UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Universitat Rovira i Virgili

Escola Tècnica Superior d'Enginyeria Química

MODELING THE REVERSE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS

Dan Mihai Libotean

Submitted to the Department of Chemical Engineering in partial fulfillment of the requirements for the degree of Doctor at the Rovira i Virgili University

Tarragona, 2007

Supervised by:

Dr. Jaume Giralt i Marcé

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

El Dr. Jaume Giralt i Marcé, Catedràtic del Departament d'Enginyeria Química de la Universitat Rovira i Virgili de Tarragona,

Fa constar que el present treball, amb el títol

MODELING THE REVERSE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS

que presenta el doctorand **DAN MIHAI LIBOTEAN** per a optar al grau de Doctor en Enginyeria Química, ha estat dut a terme sota la meva immediata direcció i que tots els resultats obtinguts són fruit del treball i l'anàlisi realitzats per l'esmentat doctorand.

I per a què es faci saber i tingui els efectes que corresponguin, signo aquesta certificació.

Tarragona, Octubre de 2007

Dr. Jaume Giralt i Marcé

Catedràtic d'Enginyeria Química

Universitat Rovira i Virgili

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

...dedicated to my parents and my wife

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Acknowledgments

I would like to express my sincere appreciation to my supervisor, Dr. Jaume Giralt, for his generous advice, guidance, trust, encouragement and patience throughout the course of developing this doctoral thesis, and to the Departament d'Enginyeria Química of Universitat Rovira i Virgili for the financial support.

I am grateful to Dr. Francesc Giralt and Dr. Yoram Cohen for their valuable and constructive ideas and comments.

I also wish to acknowledge Dr. Robert Rallo and Dr. Joan Ferrer for sharing their knowledge in neural networks, and to Dr. Gabriela Espinosa for her help and advices.

I am grateful to the members of the Transport Phenomena research group at URV, and to the members of Polymer and Separation Research Laboratory at UCLA.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Abstract

One of the more serious problems encountered in reverse osmosis (RO) water treatment processes is the occurrence of membrane fouling, which limits both operation efficiency (separation performances, water permeate flux, salt rejection) and membrane life-time. The development of general deterministic models for studying and predicting the development of fouling in full-scale reverse osmosis plants is burden due to the complexity and temporal variability of feed composition, diurnal variations, inability to realistically quantify the real-time variability of feed fouling propensity, lack of understanding of both membrane-foulants interactions and of the interplay of various fouling mechanisms. In the present study, artificial neural network (ANN)-based models were developed based on direct analysis of experimental data for predicting process operation performance. Two approaches were considered; one based on characterizing the organic compounds passage through RO membranes, and a second one based on modeling the dynamics of permeate flow and separation performances for a full-scale RO desalination plant.

Organic solute sorption, permeation and rejection by RO membranes from aqueous solutions were studied via artificial neural network-based quantitative structure-property relationships (QSPR) for a set of 50 organic compounds for polyamide and cellulose acetate membranes. The separation performance for the organic molecules was modeled based on available experimental data achieved by radioactivity measurements to determine the solute quantity in feed, permeate and sorbed by the membrane. Solute rejection was determined from a mass balance on the permeated solution volume. ANN-based QSPR models were developed for the measured organic sorbed (M) and permeated (P) fractions with the most appropriate set of molecular descriptors and membrane properties selected using three different feature selection methods. Principal component analysis and self-organizing maps pre-screening of all 50 organic compounds defined by 45 considered chemical descriptors were used to identify the models applicability domain and chemical similarities between the organic molecules. The QSPR models predicted the M and P mass fractions within the range of experimental errors of the measurements. Somewhat higher prediction errors were encountered for a few chemicals that were not well represented within the present chemical

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

domain; however, the errors were consistent with the experimental standard deviations of the measurements. The ANN-based QSPRs were validated by means of a mass balance test applied not only to the 50 organic compounds used to develop the models, but also to a set of 143 new compounds. The quality of the QSPR/NN models developed suggests that there is merit in extending the present compound database and extending the present approach to develop a comprehensive tool for assessing organic solute behavior in RO water treatment processes.

The dynamics of permeate flow rate and salt passage for a RO brackish water desalination pilot plant were captured by ANN-based models. The effects of operating parameters, feed water quality and fouling occurrence over the time evolution of the process performance were successfully modeled by a back-propagation neural network. In an alternative approach, the prediction of process performance parameters based on previous values was achieved using a Fuzzy ARTMAP analysis. The neural network models built are able to capture changes in RO process performance and can successfully be used for interpolation, as well as for extrapolation prediction, fact that can allow reasonable short time forecasting of the process time evolution. It was shown that using real-time measurements for various process and feed water quality variables, it is possible to build neural network models that allow better understanding of the onset of fouling. This is very encouraging for further development of optimization and control strategies, based on soft sensors able to anticipate process upsets.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Abbreviations and nomenclature

AM1	Austin model 1
ANN	artificial neural network
ANNIGMA	artificial neural net input gain measurement approximation
ANQ	artificial neural network quantitative structure-property relationship
ANQ/PA	ANQ for the collection of for polyamide membranes
ANQ/PACA	ANQ for the collection of for polyamide and one cellulose acetate membranes
ART	adaptive resonance theory
ARTMAP	adaptive resonance theory map
ASTM	American Society of Testing Materials
ATR-FTIR	attenuated total internal reflection Fourier transform infra-red
ATSA	adjusted total surface area
bmu	best matching unit
BSE	backward step elimination
CA	cellulose acetate
CFS	correlation feature selection
C-plane	component plane
DB	domain boundary (border)
DPM	disintegrations per minute
EPF	element permeate flow rate
FA	Fuzzy ARTMAP
GA	genetic algorithm
GPM	gallons per minute
HOMO	highest occupied molecular orbital
LOO	leave one out
LUMO	lowest unoccupied molecular orbital
MGD	million gallons per day
MOPAC	molecular orbital package
MWCO	molecular weight cut-off
NF	nanofiltration
NN	neural network
PA	polyamide
PACA	polyamide
PCA	principal components analysis
QSPR	quantitative structure-property relationship
RBFNN	radial basis function neural network
RMS	root mean square
RMSE	root mean squared error
RO	reverse osmosis
SC	scintillation cocktail
SOM	self-organizing map
SOM-DA	self-organizing map dissimilarity analysis
STM	short term memory
TCF	temperature correction factor
TDS	total dissolved solids
TFC	thin film composite
U-matrix	unified distance matrix
WEKA	Waikato environment for knowledge analysis

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Symbol	Equation	Parameter
A	(1.1)	– solvent permeability coefficient [kg·m^{-2}·kPa^{-1}·s^{-1}]
A	(2.11); (4.8)	– input pattern to ARTa module of a Fuzzy ARTMAP neural network
$ANNIGMA_{ik}$	(2.21)	– ANNIGMA score between input variable i and output variable k
B	(1.3)	– solute permeability coefficient [m·s^{-1}]
B	(4.8)	– input pattern to ARTb module of a Fuzzy ARTMAP neural network
C	(2.10)	– number of clusters in a SOM
C_b	(1.7)	– bulk solution concentration [kg·m^{-3}]
C_b	(4.4); (4.6)	– brine concentration [mg/l]
$C_{b,a}$	(4.7)	– brine concentration at actual conditions [mg/l]
$C_{b,s}$	(4.7)	– brine concentration at standard conditions [mg/l]
C_f	(1.2); (1.6)	– feed concentration [kg·m^{-3}]
C_f	(4.6)	– feed concentration [mg/l]
$C_{f,a}$	(4.7)	– feed concentration at actual conditions [mg/l]
$C_{f,s}$	(4.7)	– feed concentration at standard conditions [mg/l]
CF	(1.8)	– concentration factor
C_m	(1.7)	– concentration at membrane surface [kg·m^{-3}]
Cn	(4.1); (4.2)	– conductivity [µS·cm^{-1}]
C_p	(1.2); (1.6); (1.7)	– permeate concentration [kg·m^{-3}]
D	(2.18)	– dissimilarity measure between two maps in the SOM-DA method
D_s	(1.2); (1.3)	– solute diffusion coefficient [m^2·s^{-1}]
E	(2.9)	– objective function in *k-means* algorithm
EPF_a	(4.7)	– measured average RO element permeate flow rate [GPM]
EPF_s	(4.7)	– average RO element permeate flow rate at standard conditions [GPM]
I	(2.12) – (2.15)	– input pattern to an ART module
J_p	(1.4); (1.7)	– permeate volumetric flux [m·s^{-1}]
J_s	(1.2); (1.4)	– solute mass flux [kg·m^{-2}·s^{-1}]
J_w	(1.1); (1.4)	– solvent mass flux [kg·m^{-2}·s^{-1}]
K_{ow}		– octanol-water partition coefficient
K_s	(1.2); (1.3)	– distribution (partition) coefficient solute-membrane
LG	(2.20); (2.21)	– local gain in the ANNIGMA approach
M	(3.1)	– sorbed solute fraction
$\hat{M}$	(3.5)	– predicted sorbed solute fraction
M_1, M_2	(2.18)	– trained self organized maps
$Merit_s$	(2.16)	– performance of a feature subset S in the CFS method
P	(3.1)	– permeate solute fraction
$\hat{P}$	(3.5)	– predicted permeate solute fraction
$P_{c,a}$	(4.3)	– measured concentrate pressure [kPa]
$P_{f,a}$	(4.3)	– measured feed pressure [kPa]
$P_{p,a}$	(4.3)	– measured permeate pressure [kPa]
$P_{c,s}$	(4.3)	– concentrate pressure at standard conditions [kPa]
$P_{f,s}$	(4.3)	– feed pressure at standard conditions [kPa]
$P_{p,s}$	(4.3)	– permeate pressure at standard conditions [kPa]
Q	(2.9); (2.10)	– cluster containing similar SOM units
Q_f	(1.5)	– feed flow rate [m^3·s^{-1}]
Q_p	(1.5)	– permeate flow rate [m^3·s^{-1}]
$Q_{p,a}$	(4.3)	– measured permeate flow rate [GPM]

ISBN:978-84-691-2701-8/DL:T.386-2008

Symbol	Equation	Parameter
$Q_{p,s}$	(4.3)	– standardized permeate flow rate [GPM]
R	(1.6); (1.8)	– rejection
R		– rejected solute fraction
$\hat{R}$		– predicted rejected solute fraction
R^2		– coefficient of determination
R_c	(3.5)	– calculated rejected fraction, based on a mass balance and the predicted sorbed and permeate fractions
S_c	(2.10)	– within-cluster distance in a SOM (the sum of the distances between each pattern that lies in the cluster and the cluster centroid)
$\%SP_a$	(4.7)	– measured percent salt passage
$\%SP_s$	(4.7)	– standardized percent salt passage
T	(2.4)	– target value in a back-propagation neural network
T	(4.4); (4.5)	– temperature [K]
T_j	(2.11)	– choice function in a Fuzzy ARTMAP neural network
TCF	(4.5)	– temperature correction factor
TCF_a	(4.3); (4.7)	– temperature correction factor at actual conditions
TCF_s	(4.3); (4.7)	– temperature correction factor at standard conditions
TDS_{NaCl}	(4.1); (4.2)	– equivalent NaCl total dissolved solids [mg/l]
W	(2.2); (2.3); (2.5); (2.6); (2.20)	– weights of a back-propagation neural network
X	(2.2); (2.3); (2.6)	– input pattern to back-propagation neural network
X	(2.1)	– un-normalized variable
X'	(2.1)	– normalized variable
Y	(1.5); (1.8); 4.6)	– recovery
Y	(2.2); (2.4)	– output of one neuron in the back-propagation algorithm
a	(2.11)	– input vector, which together with its complement forms the input pattern to an ART module of a Fuzzy ARTMAP neural network
a^c	(2.11)	– complement of the input vector a
b	(2.2); (2.3); (2.6)	– bias neuron in the back-propagation neural networks
bmu	(2.7); (2.8); (2.19)	– best matching unit
bmu'	(2.19)	– second best matching unit
c_j	(2.9)	– SOM cluster centroids
d	(2.18); (2.19)	– Euclidean distance
d_{ce}	(2.10)	– distance between the centroids of two clusters
df	(2.4); (2.5)	– derivative of the activation function in back-propagation algorithm
f	(2.2); (2.3)	– back-propagation transfer function
f		– solute mass in feed
$h_{bmu,i}$	(2.7); (2.8)	– neighborhood function in the SOM algorithm
k	(1.7)	– mass transfer coefficient [$m \cdot s^{-1}$]
k	(2.16)	– number of features in a subset S, in the CFS method
l	(1.3)	– membrane thickness [m]
m	(2.7); (2.19)	– weights of a Self Organizing Map neural network
m	(3.1)	– solute mass sorbed by the membrane
n	(2.6)	– iteration number in a back-propagation learning algorithm
n_h	(3.2)	– number of neurons in the hidden layer of a back-propagation neural network
n_i	(3.2)	– number of neurons in the input layer of a back-propagation neural network
n_o	(3.2)	– number of neurons in the output layer of a back-propagation neural network

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Symbol	Equation	Parameter
n_{tr}	(3.2); (3.4)	– number of data in the training set
n_{ts}	(3.4)	– number of data in the test set
p	(3.1)	– solute mass permeating the membrane
q^2	(3.3)	– explained variance in prediction
q^2_{tr}	(3.4)	– explained variance in prediction for the training set
q^2_{ts}	(3.4)	– explained variance in prediction for the test set
r	(2.17)	– Pearson's correlation coefficient
r	(3.1)	– solute mass rejected
r_{bmu}	(2.8)	– position of *best matching unit (bmu)* in a SOM
$\overline{r}_{ff}$	(2.16)	– average absolute feature-feature intercorrelation
$\overline{r}_{ft}$	(2.16)	– average absolute feature-target correlation in the CFS method
r_i	(2.8)	– position of unit i in a SOM
t	(2.7); (2.8)	– SOM training step
w	(2.11)	– weights of a Fuzzy ARTMAP neural network
w_J	(2.13); (2.14); (2.15)	– weights of the chosen category J in a Fuzzy ARTMAP neural network
x	(2.7); (2.9); (2.18); (2.19)	– SOM input sample
$\overline{y}$	(3.3)	– average fraction value of experimental data for all compounds
y_i	(3.3); (3.4)	– experimental fraction for the compound i
$\hat{y}_i$	(3.3); (3.4)	– predicted fraction for the compound i
$\overline{y}_{tr}$	(3.4)	– average value of the experimental data for all compounds in the training set
ΔA_i	(2.20)	– variation of the i^{th} input, in the ANNIGMA approach
ΔO_k	(2.20)	– variation of the k^{th} output in the ANNIGMA approach
ΔP	(1.1)	– membrane pressure gradient [kPa]
ΔW	(2.6)	– weights change in the back-propagation learning algorithm
Δb	(2.6)	– bias change in the back-propagation learning algorithm
$\Delta\pi$	(1.1)	– osmotic pressure difference [kPa]
α	(2.6)	– momentum term in back-propagation algorithm
α	(2.7)	– adaption coefficient in a SOM algorithm
α	(2.11)	– choice parameter in a Fuzzy ARTMAP neural network
β	(2.6)	– learning rate in back-propagation algorithm
β	(2.14)	– learning rate in Fuzzy ARTMAP algorithm
δ	(2.4) – (2.6)	– error gradient in back-propagation neural network
π_b	(4.4)	– brine osmotic pressure [kPa]
π_p	(4.4)	– permeate osmotic pressure [kPa]
$\pi_{b,a}$	(4.3)	– brine osmotic pressure at actual conditions [kPa]
$\pi_{p,a}$	(4.3)	– permeate osmotic pressure at actual conditions [kPa]
$\pi_{b,s}$	(4.3)	– brine osmotic pressure at standard conditions [kPa]
$\pi_{p,s}$	(4.3)	– permeate osmotic pressure at standard conditions [kPa]
ρ	(2.13)	– vigilance parameter in a Fuzzy ARTMAP neural network
ρ_a	(2.15)	– vigilance parameter of the ARTa module in a Fuzzy ARTMAP neural network
ρ_p	(1.4)	– permeate density [$kg \cdot m^{-3}$]
σ	(1.1)	– reflection coefficient
σ	(2.8)	– variance of the Gaussian used to define the neighborhood function in the SOM algorithm

ISBN:978-84-691-2701-8/DL:T.386-2008

Table of Contents

Abstract ..... **i**

Abbreviations and nomenclature ..... **iii**

Table index ..... **ix**

Figure index ..... **x**

1. Introduction **1**

Objectives 10

2. Theoretical fundaments **13**

2.1. Fouling modeling in reverse osmosis and desalination processes *13*

2.2. Artificial neural networks *22*

Back-propagation 23

Self-Organizing Map 26

Fuzzy ARTMAP 29

2.3. Methods for selection of the most suitable set of input parameters *32*

a) Correlation Feature Selection 32

b) Self-Organizing Map Dissimilarity Analysis 34

c) Artificial Neural Net Input Gain Measurement Approximation 35

3. Quantitative structure-property relationship for organic compound rejection in reverse osmosis membranes **37**

3.1. Experimental data, pretreatment and model development *37*

RO membrane characterization studies 41

Characterization of organic compounds 43

Data conditioning and selection of compounds belonging to the same chemical domain 47

Development and quality assessment of ANN models 52

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

3.2. Results *54*

Selection of model input parameters for Independent ANQ models 55

Selection of model input parameters for Membrane-Composite ANQ models 57

Selection of model input parameters for MP-Composite ANQ models 58

Correlating input descriptors for organic chemical separation performance 58

Performances of QSPR models for solute sorption, passage and rejection 60

3.3. Validating the QSPR models *76*

4. Reverse osmosis plant performance **83**

4.1. Full-scale RO data *83*

4.2. Back-propagation approach for modeling plant performance *88*

Data preprocessing and analysis 88

Results 90

4.3. Fuzzy ARTMAP approach for modeling plant performance *100*

5. Conclusions **105**

ANNEXES **109**

ANNEX I. Chemical structures of the 50 organic compounds used for developing QSPR models 109

ANNEX II. Molecular descriptors and reverse osmosis experimental data for the 50 organic compounds used to develop QSPR. 113

ANNEX III. Internal validation performance of the ANN-based QSPRs built. 125

ANNEX IV. External validation performance of the ANN-based QSPRs built. 129

ANNEX V. List of 143 new organic compounds used for validating the QSPR models 133

ANNEX VI. Performance of training, validation and test data sets for the normalized permeate flow rate and normalize salt passage models based on back-propagation 139

References **145**

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Table index

Table 3.1. Organic compounds with available experimental data, with identification of application and/or effects.....38

Table 3.2. Properties of membranes used for experimental analysis.....41

Table 3.3. Molecular and membrane descriptors used for developing QSPR models. 45

Table 3.4. Feature selection results for the Independent ANQ, Membrane-Composite ANQ/PA and MP-Composite ANQ models. 54

Table 3.5. Comparison between the range of variation of the seven molecular descriptors selected as input to the models, for the 50 chemicals with available experimental data and for the 143 organics without experimental data. 77

Table 3.6. Mass balance relative errors for the 193 compounds, for each of the three membranes.....78

Table 4.1. Average feed composition.....85

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Figure index

Figure 1.1. Schematic illustration of reverse osmosis process. 2

Figure 1.2. Spiral wound RO membrane module. 4

Figure 1.3. Concentration polarization and particle deposition in cross flow membrane filtration. 5

Figure 1.4. Variation of concentration factor with rejection and recovery. 6

Figure 1.5. Evolution of total cost and energy consumption for seawater desalination in Spain. 8

Figure 1.6. Location of the higher capacity RO desalination plants in Spain in 2006. 8

Figure 1.7. Evolution of RO desalinated water use in Spain. 9

Figure 1.8. Methodology for developing NN-based QSPR models. 11

Figure 2.1. Multilayer neurons architecture. 23

Figure 2.2. Single neuron model. 24

Figure 2.3. Different neuron activation functions. 24

Figure 2.4. Different SOM topologies. 27

Figure 2.5. Self-Organizing Map. 28

Figure 2.6. Fuzzy ARTMAP architecture. 30

Figure 2.7. CFS algorithm for selection of the "best feature subset". 33

Figure 3.1. Schematic illustration of solute sorption, permeation and rejection by the RO membrane in the experimental dead-end filtration mode. 42

Figure 3.2. Analysis of the chemical space. 49

ISBN:978-84-691-2701-8/DL:T.386-2008

Figure 3.3. Discriminant functional group for the compounds in the three families identified by the PCA. 51

Figure 3.4. LOO cross-validation of Independent ANQ models for the polyamide BW30 membrane. 62

Figure 3.5. LOO cross-validation of Independent ANQ models for the polyamide TFCHR membrane. 64

Figure 3.6. LOO cross-validation of Independent ANQ models for the cellulose acetate CA membrane....66

Figure 3.7. External validation for the Independent ANQ models for the BW30, TFCHR and CA membranes with the descriptors selected by CFS, SOM-DA and ANNIGMA for the M, P and R fractions corresponding only to the test set compounds....68

Figure 3.8. Membrane-Composite ANQ/PACA models. Internal validation....70

Figure 3.9. LOO cross-validation of MP-Composite ANQ models for the polyamide BW30 membrane. 72

Figure 3.10. LOO cross-validation of MP-Composite ANQ models for the polyamide TFCHR membrane. 73

Figure 3.11. LOO cross-validation of MP-Composite ANQ models for the cellulose acetate CA membrane. 74

Figure 3.12. External validation for the MP-Composite ANQ models for the BW30, TFCHR and CA membranes with the descriptors selected by CFS and ANNIGMA for the M, P and R fractions corresponding only to the test set compounds....75

Figure 4.1. Diagram flow of the two-stage RO plant from Port Hueneme, California, with identification of monitored process parameters....84

Figure 4.2. Conductivity-TDS correlations for permeate and feed-brine....85

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Figure 4.3. Variability of feed water TDS during the RO plant evaluation period. 86

Figure 4.4. Time evolution of the normalized salt passage and normalized permeate flow 87

Figure 4.5. Time evolution of the pressure drop along the membrane channel, for each one of the two stages and for the overall process. 87

Figure 4.6. Identification of input and output variables used for modeling the RO plant performance. 89

Figure 4.7. Statistical analysis for determination of the optimal network structure and optimal length of time intervals based on q^2 quality indices, for modeling the normalized permeate flow rate. 91

Figure 4.8. Average relative error for the best architecture (chosen based on q^2 index) for each length of time, for the normalized permeate flow rate models. 91

Figure 4.9. Statistical analysis for determination of the optimal network structure and optimal length of time interval based on q^2 quality indices, for modeling the normalized salt passage. 93

Figure 4.10. Average relative error for the best architecture (chosen based on q^2 index) for each length of time, for the normalized salt passage. 93

Figure 4.11. Normalized permeate flow rate predictions for the model built using 7 hrs time intervals and the neural network architecture 7:5:1. 94

Figure 4.12. Normalized salt passage predictions for the model built using 7 hrs time intervals and the neural network architecture 7:5:1. 95

Figure 4.13. Comparison of normalized permeate flow predictions for different length of time intervals, for the operational period 1750-1950 hrs. 96

Figure 4.14. SOM clustering of normalized permeate flow operational patterns. 97

Figure 4.15. Representation of normalized permeate flow operational patterns in 12 clusters. 97

Figure 4.16. Comparison of normalized salt passage predictions for different length of time intervals, for the operational period 1750-1950 hrs. 99

Figure 4.17. SOM clustering of normalized salt passage operational patterns. 99

Figure 4.18. Representation of normalized salt passage operational patterns in 11 clusters. 100

Figure 4.19. Input and output data configuration for the three parallel Fuzzy ARTMAP models developed for both normalized permeate flow rate and normalized salt passage....101

Figure 4.20. Test data set predictions for the normalized permeate flow rate using the Fuzzy ARTMAP approach. 102

Figure 4.21. Test data set predictions for the normalized salt passage using the Fuzzy ARTMAP approach. 103

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

1. Introduction

In the recent years, several factors have led to the development of membrane separation technology. The most important ones are the necessity of fresh water production for drinking, domestic, agricultural, landscape or industrial uses, the requirement of higher performance level methods for waste water reclamation and reuse applications, as well as lower regulatory maximum allowed levels of contaminants. Membrane processes are often chosen in water treatment technology since these applications achieve high removals of constituents such as dissolved solids, organic carbon, inorganic ions, and regulated and unregulated organic compounds. Reverse osmosis (RO) and nanofiltration (NF) membrane processes are used around the world for potable and ultra-pure water production, chemical process separations, as well as desalination of seawater (salinity around 35 g/l) and brackish water (less salty than the seawater). Moreover, lately there has been a growing interest in the integration of such membrane technologies for municipal and industrial water treatment, since they have been recommended as suitable for cost-effective desalination and removal of a wide range of low-molecular-weight trace organic constituents [1-9]. Organic compounds of particular interest include endocrine disruptors, human and animal antibiotics, disinfection by-products, insecticides and herbicides, and various pharmaceutical drugs. Many of these compounds have been detected in natural ecosystems at bioactive concentrations [10-12].

Reverse osmosis is a pressure driven membrane separation process, used for removing low molecular weight solutes, such as inorganic salts or small organic molecules, from a solvent. It relies on the use of a semi permeable membrane, which allows solvent molecules to pass through it, impeding the pass of solutes. When two solutions of different concentrations are separated by such a membrane, the solvent from the lower concentration solution will move through the membrane into the concentrated one, in a process called osmosis. The osmotic flow is attributed to the tendency to equalize the both size's solute concentrations. However, if the liquid on one side of the membrane is pure solvent, the two concentrations can never

ISBN:978-84-691-2701-8/DL:T.386-2008

be equal. In this case, the process of osmosis continues until the chemical potentials of both solutions are equal. This happens when the pressure exerted by the concentrated solution against the membrane is high enough to prevent any further solvent flow. The hydrodynamic pressure difference between the two solutions found at chemical potential equilibrium is called the osmotic pressure difference. In a reverse osmosis process, a pressure must be applied to the concentrated solution in order to overcome the osmotic pressure and to force the solvent to cross the membrane against the concentration gradient, as represented schematically in Figure 1.1 [3].

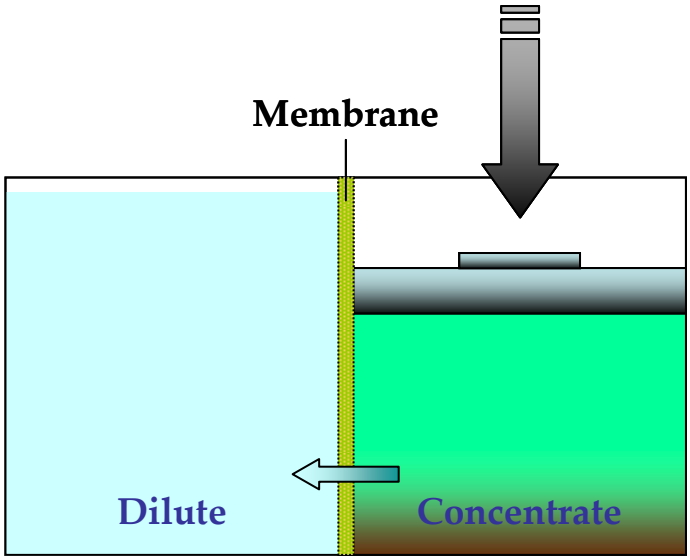


Figure 1.1. Schematic illustration of reverse osmosis process.

The solvent is driven through the membrane by pressure (convection), whereas the mass transfer of the solutes is diffusion controlled. Hereby, the permeation of the solvent through the membrane can be described using the pore flow model [13]:

$$J_w = A \cdot (\Delta P - \sigma \cdot \Delta \pi) \tag{1.1}$$

where J_w is the solvent mass flux that passes the membrane $\left[\mathrm{kg \cdot m^{-2} \cdot s^{-1}}\right]$, A is solvent permeability coefficient (characteristic for a given membrane) $\left[\mathrm{kg \cdot m^{-2} \cdot kPa^{-1} \cdot s^{-1}}\right]$, ΔP is the membrane pressure gradient $\left[\mathrm{kPa}\right]$, $\Delta \pi$ is the osmotic pressure difference $\left[\mathrm{kPa}\right]$, and σ is the reflection coefficient, which is a measure of the membrane selectivity (i.e., $\sigma = 0$, no membrane selectivity; $0 < \sigma < 1$, solute transport, not a completely semi permeable membrane; $\sigma = 1$, ideal membrane, no solute transport).

ISBN:978-84-691-2701-8/DL:T.386-2008

The solute mass flux can be described according to the solution diffusion model, in which the solute dissolves in the membrane and then diffuses through the membrane down a concentration gradient [14]:

$$J_s = \frac{D_s \cdot K_s}{l} \cdot \left(C_f - C_p\right) \tag{1.2}$$

where J_s is the solute mass flux that passes the membrane $\left[\mathrm{kg \cdot m^{-2} \cdot s^{-1}}\right]$, D_s is the solute diffusion coefficient in the membrane material $\left[\mathrm{m^2 \cdot s^{-1}}\right]$, K_s is the distribution or partition coefficient, l is the membrane thickness $[\mathrm{m}]$, and C_f, C_p are the concentrations in the feed and permeate solution $\left[\mathrm{kg \cdot m^{-3}}\right]$, respectively.

The solute permeability coefficient (B, $\left[\mathrm{m \cdot s^{-1}}\right]$), can be expressed as a function of diffusion and partition coefficients and membrane thickness, as

$$B = \frac{D_s \cdot K_s}{l} \tag{1.3}$$

The permeate volumetric flux (J_p, $\left[\mathrm{m \cdot s^{-1}}\right]$) can be calculated subsequently as the sum of the solute and solvent fluxes:

$$J_p = \frac{J_w + J_s}{\rho_p} \tag{1.4}$$

where ρ_p is the permeate density $\left[\mathrm{kg \cdot m^{-3}}\right]$.

Reverse osmosis performance can be expressed in terms of recovery and rejection. Recovery (Y) is defined as the fraction of the feed flow that passes through the membrane, as presented in Eq. (1.5). Rejection (R), defined in Eq. (1.6), expresses the extent to which a solute is rejected by the membrane.

$$Y = \frac{Q_p}{Q_f} \tag{1.5}$$

$$R = 1 - \frac{C_p}{C_f} \tag{1.6}$$

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

In Eqs. (1.5) and (1.6), Q is to the volumetric flow rate $\left[\mathrm{m}^3 \cdot \mathrm{s}^{-1}\right]$, and C denotes the concentration $\left[\mathrm{kg} \cdot \mathrm{m}^{-3}\right]$, while the subscripts p and f refer to the permeate and feed streams, respectively.

The choice of membrane material directly influences the separation efficiency, as the membrane characteristics influence the solvent and solute fluxes through the permeability coefficients. For obtaining a good efficiency, the membrane material must have high affinity for the solvent, and low affinity for the solute. The most common reverse osmosis membranes which attained the stage of economic application in water purification plants are made of cellulose acetate (CA) or polyamide (PA).

For most technical applications, RO membranes are used in cross flow design where water is flowing continuously over the membrane surface. Since the permeate flow is proportional to area of the membrane, spiral wound modules are used, obtained by rolling stacks of membranes with separating spacer mats into cylindrical shape unit. Such a configuration offers high surface area per unit volume. The salt solution is fed axially, the water permeates the membranes and flows radial toward the center of the cylindrical module where is collected in the permeate pipe, as presented in Figure 1.2.

