethanol product). The minimum purity of the ethanol product is specified as 99.7 mass%. The ethanol is concentrated in a column having maximum 31 theoretical trays. The distillate enters the membrane system where the water is separated from ethanol. The column works at 1 bar; the feed is at its boiling point. The objective function is TAC. Since in this example the permeated water is taken as waste, 70% recovery is specified only. In this case, our superstructure is a 16 × 16 grid of 1 m<sup>2</sup> modules. Our new superstructure representation is used, which excludes structural multiplicity; but the method of successive refinement is not yet applied. The optimum is found to be 151 582 US\$/year. (Fig. 14). It is worth noting that the solution time of this problem is quite high, due to the big superstructure and use of many binary variables.

This first example had already shown the usefulness of our model and representation. An extended problem with the possibility of permeate-recycling and greater recovery in the same system was also tried to solve with GAMS, but numeric problems occurred in the solution of the MILP subproblems. That is why the method of successive refinement is elaborated.

### 6.2. Example 2

In this case the feed is 180 kg/h; its composition is 80 mass%, minimum 95% recovery is specified, and the permeate can be totally or partially recycled. The objective function is TAC. Here we applied the method of successive refinement, because no solution could be obtained otherwise. In a first trial, a suboptimal network is determined with 12 cells (4 m<sup>2</sup>) in a module and seven modules in a section. The suboptimal results are shown in Fig. 15. TAC = 90 529 US\$, with a 137.4 kg/h product rate, corresponding to a total cost per unit product (TCP) of 0.0845 US\$/kg product. Recycling split ratios  $\alpha$ , with index referring to the membrane section where the stream comes from, are also shown in Fig. 15.

The suboptimal membrane system consists of seven sections, each of which contains four modules (16 m<sup>2</sup> per section). The optimal tray number is 28, the reflux ratio is 3.93 and the feed tray number is 15.

In the second turn, modules of 1 m<sup>2</sup> were applied (three cells in a module); and the number of modules in a section were varied between 12 and 19. The results are shown in Fig. 16. Here the optimal number of trays is 28, the reflux rate is 3.96, the feed tray number is 15, and 16 m<sup>2</sup> is put in each membrane section, except the last one where 15 m<sup>2</sup> is used. The

optimum in TAC is 90398 US\$, with a 137.4 kg/h product rate corresponding to a total cost per unit product (TCP) of 0.0843 US\$/kg product. These results demonstrate the effectiveness of successive refinement that enables us to solve large MINLP problems improving the optimum step by step, while slightly modifying the optimal parameters.

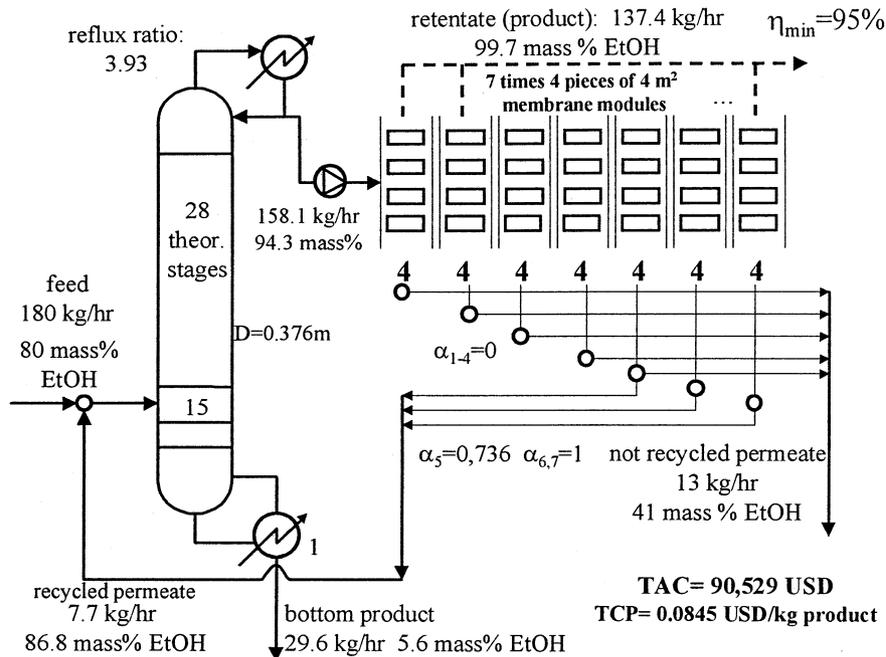


Fig. 15. Suboptimal solution of Example 2. First level of the successive refinement.