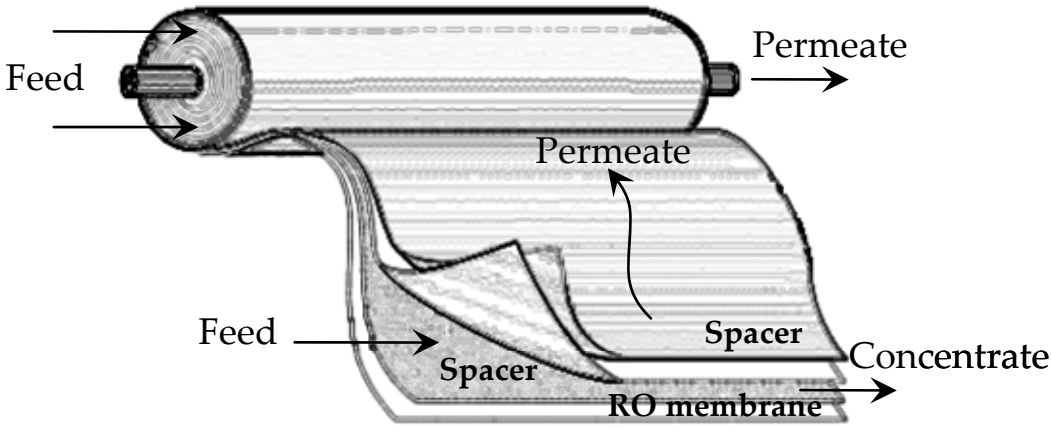


Figure 1.2. Spiral wound RO membrane module.

Membrane life-time and separation performances, quantity (i.e., water flux) and quality (i.e., salt rejection), are primarily affected by the flux inhibiting boundary layer effects, especially the phenomena of concentration polarization, fouling and scaling. When the feed solution flows over the membrane surface, as presented in Figure 1.3, the rejected species accumulate

ISBN:978-84-691-2701-8/DL:T.386-2008

next to the membrane surface forming a layer of higher concentration (C_m) than the one in the solution bulk (C_b) [15].

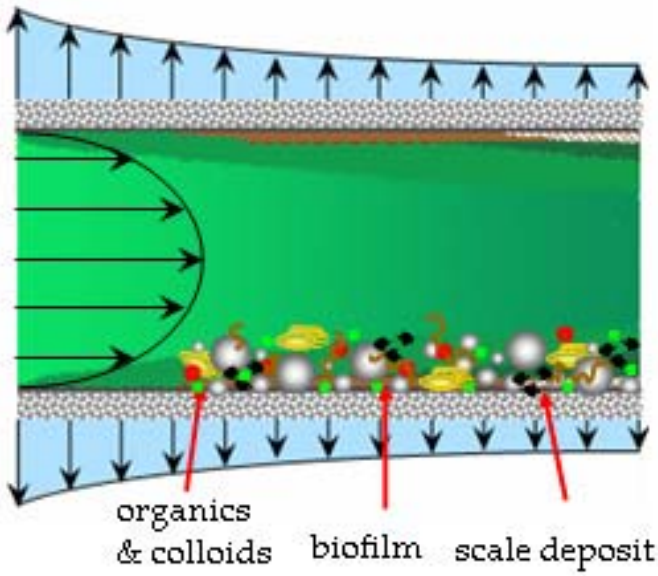


Figure 1.3. Concentration polarization and particle deposition in cross flow membrane filtration.

The occurrence of the high concentration layer near the membrane surface is called concentration polarization and expressed using Eq. (1.7) deduced from a film theory model. This phenomenon leads to an increase in the trans-membrane osmotic pressure, in the salt passage (i.e., the ratio between permeate and feed-brine concentrations), as well as the surface fouling and mineral scaling formation.

$$\frac{C_m - C_p}{C_b - C_p} = \exp\left(\frac{J_p}{k}\right) \tag{1.7}$$

In Eq. (1.7), C is the concentration $\left[\mathrm{kg \cdot m^{-3}}\right]$, with subscripts m, p and b referring to membrane surface, permeate flow and bulk solution, respectively, J_p is the permeate flux $\left[\mathrm{m \cdot s^{-1}}\right]$ and k is the mass transfer coefficient $\left[\mathrm{m \cdot s^{-1}}\right]$.

Membrane fouling is considered as a group of physical, chemical and biological effects leading to irreversible loss of membrane permeability. This phenomenon refers to the deposition of undesirable material on the membrane, that leads to the formation of one, or several layers on the membrane surface accompanied by plugging the membrane pores [16]. The main factors of fouling occurrence are the adsorption of feed components, clogging of pores, depositions of solids, crystallization and compaction of the membrane structure,

ISBN:978-84-691-2701-8/DL:T.386-2008

chemical interaction between membrane material and components of the solution, gel layer formation and bacterial growth. The foulants forming deposit on the membrane are sparingly soluble salts, dissolved organic substances, colloidal and particulate matter and microorganisms [17]. The accumulation of inorganic ions near and at the membrane surface can lead to an increase in their concentration, exceeding the solubility limits of various sparingly soluble mineral salts such as calcium carbonate, calcium sulfate and barium sulfate. These mineral salts may then crystallize directly onto the membrane surface, or can precipitate in the bulk near the membrane followed by deposition of formed crystals onto the membrane surface. This phenomenon leading to permeate flux decline and shorter membrane life time is called scaling [18].

A critical parameter that controls the deposition of undesired material onto the membrane surface is the concentration factor, defined as the ratio between the concentration of the rejected portion on the flow and the concentration of the feed flow ($CF = C_b/C_f$). In order to be economically feasible, membrane desalination processes have to be operated at high recovery levels (higher than 75%). Increasing the recovery at the high rejection level required (higher than 95%) leads to an excessive increase in the concentration factor, as presented in Eq. (1.8) and Figure 1.4, and accordingly, to an increase in the bulk concentration. Therefore, the increase of the concentration factor can lead to fouling and scaling formation.

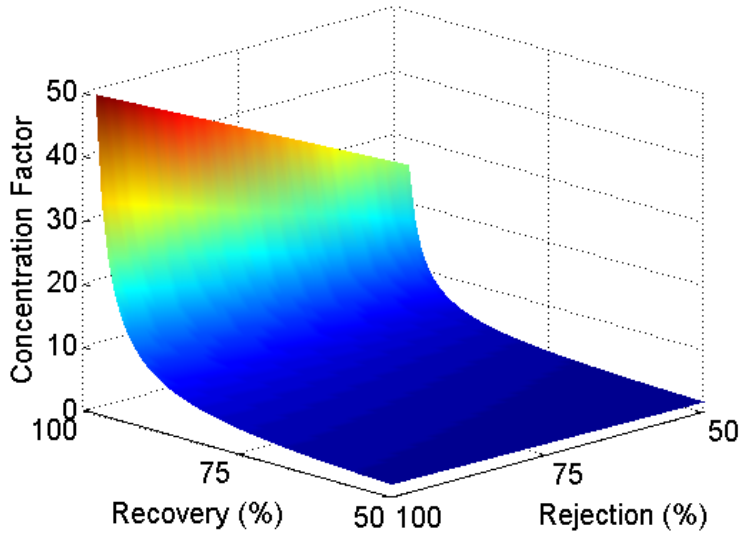


Figure 1.4. Variation of concentration factor with rejection and recovery.

$$CF = \frac{1}{1-Y} \cdot \left[1 - Y \cdot (1-R)\right] \tag{1.8}$$

ISBN:978-84-691-2701-8/DL:T.386-2008

The techniques used to reduce the concentration polarization are increasing flow rate, assembling an intensifier for turbulent flow, impulse and agitating methods, periodic depressurization of membrane tube, flow reversal, precoating of membrane surfaces and modification of membrane polymeric structure. Besides the use of all these methods for limiting the concentration polarization, fouling and scaling can be controlled by feed pretreatments and regular membrane cleaning [19].

Most of the applications of RO are in the purification of water, mainly the desalination of brackish and seawater to produce potable water, but there are also applications in food and diary industries, pharmaceutical and cosmetics production, water softening, ultra water production for electronic industries, as well as treatment of municipal and industrial wastewater and agricultural drainage water. Besides reverse osmosis or nanofiltration, there are several available technologies for desalinating water, such as distillation (multi stage flash, or multi effect evaporation), or other membrane separation techniques like electrodialysis (voltage-driven). All this techniques are comparable with respect to the produced water quality, the main difference between them being the production costs. The thermal methods, although very used, present high energy consumption and corrosion problems, due to the high operational temperatures. Comparing with the other available techniques, the quality of the feed is not so important for the thermal methods, since these systems are not susceptible for fouling. Nevertheless, scale formation still can be a problem [20]. Electrodialysis, although a membrane separation process, does not present so much risk of fouling or scaling, so it does not require a strict pretreatment of the feed water. Due to the high energy consumption which is proportional with the concentration of the feed salted solution, it is mainly used for brackish water desalination [21]. The pressure-driven membrane processes are considered to be the most promising methods for brackish and seawater desalination. They operate at ambient temperature, therefore presenting a small corrosion risk. The dimensions of equipment is smaller compared with other alternative methods, and one of the most important advantage is that even though they need high energy consumption, part of it can be recovered [20].

The desalination techniques are used in Spain since 1970s, when the first systems based on multi stage flash distillations were installed in Ceuta, Gran Canaria, Lanzarote and Fuerteventura. The total operation cost of the desalination systems, together with the energy

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

consumption presented a continuous decrease since then as presented in Figure 1.5. One reason is the development of membrane separation processes, and the continuous increase in the number of reverse osmosis-based desalination installations. In spite of the intense research carried out for improving the operation and lowering the energy consumptions of the desalination techniques, the total cost did not show a decrease in the later years. The reason is that even though the energy consumption tends to decrease, the cost of the energy presents an increasing trend [22,23].

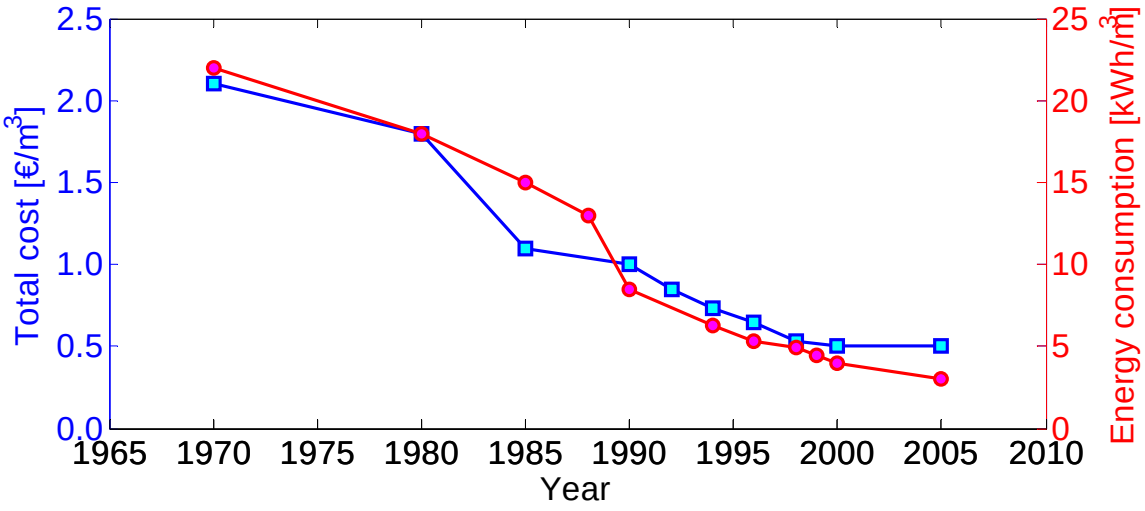


Figure 1.5. Evolution of total cost and energy consumption for seawater desalination in Spain.

Figure 1.6. Location of the higher capacity RO desalination plants in Spain in 2006.

ISBN:978-84-691-2701-8/DL:T.386-2008

The number of reverse osmosis plants used for water desalination in Spain increased a lot in the recent years. According to the Asociación Española de Desalación y Reutilización, in 2006 there were more than 900 water desalination RO plants in Spain, with the total water production capacity around $1.5 \cdot 10^6$ m^3/day. The location of the higher capacity RO desalination plants in Spain in 2006 is presented in Figure 1.6 [22].

The use of RO desalinated water increased in the recent years in Spain as presented in Figure 1.7, for all activity domains, like industrial, agricultural, or domestic use. As presented, the total use of desalinated water doubled from year 2000 to 2004, from 0.7 hm^3/day to 1.4 hm^3/day [22].

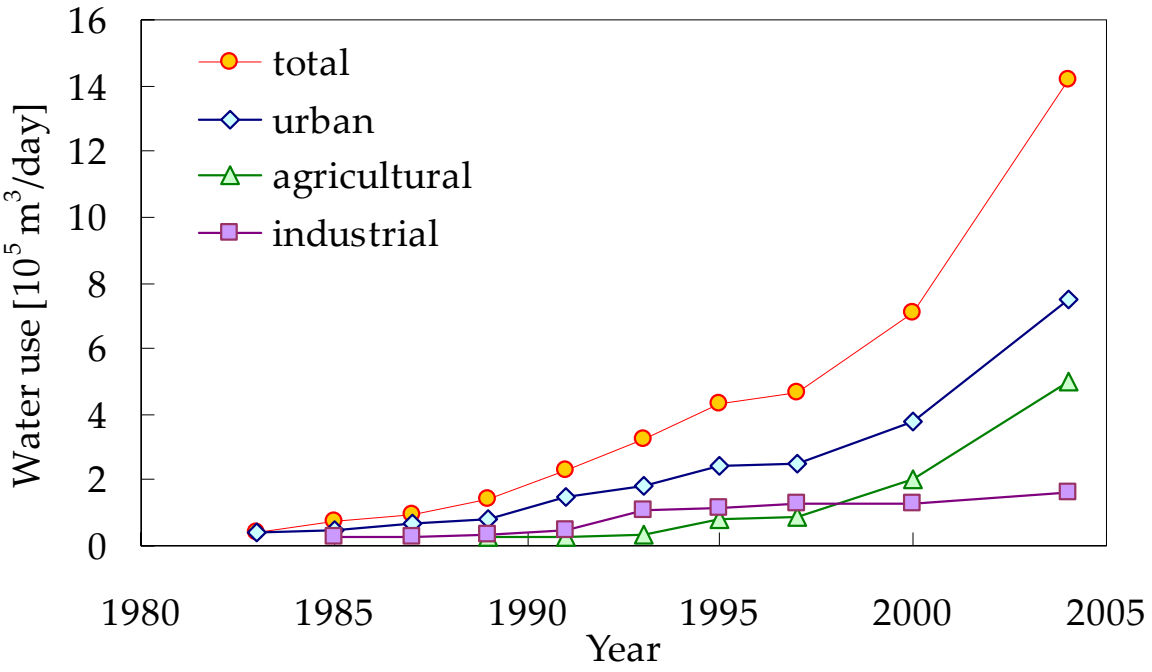


Figure 1.7. Evolution of RO desalinated water use in Spain.

To advance the efficient operation of modern RO membrane desalination plants it is necessary to establish an effective approach to model plant operation and to identify deviations in process conditions due to fouling and mineral salt scaling [15]. Ultimately there were several intents for developing theoretical approaches based on physical concepts to simulate the performance of the membrane separation processes [13-15,24-30]. Even though some of them showed some success, each attempt presented limitations, due to the complexity of the problem. Therefore, there are no general deterministic models available for predicting the development of fouling in full-scale RO plants. The major obstacles to

ISBN:978-84-691-2701-8/DL:T.386-2008

developing such predictive models are the complexity and temporal variability of feed composition, diurnal variations, the inability to realistically quantify the real-time variability of feed fouling propensity, lack of understanding of both the interplay of various fouling mechanism and the precise role of membrane surface properties and membrane interactions with various foulants and fouling precursors [30].

These drawbacks might be overcome by developing empirical models based on the direct analysis of the experimental data, and the use of artificial neural networks (ANN) seems to be a reliable option.

Objectives

The main objective of this study is to develop artificial neural network-based models for representing the RO membrane processes operation performance. In order to accomplish the presented purpose, specific sub-objectives are stated, as follows:

- To identify the molecular parameters of organic compounds and the membrane properties which determine and control the organics permeation through RO membranes, by applying different feature selection techniques.

- To establish correlations between molecular structure information, membrane properties and experimental fouling data regarding the organic compounds via artificial neural network-based quantitative structure-property relationships (QSPRs).

- To model the influence of organic compounds on the fouling processes occurring in reverse osmosis, by means of artificial neural networks.

- To describe the dynamics of a reverse osmosis plant performance, by integrating the effect of operating parameters, feed water quality and fouling phenomena occurrence on the time evolution of permeate flux and salt passage.

- To develop neural network models based on real experimental data from a full-scale RO pilot plant, to capture the plant performance evolution and to allow reasonable short term forecasting. This would allow a better understanding of the relationship between process condition and the onset of fouling, as well as the development of

ISBN:978-84-691-2701-8/DL:T.386-2008

optimization and control strategies and soft sensors able to anticipate the process upsets.

In order to achieve these objectives, two systematic approaches were demonstrated for the use of artificial neural network-based models to describe the RO process performance. The first methodology is based on developing quantitative structure-property relationships to correlate molecular properties of organic compounds and membrane parameters with experimental fouling data. A literature review indentified several studies which demonstrated the influence of molecular parameters over the organic compounds rejection by polymeric membranes. However, most of the existing approaches presented in the literature are focused on small number of compounds belonging to specific classes, and are based on describing single parameter influence over the membrane retention performances. For different types of membranes and classes of compounds considered, conflicting trends were observed between various molecular parameters and organics rejection. Therefore, ANN-based QSPRs constitute an effective approach which allows the development of general correlations considering the simultaneous influence of several molecular and membrane properties over the organics behavior when facing RO membranes.

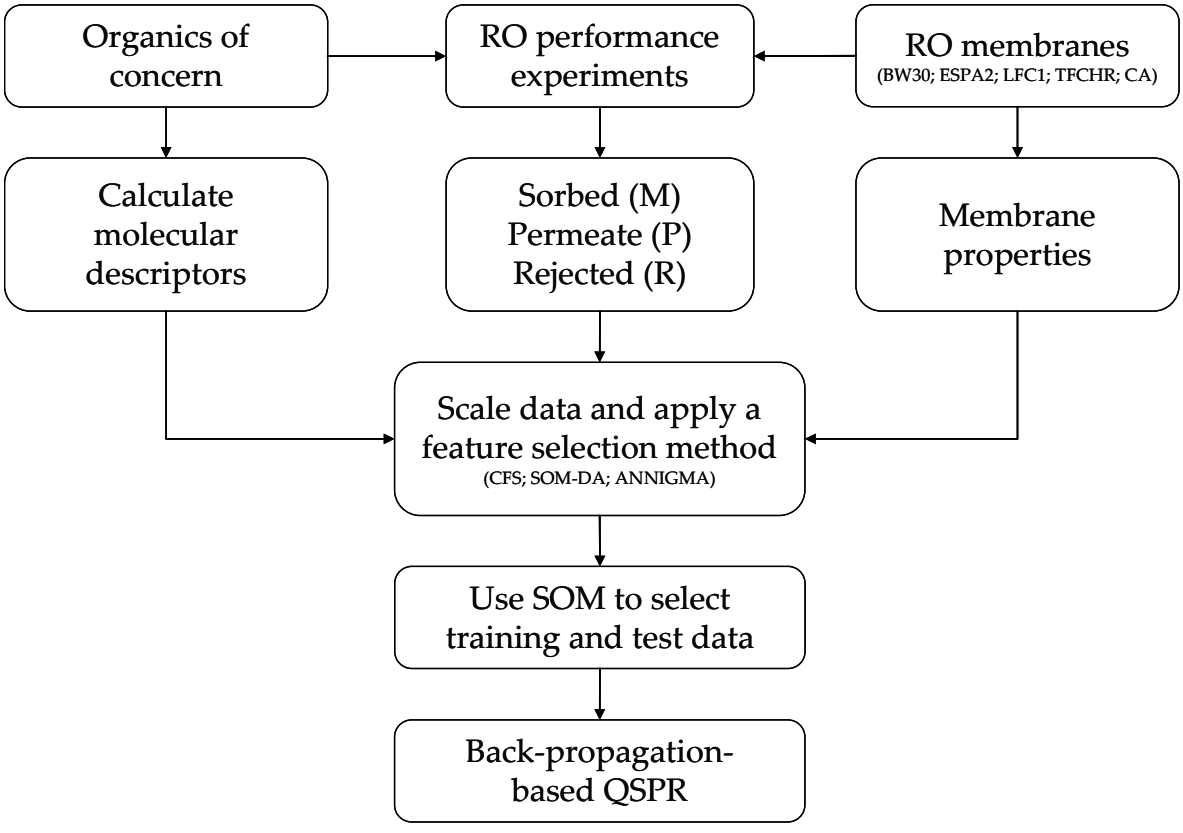


Figure 1.8. Methodology for developing NN-based QSPR models.

ISBN:978-84-691-2701-8/DL:T.386-2008

According to the methodology presented in Figure 1.8, RO experiments were performed for a number of 50 organic compounds, belonging to diverse classes and characterized by a set of 45 molecular descriptors, and 5 commercial membranes, characterized by 9 properties. The experimental data used to analyze the organics behavior in RO processes in terms of absorption, permeation and rejection, were provided by Orange County Water District, Los Angeles, California. Three different feature selection methods were used to select the smallest number of relevant input parameters among the molecular descriptors and membrane properties used to develop QSPR models for describing the passage and the sorbed fractions. The rejected fraction was further calculated from a simple mass balance using the predictions of the former two fractions.

The second methodology consists in modeling the dynamics of the reverse osmosis process performance parameters. The approaches available in literature have typically addressed the problem of system performance given a constant feed quality with ANN models used to describe permeate flux decline or variation of separation performances. A limited number of studies have explored the use of ANN to capture the dynamics of filtration processes in situations when feed quality may vary. Although it is well accepted that the ANN-based models can effectively describe the process performance variations, the approaches developed up to now proved to be successful for interpolation, but without the capability of forecasting. As previously mentioned, in order to optimize the design, operation and control of the membrane processes, is necessary to dispose of a model able to capture and forecast the process dynamics. Experimental data from a reverse osmosis brackish water desalination plant, provided by WaterEye Corporation, were used for developing ANN models. The ANN input and output variables were selected to ensure that they reveal clear information regarding RO process performance. A unique element of the present approach was the introduction of system memory effect, whereby past performance was considered in the predictions of future plant behavior. Back-propagation-based models were developed to describe the effect of operating parameters, feed water quality and fouling occurrence over the time evolution of process performance parameters. An alternative approach based on Fuzzy ARTMAP was developed for predicting process performance parameters based on previous values.

ISBN:978-84-691-2701-8/DL:T.386-2008

2. Theoretical fundaments

2.1. Fouling modeling in reverse osmosis and desalination processes

Many studies are devoted to the prediction, quantification and control of fouling, not only for the pressure-driven membrane filtration processes, but also for submerged bioreactors used for biological waste water treatment technology or even water circulating temperature controllers. Several approaches were proposed for fouling diagnosis in membrane filtration processes, like quantitative models for explaining the organic fouling based on solute properties [31-39], development of neural networks predictive models to describe the adverse impact of fouling occurrence over the process performance [40-51], or different methodologies for in-situ monitoring of these processes, e.g. the use of capacitive microsensors combined with ultrasonic time-domain reflectometry [52], or development of membrane fouling simulator [53]. Recent studies were developed to mathematically modeling the membrane fouling in submerged membrane reactors, providing more fundamental understanding of critical factors governing fouling in these systems [54,55]. A neural network-based specialized tool was developed to classify and diagnose the functioning mode of water circulation electrical controllers, successfully used for detection of simulated fouling in this system [56,57].

Studies on organic fouling of RO membranes have shown that the rejection of organic substances is governed by their physicochemical properties (e.g., molecular size, solubility, diffusivity, polarity, hydrophobicity, charge), membrane properties (e.g., permeability, pore size, surface roughness, hydrophobicity, charge), process operating conditions (e.g., flux, trans-membrane pressure, temperature, feed pH) and feed water composition [31-39,58-66]. The early work of Matsuura and Sourirajan [31] investigated the correlation of cellulose acetate rejection of 54 organic compounds (32 alcohols and phenols and 22 mono-carboxylic acids) as a function of the relative acidity of the molecule, estimated by the shift in the OH^-

ISBN:978-84-691-2701-8/DL:T.386-2008

band maximum in the IR spectra, and of the Taft number, which accounted for the effect of substituents on the polar effect of the organic molecule [67]. The rejection of alcohols and phenols was reported to decrease with increasing acidity with a steep change in rejection for the low acidity range. For mono-carboxylic acids, the rejection decreased with increased acidity (as represented by the pKa) to a minimum level, thereafter displaying increased rejection with increased acidity. The rejection decreased with increasing Taft number for alcohols, phenols and aliphatic mono-carboxylic acids, while a reverse trend was observed for substituted benzoic acids [31]. Kastelan-Kunst et al. [33] also reported that the rejection of organic compounds (2-ethoxy ethanol; 1,2-ethandiol; 2,2-dimethyl-1,3-propanediol; formaldehyde; 2-butanone; ethyl acetate; tetrahydrofuran) by FT30 cellulose acetate RO membranes, decreased linearly with increased Taft number. Van der Bruggen et al. [34] measured the rejection of four pesticides (i.e., atrazine; simazine; diuron; isoproturon) by four NF membranes (three polyamides and one polyethersulfone) and concluded that the rejection of organics of approximately the same size decreased with increasing solute dipole moment.

It is generally held that solute retention increases with increasing molecular size (which often correlates with molecular weight). However, several studies [37,38] have shown that even large molecules, such as certain endocrine disrupting compounds, can pass through RO membranes. Van den Bruggen et al. [35] correlated the rejection of 25 organics (including alcohols, ketones, esters, sugars and dyes) in NF membranes (two polyamides and two polysulfones) with solute size parameters, such as molecular weight, Stokes diameter and equivalent molar diameter (derived from molar volume), and a molecular diameter (obtained based on optimized molecular configuration). This study demonstrated that for RO and NF membranes organic solute rejection generally decreased with increasing dipole moment and increased with molecular size. Kiso et al. [37] reported that rejection of 14 pesticide by one RO membrane (polyamide) and three NF membranes (one polyamide and two polyethersulfone) increased with solute hydrophobicity as quantified by the organic solute octanol-water partition coefficient (K_{ow}). Rejection also increased with molecular weight and molecular width, i.e., a parameter computed based on the molecule projected area on a plane perpendicularly to the axis that connect the two most distant atoms [36]. In subsequent studies, using the same membranes, Kiso et al. [36,38] showed that the rejection of alcohols and saccharides increased with increased molecular width. However, no

ISBN:978-84-691-2701-8/DL:T.386-2008

significant relationship was observed between the rejection of aromatic compounds (11 alkyl phthalates and 7 mono-substituted benzenes) and the molecular size. Nevertheless, the rejection of these compounds increased with K_{ow}, with the best linear correlation ($R^2 = 0.812$) obtained for the mono-substituted benzenes. Rejection of alkyl phthalates was higher than 95% for 9 of the 11 compounds considered for membranes that displayed high NaCl rejection, irrespective of their K_{ow} values. For membranes with low NaCl rejection, high organic rejection (> 90%) was observed for compounds with $K_{ow} > 4.7$, while low organic rejection (< 40%) was obtained for compounds with $K_{ow} < 4$.

Ozaki and Li [32] evaluated for charged ultra-low pressure polyamide membranes the correlation of the rejection of 19 organic compounds (i.e., 5 alcohols; 9 phenols; acetic acid; urea; glucose; aniline and methyl chlorophenoxy acetic acid) with their molecular weight, molecular size and acid dissociation constant (pKa). At pH 5 and 9, organic solute rejection increased linearly (with $R^2 > 0.957$) with molecular weight in the range of 30-180 Daltons for 6 of the undissociated organics (i.e., methyl alcohol; ethyl alcohol; ethylene glycol; triethylene glycol; urea; glucose), excluding benzyl alcohol. Rejection also correlated linearly with molecular width ($R^2 > 0.943$) for the undissociated organics when triethylene glycol was excluded. The rejection of dissociated organics (i.e., 9 phenols; acetic acid; aniline and methyl chlorophenoxy acetic acid), however, did not correlated with neither molecular weight nor molecular width, but rejection did decrease linearly with the pKa at pH of 5, while two distinct and separable linear domains below and above pKa ≈ 7 was observed.

More recently, Kimura et al. [39] reported for a polyamide RO membrane an increased rejection with increased molecular weight for 11 organic compounds including 4 neutral endocrine disruptors (i.e., 4-phenylphenol; carbaryl; bisphenol A; 17β-estradiol), 5 pharmaceutical active compounds (i.e., phenacetine; primidone; isopropylantipyrine; carbamazepine; sulphamethoxazole), caffeine and 2-naphtol. These authors also noted, in agreement with previous studies [34], that the rejection of organic solutes of approximately the same size by a polyamide membrane decreased with increasing dipole moment. However, increased rejection with increased dipole moment was observed for the cellulose acetate membrane. Interestingly, for either the polyamide or the cellulose acetate membranes, there was no apparent correlation between organic solute rejection and the solute octanol-water partition coefficient.

ISBN:978-84-691-2701-8/DL:T.386-2008

Schutte [62] investigated the performance characteristics of two commercially available RO membranes (one cellulose acetate and one composite polyamide) with respect to rejection of 20 organic compounds including benzene, toluene, acetone, cyclohexane, 11 alkyl alcohols (methanol, ethanol, 1-propanol, 2-propanol, 1-butanol, 2-butanol, 2-methyl-1-propanol, 2-methyl-2-propanol, 1-pentanol, 1-hexanol and 1-heptanol), 7 alkyl phenols (phenol, 4-methyl phenol, 4-ethyl phenol, 2,6-dimethyl phenol, 4-n-propyl phenol, 4-isopropyl phenol and 4-n-butyl phenol). Reverse osmosis experiments were performed at three different operating pressures ranging from 1405 to 5620 kPa. The polyamide membrane rejection of linear alkyl alcohols increased with increasing molecular weight. The rejection of branched isomers was observed to be higher compared with the rejection of linear isomers of equal molecular mass. The polyamide membrane rejection of alkyl phenols, benzene and toluene increased linearly with molecular weight (the best linear correlation obtained $R^2 = 0.934$). In the case of cellulose acetate membrane no correlation was observed between the molecular weight of the considered compounds and their rejection. Moreover, cellulose acetate membrane showed lower rejection compared with the polyamide membrane. Since the organics passage through RO membranes depends on both sorption and diffusion, the solute flux was correlated with the adjusted total surface area (ATSA) of the molecules. The ATSA was calculated by adjusting the total cavity area of each molecule (parameter which gives a quantitative indication of the sorption of organic solute by the membrane), with a hydrodynamic shape factor (parameter which reflect differences in diffusion coefficient of the solutes). The logarithm of solute flux decreased linearly with increasing the adjusted total surface for both alkyl alcohols and alkyl phenols, with correlation coefficient of 0.960 and 0.940, respectively. However, it was noted that the developed correlations, consistently predicted higher solute fluxes for the branched isomers, meaning that the hydrodynamic factors considered did not account fully for the branching effect.

More recently, Bellona et al. [60] carried out a comprehensive literature review regarding factors affecting organics rejection and rejection mechanisms for NF and RO water treatment. The solute parameters identified to determine the organics rejection were molecular weight (found to be important especially for the non-charged, non-polar compounds), molecular size (length and width), molecular structure (e.g., number of methyl groups in the molecule), acid dissociation constant, hydrophobicity/hydrophilicity, polarity and diffusion coefficient. Membrane properties that affect rejection included molecular weight cut-off (MWCO),

ISBN:978-84-691-2701-8/DL:T.386-2008

desalting degree, porosity, morphology (i.e., roughness), hydrophobicity/hydrophilicity (i.e., contact angle) and surface charge (i.e., zeta potential). Moreover, feed water composition (i.e., pH, ionic strength, hardness, presence of organic matter) were also identified to influence the rejection. In addition, the authors proposed a qualitative classification of organic compounds. Ten categories were identified when grouping the organics by comparing their physico-chemical characteristics (i.e., molecular weight, acidity constant, hydrophobicity and molecular width) with membrane properties (i.e., MWCO, pore size, membrane charge) and operation parameters (i.e., pH). A general degree of rejection is given for each category, in terms of low, moderate and high.

Van der Bruggen et al. [61] extended the qualitative classification proposed by Bellona [60], using experimental data to develop a semi-quantitative approach for assessing the organics rejection. Following the classification algorithm proposed by Bellona [60], 42 organic compounds were clustered into the ten categories previously identified. Based on experimental analysis for 12 compounds and 3 RO membranes, and previous results reported in literature concerning the 42 organic compounds and 15 different RO membranes, expected rejection ranges were proposed for each category. They concluded that the categories including hydrophobic compounds are badly defined, since they include both compounds with low and high rejection. Moreover, they suggested that additional molecular, membrane and operation parameters might be considered for a full quantification of organics rejection.

Given the significant impediments in developing models based on phenomenological hypotheses to describe the dynamics of the RO processes, techniques focusing on direct analysis of experimental data were investigated. Hence, artificial neural network-based models have proved to be a viable alternative to model the plant performance variations using physically meaningful, easy and inexpensive to measure process parameters. Previous attempts of using artificial neural networks to describe dynamic filtration processes focused on modeling the permeate flux decline, or equivalent increase in total membrane resistance, as well as variations in separation performances, usually related to rejection.

ISBN:978-84-691-2701-8/DL:T.386-2008

Part of the available approaches for modeling membrane separation processes by means of ANN considered a steady-state procedure, in order to identify the influence of different process variables on the separation performances. Accordingly, Niemi et al. [48] used neural networks to simulate the reverse osmosis of aqueous ethanol and acetic acid solution, and ultrafiltration of a bleach plant effluent. Laboratory experiments considering a wide range of several process parameters (i.e., feed flow velocity, temperature, concentration and pressure) [68] were the basis of ANN models built to estimate the permeate flux and the rejection. The extreme experimental values were used in the training phase, and testing subsequently the model using the whole data set. The neural network (NN) predictions were slightly better than the ones obtained by a finely porous mass transfer model [68], reducing significantly the computational time. The influence of pressure, concentration and temperature of the feed over the permeate flow rate for a RO process using a spiral wound FilmTec SW30 membrane was investigated by Abbas and Al-Bastaki [40]. Different experimental runs were performed, varying the three feed parameters previously mentioned, while maintaining constant the feed flow rate and the permeate pressure. A 3:5:1 back-propagation neural network was trained using the experimental values measured for the extreme operating temperatures (i.e., 10 ºC and 30 ºC), and tested with the data corresponding to the intermediate temperature (i.e., 20 ºC). The predicted and experimentally determined permeate flow rates correlated linearly with the slope of the best line fit of 1.08, and the coefficient of determination $R^2 = 0.989$. However, when the experimental data corresponding to an extreme temperature value were selected for the test set, and the network was trained with the remaining data set, the model revealed poor performance. These results confirmed the expectation that ANN cannot be applied for data extrapolation (i.e., operational ranges that were not covered by the training data set).

Several studies addressed the problem of system performance evolution during the process operation. Dornier et al. [47] studied the use of neural networks for the case of raw cane sugar syrup microfiltration system, for integrating the effects of hydrodynamic conditions on the time evolution of the total hydraulic resistance of the membrane. Using a NN architecture with three inputs (i.e., time, trans-membrane pressure and crossflow velocity), two hidden layers (with 5 and 3 neurons, respectively, as resulting from an optimization process) and one output (i.e., total membrane resistance), it was showed that the best results were obtained when experiments in the centre and periphery of the parametric range were

ISBN:978-84-691-2701-8/DL:T.386-2008

used in training a model based on constant operating conditions. For this case, the total membrane resistance was predicted with the variation coefficient of 7.0% (defined as the ratio between the root mean squared error and the experimental mean value, expressed in percent), and the correlation between the predicted and experimental values characterized by $R^2 = 0.975$. The capacity of neural networks models to represent the evolution of process performance under variable operation conditions was also investigated. In this case, the network was trained with four different experimental runs with filtration time ranging from 140 to 180 hrs, and tested using three other sets when the filtration time varied from 100 to 180 hrs. Acceptable values were obtained for the variation coefficient and the coefficient of determination between the experimental and predicted membrane resistance on the whole data base (16.1% and 0.874, respectively). However, the total membrane resistance could not be well reproduced for one experimental run with a dynamic different from the ones used in training. Also, it is expected that the model can not be applied beyond the time range considered in training (i.e., maximum of 180 hrs).

Razavi et al. [49] studied the ability of neural network approach for the dynamic simulation of crossflow milk ultrafiltration under constant feed quality. Using laboratory experimental data, the permeate flux and the total hydraulic resistance were predicted as a function of operation time, pH and fat percent of the feed. A set of processing conditions was used for developing single curve simulation in order to enable the selection of optimum number and arrangement of training points. Subsequently, 6 experimental points for each set of feed quality conditions, including data corresponding to the beginning and the end of filtration period, were chosen for training a neural network model. As a result, using only 10% of experimental data for the learning base, a high accuracy model was built with the average relative error 1.06%.

Delgrange et al. [44,45] used the neural networks modeling for an ultrafiltration drinking water pilot plant to predict the hydraulic resistance and the trans-membrane pressure at the end of a filtration cycle and at the beginning of a next one. The best configuration of input parameters was found to include the turbidity of raw water, temperature and the permeate flow rate. Since process history influence the membrane performance, information from the beginning of the cycle and from the end of the previous one was considered. Prediction errors lower than 5% were obtained when modeling both cases of reversible and irreversible

ISBN:978-84-691-2701-8/DL:T.386-2008

fouling. Although turbidity was the only water quality parameter considered as input, good predictions were obtained also when feed water contained organic matter, as a result of considering history information. In a subsequent study, Delgrange-Vincent et al. [46] developed a model based on two feed-forward neural networks interconnected in a recurrent way, for predicting the productivity of an ultrafiltration pilot plant. The evolution of the total membrane resistance at the end of each operational cycle, and at the beginning of a new cycle after backwashing was predicted based on filtration operating parameters (i.e., permeate flow, filtration time), water quality parameters (i.e., turbidity, dissolved oxygen, pH, ultraviolet absorbency) and backwash operating parameters (i.e., backwash pressure and chlorine concentration). The model allowed good predictions even in the case of changing water quality and operating conditions, for both reversible and irreversible fouling, with 90% of the experimental points predicted with less than 10% of error. Shetty et al. [51] investigated the use of a neural network for predicting the time evolution of the membrane resistance in a drinking water nanofiltration process, for several configurations: flat membrane sheets, single and multiple spiral-wound elements, for both bench- and full-scale tests. Models based on back-propagation architecture, implementing a Levenberg-Marquardt learning algorithm were developed to relate influent flow rate (i.e., sum of feed water and recovery water flow rates), permeate flux, total dissolved solids (TDS) index, ultraviolet absorbance at 254 nm, pH and temperature of the feed water and operational time with the evolution of the total membrane resistance. When the recovery varied during the process operation, both feed flow rate and influent flow rate were considered as input parameters. The presence of experimental data corresponding to minimum and maximum values of each input parameter was assured in the training data set. It was shown that using only 10% of experimental points for training allowed the prediction of 93% of the data with an absolute relative error below 5%.

The dynamic rate of crossflow ultrafiltration of colloidal dispersions given a constant feed quality was predicted using ANNs by Bowen et al. [41]. The permeate flux decline was predicted based on ionic strength (i.e., measure of the average electrostatic interactions among ions in an electrolyte), zeta potential (i.e., electrostatic potential generated by the accumulation of ions at the surface of a colloidal particle that is organized into an electrical double-layer), time and applied pressure. In a first attempt, a 4:12:1 neural network was trained using 4 to 6 experimental points from the filtration profiles corresponding to extreme

ISBN:978-84-691-2701-8/DL:T.386-2008

pH and ionic strength conditions, and high, medium and low pressure values. Almost 84% of experimental data were predicted within the 10% error margin, with an average error of 5.6%. Greater prediction accuracy was obtained when training a 4:10:1 network with experimental data corresponding to extreme operating pressures for each pH-ionic strength combination, and intermediate operating pressure for one set of solution conditions. In this case, close to 95% of experimental data were predicted within the 10% error margin, with the average error of 3.6%. Chellam [42] investigated the use of ANN in simulation of transient permeate flux decline caused by polydispersed colloids during constant feed quality crossflow microfiltration. Fouling caused by three different types of rigid, stable particles with different size distribution under a wide range of hydrodynamic conditions was analyzed. The instantaneous permeate flux was modeled as a function of initial feed concentration, initial feed flux, entrance shear rate, instantaneous trans-membrane pressure and filtration time. For each one of the colloidal suspensions, an individual ANN model was trained using extreme values of input parameters. Using about 23% of the experimental data for training phase, accurate models able to predict the majority of observations (~95% of entire data set) with relative errors less than 10% were developed.

Chen and Kim [43] compared the performance of a radial basis function neural network (RBFNN), a regular multilayer feed-forward back-propagation neural network and a multiple lineal regression method for the prediction of the permeate flux decline in crossflow membrane filtration of colloidal suspension under constant feed quality [69]. The particle size of the suspended solids (SiO_2), solution pH and ionic strength, trans-membrane pressure and filtration time were used as input parameters. Training the networks with approximately 17% of the data selected to be equally spaced in time and including the extreme values of the experimental data, the best results were obtained in the case of the radial basis function neural network. In this case, 97% of test data were predicted with less than 10% of relative error, with the correlation between the predicted and experimental values characterized by $R^2 = 0.988$. Slightly worse results were obtained when using a back-propagation neural network, when 87% of test data were predicted with less than 10% of relative error, with the coefficient of determination between the measured and predicted values $R^2 = 0.958$. As expected, the worst results were obtained when using a multiple linear regression method for predicting the permeate flux decline.

ISBN:978-84-691-2701-8/DL:T.386-2008

Sahoo and Ray [50] used the same data set [69] as Chen and Kim [43] to develop a genetic algorithm (GA)-based method for searching the optimal geometry of a back-propagation neural network and a radial basis function network. The influence of training dataset size as well as the importance of scaling the data was also analyzed. The results confirmed that the models performance enhance when using a larger training dataset, and also, the use of scaled data slightly improve the performance of the models. Comparing their results with the ones obtained by Chen and Kim [43] (in terms of R-values between the predictions and the experimental values), it is concluded that the GA-optimized ANNs outperforms significantly the trial-and-error calibrated ANNs. Anyway, it is not very clear whether their performance index refers to the correlation coefficient (r), or to the coefficient of determination (R^2) like used by Chen and Kim [43], since comparing the root mean squared error (RMSE) values obtained in the two studies for the RBFNN, small improvement in models quality is seen only when using a large training dataset and scaling the data. In contrast with the conclusions of Chen and Kim [43], back-propagation neural networks provided better results than radial basis function neural networks, and this can be attributed to the use of optimum network geometry, found using the GA method.

2.2. *Artificial neural networks*

Artificial neural networks are numeric techniques able to capture and represent complex input-output relationships. They have the ability to learn linear, as well as non-linear correlative patterns between sets of input data and corresponding target values, directly from the data set that is modeled. They can also be successfully used in classification problems, since there are specific algorithms available to group the input patterns in different clusters based on similarities-dissimilarities between them. The ANN are characterized by processing units (neurons) and adjustable parameters (weights) [70].

In the ANNs approaches, data normalization is necessary before starting the training process, to ensure that the influence of the input variable in the course of model building is not biased by the magnitude of their native values, or their range of variation. The normalization technique used consist in a linear transformation of the input/output variables to the range [0,1] using the following expression:

ISBN:978-84-691-2701-8/DL:T.386-2008

$$X'_{ij} = \frac{X_{ij} - \min\left(X_j\right)}{\max\left(X_j\right) - \min\left(X_j\right)} \tag{2.1}$$

where X'_{ij} denote the normalized variable *j* for pattern *i* and $\min\left(X_j\right)$ and $\max\left(X_j\right)$ are the minimum and maximum values of that variable in the respective dataset.

For the predictive ANN algorithms used, the model performance is evaluated using the quality indices specifically defined for each particular application. Moreover, the average absolute and relative errors, standard deviation of the absolute and relative errors together with the maximum values of these errors are also reported.

Back-propagation is a neural network training method based on a forward flowing of information, and back-propagated error corrections. The back-propagation networks are usually organized in layers of neurons, as the architecture presented in the Figure 2.1.

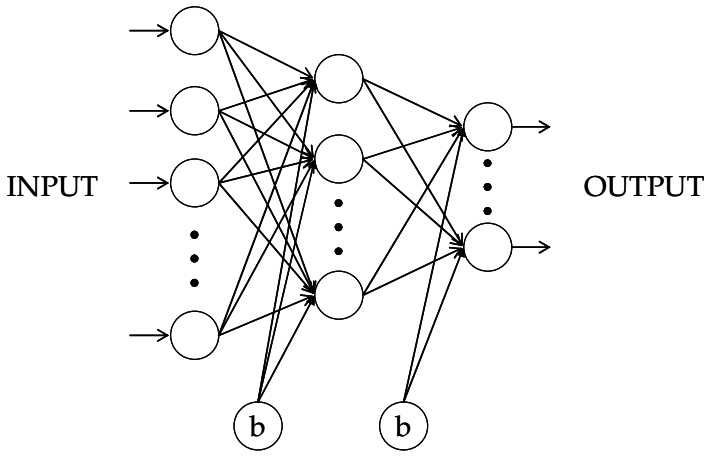


Figure 2.1. Multilayer neurons architecture.

Connections are made between the neurons of adjacent layers: a neuron is connected so that it receives signals from each neuron in the preceding layer and transmits signals to each neuron in the immediately succeeding layer. Usually, there are at least three neurons layers: an input layer which receives the input data, one or more hidden layers, and an output layer. Additionally, a bias neuron (*b*) that supplies an invariant output is connected to each neuron in the hidden and output layer [71].

ISBN:978-84-691-2701-8/DL:T.386-2008

Each processing element (neuron) receives a number of inputs, X_i. A weighted sum of these signals is calculated, using the neuron's assigned weights W_i, which is transformed by an activation function f to produce a single output signal Y, that is send to the neurons in the succeeding layer. The output of one neuron is calculated using the Eq. (2.2), as can be deduced from the sketch exposed in Figure 2.2:

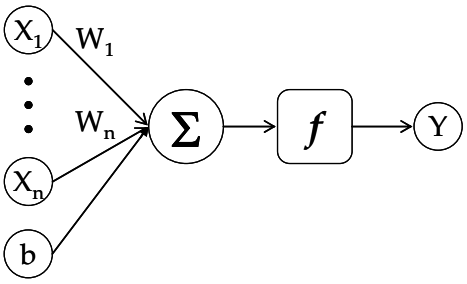


Figure 2.2. Single neuron model.

$$Y = f\left(\sum_i X_i \cdot W_i + b\right) \tag{2.2}$$

The activation function defines the output of the neuron in terms of the activity level at its input. Different expressions can be used for the neuron's activation function, like a step, sigmoid, tangent sigmoid or linear function, presented in Figure 2.3.

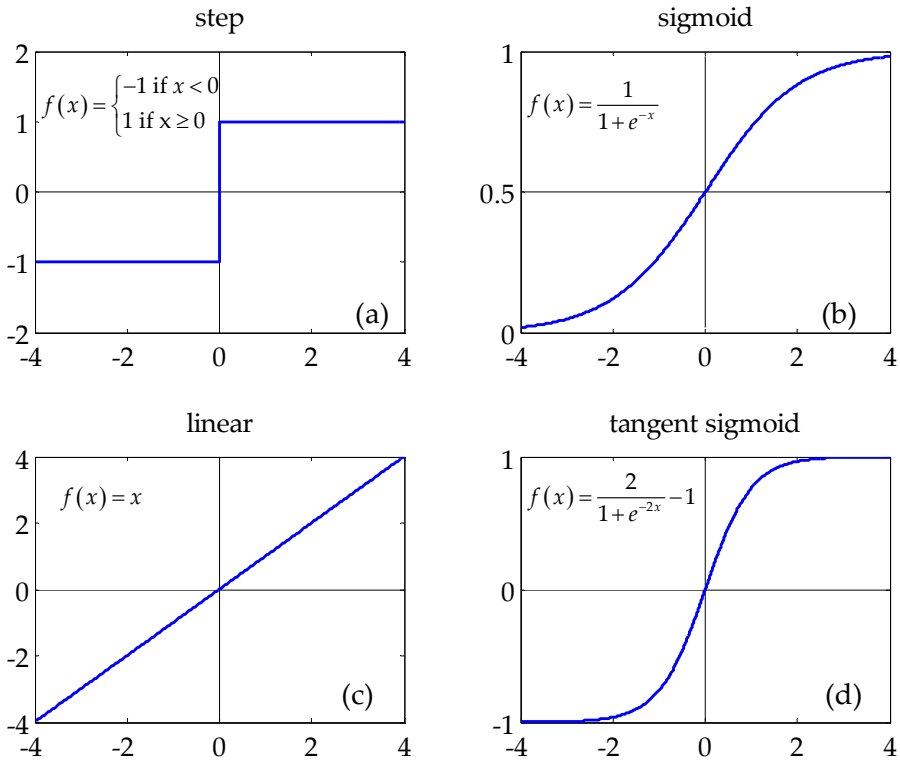


Figure 2.3. Different neuron activation functions: (a) step; (b) sigmoid; (c) linear; (d) tangent sigmoid.

ISBN:978-84-691-2701-8/DL:T.386-2008

The back-propagation training consists of two passes of computation: a *forward pass* and a *backward pass*. In the forward pass an input pattern vector is applied to the neurons in the input layer. The signals from the input layer propagate to the units in the first hidden layer, each one producing an output as described above. The outputs of these neurons are propagated using the same algorithm to units in subsequent layers until the signals reach the output layer where the actual response of the network to the input vector is obtained. Extending the formula for calculating the output of a single neuron (Eq. (2.2)) for the general case of any unit from any layer, leads to:

$$X_k^j = f^{j-1}\left(\sum_i X_i^{j-1} \cdot W_{i,k}^{j-1} + b^{j-1}\right) \tag{2.3}$$

where superscript *j* represents the layer number, while subscripts *i* and *k* represent the neurons indices in layer *j-1* and *j*, respectively. The networks' weights ($W_{i,k}^j$), that are fixed during the forward pass, are all adjusted during the backward pass in accordance with a back-propagated error signal for minimizing an error function [72].

For the output layer neurons, the error gradient is calculated based on the difference between the target value and the neuron's output (Eq. (2.4)), while for the hidden layer neurons the error gradient is determined by calculating the weighted sum of errors at the previous layer, as expressed in Eq. (2.5).

$$\delta_k^o = df^o \cdot \left(T_k - Y_k\right) \tag{2.4}$$

$$\delta_i^{j-1} = df^{j-1} \cdot \sum_k \delta_k^j \cdot W_{i,k}^j \tag{2.5}$$

In Eq. (2.4) the superscript *o* represents the output layer, δ_k^o is the error gradient of the k^{th} neuron, df^o is the derivative of the activation function, Y_k is the output of the k^{th} neuron, while T_k is the k^{th} target variable. In Eq. (2.5), δ_i^{j-1} is the error gradient of the i^{th} neuron from the layer *j-1*, and df^{j-1} is the derivative of the activation function of the layer *j-1*.

The principle used for weights adaptation is also known as generalized delta rule. Once the error gradients are evaluated for every layer, the biases and the weights are updated according to the equations [73]:

ISBN:978-84-691-2701-8/DL:T.386-2008

$$
\begin{aligned}
&W_{i,k}^{j}(n+1) = W_{i,k}^{j}(n) + \Delta W_{i,k}^{j}(n) \\
&\Delta W_{i,k}^{j}(n) = \beta \cdot \delta_{k}^{j} \cdot X_{i}^{j} + \alpha \cdot \Delta W_{i,k}^{j}(n-1) \\
&b_{i}^{j}(n+1) = b_{i}^{j}(n) + \Delta b_{i}^{j} \\
&\Delta b_{i}^{j} = \beta \cdot \delta_{k}^{j} + \alpha \cdot \Delta b_{i}^{j}(n-1)
\end{aligned}
\tag{2.6}
$$

where n is the iteration number, while α and β are two parameters characterizing the learning process. Here, β is the learning rate and α is the momentum term introduced to improve the convergence by taking into account the effect of the weights changes from the previous iteration.

A more efficient method used for weights adaptation is the Levenberg-Marquardt algorithm [74,75], which is a combination between the gradient descent rule and the Gauss-Newton method. The algorithm uses a parameter to decide the step size, which takes large values in the first iterations (equivalent with the gradient descent algorithm), and small values in the later stages (equivalent with the Gauss-Newton method). It combines the ability of both methods (i.e., convergence from any initial state in the case of gradient descent, and rapid convergence when reach the vicinity of the minimum error in the case of Gauss-Newton method) while avoiding their drawbacks [71,76].

For the learning phase, the data must be divided in two sets: the training data set, which is used to calculate the error gradients and to update the weights, and the validation data set, which allows to select the optimum number of iterations in which the networks learns general information from the training set. As the number of iterations increases, the training error drops whereas the validation data set error begins to drop, then reaches a minimum and finally increases. Continuing the learning process after the point when the validation error arrives to a minimum leads to a process called over-fitting, when the network became specific to the pattern vectors that form the training data set. After finishing the learning process, another data set (test set) is used to validate and confirm the prediction accuracy [44].

Self-Organizing Map (SOM) is a tool for visualization and classification of high-dimensional data, by implementing an orderly mapping onto a regular low dimensional grid, while preserving the relations between the input patterns [77,78]. The map (network) consist of a

ISBN:978-84-691-2701-8/DL:T.386-2008

set of units (neurons or cells), originally arrange in physical positions according to a topology function, since the map can have a rectangular or hexagonal grid, with plate, cylindrical or toroidal shape, as depicted in Figure 2.4. During the learning process, the map adapts itself to represent all the available input patterns ordered on the grid so that similar samples are close to each other and dissimilar ones far from each other [79].

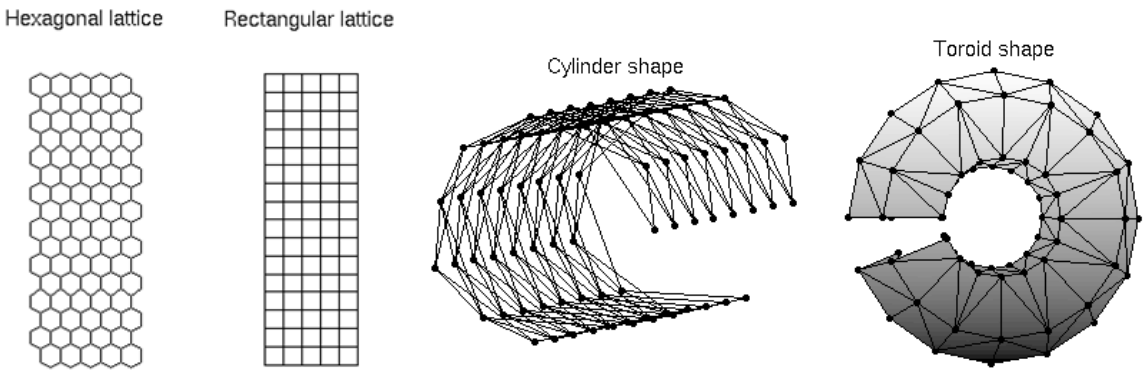


Figure 2.4. Different SOM topologies.

At the beginning of the learning process, a prototype vector (often randomly initialized) of the same dimension as the input data vectors is assigned to each map unit. At presenting a current input pattern, it is simultaneously compared with all the map's neurons, in order to express the dissimilarity between it and each prototype in terms of a general distance function, in most of the cases the Euclidian one. The best matching unit (*bmu*), which is the network's cell with the prototype most similar to the input, is selected (Figure 2.5). The next step is to update the weights of the network, by moving the best matching unit and its topological neighbors closer to the input vector in the input space. The update rule for the prototype vector of unit *i* is:

$$m_i(t+1) = m_i(t) + \alpha(t) \cdot h_{bmu,i}(t) \cdot \left[x - m_i(t)\right] \tag{2.7}$$

where m_i is the prototype vector of unit *i*, *t* is the training step, *x* is the input vector, $\alpha(t)$ is the adaptation coefficient also called the learning rate factor, and $h_{bmu,i}(t)$ is the neighborhood function, often taken to be the Gaussian function expressed in Eq. (2.8) centered on the winner unit denoted *bmu*.

$$h_{bmu,i}(t) = \exp\left(-\frac{\left\|r_{bmu} - r_i\right\|^2}{2 \cdot \sigma^2(t)}\right) \tag{2.8}$$

ISBN:978-84-691-2701-8/DL:T.386-2008

In Eq. (2.8), r_{bmu} and r_i are the positions of neurons *bmu* and *i*, respectively, on the SOM grid and $\sigma^2(t)$ is the variance of the Gaussian. Both $\alpha(t)$ and $\sigma(t)$ decrease monotonically with time, starting with higher values for the ordering phase in which a rough classification is achieved, and attaining smaller values in the tuning phase when a fine adjustment of the map is performed [77,78].

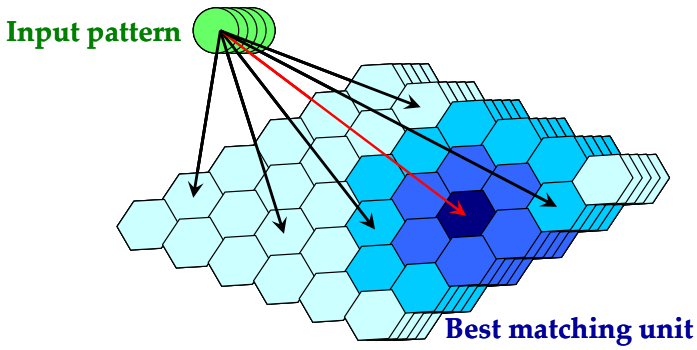


Figure 2.5. Self-Organizing Map.

This process is repeated until the classification stabilizes, i.e. no more adaptations are needed. Once the SOM has been trained, several methods can be used for visually inspect the results of the clustering process. The most widely used method is based on the unified distance matrix (U-matrix), which indicates the overall shape of the map by means of distances between prototype vectors of neighboring map units in the original grid. A graphical representation of the contribution of each variable of the input vector to the clustering process is obtained by extracting the component planes (C-planes) [80].

The number of formed clusters can be found by applying a partitioning algorithm, to a new data set that consists of the prototypes of the trained network. A partitioning algorithm organizes the data set into a number of clusters by minimizing some criterion or error function [80]. One of the simplest unsupervised algorithms to solve the clustering problem is the *k-means*. The procedure is based on defining *k* centroids (c_j), one for each cluster (Q_k), and group the data set into these clusters by minimizing the following objective function (E):

$$E = \sum_{j=1}^{k} \sum_{x \in Q_k} \left\| x - c_j \right\|^2 \tag{2.9}$$

This can be achieved by an iterative procedure that consists in associating each sample form the data set to the nearest cluster (in terms of distance to the cluster centroid), and then recalculate new centroids as the centre of gravity of the clusters resulted from the previous

ISBN:978-84-691-2701-8/DL:T.386-2008

step. The process must be repeated until the centroids do not change their location any more. The *k-means* algorithm does not necessarily find the most optimal configuration, because it is very sensitive to the initial randomly selected centroids. In order to reduce this effect, the procedure can be run multiple times [71].

The decision criterion of the partitioning algorithm, which identifies the proper number of clusters and their distribution, is a validity index. According to Davies-Bouldin index [81], the best clustering minimizes the following function:

$$f(k)=\frac{1}{C}\sum_{k=1}^{C}\max_{l\neq k}\left\{\frac{S_c(Q_k)+S_c(Q_l)}{d_{ce}(Q_k,Q_l)}\right\} \tag{2.10}$$

where *C* is the number of clusters, S_c is the within-cluster distance (the sum of the distances between each pattern that lies in the cluster and the centroid), d_{ce} is the distance between the centroids of two clusters. According to this index, the partitioning that offers the most compact clusters and well separated from each other is selected.

Fuzzy ARTMAP

ARTMAP is a class of neural network architectures designed for classification, based on Adaptive Resonance Theory (ART), which perform incremental supervised learning of recognition categories and multidimensional maps. The fuzzy ARTMAP neural network is formed by a pair of fuzzy ART modules linked by an associative memory and an internal controller, as shown in Figure 2.6 [82].

During the supervised learning, one of the ART modules (ART_a) receives a set of input patterns A, meanwhile the other module (ART_b) receives a corresponding set of input patterns B, which is the correct prediction given A. Each module performs a classification of the input received, and then a linking of the categories is realized by an associative learning network. Every input pattern to the ART modules must be presented in complement coding form including the input vector (a) and its complement (a^c), as expressed in Eq. (2.11). Therefore, the input vector *a* must previously be normalized using Eq. (2.1), so that each of its components lies in the interval [0,1].

ISBN:978-84-691-2701-8/DL:T.386-2008

$$A = \left(a, a^c\right);\ a_i^c = 1 - a_i \tag{2.11}$$

This is a normalization rule that preserves amplitude information [83].

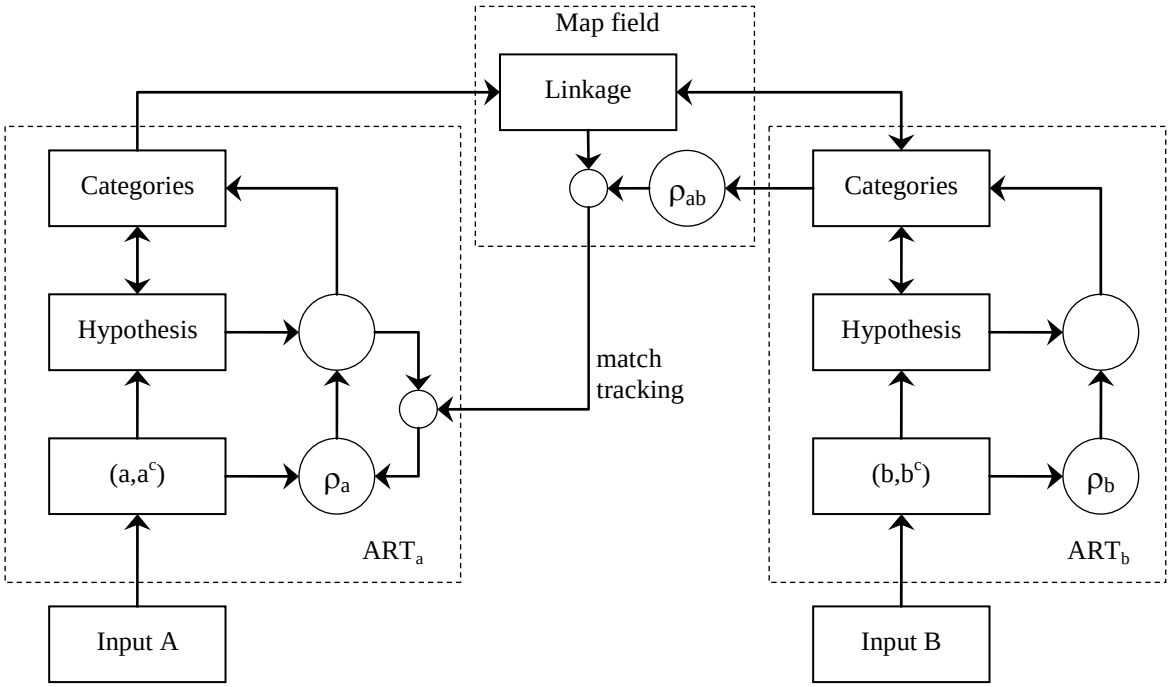


Figure 2.6. Fuzzy ARTMAP architecture.

In each ART module, the input pattern (*I*) must be asses to a category, characterized by a set of adaptive weights (*w)*. The classification procedure of fuzzy ART is based on the Fuzzy Set Theory [84]. Accordingly, the category is chosen based on a choice function, *Tj*, by comparing the input pattern with the weights of each one of the existing categories:

$$T_j(I) = \frac{\left|I \wedge w_j\right|}{\alpha + \left|w_j\right|} \tag{2.12}$$

where w_j is a vector of adaptive weights for the j^{th} category; α is the choice parameter ($\alpha > 0$), $\wedge$ is the fuzzy intersection operator defined by $\left(p \wedge q\right)_i = \min(p_i, q_i)$ for any vectors *p* and *q* having the same dimension, and the norm $|\cdot|$ is defined as the sum of the components of the vector.