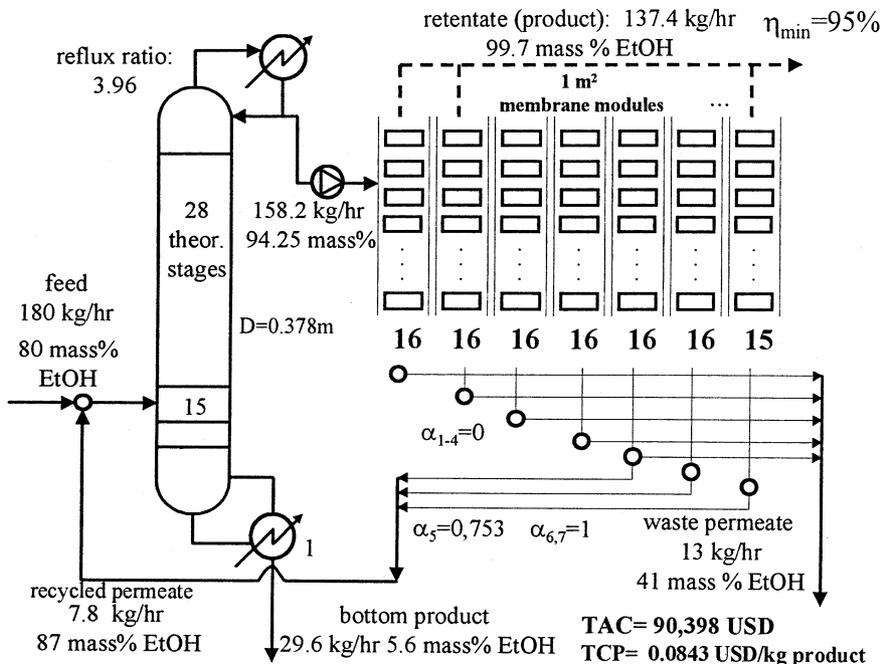


Fig. 16. Optimal solution of Example 2. Second level of the successive refinement.

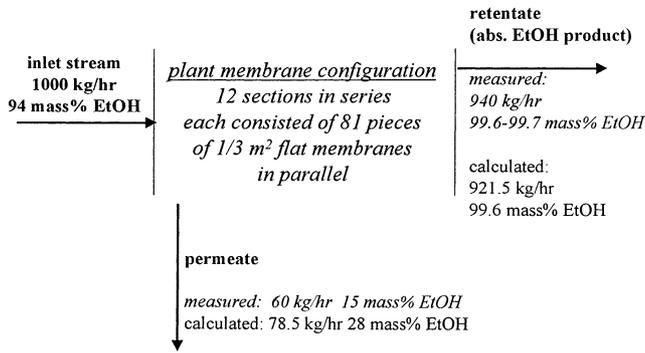


Fig. 17. Calculated and measured output stream properties for the fixed industrial inlet stream and membrane configuration.

## 7. Industrial case study

### 7.1. Applicability of our membrane model for membranes used in an existing plant

The applicability of our membrane model for a membrane type used in an industrial scale pervaporation plant has been examined first. The outlet streams and concentrations for the fixed industrial membrane structure and inlet stream were calculated using our model. The results of this calculation can be seen in Fig. 17.

The calculated retentate (product) flow rate corresponds with the measured one within 2%, while the calculated retentate concentration fits exactly to the retentate concentration measured at the plant. The higher deviance between the calculated and measured permeate concentrations and flow rates can be explained by the relatively great difference in the amount

of the permeate compared to the inlet ethanol stream. The relatively good accordance between the calculated and measured retentate flow and concentration data allows us to use our optimization technique for industrial scale applications, too. In the following chapters, industrial examples are discussed.