The category with the maximal choice function is chosen (noted by index *J)*, and thereupon is checked whether the resonance occurs. This happens if the match function of the chosen category meets the vigilance criterion:

ISBN:978-84-691-2701-8/DL:T.386-2008

$$\frac{\left|I \wedge w_J\right|}{|I|} \geq \rho \tag{2.13}$$

In Eq. (2.13) ρ is the vigilance parameter ($\rho \in [0,1]$), which controls the number of created categories and allows the implementation of desired accuracy criterion in the classification procedure. This parameter calibrates the minimum confidence that an ART module must have in a recognition category activated by an input pattern, in order to accept that category rather than search for a better one [83,85]. If the vigilance criterion is not satisfied, the mismatch reset occurs, and the next category possessing high value for the choice function is chosen. The search process continues until a category to meet the resonance criterion is found, or if this is not achieved, a new category is created. Once a category is selected for the presented input pattern, its weight vector is updated according to the Eq. (2.14), using a learning rate parameter, $\beta \in [0,1]$:

$$w_J^{new} = \beta \cdot \left(I \wedge w_J^{old}\right) + (1-\beta) \cdot w_J^{old} \tag{2.14}$$

The associative memory records the link between the classes corresponding to the input patterns presented to each ART module. The internal controller supervises if the new link is in contradiction with any other previously recorded. If no contradiction is found, the link is recorded. Otherwise, the input pattern for the ART_a module is reclassified with a larger vigilance parameter in a process called match tracking. It enables the neural network to learn about similar patterns with different consequences, by sacrificing a minimum amount of generalization in order to correct a predictive error. Therefore, the vigilance parameter is set to be slightly larger than the match function, using a small positive infinitesimal quantity (ε) [86,87].

$$\rho_a = \frac{\left|I \wedge w_J^a\right|}{|I|} + \varepsilon \tag{2.15}$$

The Fuzzy ARTMAP neural networks were designed for data classification. However, a modified architecture introduced by Giralt et al. [88] allows also the generation of an output pattern once the network is trained using the algorithm prior presented. In the predictive mode, only the category layer from ART_b module is active, and linked to ART_a to provide an output for each input pattern presented to the later module [88]. The generated output is based on the adaptive weights of the ART_b module.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

2.3. Methods for selection of the most suitable set of input parameters

In order to avoid the use of redundant information in model training, it is desirable to select the smallest number of input parameters (i.e., hereinafter termed "features"), while preserving the most relevant input information. In other words, the aim is to obtain the most suitable set of input parameters selecting each feature that provides useful information, while avoiding the duplication of information already afforded by other selected parameters [89]. There are available several feature selection methods, which can lead to different set of inputs, due to differences in the selection algorithm. There are available two common approaches: *filters*, that evaluate features based on general characteristics of the data, independent of any particular algorithm, and *wrappers* that employ a statistical re-sampling technique using the actual target learning algorithm to estimate the accuracy of feature subsets [90,91].

To assure that none of the relevant input parameters are overlooked, three different input variable selection methods (two filters and one wrapper) are utilized: a) Waikato Environment for Knowledge Analysis Correlation Feature Selection (WEKA-CFS; [91]), b) Self-Organizing Map Dissimilarity Analysis (SOM-DA; [92]), and c) Artificial Neural Net Input Gain Measurement Approximation (ANNIGMA, [93]).

a) Correlation Feature Selection (CFS) method used is the one included in the WEKA software package. WEKA is a collection of machine learning algorithms that provide a general purpose environment for automatic classification, regression, clustering, and feature selection, including algorithms for modeling such as decision trees, rule sets and linear discriminants, as well as pre-processing data methods like discretization, normalization and feature selection. Feature selection schemes include fast filtering as well as wrapper approaches, with the evaluation measures based on correlation and entropy-based criteria [94].

The correlation feature selection method [91] uses a search algorithm along with a function to evaluate the "merit" of each feature subset, in order to select features that are highly correlated with the desired target, but low correlated with other previously selected parameters [95]. The heuristic by which CFS measures the performance of a feature subset

ISBN:978-84-691-2701-8/DL:T.386-2008

takes into account the usefulness of individual features for predicting the target variable, along with the level of intercorrelation among them:

$$Merit_S = \frac{k \cdot \overline{r}_{ft}}{\sqrt{k + k \cdot (k-1) \cdot \overline{r}_{ff}}} \tag{2.16}$$

where $Merit_S$ is the heuristic performance of a feature subset S containing k features, $\overline{r}_{ft}$ the average absolute feature-target correlation, and $\overline{r}_{ff}$ the average absolute feature-feature intercorrelation. The Pearson's correlation coefficient between two n-size arrays x and y, with the average values $\overline{x}$ and $\overline{y}$, respectively, can be defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{2.17}$$

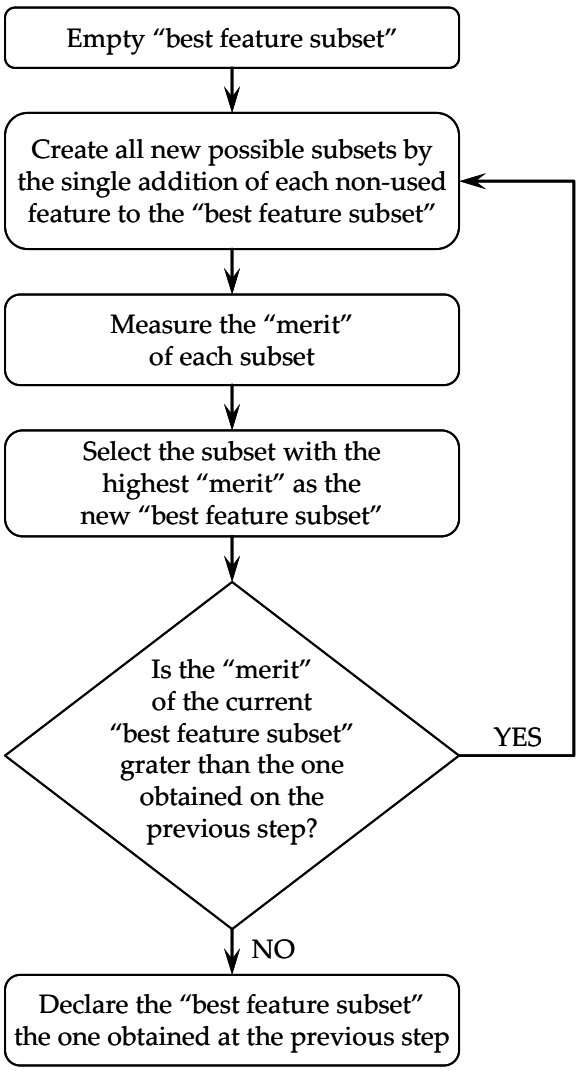


Figure 2.7. CFS algorithm for selection of the "best feature subset".

ISBN:978-84-691-2701-8/DL:T.386-2008

Being based on a forward selection algorithm, the search starts having the "best feature subset" as an empty set of features and generates all possible single feature expansions. In an iterative process, the subset with the highest performance is chosen and is expanded by creating all possible subsets by adding single non-used features. If expanding a subset results in no "merit" improvement, the search terminates and the best unexpanded subset is returned as the "best feature subset" (Figure 2.7) [91,96].

b) Self-Organizing Map Dissimilarity Analysis feature selection approach proposed by Rallo et al. [92] is based on the projection of all the candidate subsets of variables on the space generated by a SOM. An indicator of the relevance of each subset with respect to the target variable is obtained by comparing the generated maps based on the dissimilarity measure proposed by Kaski and Lagus [97]:

$$D(M_1, M_2) = \frac{\sum_{i=1}^{n} \left[\frac{\left| d_{M_1}(x_i) - d_{M_2}(x_i) \right|}{d_{M_1}(x_i) + d_{M_2}(x_i)} \right]}{n} \tag{2.18}$$

In the above equation, *D* is the dissimilarity measure between two maps, M_1 and M_2, and $d(x_i)$ is the Euclidean distance over the map, from the input sample x_i to its second best matching unit, denoted by $bmu'(x_i)$, passing first from x_i to the best matching unit, denoted by $bmu(x_i)$. For each input sample *x*, the distance is calculated considering the shortest continuous path (passing through neighbor units) between *bmu(x)* and *bmu'(x)*:

$$d(x) = \left\| x - m_{bmu(x)} \right\| + \min_i \sum_{k=0}^{I_i(bmu'(x))-1} \left\| m_{I_i(k)} - m_{I_i(k+1)} \right\| \tag{2.19}$$

where $I_i(k)$ denotes the index of a k^{th} unit on the i^{th} path along the map grid from unit *bmu(x)* to *bmu'(x)*, and *m* represents the weight vector associated with each cell unit.

The application of the approach consisted of using all the available data (molecular descriptors and target variables) to build the SOM. Next, the C-planes are extracted and the U-matrix is computed. The process of selection of relevant variables starts by the identification of the redundant ones, using a redundancy index defined by Rallo et al. [83] that takes into account the correlation between variables and their representation over the

ISBN:978-84-691-2701-8/DL:T.386-2008

map (C-planes and U-matrix). The variables which present the redundancy index greater than a certain threshold (≈0.95-0.98) are discarded from the dataset, since they provide similar information with that of other variables. After the removal of the redundant variables, a new SOM is created, followed by the extraction of the C-planes. These planes are classified in several clusters, by using a SOM approach followed by a Davies-Bouldin index procedure [81] for determining the optimal clusters configuration.

The next step in the selection of the "best feature subset" is the identification of an initial set to start the search procedure. The starting point for the search is determined by choosing a representative variable for each cluster, assuring in this way the presence of non-repetitive information in the initial set. To avoid the inclusion of irrelevant features in the initial set, only those variables that present a correlation with the target higher than the average correlation for the whole set of variables are considered. The iterative process of finding the "best feature subset" consists in the addition of the rest of variables, one by one, in the decreasing order of their absolute correlation with the target variable. For each subset of descriptors formed in this way, a new SOM is obtained. The dissimilarities between all the maps obtained are computed using Eq. (2.18), and the configuration which presents a minimum average dissimilarity between the corresponding map and all other maps is selected as the "best feature subset". The smallest average dissimilarity value indicates the maximum coherence and compactness of the information represented by those particular maps [92].

c) Artificial Neural Net Input Gain Measurement Approximation is a feature selection approach proposed by Hsu et al. [93]. The method is based on ranking the features by relevance based on the weights associated to each one by a back-propagation neural network. The reasoning behind this algorithm is that the neural networks' weights represent a measure of the gain of the input signal to the output node. An input will strongly affect the output if it has associated high weights. Therefore, a network training algorithm intends to reduce the weights of an irrelevant input and to increase the weights of a relevant one. An evaluation index is defined based on the network weights, to asses the influence of each input over each output,

ISBN:978-84-691-2701-8/DL:T.386-2008

$$LG_{ik} = \left| \frac{\Delta O_k}{\Delta A_i} \right| = \sum_{j}^{n_h} \left| W_{ij}^1 \cdot W_{jk}^2 \right| \tag{2.20}$$

where *LG* is the local gain between the input A_i and the output O_k, n_h is the number of neurons in the hidden layer, W^1 and W^2 are the weights associated with the neurons in the hidden and output layers, respectively, and the subscripts *i*, *j* and *k* refers to the neurons indices in the input, hidden and output layers, respectively. The ANNIGMA score is further calculated by normalizing the local gain to a scale of 100.

$$ANNIGMA_{ik} = \frac{LG_{ik}}{\max\left(LG_k\right)} \cdot 100 \tag{2.21}$$

This feature selection method is based on a backward stepwise elimination (BSE) wrapper algorithm. The feature selection process starts with the complete set of original variables and removes features from candidate subsets during the search. A large number of irrelevant features are eliminated in early iterations while a fine adjustment is performed in the subsequent iterations. When the performance degrades, the best of the discarded features are brought back into the candidate subset.

In contrast with previous two feature selection methods presented, ANNIGMA offers the possibility to rank the features according to their relevance with respect to multiple targets.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

3. Quantitative structure-property relationship for organic compound rejection in reverse osmosis membranes

3.1. Experimental data, pretreatment and model development

The existing literature on organic solute rejection by RO and NF membranes summarized in Section 2.1 reveals that while rejection depends on molecular parameters, conflicting trends still exist. The aforementioned studies have mostly focused on the correlation of rejection with a few molecular properties for a small number of compounds belonging to narrow chemical classes. Clearly it would be beneficial to develop predictive models based on a detailed mechanistic understanding of the reasons for the observed organic solute rejection levels (or passage) as a function of the properties of the solute and the membrane. Nevertheless, this is a daunting task given the large number of current and future organics (and compound classes) that may be of concern in municipal and industrial wastewaters. An alternative approach is to develop quantitative structure-property relationship models that consider the simultaneous correlation of organic solute rejection with multiple molecular parameters for the membranes considered, with the potential for being applied to a broad-range of compound classes. In this regard, artificial neural networks offer a unique capability for building multi-parameter QSPRs with wide applicability domains. ANN-based QSPRs have been proposed for estimation of different physicochemical properties [98-104], as well as biological activity, pharmacological or toxicological properties [92,105-109]. Therefore, the potential application of ANN-based QSPR models for the analysis and prediction of organic solute rejection by RO membranes has been explored using experimental RO performance data for fifty different organic compounds and five different commercial RO membranes. The feature selection approaches presented in Section 2.3 have been applied to select the most appropriate model input variables to correlate and estimate, using ANN-based QSPR models, the sorption, passage, and rejection of organic compounds by RO membranes.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Table 3.1. Organic compounds with available experimental data, with identification of application and/or effects.

Family[a)]	CAS	Name	Compound class, known use and/or toxicity endpoint
A	15972-60-8	2-Chloro-2',6'-diethyl-N-(methoxymethyl)acetanilide (Alachlor)	Endocrine disruptor
A	71-43-2	Benzene	Fuel hydrocarbon-Carcinogen
A	80-05-7	2,2-bis(4-Hydroxyphenyl)propane (Bisphenol A)	Estrogenic/antiandrogen household waste water product
A	58-08-2	1,3,7-Trimethyl-2,6-dioxo-1,2,3,6-tetrahydropurine (Caffeine)	Pharmaceutical human drug
A	2921-88-2	O,O-diethyl O-(3,5,6-trichloro-2-pyridinyl) phosphorothioic acid (Clorpyrifos)	Insecticide-Industrial/household waste water product
A	57-88-5	(3beta)-Cholest-5-en-3-ol (Cholesterol)	Pharmaceutical sex/steroid hormone-Fecal indicator
A	51481-61-9	2-Cyano-1-methyl-3-(2-(((5-methylimidazol-4-yl)methyl)thio)ethyl)guanidine (Cimetidine)	Pharmaceutical human drug
A	76-57-3	3-o-methylmorphine monohydrate (Codeine)	Pharmaceutical human drug
A	120-83-2	2,4-Dichlorophenol	Algicide, antihelmintic, bactericid, agricultural fungicide
A	94-75-7	2,4-Dichlorophenoxyacetic acid	Endocrine disruptor
A	84-66-2	1,2-Benzenedicarboxylic acid diethyl ester (Diethylphthalate)	Plasticizer-Industrial/household waste water product
A	56-53-1	3,4-bis(p-Hydroxyphenyl)-3-hexene (Diethylstilbestrol)	Pharmaceutical-Estrogen-Carcinogen
A	121-14-2	2,4-Dinitrotoluene	Production of isocyanate and explosives-Carcinogen
A	57-91-0	17a Estradiol	Pharmaceutical-Estrogen-Sex/steroid hormone
A	53-16-7	1,3,5(10)-estratrien-3-ol-17-one (Estrone)	Pharmaceutical-Sex/steroid hormone
A	100-41-4	Ethylbenzene	Fuel hydrocarbon
A	71-00-1	2-Amino-3-(3H-imidazol-4-yl)propanoic acid (Histidine)	Amino acid
A	15687-27-1	2-[4-(2-Methylpropyl)phenyl]propanoic acid (Ibuprofen)	Non-steroidal anti-inflammatory drug
A	58-89-9	1,2,3,4,5,6-Hexachlorocyclohexane (Lindane)	Insecticide
A	298-00-0	O,O-Diethyl-O-4-nitro-phenylthiophosphate (Methyl parathion)	Insecticide
A	98-95-3	Nitrobenzene	Solvent and mild oxidizing agent
A	104-40-5	4-Nonylphenol	Surfactant-Endocrine disruptor
A	87-86-5	2,3,4,5,6 Pentachlorophenol	Endocrine disruptor
A	108-95-2	Phenol	Phenolic compound
A	85-44-9	1,2-Benzenedicarboxylic anhydride (Phthalic anhydride)	Plasticizer-Industrial/household waste water product
A	57-83-0	Pregn-4-ene-3,20-dione (Progesterone)	Pharmaceutical-Sex/steroid hormone
A	19466-47-8	beta-Sitostanol-n-hydrate	Plant sterol-Endocrine disruptor
A	58-22-0	17b-Hydroxy-4-androsten-3-one (Testosterone)	Hormone
A	108-88-3	Toluene	Solvent-Carcinogen
A	85-01-8	Phenanthrene	Polycyclic aromatic hydrocarbon

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Table 3.1. Organic compounds with available experimental data, with identification of application and/or effects (continuation).

Family[a)]	CAS	Name	Compound class, known use and/or toxicity endpoint
B	56-41-7	2-Aminopropanoic acid (Alanine)	Amino acid
B	70-47-3	2-Amino-3-carbamoylpropanoic acid (Asparagine)	Amino acid
B	56-84-8	2-Aminobutanedioic acid (Aspartic acid)	Amino acid
B	52-90-4	2-Amino-3-mercaptopropanoic acid (Cysteine)	Amino acid
B	79-43-6	2,2-Dichloroacetic acid	Disinfect byproduct
B	124-40-3	N,N-dimethylamine	Raw material, or solvent in synthesis
B	56-40-6	Aminoethanoic acid (Glycine)	Amino acid
B	56-87-1	(S)-2,6-Diaminohexanoic acid (Lysine)	Amino acid
B	63-68-3	(S)-2-Amino-4-(methylsulfanyl)-butanoic acid (Methionine)	Amino acid
B	62-75-9	N-nitroso dimethyl amine	Carcinogen
B	75-65-0	tert-Butyl alcohol	Alcohol-Industrial solvent
B	72-19-5	(2S,3R)-2-Amino-3-hydroxybutanoic acid (Threonine)	Amino acid
B	76-03-9	Trichloroacetic acid	Disinfection byproduct
B	57-13-6	Urea	Fertilizer
B	72-18-4	(S)-2-Amino-3-methyl-butanoic acid (Valine)	Amino acid
B	127-18-4	1,1,2,2-Tetrachloroethylene	Industrial chlorinated solvent
DB	85721-33-1	1-Cyclopropyl-6-fluoro-1,4-dihydro-4-oxo-7-(1-piperazinyl)-3-quinolinecarboxylic acid (Ciprofloxacin)	Pharmaceutical human/veterinary antibiotic
DB	564-25-0	4-(Dimethylamino)-1,4,4a,5,5a,6,11,12a-octahydro-3,5,10,12,12a-pentahydroxy-6-methyl-1,11-dioxo-2-naphthacenecarboxamide monohydrate (Doxycycline)	Pharmaceutical human/veterinary antibiotic
DB	60-00-4	Ethylenediaminetetraacetic acid	Chelating agent
DB	60-54-8	Tetracycline	Antibiotic

[a)] Family of compound as identified in Figure 3.2a. A – Family A; B – Family B; DB – Domain Border.

The experimental data used for analyzing the RO membranes performance with respect to the organic compounds in terms of sorption, passage and rejection, were provided by Orange County Water District, Los Angeles, California [110]. The set of 50 compounds listed in Table 3.1 mostly of public health concern, was selected for a detailed experimental RO study. The selection was made based on an interrogation of several available databases regarding monitoring rules for contaminants and toxic substances, including the U.S. Geological Survey Toxic Substances Hydrology Program [111], U.S. Environmental Protection Agency Unregulated Contaminant Monitoring Rule [112], U.S. Environmental Protection Agency Announcement of the Drinking Water Contaminant List [113], and the California Department of Health Services Unregulated Chemicals Requiring Monitoring [114]. The list of compounds includes endocrine disruptors, pharmaceutically active compounds, antibiotics and antimicrobial agents, neuroactive drugs, insecticides, herbicides, pesticides, disinfection byproducts, solvents, industrial pollutants and fuel hydrocarbons. Several amino acids were also considered to broaden the range of molecular properties variations.

Five commercial RO membranes, four polyamides (BW30, ESPA2, LFC1, TFCHR) and one cellulose acetate (CA), whose properties are listed in Table 3.2, were selected for a detailed experimental evaluation of their performance expressed as sorption, passage and rejection with respect to the selected organic compounds. Membrane properties used to characterize the selected RO membranes include contact angle, zeta potential at pH=7 and zeta potential slope (at the pH range of 5-7), root-mean-square (RMS) surface roughness and specific water flux. Additional information for the polyamide membranes include the polyamide layer thickness, two COO^-/Amide ratios and the OH^-/Amide ratio derived from attenuated total internal reflection Fourier transform infra-red (ATR-FTIR) spectroscopic measurements. These four polyamide membrane parameters are unitless relative indices based on ratios between the absorption at different wavelengths corresponding to the presence in the membrane of carboxyl group (1415 cm^{-1}), amide I bonds (1665 cm^{-1}), amide II bonds (1542 cm^{-1}), hydroxyl group (3400 cm^{-1}) and polysulfone membrane support layer (874 cm^{-1}). The contact angles along with the zeta potential are typically used as indicators of the degree of membrane hydrophilicity. The RMS surface roughness is also reported as a surrogate measure that indicates possible differences in sorption surface area. The polyamide layer thickness directly affects membrane transport resistance in the polyamide membranes.

ISBN:978-84-691-2701-8/DL:T.386-2008

Table 3.2. Properties of membranes used for experimental analysis.

Membrane Properties	BW30	ESPA2	LFC1	TFCHR	CA
Contact angle [degrees]	61.5	61.3	61.7	61.5	66.2
Zeta potential [mV]	-12.8	-26.0	-17.3	-16.3	-22.4
Zeta potential slope (pH 5-7)	-2.67	-5.00	-1.03	-1.61	-0.62
COO^-/Amide I ratio	0.46	0.31	0.43	0.33	-
COO^-/Amide II ratio	0.42	0.27	0.42	0.33	-
OH^-/Amide I ratio	2.09	0.53	1.37	0.80	-
Polyamide thickness	1.30	1.31	1.19	0.69	-
Roughness [nm]	82.9	90.9	111.5	48.6	44.6
Specific water flux $[m^3 \cdot m^{-2} \cdot s^{-1} \cdot kPa^{-1} \cdot 10^8]$	1.03	1.44	1.44	1.23	0.34

BW30 – Thin Film Composite (TFC) brackish water RO membrane (DOW Filmtec); ESPA2 – TFC brackish water RO membrane (Hydranautics); LFC1 – TFC low fouling brackish water RO membrane (Hydranautics); TFCHR – TFC high rejection RO membrane (Koch Membrane Systems); CA – Cellulose acetate brackish water RO membrane (Osmonics).

RO membrane characterization studies

The organic compounds used, with purity > 99%, were stored either at 4 ºC or –20 ºC (depending on the compound) for a minimal period of time (typically less than one week) prior to assay to lessen the opportunity for post-manufacture chemical changes. Compounds labeled with ^{14}C were chosen preferentially over compounds labeled with ^{3}H to reduce the possibility of radiolysis during storage and to suppress ^{3}H proton exchange with water during interaction with the membrane [115]. Only four compounds labeled with ^{3}H were used. These were cimetidine (51481-61-9), beta-Sitostanol-n-hydrate (19466-47-8), doxycycline (564-25-0) and tetracycline (60-54-8).

Membrane characterization tests consisted of the determination of solute permeation and sorption thereby enabling calculation of rejection in a series of dead-end membrane filtration experiments carried out in the apparatus depicted Figure 3.1. Solute mass in the feed, collected permeate and sorbed by the membrane was determined based on measurements of the radioactivity of the feed, permeate and the membrane itself. Solute mass rejected by the membrane was determined by the difference between the solute mass in the feed charge and the sum of the mass accumulated on the membrane plus the organic compound mass in the permeate.

Membrane performance studies were carried out using a small dead-end stainless-steel/Teflon pressure filtration cell (VWR, Bristol, CN), which supported the membrane

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

coupon (1.25 cm diameter) on a perforated stainless steel disk with the feed surface sealed with a Teflon O-ring. Membrane samples measuring 10.1x15.2 cm were preconditioned under crossflow conditions in a plate-and-frame stainless steel RO cell at a pressure of 1034 kPa for 16 hrs using 1 μohm-cm deionized water to hydrate and clean the membranes. Following preconditioning, circular 1.25 cm diameter coupons of membrane were cut for use in a high pressure dead-end filtration cell drawn schematically in Figure 3.1. These conditioned membrane coupons were stored in 17 Mohm-cm ASTM I ultrapure water at 4 ºC for no more than one week before use.

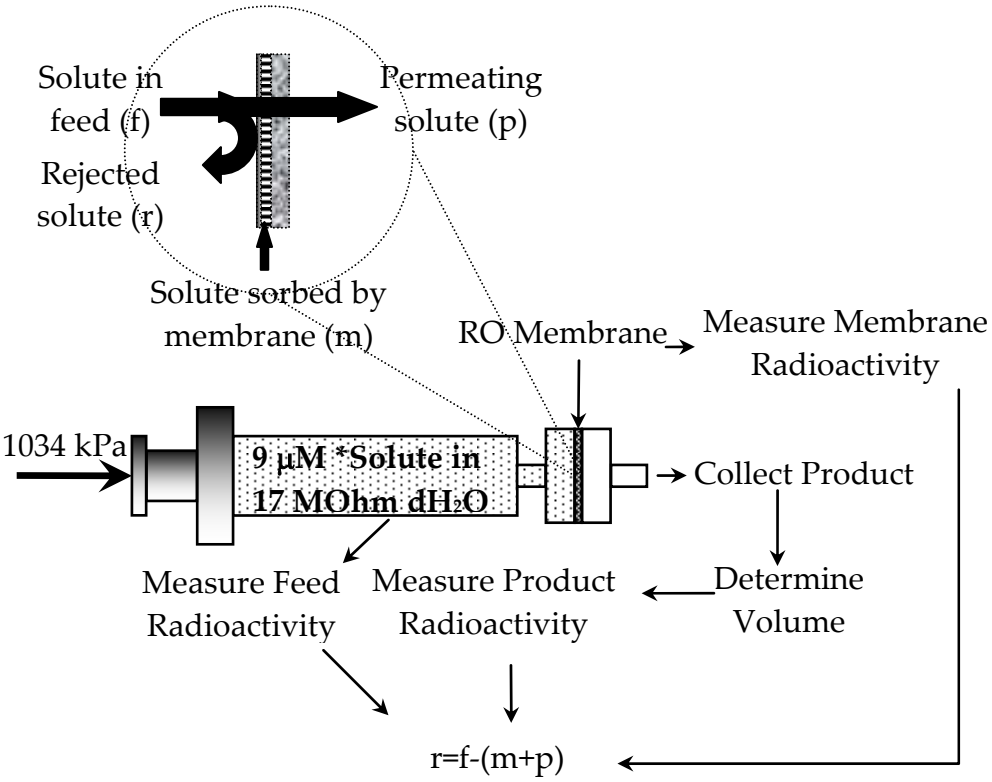


Figure 3.1. Schematic illustration of solute sorption, permeation and rejection by the RO membrane in the experimental dead-end filtration mode.

Prior to each experiment, the feed side of the pressure cell (Figure 3.1) was filled with 5 ml feed solution, prepared using ultrapure water, with the target organic at concentration of about 9 μM, resulting in typically 10^5 – 10^6 disintegrations per minute (DPM) of the radiolabeled (^{14}C or ^{3}H) test compound. At this concentration, the effects of concentration polarization on the osmotic pressure were expected to be relatively low, despite the dead-end filtration mode of operation. All the experiments were carried out at 1034 kPa and 24 ºC with the feed solution pH adjusted to 7 using HCl or NaOH. A minimum of five replicate membrane performance measurements were performed with each membrane-solute

ISBN:978-84-691-2701-8/DL:T.386-2008

combination. All pressure cell components were thoroughly cleaned and decontaminated prior to each experiment with a radiodecontamination solution (Radiacwash #005-400, Biodex Medical Systems, Inc., Shirley, MA), followed by detergent cleaning to remove organic contaminants (Micro-90, International Products Corporation, Burlington, NJ). All system components were subsequently washed with deionized water (1 μohm-cm deionized water) and subsequently soaked in water for a minimum of 1 hr. Prior to use all system components were scrubbed with a nylon bristle brush, rinsed with deionized water followed by rinsing with 70% laboratory grade denatured ethanol, an additional rinse with deionized water and finally drying in air.

Permeate product was collected in a 10 ml of scintillation *cocktail* (SC) solution (Optifluor, Packard Instrument Company, Meriden, CT) in a 22 ml scintillation vial, through a 18-gauge hypodermic needle attached to the pressure cell product side. Once a permeate volume of approximately 0.5 ml was collected (and weighted to precision of ± 0.005 g), the membrane coupon was removed and rinsed by sequentially immersing and swishing in three 400 ml beakers containing 350 ml of 17 Mohm ASTM I grade ultrapure water. Excess solution was wicked away from the membrane surface using an adsorbent paper and the membrane was then immersed into a 22 ml scintillation vial containing 10 ml of the SC solution. Membrane samples were incubated overnight in order to facilitate permeation of the *cocktail* into the membrane material. The above procedure yielded higher than 99% recovery of membrane-retained (i.e., sorbed) organics. Scintillation vials containing feed, permeate and membrane samples were analyzed using a scintillation counter (Wallac LKB 1219 Rackbeta Liquid Scintillation Counter, Perkin-Elmer, Shelton, CT). Quench and counting efficiency were corrected using the external sample channel ratio method with ^{226}Ra as the external standard to yield a DPM measurement which was corrected for background DPM measured for a 10 ml of a reference SC solution.

Characterization of organic compounds

Molecular descriptors were derived from molecular calculations given the chemical structures of the selected compounds listed in Table 3.1. Molecular structures, presented in ANNEX I, were first drawn using ACD/ChemSketch 8.00 (Advance Chemistry Development

Inc.) [116] and converted to three dimensional structures using the CAChe Software (Oxford Molecular Ltd.) [117]. The geometry of the three dimensional structures for the water dissolved compounds were subsequently optimized using the molecular orbital package (MOPAC) with the AM1 (Austin Model) Hamiltonian. MOPAC is a semi-empirical quantum-mechanical computational tool that uses the presence and positions of electrons between atoms to compute and minimize an energy related to the heat of formation, by solving the Schrödinger equation for the best molecular orbitals and geometry of the chemical molecule considered [118]. The AM1 Hamiltonian is an operator in the Schrödinger equation that describes the energy of the electrons and nuclei in the molecule, based on the modified neglect of differential diatomic overlap approximation [119,120].

The initial set of 45 molecular descriptors (Table 3.3) was selected to ensure inclusion of the major descriptors that have been shown effective for neural network-based correlations of chemical properties such as aqueous solubility [100], octanol-water partition coefficient [102], infinite-dilution activity coefficient [104], critical properties [99], vapor pressure [101] and Henry's law constant [103], in addition to those correlating descriptors reported in previous studies of organic solute rejection by RO membranes [31-39]. The selected chemical descriptors included constitutional, topological, geometrical, electrostatic and quantum chemical parameters [120].

The constitutional descriptors included the number of atoms in the solute molecule, bond counts (single bonds and double bonds), number of rings, size of the smallest and the largest ring, and molecular weight. The bonds count excluded ionic bonds, and the coordinate bonds were counted as simple bonds. Molecular topological descriptors included three connectivity indices [121,122] of orders 0, 1 and 2, three valence connectivity indices [121,122] of orders 0, 1 and 2, and three κ (kappa) shape indices of orders 1, 2 and 3 [123]. Molecular connectivity indices encode two-dimensional structural information into numerical values based on a molecular structure which is expressed topologically by a hydrogen-suppressed graph. The connectivity indices are the valence weighted counts of the connected subgraphs. The zeroth order term (atomic) is related to the degree of branching and size of the molecule expressed as the number of non-hydrogen atoms. The first order term (bond) represents a dissection of the molecular skeleton into "two contiguous bond" fragments. The second order (path) is a weighted count of four atoms (three-bond) fragment representing the

ISBN:978-84-691-2701-8/DL:T.386-2008

potential of rotation around the central bond. The first order kappa shape index quantifies the number of cycles in the chemical compound, the second order kappa shape index quantifies the degree of linearity or star-likeness of the chemical, and the third order kappa shape index quantifies the degree of branching toward the center of the chemical.