### 7.2. Base case

An industrial hybrid dehydration plant with fixed industrial inlet stream and membrane configuration was first optimized. The results of this optimization, which can be seen in Fig. 18, serve as base case for the later comparisons. It is important to emphasize that just the distillation column and the recycle streams were optimally designed; the membrane configuration is kept identical to the configuration of the existing plant. The existing plant contains 12 sections in series each consisted of 81 pieces of 1/3 m<sup>2</sup> flat membranes in parallel. Total permeate recycling is used at the base case and 97.5% ethanol yield is specified which means that 97.5% of the total amount of inlet ethanol is withdrawn in form of absolute ethanol in the product stream. The TAC of the base case is 373820 US\$, of which 219472 US\$ is the operating cost of the distillation column. The column works at a reflux ratio of 3.262.

### 7.3. Optimally designed hybrid system at 97.5% ethanol yield

Second, optimal design for all the distillation column, pervaporation membranes and recycle streams were

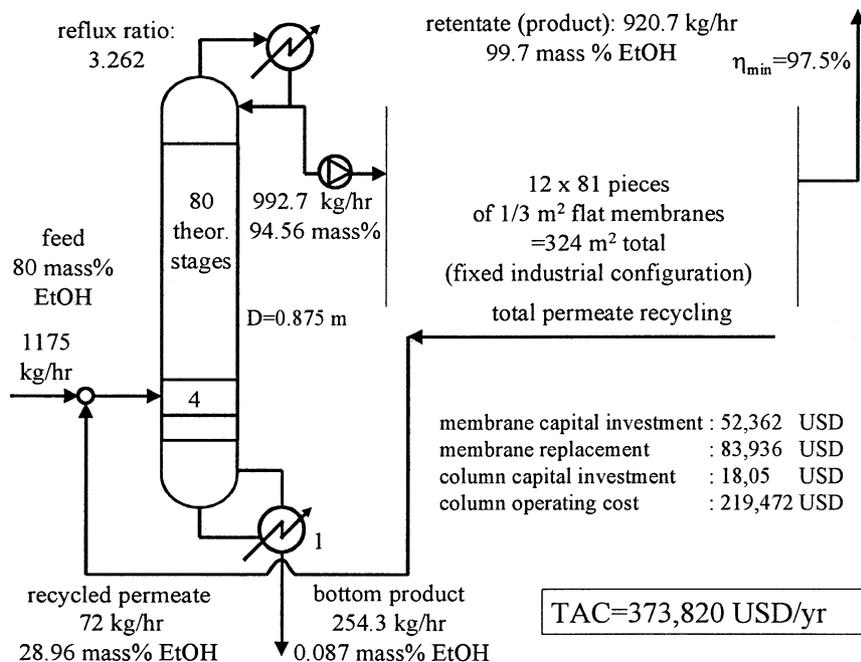


Fig. 18. Base case. Optimized hybrid ethanol dehydration plant with fixed industrial inlet stream and membrane configuration.

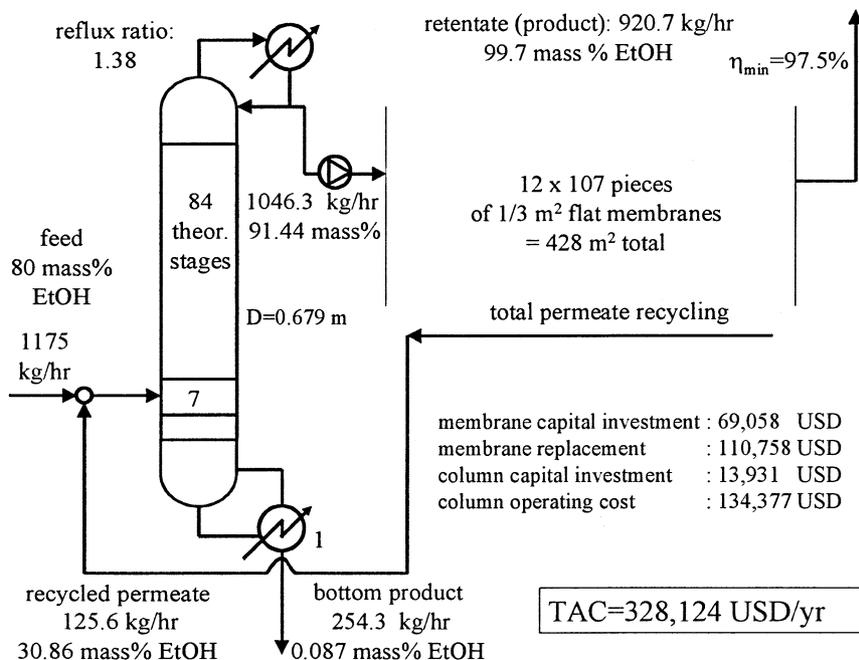


Fig. 19. Optimally designed hybrid system at 97.5% ethanol yield.

carried out. Still fixed total permeate recycling is used. The result of this design is shown in Fig. 19. The TAC of the hybrid system is 328 124 US\$. It can be seen that the total annual cost of the plant can be decreased by 12.2%. This saving is due to the increased overall membrane surface (from the industrial 324 to 428 m<sup>2</sup>), that allows to decrease the reflux ratio in the column from 3.262 to 1.38. It is worth mentioning that the inlet ethanol concentration to the pervaporation system dropped from 94.56 to 91.44%.

#### 7.4. Influence of the specified ethanol yield on the TAC

In the industrial practice (base case), 97.5% ethanol yield is set up. Depending on environmental regulations or plant conditions, however, greater or smaller ethanol yields could also be required. Because of this possibility, optimizations with fixed 95 and 99% ethanol yields were also carried out. The effect of the specified ethanol yield on the TAC of the hybrid plant is illustrated in Figs. 20 and 21. The increase in TAC with increasing yield is due to both changing the reflux ratio and overall membrane surface. It can be seen that the optimized system is less sensitive to the change of the specified ethanol yield. Changing the yield specification from 97.5 to 99% results in just 1.5% increase of TAC in case of the optimized system. The same increase in the yield at the industrial base case entails 4% increase in TAC. Decreasing the yield specification from 97.5 to 95% results in 5 and 3% decrease in the TAC at the industrial base case and at the optimized system, respectively. This is due to the fact that using the fixed

industrial membrane configuration, the yield specification can be fulfilled mostly by adjusting the reflux ratio of the distillation column.

#### 7.5. Partial permeate recycling

The superstructure is formulated in a way that enables partial recycling of the permeate streams. In spite of this opportunity, allowing partial recycling results in designs with total recycling, in all the industrial cases. On the other hand, partial recycling is found to be optimal only when the specific membrane investment cost is decreased by approximately 50% and the alcohol yield is set up for only 95%. However, this design alternative results in just 1.2% saving in TAC, compared to the optimal design with fixed total permeates recycling. This is due to the decreased mass load of the distillation column and to the somewhat less diluted feed of the membrane subsystem. It can be concluded that total permeate recycling is always more profitable when realistic ethanol yields are set up.

#### 7.6. Sensitivity analysis on overall membrane surface

It has been shown above, that in case of 97.5% ethanol yield, the total annual cost of the plant can be decreased by 12.2% by increasing the total membrane surface from 324 to 428 m<sup>2</sup>. Since the total membrane surface has a great influence on the total annual costs, additional designs were carried out at fixed 360 and 480 m<sup>2</sup> membrane surfaces. The dependence of the TAC and of the reflux ratio on the overall membrane surface

is shown in Fig. 22. Fig. 22 clearly shows that an overestimated total membrane surface results in less increase in the TAC compared to the case of using less membrane surface than the optimal. This is also because of the fact that the operating cost of the distillation column is high and so the TAC is quite sensitive to the reflux ratio of the column.

7.7. Sensitivity analysis on the membrane replacement cost

So far, the replacement cost of 1 m<sup>2</sup> of pervaporation

membranes had been taken 775 US\$. It is, however, interesting to examine how the membrane replacement cost influences the design and operational parameters of the optimal hybrid system. The price of the membranes may decrease with the development of the membrane technology or may increase because of some economical reasons. In order to explore the expectable effects of this possibility, optimal designs were carried out using different membrane replacement costs, varying from 40 to 120% around the original price. According to the results, the cost of the membranes in the investigated price interval does not considerably affect