Table 3.3. Molecular and membrane descriptors used for developing QSPR models.

Molecular descriptors and membrane properties	
1 – Atom count (all atoms)	*29 – HOMO energy [eV]*
2 – Bond count (all bonds)	*30 – LUMO energy [eV]*
3 – Bond count (single bonds)	*31 – Dielectric energy [kcal/mol]*
4 – Bond count (double bonds)	*32 – Steric energy [kcal/mol]*
5 – Ring count (all rings)	*33 – Heat of formation [kcal/mol]*
6 – Size of smallest ring	*34 – One term energy: electron-electron repulsion [eV]*
7 – Size of largest ring	*35 – One term energy: electron-nuclear attraction [eV]*
8 – Molecular weight [Da]	36 – One term energy: total energy[eV]
9 – Connectivity index order 0	37 – Two center energy: electron-electron repulsion [eV]
10 – Connectivity index order 1	38 – Two center energy: electron-nuclear attraction [eV]
11 – Connectivity index order 2	39 – Two center energy: nuclear-nuclear repulsion [eV]
12 – Valence connectivity index order 0	40 – Two center energy: total electrostatic [eV]
13 – Valence connectivity index order 1	41 – Two center energy: resonance [eV]
14 – Valence connectivity index order 2	42 – Two center energy: exchange [eV]
15 – Shape index kappa1	43 – Two center energy: total energy [eV]
16 – Shape index kappa2	44 – Total energy [eV]
17 – Shape index kappa3	45 – Molar refractivity
18 – Moment of inertia A [10^{-40}g·cm^2]	46 – Contact angle [degrees]
19 – Moment of inertia B [10^{-40}g·cm^2]	*47 – Zeta potential [mV]*
20 – Moment of inertia C [10^{-40}g·cm^2]	*48 – Zeta potential slope (pH 5-7)*
21 – Solvent accessibility surface area [Å^2]	49 – COO^-/Amide I ratio
22 – Polarizability [Å^3]	50 – COO^-/Amide II ratio
23 – Dipole moment [Debye]	51 – OH^-/Amide I ratio
24 – Dipole vector X [Debye]	*52 – Polyamide thickness*
25 – Dipole vector Y [Debye]	53 – Roughness [nm]
26 – Dipole vector Z [Debye]	*54 – Specific water flux [$m^3 \cdot m^{-2} \cdot s^{-1} \cdot kPa^{-1} \cdot 10^8$]*
27 – Dipole point-charge [Debye]	
28 – Dipole hybridization [Debye]	

Variables from 1 to 45 represent molecular descriptors, while variables from 46 to 54 are properties of the membranes. Variables 49 to 52 refer only to the polyamide membranes. In italic font are presented molecular/membrane descriptors selected at least for one model.

The geometrical descriptors were the moments of inertia (A, B and C) and the solvent accessibility surface area. The moments of inertia characterize the mass distribution in the molecule and the susceptibility of the molecule to different rotational transitions. Each moment of inertia is defined with respect to a specific rotational axis. The solvent accessibility surface area is the molecular surface area that is accessible for contact with a sphere of 1.4 Å^2 which approximates the radius of a water molecule [124].

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

The electrostatic descriptors [120] were the polarizability, dipole moment, dipole vectors (X, Y and Z), dipole point-charge and dipole hybridization. The polarizability represents the response of electron distribution to an externally-applied static electric field. The dipole moment accounts for the internal separation of the positive and negative charges in a molecule, being a sum of two terms: one term corresponding to the non-uniform distribution of the electrons in bonds (dipole point-charge), and the second term to the influence of the atoms hybridization (dipole hybridization). The dipole vectors provide information regarding the spatial orientation of the charge distribution [125].

Quantum chemical descriptors included 15 energy descriptors, heat of formation and molar refractivity [120]. The quantum total energy parameter is defined as the sum of one-center and two-center energy terms which were considered as additional potential chemical descriptors. The one-center energy terms include electron-electron repulsion and electron-nuclear attraction. The two-center energy terms include resonance energy, exchange energy, electron-electron repulsion, electron-nuclear attraction, and nuclear-nuclear repulsion. The total electrostatic (or Coulombic) interaction is equal to the sum of the following two-center energy terms: electron-electron repulsion, electron-nuclear attraction and nuclear-nuclear repulsion. The resonance energy corresponds to the difference in delocalized pi electrons and localized pi electron in a double bond. The exchanged energy involves two electrons where the energy of attraction is between the nuclei and the overlap charge in the bond. HOMO energy is the energy required to remove an electron from the highest unoccupied molecular orbital, while the LUMO energy is the energy gained when an electron is added to the lowest unoccupied molecular orbital. The heat of formation is the energy released or used when a molecule is formed from elements in their standard state. The steric energy is a summation of the energy terms for all included bonds, angles and torsions, taking into account also the non-bonded interactions (e.g., van der Waals and electrostatic interactions). The dielectric energy is the stabilizing portion of the total energy of a molecule that results from screening the charges in the molecule by a dielectric. Molecular refractivity is related to the refractive index, molecular weight and density [126].

The initial set of descriptors (Table 3.3) used as inputs in the QSPR models included 45 molecular solute descriptors and 9 membrane properties. The output variables are considered the membrane performance parameters, which included the solute mass in the

ISBN:978-84-691-2701-8/DL:T.386-2008

permeate (p) and sorbed by the membrane (m) for a given permeate volume collected. These performance parameters were converted into mass fractions:

$$M = \frac{m}{p+m+r};\ P = \frac{p}{p+m+r} \tag{3.1}$$

It is also noted that the above mass fractions can also be considered as the fractions of the fluxes of solute permeation and sorption per membrane surface area, relative to the total additive solute mass flux over the permeation period. The rejected fraction R was then calculated from a simple mass balance, i.e., $R = 1-(M+P)$.

Data conditioning and selection of compounds belonging to the same chemical domain

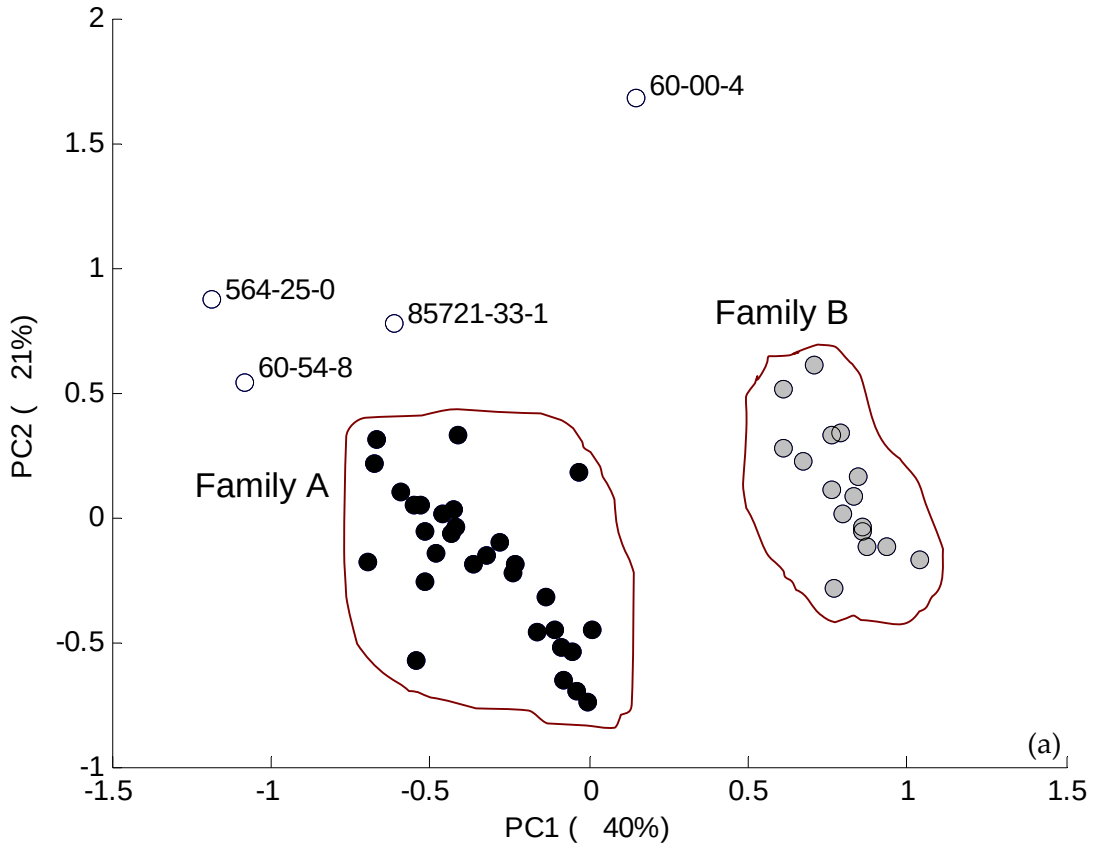
Prior to model development, all input and output parameters (i.e., solute and membrane descriptors and solute fractions) were normalized in the range [0,1] using Eq. (2.1). Further, the exploration of the chemical space for defining the model application domain is necessary. Chemicals, such as those listed in Table 3.1, are usually characterized in terms of molecular descriptors by using different approaches. For example, descriptor value ranges, principal component ranges, geometric methods based on the convex hull, distance-based methods, and probability density modeling methods can be applied [127]. The principal components analysis (PCA)-based approach, which uses the orthogonal coordinate system defined by the principal components, is one of the most widely adopted approaches. A 2D projection onto the space spanned by the two first principal components usually provides adequate information about the distribution of data in the input space. On the other hand, the K-means clustering of a Self-Organizing Map (SOM) built to classify the considered chemicals is a suitable alternative to PCA. First, SOM is a topology preserving projection method which permits visualization of the data space in a 2D plot. Second, the SOM clustering process uses Euclidean distances between vectors formed by compounds' chemical descriptors to compute the similarity between chemicals in the dataset. Finally, SOM approaches the point probability density of the input space in a way that more units are placed in regions of the input space where data points are dense and fewer units where density is sparse.

The PCA and SOM results for 50 chemicals listed in Table 3.1 are shown in Figure 3.2. Each compound in these plots is represented by a 45-dimensional vector formed by all molecular

ISBN:978-84-691-2701-8/DL:T.386-2008

descriptors listed in Table 3.3. The PCA projection results (Figure 3.2a) suggest the presence of two chemical families. Family A with the first 30 chemicals and Family B with the following 16 chemicals listed in Table 3.1. Figure 3.2a also identifies four chemicals, with their CAS numbers indicated, which are located closer to the boundaries of the chemical domain (DB chemicals; see also Table 3.1) and thus will significantly influence any model developed. Figure 3.2b shows the K-means classification of the SOM prototype vectors representing the clusters obtained after classifying all 50 chemicals that are also represented by vectors of descriptors. Ten coherent chemical families (in terms of molecular descriptors) can be identified from the clustering of SOM prototypes in Figure 3.2b.

The PCA discrimination between chemicals in Families A and B (Table 3.1) is mainly accomplished by the occurrence of aromatic rings in the former or of amino functional groups in the latter. Family B contains chemicals without rings in their molecular structure. Moreover, it includes 9 of the 10 amino-acids listed in Table 3.1, the exception being Histidine (71-00-1) which belongs to Family A because it is an amino-acid with an imidazol aromatic ring in its molecule. Family B also includes three amines, two acids, one alcohol and one halogenated compound. It should also be noted that the 16 chemicals of Family B (Table 3.1 and Figure 3.2a) constitute class 5 in the SOM classification depicted in Figure 3.2b.



ISBN:978-84-691-2701-8/DL:T.386-2008

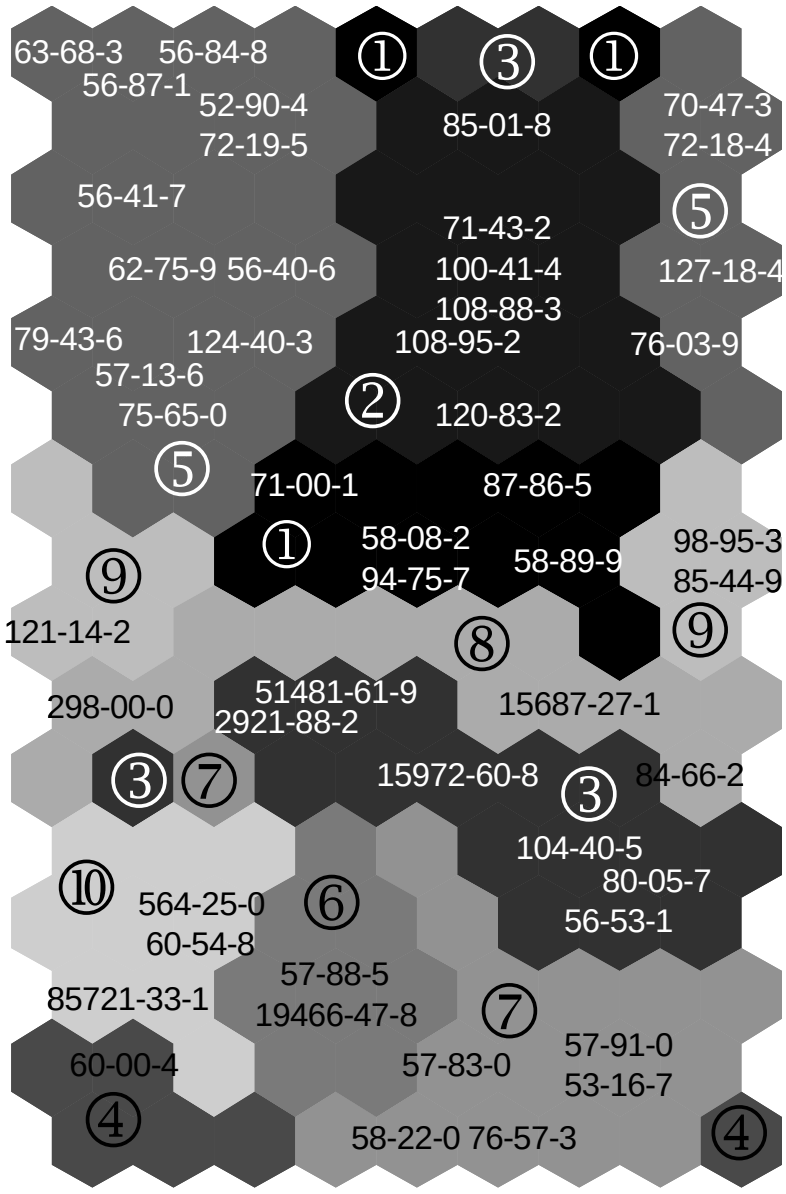


Figure 3.2. Analysis of the chemical space by means of (a) PCA; (b) SOM.

Of the chemicals near the domain boundary (DB), ethylenediaminetetraacetic acid (60-00-4) is unique from the molecular structure viewpoint since it constitutes a single class in the SOM (class 4), i.e., it is not structurally similar to any of the other chemicals in Table 3.1. The antibiotics tetracycline, doxycycline and ciprofloxacin (60-54-8, 564-25-0, 85721-33-1), previously detected at the domain borders by the PCA-based approach, form another coherent and separate SOM class (class 10 in Figure 3.2b). These three antibiotics are located in the neighborhood of ethylenediaminetetraacetic acid (60-00-4). Thus, the PCA and SOM classifications complement each other in the characterization of the chemical domain explored in the current study with respect to organic chemicals passages through RO membranes.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

A more detailed understanding of chemical domain of the current 50 chemicals can be obtained from the examination of the functional groups that best discriminate between the three families of compounds A, B, and DB in Table 3.1, as suggested elsewhere [128]. This functional group analysis is summarized in the histogram depicted in Figure 3.3.

The more characteristic functional groups of Family A are nCq (number of total quaternary sp3 C), nCrq (number of ring quaternary sp3 C), nN+ (number of positive charged N), nArNO2 (number of aromatic nitro groups), nArCOOR (number of aromatic esters), nArOR (number of aromatic ethers), nPO4 (number of phosphates/thiophosphates), nImidazoles (number of Imidazoles), nRCONR2 (number of aliphatic tertiary amides), nN=C-N< (number of amidine derivatives), nC(=N)N2 (number of guanidine derivatives), nNq (number of quaternary N), nN(CO)2 (number of imides [thio-]), nROR (number of aliphatic ethers), nO(C=O)2 (number of anhydrides [thio-]), nCH2RX (number of CH2RX), nCXr (number of X on ring sp3 C), and nPyridines (number of Pyridines). For Family B, the more characteristic functional groups are nR=Cp (number of terminal primary sp2 C), nRNNOx (number of aliphatic N-nitroso groups), nSH (number of thiols), nCHRX2 (number of CHRX2), nR=CX2 (number of R=CX2) and nCRX3 (number of CRX3). For the DB chemicals, the more characteristic functional groups are nArCO (number of aromatic ketones) and nArNR2 (number of aromatic tertiary amines).

The above suggests that the selected compounds are similar in terms of functional groups that are both coherent with the families identified by PCA and SOM analyses, and match the selection criteria. For example, chemicals that are of public health concern are included in Family A, amino acids in Family B and antibiotics in the DB compounds class. In addition, the above classification and domain characterization results indicate that the majority of the 50 chemicals reasonably span the chemical space. Since the data set is very small from a QSPR development point of view, all chemicals have been considered in the current model building, even though higher prediction errors are expected for under-represented compounds.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

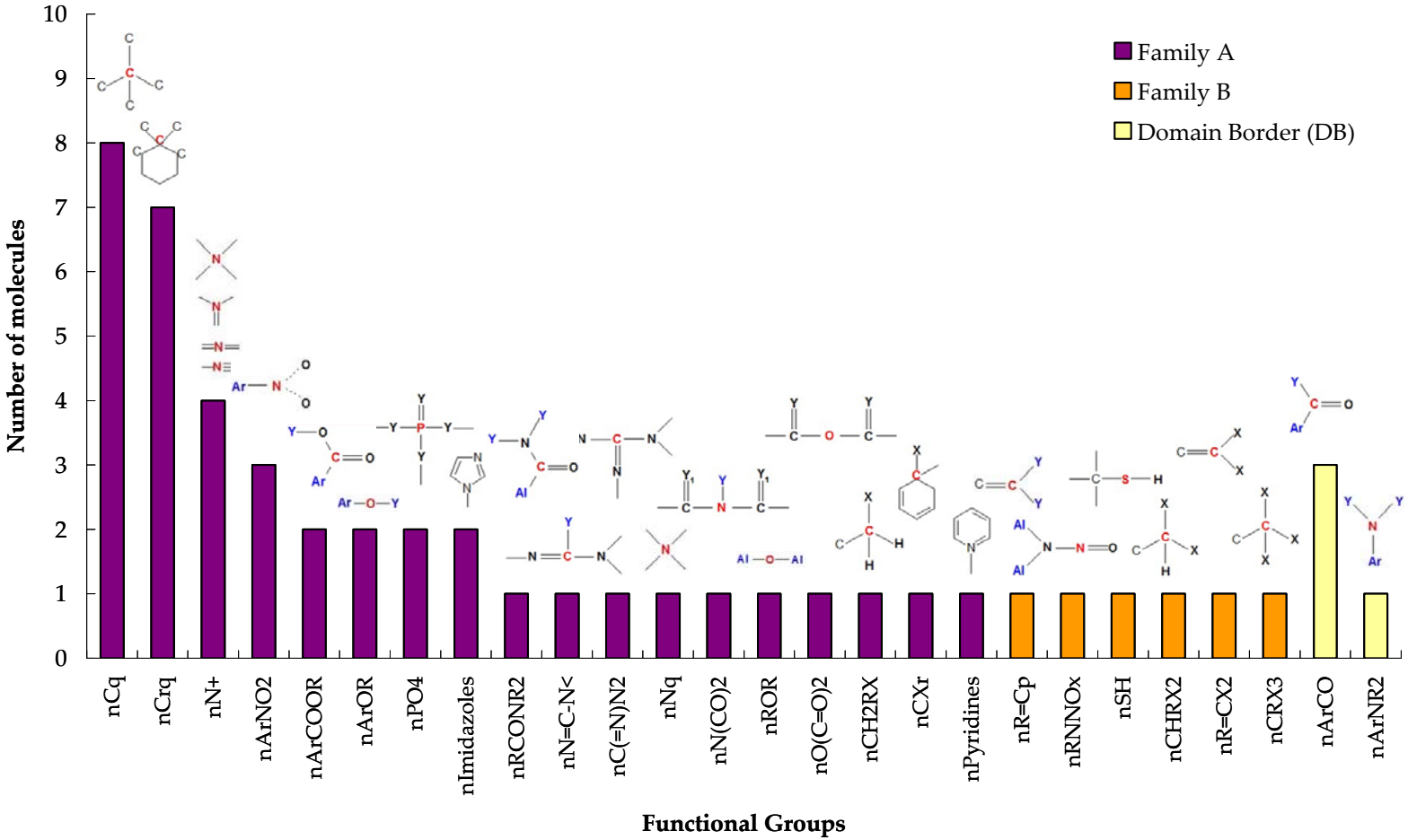


Figure 3.3. Discriminant functional group for the compounds in the three families identified by the PCA. Functional groups abbreviations taken from: http://www.talete.mi.it/help/dragon_help/index.html?FunctionalGroupCounts11.

ISBN:978-84-691-2701-8/DL:T.386-2008

Development and quality assessment of ANN models

Several artificial neural networks-based QSPR models were developed to analyze the influence of the chemical structure on the sorption (M), passage (P) and rejection (R) of organic compounds determined experimentally for four polyamide and one cellulose acetate RO membranes. The models were developed based on back-propagation architecture with one input layer, one hidden layer and one output layer. The linear transfer function was utilized for the input and output layers and a hyperbolic tangent transfer function was used for the hidden layer (see Figure 2.3) [71]. For each model that was generated, the network architecture was established with the condition that the total number of connections between network's neurons would not exceed the total number of input data points. This condition was specified as

$$n_h = \min\left(n_h^{\max}; 2 \cdot n_i - 1\right);\ n_h^{\max} \le \frac{n_{tr} - n_o}{1 + n_i + n_o} \tag{3.2}$$

where n_i, n_h, and n_o are the number of neurons in the input layer, hidden layer and output layer respectively, and n_{tr} is the number of data in the training set.

Two types of analyses were carried out, the first based on internal validation, with a leave-one-out (LOO) cross-validation procedure [129], and the second one consisted on an external validation with an independent set of test compounds that were not used for model training [129,130]. For both cases, the average absolute and relative errors, standard deviations and maximum value of these errors were also computed.

For the internal model validation, each one of the 50 chemicals in Table 3.1 was individually and sequentially eliminated from the data set and the remaining (50-1) compounds used to train 50 different models. The cross-validation explained variance in the prediction index, q^2, was then calculated for all the individually predicted mass fractions using the 50 models [129],

$$q^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{3.3}$$

ISBN:978-84-691-2701-8/DL:T.386-2008

where y_i and $\hat{y}_i$ refers to the experimental and predicted mass fraction for the compound *i*, $\overline{y}$ is the average fraction value of experimental data for all n compounds and q^2 is the explained variance in prediction index, which varies from 0 to 1. A low value of q^2 in the LOO test typically indicates a model with low internal predictive ability and low robustness or ability to avoid the influence of outliers [130]. However, the converse does not necessarily hold, since it has been shown that a high value of q^2 obtained for internal validation is an insufficient criterion for a QSPR model to be highly predictive, especially when the number of descriptors is approaching or is higher than the number of compounds [130]. Therefore model testing by external validation is also needed, i.e., by using an external data set not used to train the model.

Accordingly, external validation of model quality with separate but complementary training and test sets was evaluated with the following two indices:

$$q_{tr}^2 = 1 - \frac{\sum_{i=1}^{n_{tr}} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n_{tr}} \left(y_i - \overline{y}_{tr}\right)^2};\ q_{ts}^2 = 1 - \frac{\sum_{i=1}^{n_{ts}} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n_{ts}} \left(y_i - \overline{y}_{tr}\right)^2} \tag{3.4}$$

where q_{tr}^2 and q_{ts}^2 are the training and test set explained variance in prediction, respectively, and $\overline{y}_{tr}$ is the average value of the experimental data belonging to the training set. The number of samples in the training and test set are represented by n_{tr} and n_{ts}, respectively [129]. A clustering SOM-based algorithm [77] was used to divide the chemical data set (using the best feature subset and target variable), for each selected network architecture, into consistent training and test sets. In the present approach, the compound nearest to the centroid of each SOM cell was taken to be as most representative of that map unit. The representative compounds of the six cells with the higher number of hits (i.e., number of molecules allocated to each cell) were selected for the test set (i.e., six compounds), with the remaining compounds (44) assigned to the training set. The above procedure assured that the training set contained data that were reasonably representative of the entire chemical domain.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

3.2. Results

Three types of artificial neural networks QSPR models were developed for the sorbed fraction (M) and passage fraction (P): (a) independent ANN-QSPR models for each of the solute fractions and membranes used (hereinafter termed Independent ANQ models), (b) specific ANN-QSPR models for each of the solute fractions for the composite collection of several membranes (hereinafter termed Membrane-Composite ANQ models), and (c) ANN-QSPR models that simultaneously considered the two solute fractions for each specific membrane (hereinafter termed MP-Composite ANQ models). The predicted rejection fraction (R) was calculated from a simple mass balance, i.e., $R = 1 - \left(\hat{M} + \hat{P} \right)$ where $\hat{M}$ and $\hat{P}$ were estimated from the ANN-QSPR models. The models were built using the most suitable set of input descriptors (Table 3.4) selected from the initial set of indices (Table 3.3) by the three feature selection methods presented in Section 2.3.

Table 3.4. Feature selection results for the Independent ANQ, Membrane-Composite ANQ/PA and MP-Composite ANQ models.

Independent ANQ models			
Membrane	**FS method**	**M fraction**	**P fraction**
CA	CFS	6 28 30	6 8 13 14 28 30
	SOM-DA	4 5 6 8 17 23 24 26 28 30 31 32 33	1 4 5 6 8 9 16 19 21 27 28 30 32 35
	ANNIGMA	4 6 7 23 26 27 28 29 35	6 7 8 23 25 27 28 29 30
BW30	CFS	6 23 25 28 33	7 8 16 21 24 30
	SOM-DA	4 6 16 17 23 24 25 26 27 28 30 31 33	1 4 5 6 8 9 14 16 25 26 29 30 35
	ANNIGMA	5 6 7 16 23 24 25 29 30 33	6 7 8 21 23 27 33 34
ESPA2	CFS	6 23 25 28 33	6 7 8 13 16 21 24 30
	SOM-DA	4 5 6 17 23 25 26 28 29 30 31 32 33	1 4 5 6 8 9 14 16 27 28 29 30 32 35
	ANNIGMA	5 6 7 16 18 23 24 25 29 33	6 7 12 13 21 23 26 27 32 33
LFC1	CFS	6 17 23 25 28 33	6 8 16 21 24 30
	SOM-DA	4 5 6 16 17 23 24 25 26 27 28 29 30 31 33	1 4 5 6 8 9 15 16 17 24 29 30 35
	ANNIGMA	6 7 14 16 18 23 25 27 29 30 31 33	6 7 8 23 24 25 27 32 33 34
TFCHR	CFS	6 23 25 28 33	6 7 8 13 16 21 24 30
	SOM-DA	4 5 6 17 23 24 25 26 28 29 30 31 33	1 4 5 6 8 9 14 16 27 29 30 32 35
	ANNIGMA	5 6 7 16 23 24 25 29 30	6 7 21 24 27 28 31 32 33
Membrane-Composite ANQ/PA (BW30, ESPA2, LFC1, TFCHR) models			
Membrane	**FS method**	**M fraction**	**P fraction**
PA	SOM-DA	4 5 6 17 24 28 29 30 31 32 33 47 48 52	4 5 6 9 16 18 19 25 28 29 30 31 32 47
MP-Composite ANQ models			
Membrane	**FS method**	**M and P fraction**	
CA	ANNIGMA	4 6 8 23 26 27 28 29 30	
BW30	ANNIGMA	5 6 7 23 24 25 27 28 29 33	
ESPA2	ANNIGMA	5 6 7 23 24 25 27 28 29 33	
LFC1	ANNIGMA	5 6 7 8 16 23 24 25 27 28 29 30 31 33	
TFCHR	ANNIGMA	5 6 7 23 24 25 27 28 29 32 33	

Indices correspond to molecular and membrane descriptors, as identified in Table 3.3.

ISBN:978-84-691-2701-8/DL:T.386-2008

It should be noted that for building Membrane-Composite ANQ models, only SOM-DA selected both molecular and membrane descriptors. In Table 3.4 are presented the input parameters selected by SOM-DA for developing models for the collection of all four polyamide membranes (termed Membrane-Composite ANQ/PA models). The same parameters were utilized also for developing models for the collection of all five membranes used (termed Membrane-Composite ANQ/PACA models), adding for each fraction model the membrane Specific Water Flux among the model inputs. In the case of M fraction Membrane-Composite ANQ/PACA model, given the fact that polyamide thickness is one of the selected inputs and this parameter refers only to the polyamide membranes, the value zero was used for the cellulose acetate membrane. As presented in Section 2.3, among the three feature selection methods used, only ANNIGMA offers the possibility to rank input features according to their relevance with respect to multiple targets. Therefore, the input parameters selected using this method for developing MP-Composite ANQ models are presented in Table 3.4. However, MP-Composite ANQ models were built also by using as input parameters the union of molecular descriptors selected by CFS for the Independent ANQ models for the M and P fractions. Low model quality is expected in a similar analysis performed using the molecular descriptors selected by SOM-DA due to the increased number of input parameters chosen by this method which would lead to a small number of neurons in the hidden layer as calculated from Eq. (3.2). Therefore, such analysis is not performed.

Selection of model input parameters for Independent ANQ models

Generally, the SOM-DA and ANNIGMA offered a more ample selection comparing with CFS. The SOM-DA always selected the largest number of features for all models considered, because of the specific criteria used by this method to reduce the number of input parameters. In the SOM-DA approach, descriptors are sorted in a decreasing order of importance of influencing the topological organization of the target variable in the SOM map that accounts for chemical similarity.

A very good agreement between the three feature selection methods was observed. Table 3.4 shows that for every membrane, the set of input descriptors selected by the CFS method for the M fraction model was a subset of those selected by the SOM-DA method. However, not

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

all the descriptors selected by CFS for the M fraction models were contained also in the descriptor sets selected by ANNIGMA. The same was concluded after comparing the descriptors selected by ANNIGMA with the ones selected by SOM-DA. Nevertheless, the descriptors selected only by one method have corresponding descriptors belonging to the same descriptor class (Table 3.3) in the subsets selected by the other two methods. For example, all five molecular descriptors selected by CFS for the M fraction model for the BW30 membrane (descriptors 6, 23, 25, 28 and 33) belong also to the subset selected by SOM-DA. Four of them (i.e., 6 23, 25 and 33) belong also to the subset selected by ANNIGMA, while the remaining electrostatic descriptor (i.e., 28) was replaced by the latter method by another electrostatic descriptor (i.e., 24). Seven molecular descriptors were selected by both ANNIGMA and SOM-DA methods: 6, 16, 23, 24, 25, 30 and 33. The constitutional descriptors 5 and 7, selected only by ANNIGMA method, were replaced in the subset selected by SOM-DA by another topological descriptor (i.e., 4). Also, the quantum chemical descriptor 29, selected by ANNIGMA method, was replaced by the SOM-DA method with descriptor 31, which belongs also to the quantum chemical class.