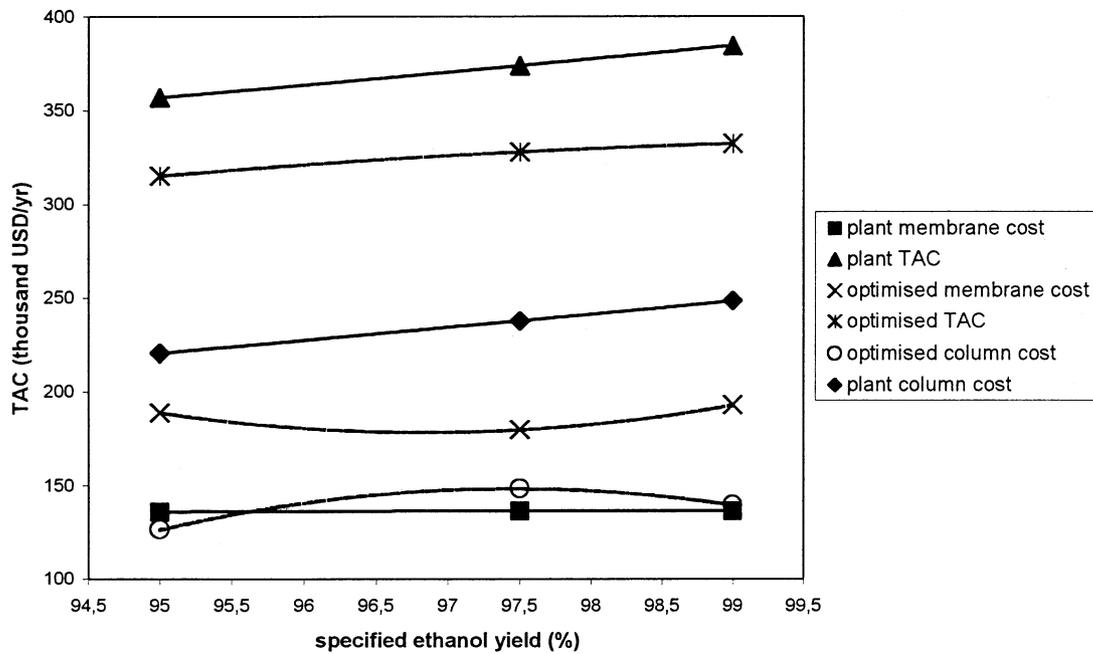


Fig. 20. Influence of the specified ethanol yield on the TAC. Optimized system vs. plant existing in the industry.

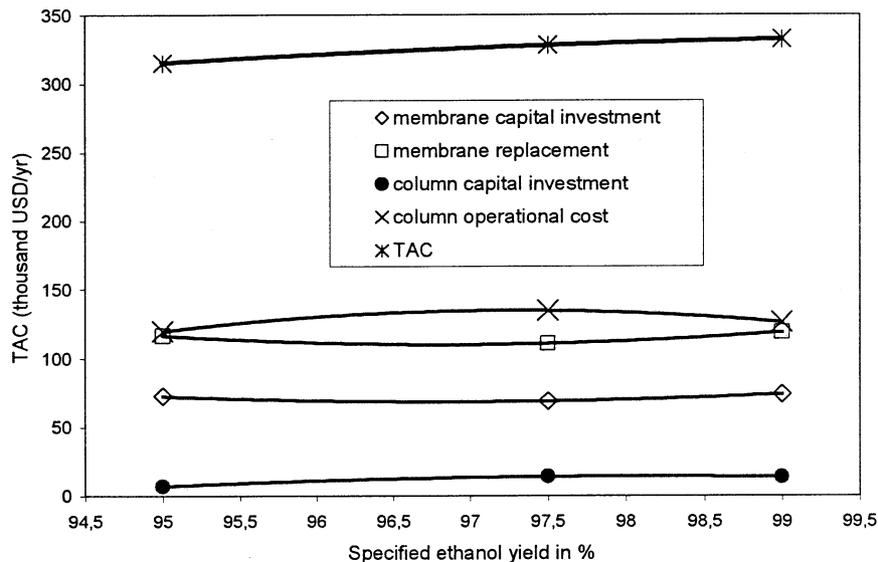


Fig. 21. Influence of the specified ethanol yield on the TAC. Optimized systems only.

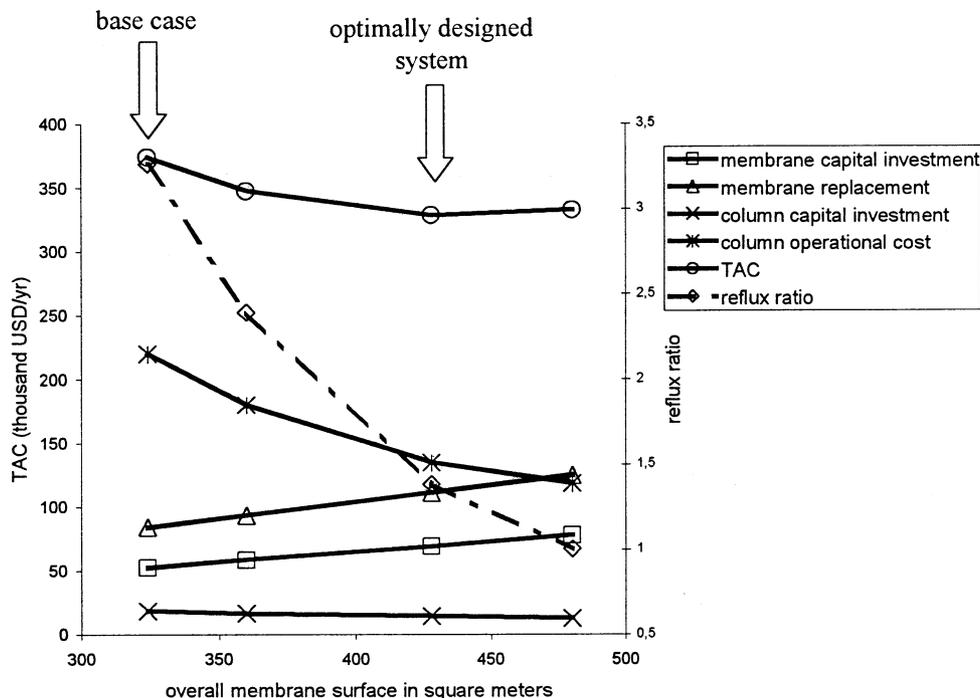


Fig. 22. Dependence of the TAC and the reflux ratio on the overall membrane surface at 97.5% ethanol yield.

the design and operational parameters of the optimal hybrid system. This is because, all in these cases, the membrane inlet concentrations are in the range of 9.7–9.9 mass% water, which is a narrow range. It is interesting to see that this range is situated near the constraint we applied regarding the membrane's toleration of concentrated water versus its expected lifetime. Hence, the pervaporation unit works at the same conditions, irrespective of the cost of the membranes.

## 8. Conclusions

Ethanol dehydration systems consisting of a distillation column and a pervaporation system are modeled and optimized using MINLP. The differential equations of a membrane module are numerically integrated. Quadratic and exponential regression, as well as metric (Shepard) and linear interpolation are studied for approximating the integral membrane model. A special form of exponential regression is selected due to its easy applicability and its relative good approximation properties. A method for radically decreasing the number of equivalent structures covered in the superstructure is suggested and applied. A successive refinement method with non-increasing number of binary variables, in order to avoid dimensional problems in solving MINLP, is developed and successfully applied.

The optimal design and operating parameters in-

cluding number of trays, feed location, reflux ratio, number of membrane sections in series and the number of membrane modules in each section are determined. Optimal structures with and without permeate recycle to the distillation column are presented.

Based on our industrial scale calculations 12% savings in the total annual cost can be achieved compared to an existing plant, by applying 32% additional membrane surface, by a radical decrease of the reflux ratio (3.3 to 1.4) in the column and by producing less concentrated alcohol in the distillate. According to our sensitivity analysis, the replacement cost of the membranes does not significantly influence the parameters of the system. This is because, in all the studied cases, the distillate concentration is situated in a narrow range of 9.7–9.9 mass% water, near the constraint we applied regarding the membrane's toleration of concentrated water versus its lifetime. In all the realistic cases, total recycling of the permeate flow proved to be optimal. On the other hand, partial recycling results in slightly cheaper designs in case of low alcohol yields and radically cheaper membranes.

## Acknowledgements

This work is supported by OTKA (Hung. Sci. Res. Fund) grants No. T030176, T021302 and F035085. We are grateful to Mrs Marta Lazi Lang for her valuable advise.