Similar conclusion is achieved in the case of P fraction models. For example, three of the molecular descriptors selected by CFS for modeling the P fraction in the case of BW30 membrane (i.e., 8, 16 and 30) were selected also by SOM-DA. The constitutional descriptor 7, selected by CFS, was replaced in the subset selected by SOM-DA by four other constitutional descriptors (i.e., 1, 4, 5 and 6). Also, the electrostatic descriptor 24 selected by CFS was replaced by the electrostatic descriptors 25 and 26 in the subset selected by SOM-DA. For the same membrane, three molecular descriptors were selected by both CFS and ANNIGMA (i.e., 7, 8 and 21). The electrostatic descriptor 24 and the quantum chemical descriptor 30 selected by CFS, were replaced by the electrostatic descriptors 23 and 27, and by quantum chemical descriptors 33 and 34, respectively, in the subset selected by ANNIGMA. Two molecular descriptors were selected by both ANNIGMA and SOM-DA (i.e., 6 and 8). The constitutional descriptor 7 selected by ANNIGMA was replace in the subset selected by SOM-DA by other three constitutional descriptors (i.e., 1, 4 and 5). Also, the electrostatic descriptors 23 and 27, together with the quantum chemical descriptors 33 and 34 selected by ANNIGMA, were replaced in the subset selected by SOM-DA by electrostatic descriptors 25 and 26 and quantum chemical descriptors 29, 30 and 35, respectively.

ISBN:978-84-691-2701-8/DL:T.386-2008

A close examination of molecular features selected in Table 3.4 reveal descriptor selection similarities between the polyamide and cellulose acetate membranes. For example, comparing the input sets selected with the CFS method, molecular descriptors 6 and 28 were commonly selected for all five membranes for the M fraction model. Similarly, molecular descriptors 8 and 30 were commonly selected for all five membranes for predicting the P fraction. However, certain differences were also observed. For example, for the M fraction prediction, molecular descriptors 23, 25 and 33 were selected by the CFS method only for the polyamide membranes, while molecular descriptor 30 was selected only for the cellulose acetate membrane. For the P fraction, molecular descriptors 16, 21 and 24 were selected only for the PA membranes, while molecular descriptors 14 and 28 were selected only for the CA membrane. Similarly, with the SOM-DA method, molecular descriptors 4, 6, 17, 23, 26, 28, 30, 31 and 33 were selected for all five membranes for the M fraction, while molecular descriptors 1, 4, 5, 6, 8, 9, 16, 30 and 35 were selected for all five membranes for the P fraction. It should be noted that molecular descriptor 8 was selected for the M fraction only for the CA membrane, while molecular descriptor 25 was selected only for the four PA membranes. Molecular descriptors 19 and 21 were selected only for the P fraction and CA membrane, while molecular descriptor 29 was selected only for the four PA membranes. ANNIGMA selected for each one of the five membranes the molecular descriptors 6, 7, 23 and 29 for the M fraction, and molecular descriptors 6, 7 and 27 for the P fraction, respectively. For the M fraction, molecular descriptors 16 and 25 were selected only for the four PA membranes, while molecular descriptors 4, 26, 28 and 35 were selected only for the CA membrane. In the case of the P fraction, molecular descriptor 33 was selected for all four PA membranes, and molecular descriptors 28 and 29 were selected only for the CA membrane. The above results are consistent with the expectation that the significance of specific solute chemical descriptors for the prediction of solute permeation and sorption (i.e., P and M fractions) should also vary with membrane properties.

Selection of model input parameters for Membrane-Composite ANQ models

Among the input parameters selected for the M fraction Membrane-Composite ANQ/PA model, molecular descriptors 4, 6, 17, 28, 30, 31 and 33 were selected also for M fraction Independent ANQ models of each one of the four PA membranes. In the case of P fraction, molecular descriptors 4, 5, 6, 9, 16, 29 and 30 selected for the Membrane-Composite ANQ/PA

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

model were selected also for the Independent ANQ models for every PA membrane. These parameters were also selected by SOM-DA in the most suitable set of input parameters for the Independent ANQ models built for the CA membrane. Therefore, the input parameters presented in Table 3.4, as selected for the composite collection of all four PA membranes, were also used for developing Membrane-Composite ANQ/PACA models. The most suitable set of input descriptors for building Membrane-Composite ANQ models for the PA membranes included membrane descriptors 47, 48 and 52 for the M fraction model and 47 for the P fraction model. For the Membrane-Composite ANQ/PACA models, the membrane specific water flux was also added among the model inputs.

Selection of model input parameters for MP-Composite ANQ models

Molecular descriptors selected by ANNIGMA for the MP-Composite ANQ models for each membrane are a subset of the union of descriptors selected by the same method for the M and P fraction Independent ANQ models. Molecular descriptors 5, 6, 7, 23, 24, 25, 27, 28, 29 and 33 were selected for all four polyamide membranes. Among these, molecular descriptors 6, 23, 27, 28 and 29 were selected also for the cellulose acetate membrane. Molecular descriptors 5, 7 24, 25 and 33 were selected only for the PA membranes, while molecular descriptors 4 and 26 were selected only for the CA membrane.

Correlating input descriptors for organic chemical separation performance

The most relevant molecular descriptors that characterize membrane performance in terms of organic solute passage and sorption, and calculated rejection, can be identified via analysis of the frequency of occurrence of the different molecular descriptors in the optimal input sets selected by the CFS, SOM-DA and ANNIGMA feature selection methods (Table 3.4) for the Independent ANQ models. Accordingly, the molecular descriptors identified as most relevant for correlating solute sorption (M fraction) are the size of the smallest ring (6), dipole moment (23), dipole hybridization (28), LUMO energy (30) and heat of formation (33). In addition, the dipole vector Y (25) was also selected as relevant for correlating solute sorption by the polyamide membranes. The most influential molecular descriptors for correlating solute passage (P fraction) for either the polyamide or cellulose acetate

ISBN:978-84-691-2701-8/DL:T.386-2008

membranes are the size of the smallest ring (6), molecular weight (8), shape index kappa 2 (16) and LUMO energy (30). For the cellulose acetate membrane, dipole hybridization (28) was selected as an additional parameter to characterize the P fraction.

The current identification of molecular descriptors as most relevant for describing organic passage, sorption rejection by RO membranes is in general agreement with previous studies. For example, previously has been reported that molecular size and steric effects influence organics rejection [32,35,37-39]. Specifically, descriptors selected in the present approach which characterize molecular size and steric effects included, for example, molecular weight (8), shape index kappa 2 (16), moment of inertia B (19). Other selected descriptors are the size of the smallest ring (6) and the heat of formation (33). The selection of the former is consistent with the fact that 70% of the compounds in the study set, those pertaining to Family A in Figure 3.2a and Table 3.1, contain at least one aromatic ring. Selection of the heat of formation (33) can also be rationalized by the fact that this parameter is related, among other factors, to molecular size and molecular bonds stability in relation to structural complexity.

The current feature selection methods also identified molecular dipole parameters, such as dipole moment (23), dipole vector Y (25) and dipole hybridization (28), in addition to the LUMO Energy (30), as relevant molecular information for organic compounds passage through RO membranes. The identification of dipole moment descriptors is consistent with previous studies [31,33,34,39] that have suggested the importance of the dipole moment as a factor affecting solute-RO membrane electrostatic interactions [35]. Previous studies have also suggested that the rejection of organic compounds is strongly influenced by surface hydrophobic/hydrophilic interactions that have been typically correlated with the solute octanol-water partition coefficient [36-38]. It is emphasized that the octanol-water partition coefficient (K_{ow}) is not a fundamental molecular parameter and thus it was not explicitly considered in the present initial set of descriptors. However, a number of the molecular descriptors identified in Table 3.4 as relevant for organic passage and sorption, i.e., molecular weight (8), dipole moment (23) and dipole hybridization (28), have also been previously identified as relevant molecular descriptors for the prediction of K_{ow} [102].

The polyamide membrane properties that were identified by the feature selection methods as being significant correlating parameters for organic compound sorption and passage were

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

the zeta potential (47), zeta potential slope (pH 5-7) (48) and polyamide thickness (52). The first two parameters are associated with membrane charge, while the latter affects membrane permeability. The specific water flux (54), used to differentiate between polyamide and cellulose acetate membranes, is also related with the membrane permeability and to the membrane pore size. Surface charge, membrane permeability and membrane pore size have been reported to affect organic rejection by RO membranes [32,58,60].

Performances of QSPR models for solute sorption, passage and rejection

The performances of all QSPRs developed are presented in ANNEX III for the internal LOO cross-validation and in ANNEX IV for the independent test set compounds used in the external validation. Given the fact that models performance for the four polyamide membranes were similar as determined by both internal and external validation methods, in the next sections are presented and discussed in detailed the results for only two of them (BW30 and TFCHR), together with the results for the cellulose acetate membrane. The selected descriptors that best explain the chemical behavior for the BW30 and LFC1 membranes, as well as for the TFCHR and ESPA2 membranes, are almost coincident, as presented in Table 3.4. Furthermore, the ranges of experimental M, P and calculated R fractions for the BW30 and TFCHR are representative for the considered polyamide membranes, as seen in ANNEX II. For brevity of reporting, the average relative errors are presented in parenthesis, just after the corresponding absolute values. It should be noted that these error calculations exclude mass fraction values that are equal to zero or that could be considered zero based on the average standard deviation of the experimental measurements for the data set under consideration.

Independent ANQ models

The M and P mass fractions predicted by the LOO internal validation Independent ANQ models, together with the calculated R fraction, are depicted in Figures 3.4 - 3.6 for the BW30, TFCHR and CA membranes, respectively. The external validation predictions for the M and P fractions, together with the calculated R fractions, are plotted in Figure 3.7 for the same

ISBN:978-84-691-2701-8/DL:T.386-2008

three membranes. All figures include the results obtained with the sets of descriptors selected by the CFS, SOM-DA and ANNIGMA methods (Table 3.4). All Independent ANQ models developed, including those for the ESPA2 and LFC1 membranes, showed an explained variance in the prediction of M and P fractions, and calculated R fraction, higher than 0.975 for internal LOO cross-validation. As expected, the explained variance for external test set validation presented lower values. However, in most of the cases the explained variance in prediction exceeded 0.900, which also indicates a remarkable model performance. The worst results were obtained for the calculated R fraction for CA membrane using the descriptors selected by CFS, which can be attributed to the small number of descriptors selected for the M fraction model for this membrane-feature selection combination (Table 3.4). The same reason is the cause of the lower model performance observed in Figure 3.7g. The internal validation average absolute errors for all predicted fractions were up to 0.020 (average relative error of 12.4%). In the case of external validation, the average absolute errors increased up to 0.077 (70.9%), except for the CA membrane models with descriptors selected by the CFS method which approximately doubled the average absolute deviation.

Internal validation with LOO models. In order to explore the adequacy of the selected chemical descriptors and to confirm their proper identification, internal LOO validation analysis was carried for Independent models for the M and P fractions, as their governing mechanisms respond to different solute/membrane interactions. The LOO validation for the M and P models and for the calculated R fraction, as shown in Figures 3.4 and 3.5 for the BW30 and the TFCHR membranes, revealed good performance. Explained variance in prediction higher than 0.976, and average absolute errors smaller than 0.017 (7.0%) were obtained for all predicted mass fractions for the BW30 and TFCHR membranes. Slightly higher average errors were obtained for the CA membrane models (Figure 3.6), with the highest average relative error of 12.4% when the chemical descriptors were selected by the CFS method. The internal validation maximum absolute errors were as high as 0.230 (270.2%), indicating the presence of outliers, particularly for the CA membrane case.

Predicted M and P fractions for the BW30 membrane with LOO models are in good agreement with the measured organic fractions as is evident in Figure 3.4. Performance of the M and P models based on CFS selected descriptors (Figure 3.4a,d,g) was with average absolute errors of 0.006 (5.1%) and 0.005 (5.2%), with standard deviations of 0.009 (11.8%)

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

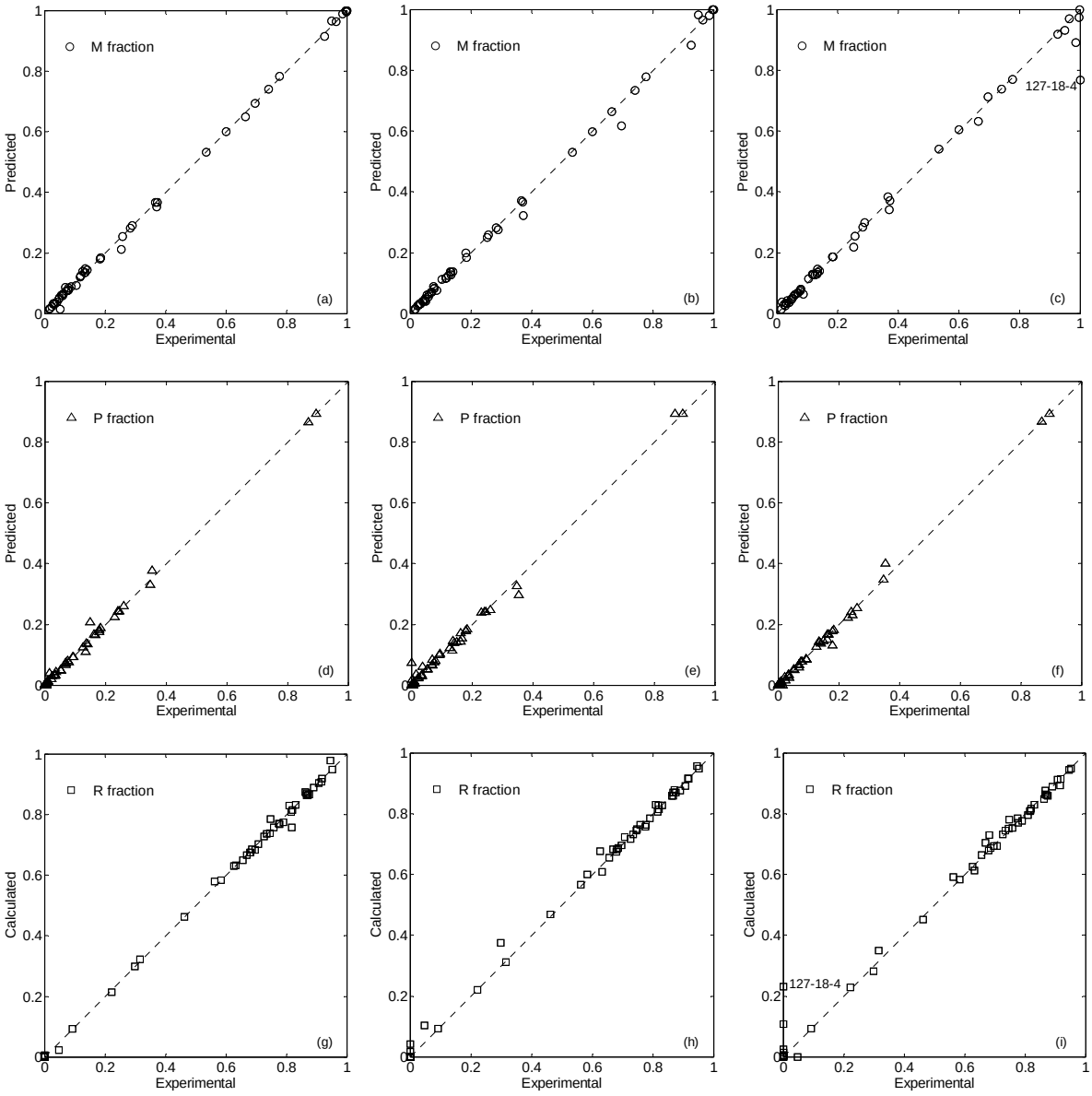


Figure 3.4. LOO cross-validation of Independent ANQ models for the polyamide BW30 membrane. Predicted M fractions with (a) CFS, (b) SOM-DA and (c) ANNIGMA descriptors; Predicted P fractions with (d) CFS, (e) SOM-DA and (f) ANNIGMA descriptors; Calculated R fractions from predicted M and P fractions with (g) CFS, (h) SOM-DA and (i) ANNIGMA.

and 0.010 (8.5%), respectively. For the calculated R fraction, the average absolute error was 0.007 (1.2%), with a standard deviation of 0.011 (1.6%). When the input molecular descriptors to the LOO models were selected by the SOM-DA method, both M and P fractions (Figure 3.4b,e,h) were predicted with essentially the same average absolute errors of 0.008 (4.2% for M and 7.0% for P), with corresponding standard deviations of 0.014 (5.3% for M and 12.8% for P). For the calculated R fraction, the average absolute error was 0.010 (1.8%), with a standard deviation of 0.016 (4.2%). Similar results were obtained also when using the input descriptors selected by ANNIGMA (Figure 3.4c,f,i), when the average absolute errors for the

ISBN:978-84-691-2701-8/DL:T.386-2008

M and P fractions models were 0.013 (4.2%) and 0.006 (6.2%), respectively, with the corresponding standard deviations of 0.035 (5.9%) and 0.010 (8.3%). The average absolute error for the calculated R fraction was 0.017 (1.8%) with a standard deviation of 0.036 (2.4%).

Comparison of Figures 3.4 and 3.5 indicates that the LOO models, for the M and P fractions, built independently for the BW30 and TFCHR polyamide membranes perform equally well. Performance of the M and P models for the TFCHR membrane based on CFS selected descriptors (Figure 3.5a,d,g) was with average absolute errors of 0.010 (6.3%) and 0.004 (3.3%), with standard deviations of 0.018 (15.0%) and 0.006 (3.6%), respectively. For the calculated R fraction, the average absolute error was 0.012 (2.4%), with a standard deviation of 0.018 (3.0%). When the input molecular descriptors were selected by SOM-DA (Figure 3.5b,e,h), model performance for the M and P fractions for the TFCHR membrane were with average absolute error of 0.006 (3.1%) and 0.007 (3.2%), with standard deviations of 0.007 (3.4%) and 0.027 (4.8%), respectively. In this case, the average absolute error for the calculated R fraction was 0.010 (1.2%) with a standard deviation of 0.025 (1.3%). When the input molecular descriptors were selected by the ANNIGMA method, the M and P fractions (Figure 3.5b,e,h) were predicted with average absolute errors of 0.012 (6.6%) and 0.004 (3.8%), respectively, with corresponding standard deviations of 0.027 (14.8%) and 0.008 (4.6%). For the calculated R fraction the average absolute error was 0.015 (1.8%) with a standard deviation of 0.027 (3.0%).

Organic compounds presenting high deviation between the predicted and experimental mass fractions (Figures 3.4 and 3.5) can be considered outliers. For example, for M fraction model build with descriptors selected by ANNIGMA for the BW30 polyamide RO membrane (Figure 3.4c), the M and P fractions for 1,1,2,2-tetrachloroethylene (127-18-4) presented an absolute deviation of 0.230 (23.0%). Figure 3.2b shows that this compound is classified alone in its SOM unit. Moreover, the distance from this compound to the center of its unit is higher than the average map topographic distance. As expected, this compound presented a high absolute deviation also for the calculated R fraction. For the second polyamide membrane (TFCHR), two outliers are revealed in Figure 3.5: N-nitroso dimethyl amine (62-75-9) in the case of the P model fraction built using the molecular descriptors selected by SOM-DA, and 1,1,2,2-tetrachloroethylene (127-18-4) for the M fraction model

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

developed based on the ANNIGMA selected descriptors. Figure 3.2b shows that N nitroso dimethyl amine (62-75-9) is also classified alone in its SOM unit.

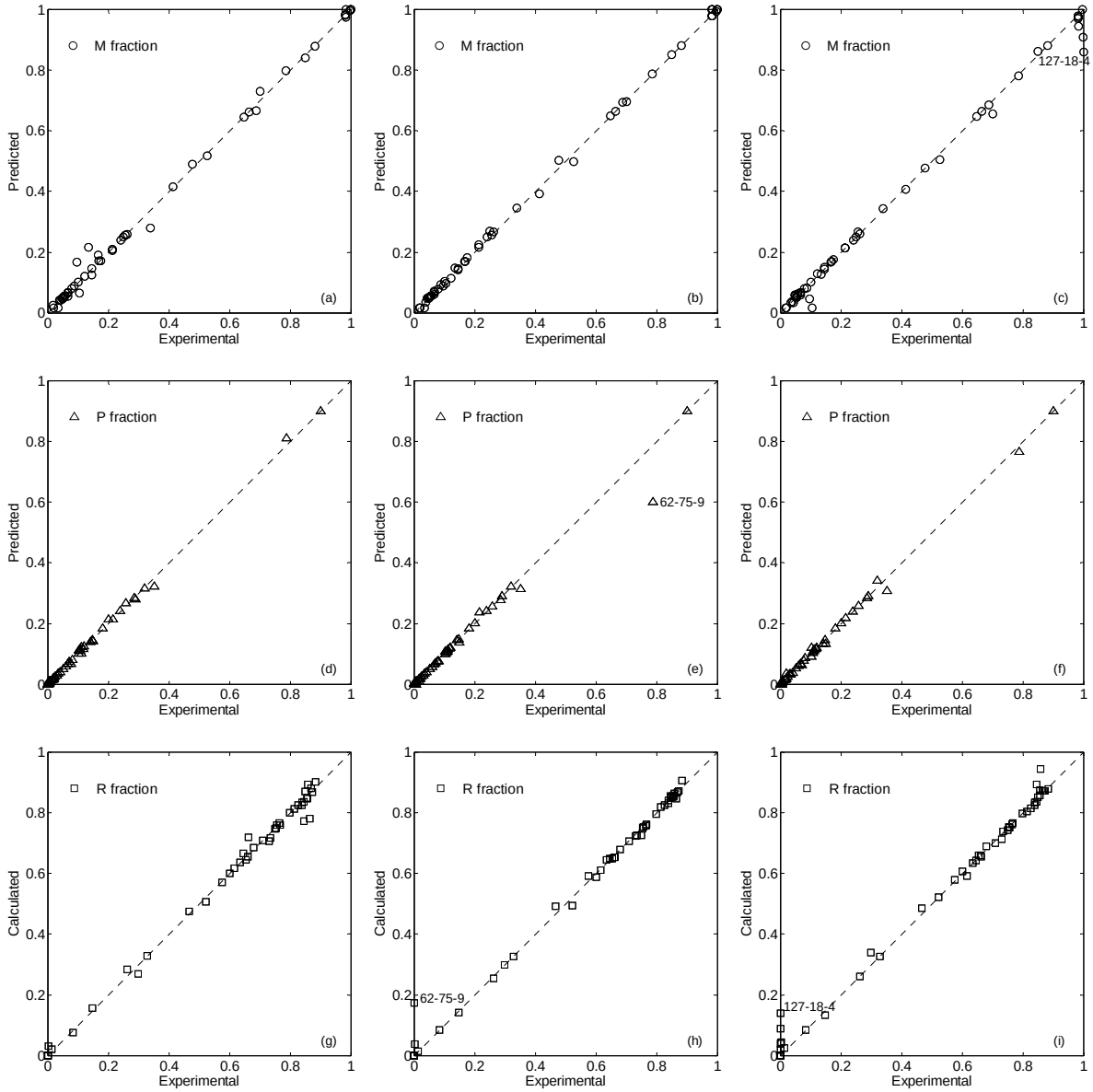


Figure 3.5. LOO cross-validation of Independent ANQ models for the polyamide TFCHR membrane. Predicted M fractions with (a) CFS, (b) SOM-DA and (c) ANNIGMA descriptors; Predicted P fractions with (d) CFS, (e) SOM-DA and (f) ANNIGMA descriptors; Calculated R fractions from predicted M and P fractions with (g) CFS, (h) SOM-DA and (i) ANNIGMA.

The results reported in Figures 3.4 and 3.5 for the two polyamide membranes are coherent in terms of the applicability domain of current models as determined by the chemical information contained in the dataset. LOO models built for the CA membrane yield predictions with higher deviations than those for the polyamide membranes. The best performance for this membrane was obtained when using the molecular descriptors selected by ANNIGMA (Figure 3.6c,f,i). The M and P fractions were predicted with average absolute

ISBN:978-84-691-2701-8/DL:T.386-2008

errors of 0.014 (7.4%) and 0.009 (2.1%), respectively, with corresponding standard deviations of 0.035 (17.9%) and 0.020 (3.8%). For the calculated R fraction, the average absolute error was 0.014 (9.8%) with the standard deviation of 0.035 (20.6%). For these models, two compounds act like outliers: ibuprofen (15687-27-1) in the M fraction model with an absolute deviation of 0.217 (105.8%), and 2,4 dichlorophenol (120-83-2) in the P fraction model presenting an absolute deviation of 0.124. Both compounds are allocated to single map units in the SOM classification presented in Figure 3.2b. The predicted M and P fractions in the CA membrane with LOO models based on the SOM-DA selected descriptors (Figure 3.6b,e,h) are in agreement with measurements with similar absolute average errors of 0.014 (8.2% for M and 3.0% for P) and standard deviations of 0.021 (13.5%) and 0.028 (3.9%), for the M and P fractions, respectively. The average absolute error for the calculated R fraction was 0.020 (11.6%) with the standard deviation of 0.030 (18.5%). In the M fraction model, the prediction of lindane (58-89-9) was observed to deviate significantly from the experimental value with 0.124 (12.6%). For the P fraction model, three compounds act like outliers: 2,4 dichlorophenol (120-83-2), cimetidine (51481-61-9) and ibuprofen (15687-27-1). These four compounds are allocated to single map units in the SOM classification presented in Figure 3.2b. The three CFS selected descriptors (Table 3.4) did not provide sufficient information for the LOO M fraction model developed for the CA membrane (Figure 3.6a) which is partially the reason for the large average absolute error of 0.018 (12.4%) and corresponding standard deviation of 0.030 (40.4%) observed in this case. However, in this case only the prediction of 2,4 dichlorophenoxyacetic acid (94-75-7) was observed to deviate significantly from its experimental value with 0.143 (270.2%). The performance of the predicted P fraction was with average absolute error of 0.012 (3.3%) with the standard deviation of 0.020 (4.4%), while the calculated R fraction presented the average absolute error of 0.020 (11.1%) with the standard deviation of 0.030 (19.0%). In this case, ibuprofen (15687-27-1) appeared also as an outlier, in spite of the fact that its predictions did not presented significant deviation from the experimental value in neither M fraction, nor P fraction models.

Figures 3.4, 3.5 and 3.6 illustrate that it is possible to describe the RO membrane performance with respect to organic compounds with the proper selection of molecular information (Table 3.4). Good agreement was obtained between the predicted and measured fractions over the entire experimental mass fraction range (i.e., [0,1]).

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

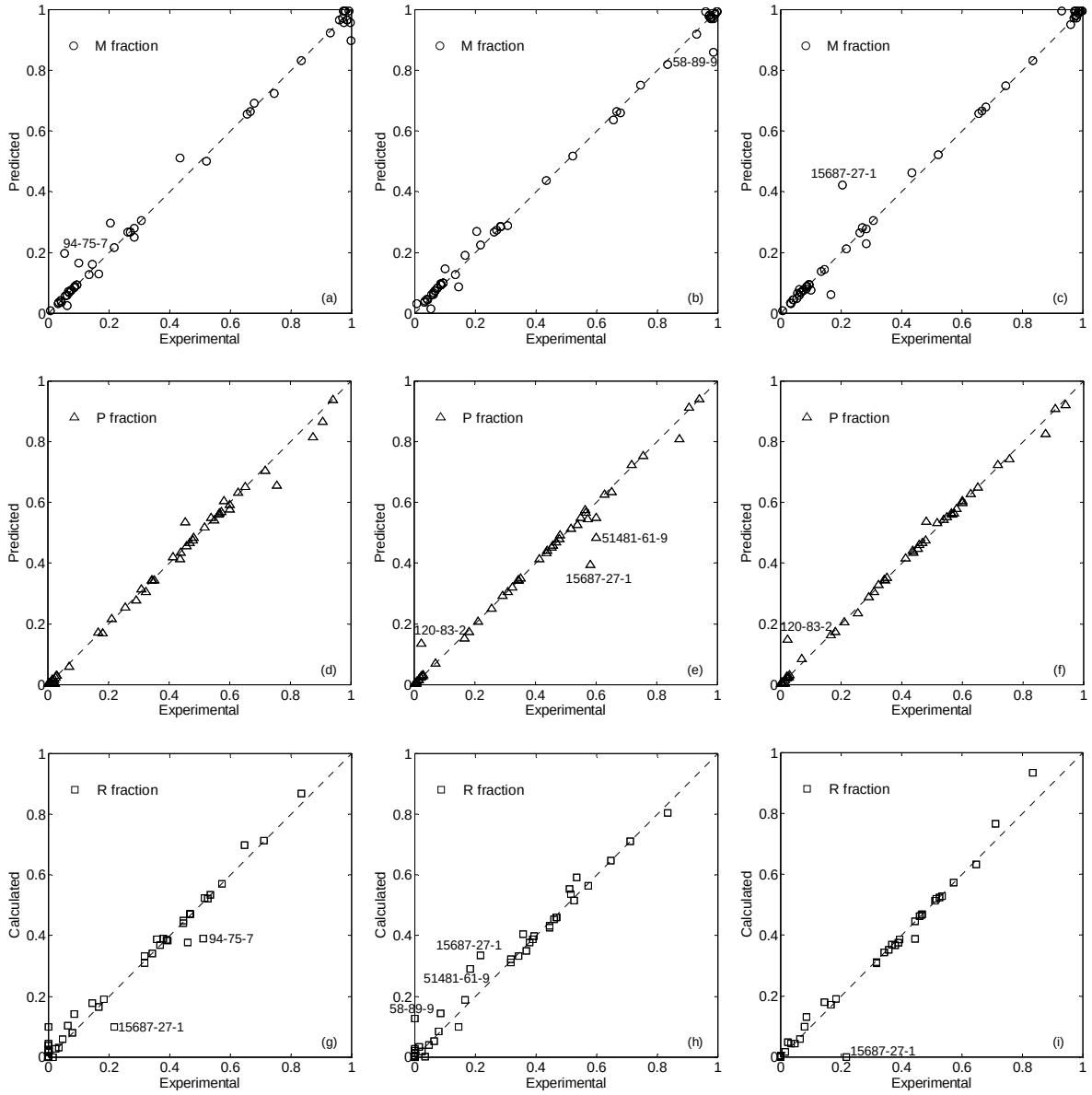


Figure 3.6. LOO cross-validation of Independent ANQ models for the cellulose acetate CA membrane. Predicted M fractions with (a) CFS, (b) SOM-DA and (c) ANNIGMA descriptors; Predicted P fractions with (d) CFS, (e) SOM-DA and (f) ANNIGMA descriptors; Calculated R fractions from predicted M and P fractions with (g) CFS, (h) SOM-DA and (i) ANNIGMA.

External validation of QSPR models. External validation is more demanding that the LOO cross-validation models discussed previously, particularly for small datasets as the one presented in Table 3.1, since the former is performed with never seen before test compounds while the latter maximizes the amount of information used for training (49 compounds in this case) and minimizes the information used for testing (1 compound) in several consecutive models (50 in this case). The acceptable compactness of the chemical space in Figure 3.2 justifies the application of an external validation, which was carried out by

ISBN:978-84-691-2701-8/DL:T.386-2008

dividing the small data set of 50 compounds into 44 compounds for training the M and P fractions QSPR models and 6 for model testing, following the SOM procedure outlined in Section 3.1. Training and test compounds were different for all M and P models, even for the same membrane. Thus, the total number of test compounds for the calculated R fractions (i.e., $R = 1 - \left(\hat{M} + \hat{P} \right)$) was always larger than 6 and at most equal to 12, including compound pairs of either test M-test P, test M-train P or train M-test P.

The predicted M, P and calculated R fractions for the test compounds are compared with experimental measurements in Figure 3.7. The Independent QSPR models developed with descriptors selected by CFS, SOM-DA or ANNIGMA method, showed explained variance in prediction indices for the test set compounds higher than 0.874, for the three membranes selected for detailed analysis. An exception is the CA membrane M fraction model based on the CFS selected descriptors, when a lower explained variance in prediction of 0.828 was obtained. This is also the cause of the drastically decrease to $q^2 \approx 0.331$ in the case of the calculated R fraction. Except the latter case, the values compared very well with the ones obtained for the LOO cross-validation, especially considering the heterogeneous nature and the small number of 44 training compounds.