## Appendix A. Nomenclature

		$J_0$	membrane module inlet flow rate (kg/h)
		$J_T$	membrane module retentate flow rate (kg/h)
<i>Superstructure (letters appearing in the figures)</i>			
F	feed flow rate of the hybrid system	$T_0$	membrane module inlet temperature (°C)
RF	feed flow rate of the distillation column	$T_T$	membrane module retentate temperature (°C)
N	number of theoretical trays in the column	$J$	retentate flow rate (kg/h)
P1	flow rate of the column top product	$J'$	permeate flux (kg/m <sup>2</sup> h)
P2	flow rate of the column bottom product	$z$	membrane length (m)
P3	flow rate of the product (dehydrated ethanol)	$H$	permeate vapor enthalpy (kJ/mol)
P4	flow rate of the recycled permeate	$H'$	retentate vapor enthalpy (J/mol)
$\alpha_I$	recycle ratio of the permeate at section I	$h$	liquid enthalpy (kJ/mol)
$\eta$	specified ethanol yield	$p$	perimeter of the membrane (m)
<i>Distillation column costing</i>			
$C_{\text{shell}}$	installed cost of the column shell (US\$)	<i>Metric interpolation</i>	
$C_{\text{trays}}$	installed cost of the valve trays (US\$)	$i$	index for the basis points
M&S	Marshall and Swift index	$P_1(J_i, c_i)$	the $i$ th basis point
$D$	diameter of the column (m)	$P = (J, c)$	point where the surface is to be approximated
$H$	height of the column (m)	$r_i(P) =  P - P_i $	the Euclidean distance of the point $P_i$ from the point $P = (J, c)$
$A$	cross section of the column (m <sup>2</sup> )	$\hat{c}_T(P)$	interpolated value at the point $P = (J, c)$
$R$	reflux ratio	$\alpha$	tuning parameter of the metric interpolation
$M_D$	average molecular weight of distillate (kg/mol)	<i>Linear interpolation</i>	
$F_c^D$	distillate flowrate of component $c$ (mol/s)	$i$	index of the basis points
$\rho_m^D$	molar density of the distillate	$\mathbf{x} = (x_0, \dots, x_i$	equidistant domain of the basis points
$T$	temperature of the distillate (K)	$= x_0 + \Delta i, \dots, x_N \}$	
$\Delta T$	temperature difference	$\mathbf{f} = \{f_0 = f(x_0), \dots, f_i$	function values at the basis points
<i>Membrane model</i>			
$c$	retentate water concentration (mass%)	$= f(x_i), \dots, f_N \}$	
$c'$	permeate water concentration (mass%)	$k = (x - x_0) / \Delta$	distance of $x$ from $x_0$ in units of $\Delta$
$c_0$	membrane module inlet concentration (mass%)	$\mathbf{A} = \{y_0^A, \dots, y_i^A, \dots, y_N^A\}$	binary vector that is applied to define the location of the actual $x$ point
$c_T$	membrane module retentate concentration (mass%)	$\mathbf{B} = \{y_0^B, \dots, y_i^B, \dots, y_N^B\}$	binary vector that is applied to define the location of the actual $x$ point
		$y_i^A$	binary elements of the vector $\mathbf{A}$
		$y_i^B$	binary elements of the vector $\mathbf{B}$

**References**

- [1] S. Widagdo, W.D. Seider, *AIChE J.* 42 (1996) 96–130.
- [2] Z. Lelkes, et al., *AIChE J.* 44 (1998) 810–822.
- [3] J. Neel, *Membrane Separation Technology. Principles and Applications*, Elsevier, Amsterdam, 1995 (Chapter 5).
- [4] B.K. Srinivas, M.M. El-Halwagi, *Computers Chem. Eng.* 17 (1993) 957–970.
- [5] J. Viswanathan, I.E. Grossmann, *Computers Chem. Eng.* 17 (1993) 949–955.
- [6] U. Sander, P. Soukup, *J. Membr. Sci.* 36 (1998) 463–475.
- [7] A. Brook et al. *GAMS. A User's Guide*. Release 2.25, Boyd and Fraser, USA, 1992.
- [8] V.N. Stabnikov, et al., *Pishch. Prom. (Kiev)* 15 (1972) 49.
- [9] J. Gmehling, U. Onken, *Vapor–liquid equilibrium data collection*, vol. I, Part 1, Verlag + Druckerei Friedrich Bischoff, Frankfurt, 1977.
- [10] Z. Novak, et al., *Computers Chem. Eng.* 20 (1996) 1425–1440.
- [11] Z. Szitkai, Z. Lelkes, E. Rev, Z. Fonyo, in: S. Pierucci (Ed.), *Proceedings of ESCAPE 10*, Elsevier, Amsterdam, 2000, pp. 643–648.
- [12] W.J. Gordon, J.A. Wixom, *J. Math. Comp.* 32 (1) (1978) 253–264.