Evaluation of the M and P fraction models with the external data test set, is shown in Figure 3.7a for the BW30 membrane based on the models built with the CFS selected molecular descriptors. The absolute average errors for the M and P fraction models are 0.034 (17.6%) and 0.024 (42.6%), respectively, with corresponding standard deviations of 0.040 (14.1%) and 0.015 (49.8%). Deviations of the same order of magnitude were also observed for the calculated R fractions: average absolute error of 0.048 (7.5%) with the standard deviation of 0.040 (4.7%). Predicted M and P fractions for the same BW30 membrane, with models developed using descriptors selected by SOM-DA method (Figure 3.7b), reveal comparable behavior, with the average absolute errors obtained for the predicted M and P fraction models respectively of 0.066 (70.9%) and 0.018 (44.5%), with standard deviations of 0.064 (88.2%) and 0.021 (70.2%). For the calculated R fraction, the average absolute error was 0.044 (7.8%) with the standard deviation of 0.054 (9.4%). Similar results were obtained also when using the molecular descriptors selected by ANNIGMA (Figure 3.7c). The average absolute errors for the predicted M and P fractions and calculated R fraction were of 0.053 (16.5%),

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

0.013 (8.8%) and 0.059 (7.1%), respectively, with the corresponding standard deviations of 0.035 (12.1%), 0.007 (4.9%) and 0.039 (5.2%).

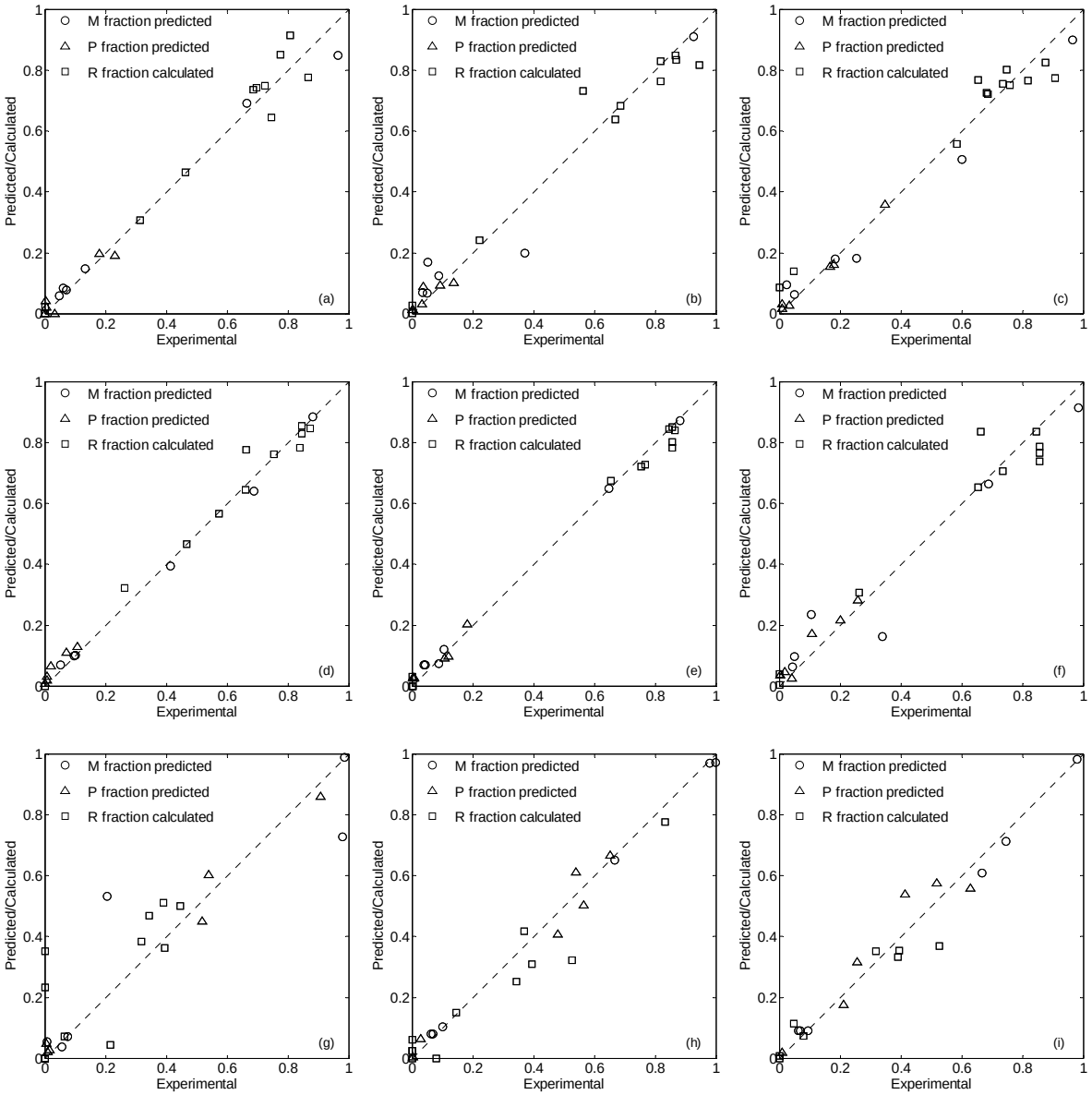


Figure 3.7. External validation for the Independent ANQ models for the BW30, TFCHR and CA membranes with the descriptors selected by CFS, SOM-DA and ANNIGMA for the M, P and R fractions corresponding only to the test set compounds. BW30 with (a) CFS, (b) SOM-DA and (c) ANNIGMA; TFCHR with (d) CFS, (e) SOM-DA and (f) ANNIGMA; CA with (g) CFS, (h) SOM-DA and (i) ANNIGMA.

Comparable performance was obtained for the TFCHR polyamide membrane (Figure 3.7d,e,f). Average absolute errors for the predicted M and P fraction models and calculated R fraction with descriptors selected by CFS (Figure 3.7d) were 0.015 (8.7%) with standard deviation of 0.017 (13.8%), 0.025 (38.5%) with standard deviation of 0.015 (25.2%) and 0.027 (5.9%) with standard deviation of 0.035 (8.0%), respectively. For the models built with

ISBN:978-84-691-2701-8/DL:T.386-2008

descriptors selected by SOM-DA method (Figure 3.7e), the average absolute errors were 0.017 (20.2%) with standard deviation of 0.012 (27.2%) for the M fraction and 0.021 (15.9%) with standard deviation of 0.003 (3.3%) for the P fraction. For the calculated R fraction, an average absolute error of 0.025 (3.7%) with the standard deviation of 0.023 (2.8%) was obtained. When using the ANNIGMA selected descriptors (Figure 3.7f), the average absolute errors obtained for the M, P and R fractions were 0.077 (55.1%), 0.030 (28.3%) and 0.057 (10.1%), respectively, with the corresponding standard deviations of 0.063 (48.3%), 0.019 (25.1%) and 0.056 (9.1%).

As in the LOO cross-validation models, the worst external validation results were obtained for the cellulose acetate membrane (Figure 3.7g,h,i). The poorer performance of the models built using the descriptors selected by the CFS method (Figure 3.7g) could be attributed, in part, to the reduced number of descriptors (i.e., 3) selected in this case for M fraction. As a result the chemical information provided to the QSPR model was insufficient and thus average absolute errors for predicted M fraction were as high as 0.112 (44.3%), with a standard deviation of 0.135 (67.3%). Lower deviations of 0.041 (10.1%), with a standard deviation of 0.025 (4.5%), were obtained for the predicted P fractions for this membrane. Consequently, the calculated R fraction also showed a high average absolute error of 0.101 (33.9%) with the standard deviation of 0.113 (29.6%). As expected, model predictions improved significantly when the M and P fraction models were developed using the SOM-DA or ANNIGMA selected descriptors (Figure 3.7h,i). In these cases, the average absolute errors obtained for the M fraction models were of 0.012 (8.5%) and 0.025 (15.6%), respectively, with the corresponding standard deviations of 0.008 (10.6%) and 0.019 (18.4%). For the P fraction models, the average absolute errors were 0.043 (10.4%) with standard deviation of 0.030 (5.3%) when SOM-DA selected descriptors were used, respectively 0.060 (18.7%) with standard deviation of 0.039 (8.5%) when ANNIGMA selected descriptors were used. It should be noted also that the average experimental standard deviation of P fractions for the CA membrane (0.036) is higher than the ones obtained for the BW30 and TFCHR polyamide membranes (0.023). It is also emphasized that the fact that the experimental P fraction data for the CA membrane covered the entire [0-1] range, as opposed to the smaller ranges for the organic passage fractions for the PA membranes, is partially responsible for the poorer performance of the models developed for the CA membranes.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Membrane-Composite ANQ models

The M and P fractions predicted by the QSPR models built for the composite collection of all five membranes considered, together with the calculated R fractions, are depicted in Figure 3.8, for both internal and external validation. It should be noted that in Figure 3.8 each compound is represented by five different points, corresponding to the five RO membranes. Therefore, in the present case the input database is larger, including the experimental data for the 50 organic compounds and all five membranes considered (i.e., in this case there are 248 available experimental points). However, during the LOO cross-validation were developed only 50 models, in each one selecting the experimental data corresponding to one molecule as test set, and using the rest of the data for training. For the external validation, 7 compounds were selected for testing the M fraction model (representing 33 points), while the rest of 43 compounds were used for training (representing 215 points). For the P fraction model, 6 compounds were selected for testing (representing 30 points), while the rest of 44 compounds were selecting for testing (representing 218 points). The total number of test compounds for the calculated R fraction was 12.

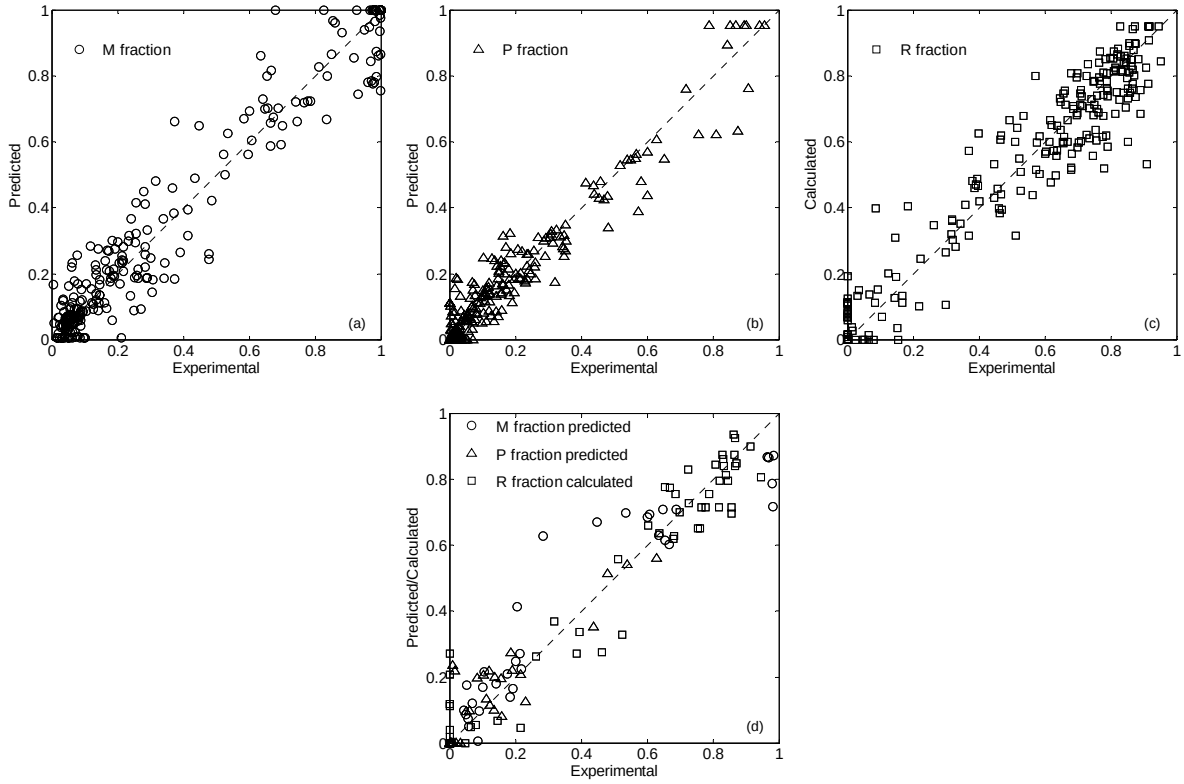


Figure 3.8. Membrane-Composite ANQ/PACA models. Internal validation for the (a) predicted M fraction, (b) predicted P fraction and (c) calculated R fraction. External validation (d) M, P and R fractions corresponding only to the test set compounds.

ISBN:978-84-691-2701-8/DL:T.386-2008

For the LOO cross-validation (Figure 3.8a,b,c), the explained variance in prediction for the M and P fraction models and calculated R fraction were higher than 0.933. The average absolute errors for the M, P and R fractions were 0.060 (34.9%), 0.038 (33.5%) and 0.067 (20.1%), respectively, with the corresponding standard deviations of 0.061 (45.6%), 0.043 (39.0%) and 0.061 (36.7%). In the case of external test set validation (Figure 3.8d), the explained variance in prediction indices decreased as low as 0.793, as expected. Also, the average absolute errors for the M and P fraction models increased to 0.088 (43.4%) and 0.052 (45.2%), respectively, with the corresponding standard deviations of 0.081 (51.6%) and 0.057 (39.5%). For the calculated R fraction, the average absolute error was 0.065 (13.4%) with the standard deviation of 0.062 (15.5%).

The above results demonstrate that the development of composite models for a collection of membranes is feasible if a sufficiently large number of membrane characteristics and data are available.

MP-Composite ANQ models

QSPRs that simultaneously considered the two experimental solute fractions are thought to better capture the membrane performance with respect to organic compounds separation. The reasoning behind this consideration is that in these models, molecular descriptors for all possible solute-membrane interaction are taken into account at once. However, the results obtained for the MP-Composite ANQ models are not much better than the ones obtained for the Independent ANQ models. This is mainly attributed to the small size of the data set. Since in the MP-Composite ANQ models the number of output variables increases (i.e., 2 outputs), in order to keep the number of connections below the number of experimental data contained in the training set, the number of hidden neurons would have to decrease according to Eq. (3.2). All MP-Composite models built, including those for ESPA2 and LFC1 membranes, presented an explained variance in prediction in the M and P fractions and calculated R fraction higher than 0.842 for the internal validation with the average absolute errors up to 0.054 (34.7%). The model performance decreased, as expected, in the case of external test set validation. The explained variance in prediction decreased down to 0.692, while the average absolute errors increased up to 0.145(109.8%).

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Internal validation with LOO models. The M and P fractions predicted by the LOO internal validation MP-Composite ANQ models for the BW30, TFCHR and CA membranes, together with the calculated R fractions are depicted in Figures 3.9, 3.10 and 3.11, respectively.

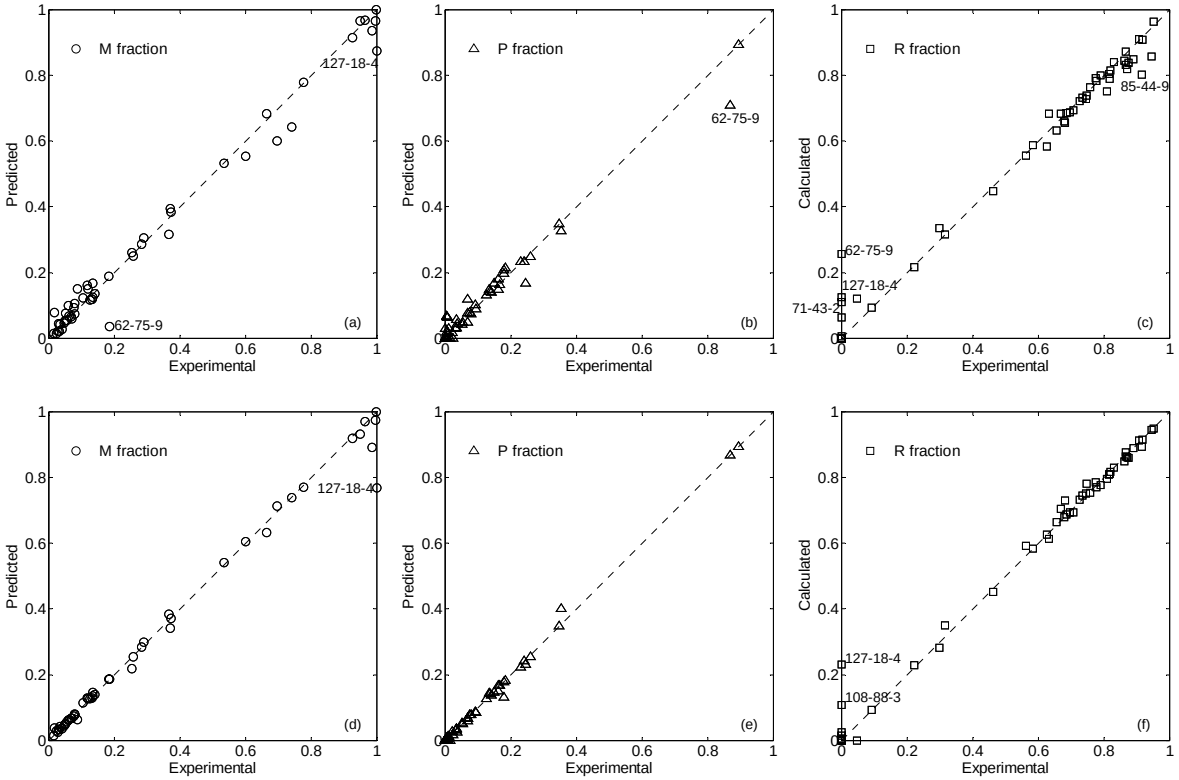


Figure 3.9. LOO cross-validation of MP-Composite ANQ models for the polyamide BW30 membrane. (a) predicted M fraction, (b) predicted P fraction, (c) calculated R fraction with CFS descriptors; (d) predicted M fraction, (e) predicted P fraction, (f) calculated R fraction with ANNIGMA descriptors.

The predicted M and P fractions and calculated R fraction for the BW30 membrane with the CFS selected descriptors (Figure 3.9a,b,c), were in good agreement with the corresponding experimental values. The average absolute errors for the M, P and R fractions were 0.025 (14.8%), 0.018 (16.3%) and 0.029 (2.9%), respectively, with the corresponding standard deviations of 0.032 (19.4%), 0.027 (22.5%) and 0.045 (3.1%). Similar results were obtained when using the molecular descriptors selected by ANNIGMA (Figure 3.9d,e,f). In this case, the model performance of the M, P and R fractions was with average absolute errors of 0.030 (13.2%), 0.022 (13.7%) and 0.028 (5.4%), respectively, with the corresponding standard deviations of 0.059 (19.2%), 0.038 (15.4%) and 0.038 (15.6%).

Comparable results were obtained also for the TFCHR membrane. When using the union of the molecular descriptors selected by CFS for the Independent ANQ models built for the M and P fractions (Figure 3.10a,b,c), the average absolute errors were 0.035 (13.6%) with a

ISBN:978-84-691-2701-8/DL:T.386-2008

standard deviation of 0.067 (21.4%) for the M fraction, 0.029 (17.5%) with a standard deviation of 0.064 (18.9%) for the P fractions, and 0.033 (7.3%) with a standard deviation of 0.051 (16.8%) for the R fraction, respectively. The performance of the MP-Composite ANQ models based on ANNIGMA selected descriptors (Figure 3.10d,e,f) were 0.027 (15.1%) with a standard deviation of 0.041 (21.3%) for the predicted M fraction, 0.023 (28.6%) with a standard deviation of 0.029 (44.2%) for the predicted P fraction and 0.029 (4.5%) with a standard deviation of 0.035 (4.1%).

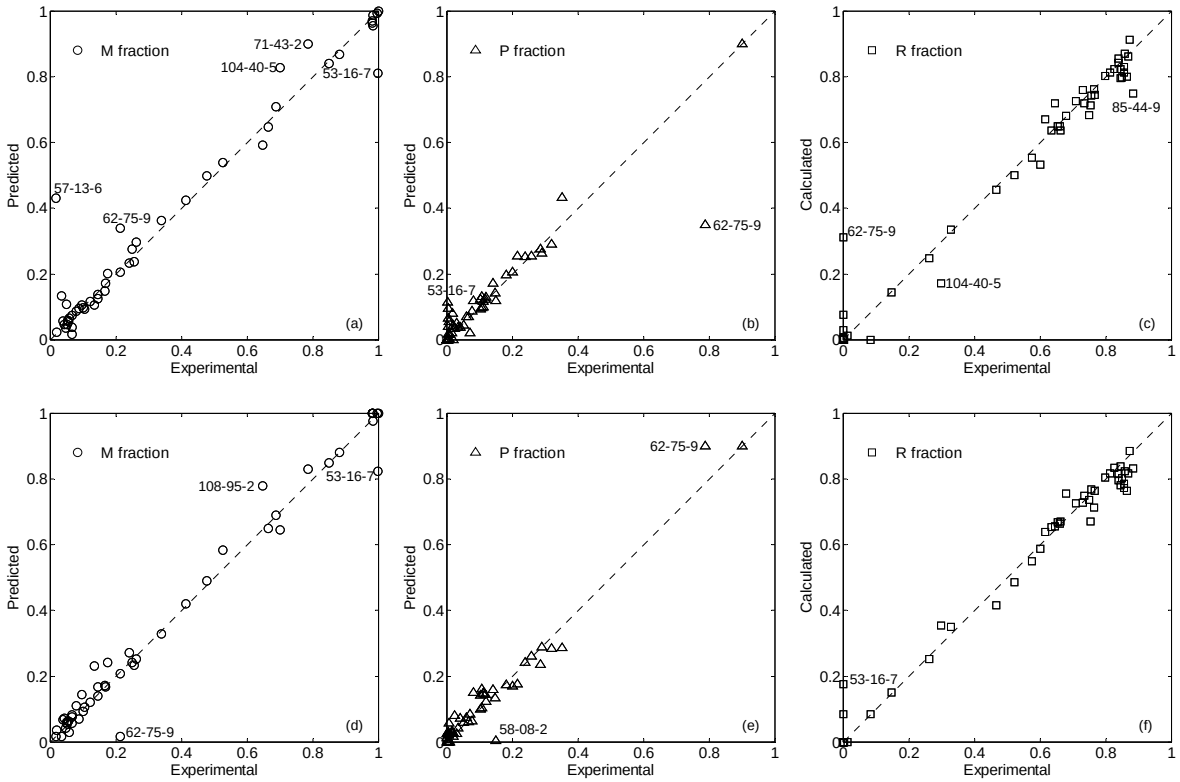


Figure 3.10. LOO cross-validation of MP-Composite ANQ models for the polyamide TFCHR membrane. (a) predicted M fraction, (b) predicted P fraction, (c) calculated R fraction with CFS descriptors; (d) predicted M fraction, (e) predicted P fraction, (f) calculated R fraction with ANNIGMA descriptors.

Internal validation analysis carried out for the CA membrane (Figure 3.11) yield predictions with slightly higher deviations than those for the polyamide membranes previously presented. When CFS selected descriptors were used (Figure 3.11a,b,c), the M and P fraction were predicted with average absolute errors of 0.033 (26.7%) the M fraction and 0.041 (10.7%) the P fraction. The standard deviations of the average absolute errors for these cases were 0.049 (54.3%) for the M fraction and 0.048 (10.1%) for the P fraction. For the calculated R fraction, the average absolute error was 0.051 (30.9%) with the standard deviation of 0.062 (43.3%). The performance of predicted M and P fractions and calculated R fraction when

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

ANNIGMA selected descriptors were used (Figure 3.11d,e,f) was with average absolute error of 0.030 (28.2%), 0.045 (12.3%) and 0.054 (34.7%), respectively, with the corresponding standard deviation of 0.032 (54.5%), 0.050 (11.3%) and 0.059 (50.5%).

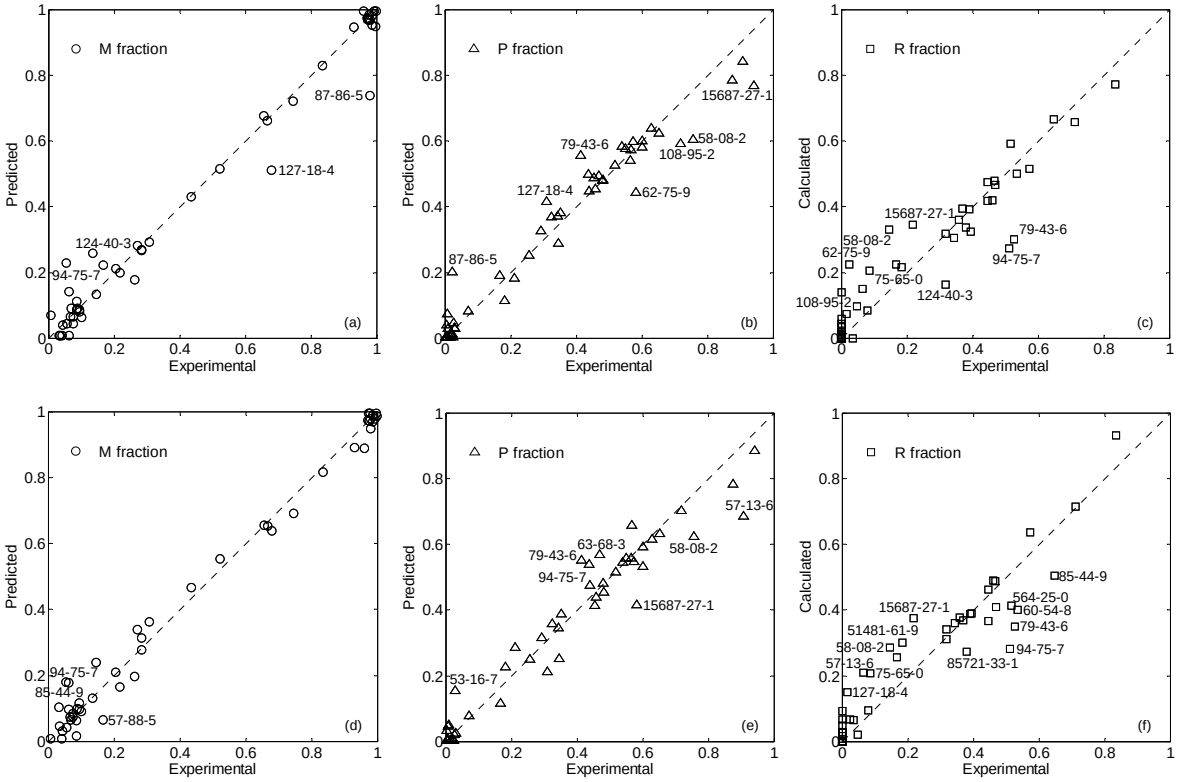


Figure 3.11. LOO cross-validation of MP-Composite ANQ models for the cellulose acetate CA membrane. (a) predicted M fraction, (b) predicted P fraction, (c) calculated R fraction with CFS descriptors; (d) predicted M fraction, (e) predicted P fraction, (f) calculated R fraction with ANNIGMA descriptors.

Most of the organic compounds presenting high deviation between the experimental values and the Independent ANQ models predictions were identified as outliers also in the MP-Composite ANQ models: 1,1,2,2-tetrachloroethylene (127-18-4), N-nitroso dimethyl amine (62-75-9), 1,4 dichlorophenoxyacetic acid (94-75-7), ibuprofen (15687-27-1) and cimetidine (51481-61-9). Moreover, compounds identified at the domain border by the PCA-based approach (Figure 3.2a), like tetracycline (60-54-8), doxycycline (564-25-0) or ciprofloxacin (85721-33-1) presented high absolute errors in the MP-Composite ANQ models. Several other compounds could not be properly described for at least one membrane-feature selection combination by the MP-Composite models.

External validation. The M and P fraction of the test set compounds for the BW30 membrane were predicted using the CFS selected descriptors (Figure 3.12a) with average absolute errors

ISBN:978-84-691-2701-8/DL:T.386-2008

of 0.076 (51.0%) with a standard deviation of 0.053 (31.2%) and 0.018 (18.2%) with a standard deviation of 0.017 (14.5%), respectively. For the calculated R fraction, the average absolute error was 0.085 (12.5%) with the standard deviation of 0.052 (9.7%). Slightly better results were obtained for the same membrane when using the ANNIGMA selected descriptors (Figure 3.12d). In this case, the average absolute errors were 0.052 (36.9%) with standard deviation of 0.042 (56.1%) for the predicted M fraction, 0.039 (53.8%) with standard deviation of 0.019 (71.1%) for the predicted P fraction, and 0.035 (4.9%) with standard deviation of 0.031 (4.7%) for the calculated R fraction.

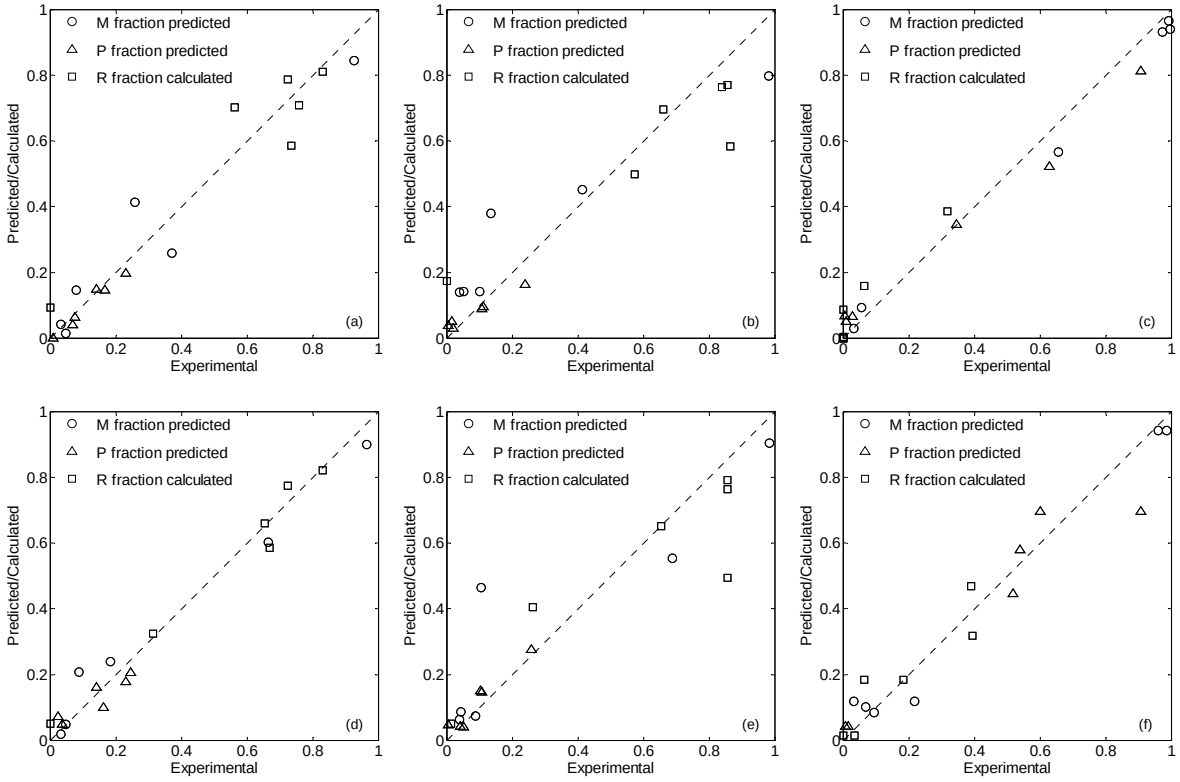


Figure 3.12. External validation for the MP-Composite ANQ models for the BW30, TFCHR and CA membranes with the descriptors selected by CFS and ANNIGMA for the M, P and R fractions corresponding only to the test set compounds. (a) BW30, (b) TFCHR and (c) CA with CFS descriptors; (d) BW30, (e) TFCHR and (f) CA with ANNIGMA descriptors.

Worst results were obtained when modeling the TFCHR membrane performance. When CFS selected descriptors were used (Figure 3.12b), the average absolute errors for the predicted M and P fractions and calculated R fraction were 0.117 (85.8%), 0.031 (21.3%) and 0.120 (13.9%), respectively, with the corresponding standard deviations of 0.082 (86.9%), 0.024 (9.3%) and 0.091 (10.7%). When ANNIGMA selected descriptors were used (Figure 3.12e), the average absolute errors were 0.109 (99.3%) with standard deviation of 0.131 (144.2%) for the predicted M fraction, 0.026 (22.4%) with standard deviation of 0.019 (19.0%) for the predicted

P fraction and 0.116 (23.0%) with standard deviation of 0.129 (23.8%) for the calculated R fraction.

For the CA membrane, when the CFS selected descriptors were used (Figure 3.12c), the average absolute errors were 0.041 (16.3%) with standard deviation of 0.029 (23.7%) for the predicted M fraction, 0.057 (9.2%) with standard deviation of 0.038 (8.2%) for the predicted P fraction and 0.042 (86.6%) with standard deviation of 0.046 (91.6%) for the calculated R fraction. When the ANNIGMA selected descriptors were used (Figure 3.12f), the average absolute errors for the predicted M and P fractions and calculated R fraction were 0.048 (65.1%), 0.080 (15.2%) and 0.052 (58.6%), respectively, with the corresponding standard deviations of 0.037 (107.3%), 0.068 (6.4%) and 0.048 (90.7%), respectively.

The LOO cross-validation and the external test set results obtained for the MP-Composite ANQ models, revealed slightly better results when using the molecular descriptors selected by ANNIGMA considering simultaneously the two experimental fractions, compared with the case of using reunion of molecular descriptors selected by CFS for the Independent ANQ models for the M and P fraction.

3.3. Validating the QSPR models

The validation of the current models was completed with the entirely new set of 143 compounds listed in ANNEX V, for which no experimental M and P mass fractions were available but that could be tested for mass balance. A valid set of models would require that the predicted sorbed, passed and rejected mass fractions for each compound ($\hat{M}$, $\hat{P}$ and $\hat{R}$, respectively) close the mass balance: $\hat{M}+\hat{P}+\hat{R}=1$. Accordingly, models for the experimental M and P fractions, and for the R fraction calculated from predicted M and P values to close the mass balance (i.e., $R_c=1-\left(\hat{M}+\hat{P}\right)$), were developed by using the entire set of experimentally screened 50 organic compounds Table 3.1, and tested with the additional 143 compounds of public health concern that were not experimentally characterized. This test set includes compounds such as endocrine disruptors, pharmaceutical active compounds, antibiotics and antimicrobial agents, neuroactive drugs, insecticides, herbicides, pesticides, industrial pollutants, fuel hydrocarbons and amino acids

ISBN:978-84-691-2701-8/DL:T.386-2008

[111-114]. The mass balance test was developed for the three membranes whose modeling results are presented in Section 3.2, using the following seven molecular descriptors: dipole moment, dipole vector X, dipole vector Y, dipole hybridization, heat of formation, size of smallest ring, and shape index kappa2. These molecular descriptors are among the ones identified in Section 3.2 as the most relevant to characterize the organic solute passage, sorbtion and rejection by RO membranes, based on the frequency of occurrence of different molecular descriptors in the optimal input sets selected by the three feature selection methods for the Independent ANQ models.

Table 3.5 shows that the new 143 compounds span slightly larger ranges of variation in the values of the molecular descriptors compared to those for the 50 organic compounds used to develop the three QSPR mass fraction models. Therefore, for model development, the mass fractions were normalized with the corresponding minimum and maximum mass fraction values in the training set of 50 chemicals, while the seven input descriptors listed in Table 3.5 were normalized with the minimum and maximum values of the complete set of 193 chemicals. This assured the possibility to extrapolate predictions beyond the chemical domain of the 50 training chemicals.

Table 3.5. Comparison between the range of variation of the seven molecular descriptors selected as input to the models, for the 50 chemicals with available experimental data and for the 143 organics without experimental data.

Molecular descriptor	Range for the 50 compounds with experimental data		Range for the 143 compounds without experimental data	
	min	Max	min	Max
Dipole moment	0.0	15.8	0.0	28.6
Dipole vector X	-8.6	10.9	-10.5	23.6
Dipole vector Y	-11.4	13.7	-15.2	13.7
Dipole hybridization	0.0	2.8	0.0	3.4
Heat of formation	-368.9	103.5	-755.0	249.2
Size of smallest ring	0.0	6.0	0.0	7.0
Shape index kappa2	1.0	10.7	1.0	25.7

The neural network architecture 7:5:1 was used for each one of the three mass fraction models, established using the conditions specified in Eq (3.2). The ability of the models to close the mass balance was assessed by computing the relative error for each one of the 193 chemicals,

$$\varepsilon(\%) = \frac{1 - \left(\hat{M} + \hat{P} + \hat{R}_c\right)}{1} \cdot 100 \qquad (3.5)$$

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Table 3.6. Mass balance relative errors for the 193 compounds, for each of the three membranes.

Index	CAS	BW30 [%]	TFCHR [%]	CA [%]	Index	CAS	BW30 [%]	TFCHR [%]	CA [%]
1	100-41-4	3.5	6.2	3.7	51	6804-07-5	7.6	1.8	9.9
2	104-40-5	0.9	13.3	2.5	52	100-75-4	6.0	4.0	26.1
3	108-88-3	0.5	16.0	0.0	53	1031-07-8	12.8	5.0	16.7
4	108-95-2	0.1	16.2	0.8	54	103-23-1	83.4	24.3	7.8
5	120-83-2	0.2	9.5	1.3	55	103-90-2	6.2	0.6	14.2
6	121-14-2	3.2	2.6	17.4	56	106-44-5	1.3	5.8	11.4
7	124-40-3	2.0	3.9	11.4	57	106-46-7	0.6	12.0	1.5
8	127-18-4	9.4	20.1	13.3	58	108-67-8	2.0	10.2	3.1
9	15687-27-1	1.1	3.0	8.3	59	108-86-1	2.2	7.9	8.3
10	19466-47-8	2.7	5.3	4.8	60	1141-38-4	14.9	6.5	0.6
11	2921-88-2	2.8	1.3	2.3	61	115-29-7	9.4	0.2	12.0
12	298-00-0	2.5	0.4	0.0	62	115-32-2	3.0	6.8	1.7
13	51481-61-9	0.4	0.1	0.5	63	115-86-6	12.5	15.6	35.5
14	52-90-4	4.3	4.9	5.3	64	115-96-8	44.4	90.8	22.7
15	53-16-7	0.8	1.1	0.6	65	117-81-7	15.1	17.7	7.2
16	56-40-6	2.3	1.8	8.8	66	117-84-0	15.1	9.2	0.7
17	56-41-7	2.7	11.4	2.8	67	118-74-1	5.4	4.5	4.6
18	564-25-0	5.8	0.4	4.2	68	120-12-7	3.8	2.0	2.6
19	56-53-1	1.8	1.2	0.0	69	121-82-4	35.8	15.7	47.1
20	56-84-8	1.9	4.7	5.8	70	122-11-2	4.5	1.8	54.0
21	56-87-1	1.6	6.3	0.7	71	122-34-9	6.6	1.6	4.1
22	57-13-6	0.4	6.6	2.6	72	124-48-1	10.4	1.9	9.1
23	57-83-0	4.9	5.0	1.1	73	127-79-7	3.1	16.2	11.2
24	57-88-5	1.2	2.7	1.1	74	12789-03-6	6.7	1.4	5.7
25	57-91-0	10.6	3.1	3.6	75	128-37-0	5.7	5.0	5.7
26	58-08-2	5.0	4.9	0.8	76	128-39-2	7.6	2.7	1.7
27	58-22-0	4.4	2.2	5.8	77	129-00-0	3.6	1.5	0.7
28	58-89-9	1.0	4.4	11.3	78	13071-79-9	4.5	10.3	7.7
29	60-00-4	2.0	4.1	4.2	79	134-62-3	3.9	5.8	0.0
30	60-54-8	2.2	4.3	1.0	80	136-85-6	4.4	12.1	4.2
31	62-75-9	1.0	4.7	2.4	81	139-13-9	4.0	13.5	13.5
32	63-68-3	2.1	6.3	0.5	82	1401-69-0	13.7	72.3	54.7
33	70-47-3	3.0	1.4	5.9	83	143545-90-8	83.3	2.0	48.3
34	71-00-1	2.6	8.1	1.9	84	144-82-1	43.8	48.7	11.0
35	71-43-2	1.4	15.3	1.0	85	154-21-2	0.8	3.0	14.9
36	72-18-4	3.6	0.1	1.5	86	1610-18-0	0.5	9.7	0.0
37	72-19-5	6.5	0.5	3.3	87	1634-04-4	4.9	0.3	0.2
38	75-65-0	0.7	1.1	7.3	88	1646-88-4	1.9	9.9	21.1
39	76-03-9	2.6	0.3	3.0	89	16655-82-6	13.2	13.2	18.1
40	76-57-3	0.9	3.8	2.1	90	1672-46-4	4.7	4.9	4.9
41	79-43-6	1.2	1.3	2.7	91	16752-77-5	8.3	5.2	1.8
42	80-05-7	3.0	3.3	1.0	92	1836-75-5	6.4	5.4	0.0
43	84-66-2	0.7	4.3	2.8	93	18559-94-9	4.9	11.4	9.9
44	85-01-8	3.8	1.6	3.4	94	1912-24-9	1.9	4.4	0.0
45	85-44-9	3.8	6.3	5.3	95	206-44-0	0.4	13.6	0.0
46	85721-33-1	5.2	1.2	0.4	96	20830-75-5	14.3	72.2	62.4
47	87-86-5	2.5	1.4	3.3	97	21087-64-9	6.0	6.9	26.9
48	94-75-7	3.0	0.5	3.4	98	2136-79-0	10.4	5.5	4.8
49	98-95-3	2.5	3.1	1.8	99	2169-87-1	1.8	19.1	16.6
50	15972-60-8	1.6	3.8	2.5	100	2212-67-1	19.2	2.6	13.3

ISBN:978-84-691-2701-8/DL:T.386-2008

Table 3.6. Mass balance relative errors for the 193 compounds, for each of the three membranes.

Index	CAS	BW30 [%]	TFCHR [%]	CA [%]	Index	CAS	BW30 [%]	TFCHR [%]	CA [%]
101	2385-85-5	5.0	8.5	4.2	148	637-92-3	9.4	10.6	1.4
102	25013-16-5	10.7	5.2	14.1	149	63-91-2	16.8	15.3	32.8
103	25812-30-0	16.7	0.4	15.8	150	64285-06-9	6.4	3.7	6.5
104	26638-19-7	1.1	36.5	7.1	151	657-24-9	49.4	52.1	18.3
105	27304-13-8	4.2	34.4	16.7	152	66357-35-5	0.9	3.7	22.8
106	298-04-4	11.9	3.2	7.7	153	67-66-3	0.6	17.9	7.2
107	3018-12-0	27.7	10.4	7.6	154	67708-83-2	34.4	15.8	1.9
108	302-17-0	24.5	7.0	5.0	155	68-22-4	0.7	1.3	5.2
109	309-00-2	0.4	3.1	11.9	156	70458-96-7	2.2	7.7	50.1
110	3252-43-5	21.1	17.2	3.8	157	719-22-2	2.7	5.4	4.1
111	330-54-1	4.1	2.5	10.3	158	72-14-0	58.6	14.7	11.2
112	330-55-2	2.7	5.0	3.8	159	72-33-3	1.9	2.0	9.6
113	333-41-5	5.8	5.2	7.3	160	723-46-6	6.2	24.3	38.7
114	3380-34-5	6.4	0.8	3.1	161	72-43-5	1.2	13.1	2.2
115	34256-82-1	3.6	8.5	3.8	162	72-54-8	3.4	5.2	0.0
116	35523-89-8	82.0	64.2	48.3	163	72-55-9	4.4	7.6	0.0
117	42399-41-7	9.8	16.1	6.7	164	738-70-5	1.5	10.9	13.1
118	474-86-2	7.1	3.1	4.9	165	74-83-9	11.8	23.5	5.2
119	486-56-6	1.7	3.0	24.6	166	74-95-3	8.8	2.3	10.4
120	50-27-1	5.6	2.7	6.1	167	74-97-5	5.9	5.3	3.0
121	50-29-3	2.6	1.9	0.0	168	75-09-2	29.1	28.0	8.3
122	50-32-8	0.8	5.6	0.0	169	75-25-2	12.5	1.3	5.8
123	5103-71-9	3.6	0.0	11.3	170	75-27-4	10.7	0.6	3.6
124	51218-45-2	2.4	11.9	5.5	171	75-71-8	17.6	62.3	11.4
125	51-28-5	9.5	4.9	8.2	172	759-94-4	22.3	4.9	17.8
126	513-88-2	9.4	43.6	7.2	173	7601-90-3	50.5	39.2	5.2
127	517-04-4	12.7	1.2	4.2	174	76420-72-9	0.0	0.7	9.6
128	517-09-9	8.2	1.8	10.3	175	76-44-8	7.5	0.4	12.3
129	53-41-8	12.8	0.3	7.7	176	79-01-6	9.6	12.5	7.2
130	54910-89-3	12.4	4.1	30.7	177	79-34-5	10.4	6.2	11.1
131	55-18-5	4.3	11.8	0.7	178	79-57-2	0.3	0.8	51.1
132	5589-96-8	4.6	3.1	5.0	179	80-32-0	4.3	11.1	0.6
133	56-45-1	3.2	4.1	11.5	180	83463-62-1	26.7	7.7	9.0
134	57-62-5	17.6	6.7	5.5	181	84-74-2	7.6	14.5	4.6
135	57-68-1	10.1	6.5	0.3	182	87-68-3	8.4	0.2	27.5
136	5902-51-2	7.9	10.0	13.5	183	924-16-3	57.7	5.9	10.6
137	59-89-2	6.0	2.2	11.1	184	930-55-2	2.2	5.3	1.2
138	60-57-1	14.9	24.1	5.5	185	93106-60-6	28.2	14.7	11.3
139	606-20-2	2.1	1.5	4.0	186	93-76-5	4.3	4.0	15.4
140	608-73-1	2.8	8.5	6.1	187	944-22-9	2.8	0.1	3.3
141	611-59-6	10.0	5.1	17.0	188	95-47-6	1.7	15.6	1.0
142	61-82-5	1.5	11.1	14.6	189	95-48-7	1.8	5.8	14.6
143	61869-08-7	12.3	4.5	2.3	190	95-50-1	4.4	4.6	0.6
144	61-90-5	12.7	17.5	3.4	191	95-63-6	2.9	12.4	2.0
145	621-64-7	16.2	29.3	0.3	192	994-05-8	9.7	11.2	1.7
146	631-64-1	17.3	32.5	4.0	193	99-87-6	5.1	4.3	5.3
147	63-25-2	2.9	4.7	18.0					

Organic compounds from 1 to 50 were used for developing the models, while chemicals from 51 to 193 were used for test.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

The mass balance errors obtained for each one of the three membranes modeled and for all 193 compounds are presented in Table 3.6. Without prescreening the new compounds with respect to the model applicability domain corresponding to the 50 compounds with experimental data, the current predictions showed that the mass balance was fulfilled in most cases. In the case of the BW30 models, the mass balance average relative error for the 50 organics with experimental data was 2.7%, with a standard deviation of 2.1%, while for the 143 new compounds the error increased to 11.4% with the standard deviation of 15.3%. For this membrane, the mass balance was closed with relative errors smaller than 5% for 105 chemicals, 39 compounds presented mass balance relative errors between 5 and 10%, 22 compounds between 10 and 15%, 10 compounds between 15 and 20%, and 3 compounds between 20 and 25%. Only 15 chemicals deviated more than 25% in the mass balance closure.

For the TFCHR membrane models, the average relative error for the 50 organics with experimental data was 4.7% with the standard deviation of 4.7%, while for the 143 new compounds the error increased to 11.4% with the standard deviation of 15.2%. In this case, the mass balance was closed with relative errors smaller than 5% for 92 chemicals, 43 compounds presented mass balance relative errors between 5 and 10%, 24 compounds between 10 and 15%, 15 compounds between 15 and 20%, 5 compounds between 20 and 25%, while only 14 chemicals presented mass balance deviations higher than 25%.

In the case of the CA membrane models, the mass balance relative error for the 50 organic compounds used for the models development was of 3.7% with standard deviation of 3.7%. The average relative error increased in the case of 143 new organic compounds to 11.0%, with the standard deviation of 12.6%. For this membrane, the mass balance was closed with relative errors smaller than 5% for 91 chemicals, 43 compounds presented mass balance relative errors between 5 and 10%, 29 compounds between 10 and 15%, 11 compounds between 15 and 20%, 4 compounds between 20 and 25%, while only 15 chemicals presented mass balance deviations higher than 25%.

The examination of the chemicals presenting mass balance relative error higher than 25% at least for one membrane leads to a list including 33 compounds. A SOM analysis revealed that these compounds are not well represented by the set of 50 chemicals used to train the models. When the 143 new chemicals were presented to an 8x7 SOM generated with the 50 organic compounds characterized by the seven descriptors listed in Table 3.5, 11 of these 33

ISBN:978-84-691-2701-8/DL:T.386-2008

chemicals were allocated to empty and isolated map units. Another 19 of them, even though allocated to occupied units, presented large classification errors, denoting low similarities with the chemicals used to train the map. The remaining 3 compounds among the 33 with the largest mass balance errors, failed only in the mass balance test for the BW30 membrane, presenting however relative deviations close to 25%.

The current mass balance results show that simple non-linear algorithms, such as back-propagation neural networks, can quantitatively assess the rejection of organic compounds in RO membranes if appropriate chemical information is considered. Even though the mass balance test is not a sufficient condition for validating the QSPR models developed, definitely it is a necessary one. Moreover, it is the only validation method that can be applied in the presented approach. According to this test, for each one of the three membranes the majority of the compounds presented a relative error lower than 25% for the mass balance based on the predicted mass fractions. In should be noted that these results were obtained based on all 193 organic molecules, without a priori screening the 143 new compounds with respect to the applicability domain for which the QSPR models were developed. Therefore, the mass balance test results provide a reasonable indication of the applicability of the approach presented in this work.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

ISBN:978-84-691-2701-8/DL:T.386-2008

4. Reverse osmosis plant performance

4.1. Full-scale RO data

As presented in Section 2.1, ANN-based models can effectively describe membrane process performance with respect to the dynamics of both flux and separations performance. ANN models developed in previous studies were based on training the model with a certain fraction of experimental data inter-dispersed through the complete dataset and including extreme values. Therefore, the resulting models were successful for data interpolation (i.e., predictions for an input variable range for which the ANN model was trained) but without the capability of forecasting (i.e., future time predictions of performance for time periods that were not covered by the training data set). The ability to forecast membrane plant performance, even for short future steps, would provide additional flexibility for integrated process control strategy and could be used to signal the need for remedial action (e.g., membrane cleaning, adjustment of process variables such as pressure and flow rates). Although ANN approach is data-driven and therefore results in plant-specific application, such an approach has the advantage of capturing the unique aspects of the plant that include operational behavior of plant equipment (e.g., pumps, valves, monitoring devices and control system), process elements (i.e., membrane modules, feed pretreatment modules), plant configuration, as well as feed quality variations. Taking into account the aforementioned advantages, a NN model of RO plant performance was developed, capable of describing temporal variations in permeate flow and salt passage, as well as for short-term forecasting. In the present approach the use of process information backward in time is utilized along with feed process variables to forecast plant performance, suggesting the possibility of using such an approach for RO process control and process fault identification.

The experimental data used for building the ANN models was provided by the WaterEye Corporation [131]. The 1 MGD (million gallons per day) RO brackish water desalination plant represented schematically in Figure 4.1 is located at Port Hueneme, California,

ISBN:978-84-691-2701-8/DL:T.386-2008

operated at 75% recovery. The first and second stages contain 9 and 5 membrane modules, respectively. The monitored plant parameters for the feed stream included flow rate, conductivity, feed pressure, pH and temperature. Permeate and concentrate (i.e., brine) monitored parameters included flow rate, conductivity and pressure. The inter-stage pressure was also monitored. Real time data of the above process parameters were collected every 10 minutes for a period of about 3 months.

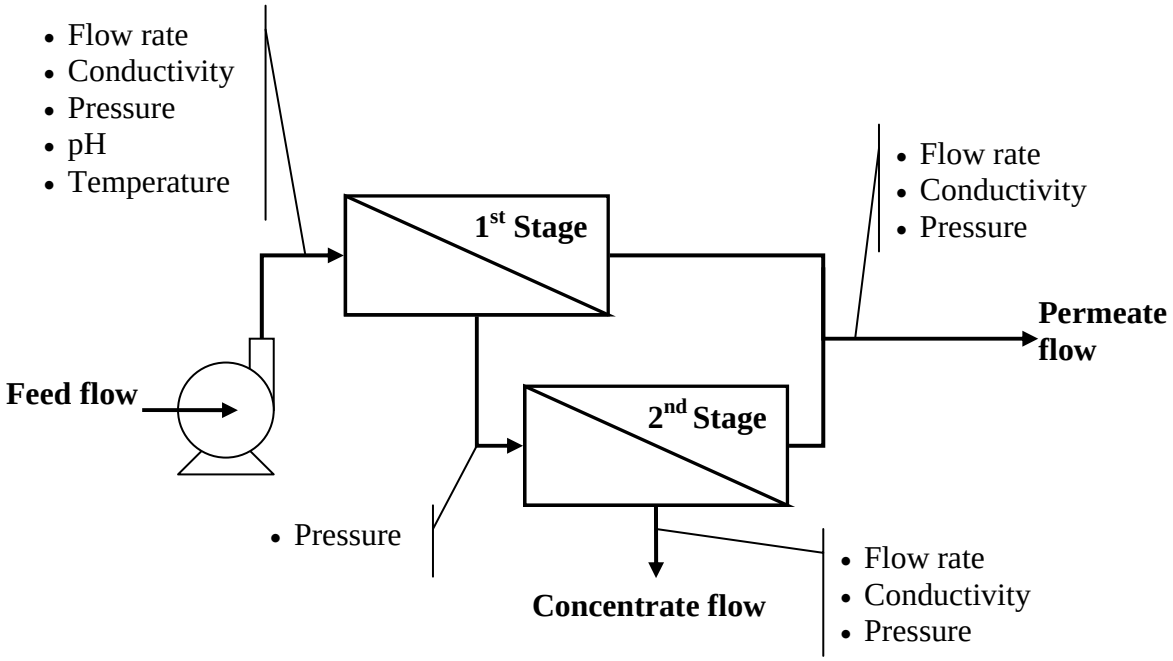


Figure 4.1. Diagram flow of the two-stage RO plant from Port Hueneme, California, with identification of monitored process parameters.

The composition of the feed, permeate and concentrate streams are not measured in real-time in commercial RO plants, but instead, conductivity measurements are reported as surrogates for salt concentration. The common approach is to correlate conductivity with the total dissolved solids concentration (TDS) using sodium chloride as the correlating salt. Such correlations can be obtained either from experimental measurements for the actual range of salt compositions of interest or based on thermodynamic multi-electrolyte calculations. In the present analysis the concentrations for both feed-brine and permeate flows were expressed in terms of [mg/l] of total dissolved solids based on conductivity [μS/cm] – TDS [mg/l] correlations derived from multi-electrolyte calculations using the OLI Analyzer software [132]. The correlations were developed based on ionic composition of the feed presented in Table 4.1, calculating the conductivity that would result from various levels of concentrations of this feed water and correspondingly the production of permeate for various levels of salt

ISBN:978-84-691-2701-8/DL:T.386-2008

passage. Based on the resulting conductivity, the equivalent NaCl TDS [mg/l] was calculated and this was utilized to arrive at the following correlations, presented also in Figure 4.2:

Permeate: $TDS_{NaCl} = 0.3455 \cdot (Cn)^{1.0169}$ (4.1)

Feed-brine: $TDS_{NaCl} = 0.1409 \cdot (Cn)^{1.1567}$ (4.2)

where Cn is the conductivity [μS/cm]. Both correlations expressed in Eqs. (4.1) and (4.2) presented coefficients of determination (R^2) higher than 0.999, and average relative errors of $8.11 \cdot 10^{-3}$%, $1.69 \cdot 10^{-2}$%, respectively.

Table 4.1. Average feed composition.

Ion/Specie	mg/l	Ion/Specie	mg/l	Ion/Specie	mg/l	Ion/Specie	mg/l
SiO_2	28	Ca^{2+}	142.6	HCO_3^-	261	NO_3^-	1.46
CO_2	8.7	Mg^{2+}	39.6	Cl^-	46.2	CO_3^{2-}	0.83
Na^+	97.4	Fe^{3+}	0.1	F^-	0.4	TDS	1071.4
K^+	4.6	Ba^{2+}	0.03	SO_4^{2-}	445	pH	7.68

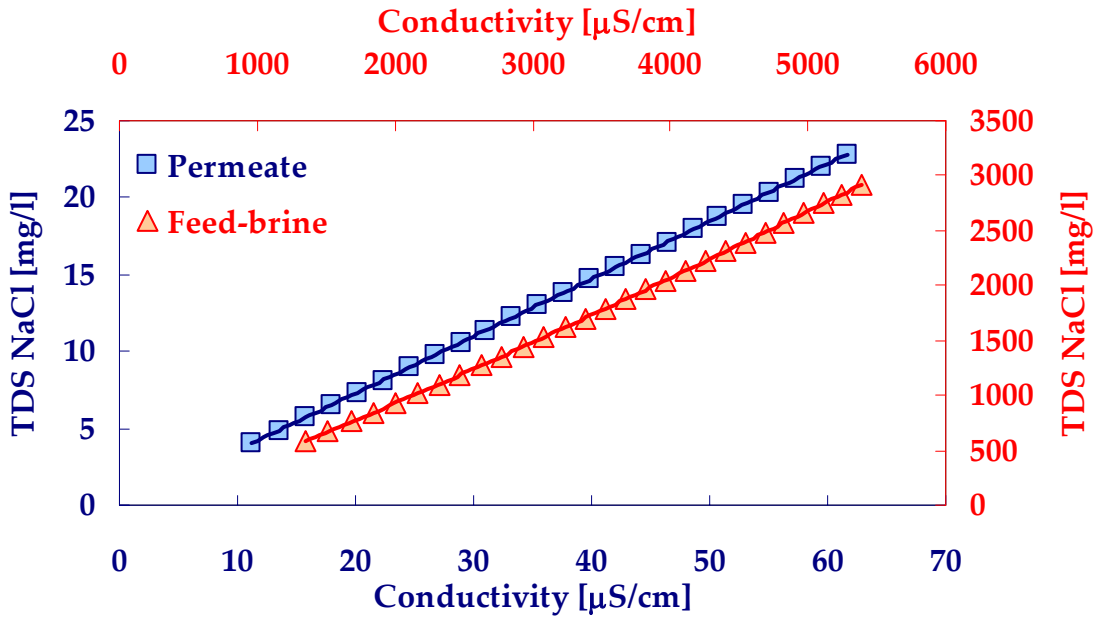


Figure 4.2. Conductivity-TDS correlations for permeate and feed-brine.

Feed quality and operational parameters can vary during plant operation, as presented in Figure 4.3 for the considered operational period, and this typically results in variations in permeate flow rate and salt passage. The occurrence of undesired phenomena like membrane fouling or scaling affects also the permeate flow rate and salt passage time evolution. Therefore, in order to effectively evaluate the plant performance, it is necessary to compare permeate flow and salt passage rates at a standard reference condition. The standardization method presented in ASTM 4516-00 [133] was used to normalize the

ISBN:978-84-691-2701-8/DL:T.386-2008

permeate flow rate and the salt passage with respect to temperature, pressure and osmotic pressure. The reference (standard) condition was set corresponding to the process parameter measurements for the first monitoring point.

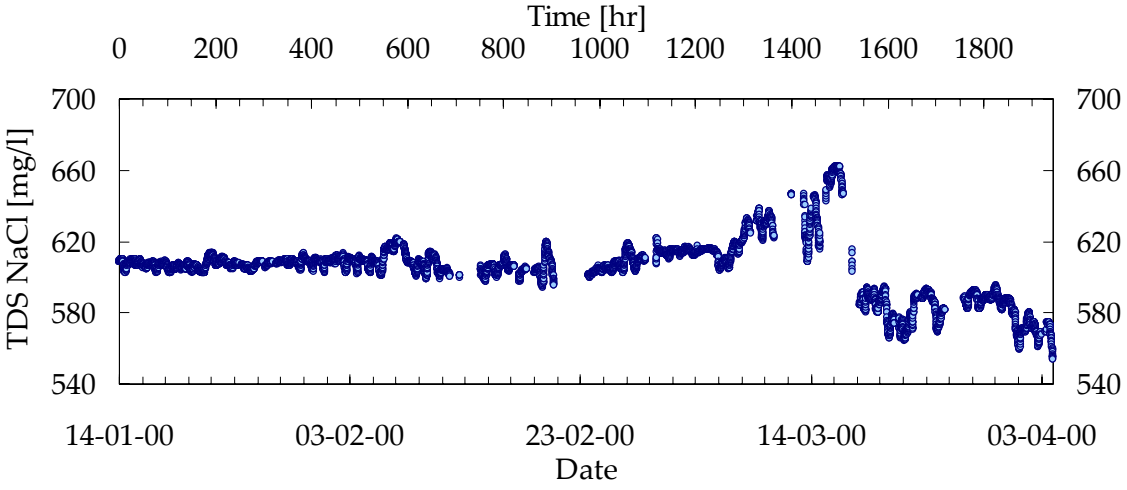


Figure 4.3. Variability of feed water TDS during the RO plant evaluation period.

According to the above ASTM procedure, the standardized permeate flow was calculated from

$$Q_{p,s} = \frac{\left[P_{f,s} - \frac{P_{f,s} - P_{c,s}}{2} - P_{p,s} - \pi_{b,s} + \pi_{p,s}\right] \cdot \left[TCF_s\right]}{\left[P_{f,a} - \frac{P_{f,a} - P_{c,a}}{2} - P_{p,a} - \pi_{b,a} + \pi_{p,a}\right] \cdot \left[TCF_a\right]} \cdot Q_{p,a} \tag{4.3}$$

where Q_p is the permeate flow rate [GPM], P_f, P_c and P_p are the feed, concentrate and permeate pressures [kPa], π_b and π_p are the brine and permeate osmotic pressures [kPa] estimated by Eq. (4.4), and *TCF* is the temperature correction factor calculated using Eq. (4.5) [131]. The subscripts *a* and *s* refer to the actual and the standard conditions, respectively.

$$\pi_b = 0.2654 \cdot C_b \cdot \frac{T}{1000 - \frac{C_b}{1000}};\ \pi_p = 0.05 \cdot \pi_b \tag{4.4}$$

$$TCF = \exp\left[3020 \cdot (1/298.15 - 1/T)\right] \tag{4.5}$$

In Eqs. (4.4) and (4.5), T is the absolute temperature [K], and C_b is the brine concentration [mg/l] expressed as a log-mean average, and calculated in terms of the recovery Y (i.e., permeate to feed flow rates ratio) and feed concentration C_f [mg/l], according to

$$C_b = C_f \cdot \ln\left[(1/(1-Y))/Y\right] \tag{4.6}$$

ISBN:978-84-691-2701-8/DL:T.386-2008

The standardized percent salt passage (*%SP*) for the RO process is calculated as

$$\%SP_s = \frac{\left[EPF_a\right]}{\left[EPF_s\right]} \cdot \frac{\left[TCF_a\right]}{\left[TCF_s\right]} \cdot \frac{\left[C_{b,s}\right]}{\left[C_{b,a}\right]} \cdot \frac{\left[C_{f,a}\right]}{\left[C_{f,s}\right]} \cdot \%SP_a \tag{4.7}$$

where *EPF* is the average RO element permeate flow rate [GPM].

The time evolutions of the standardized permeate flow and salt passage for the considered operational period, calculated according to the ASTM 4516-00 specification expressed in Eqs. (4.3)-(4.7), are represented in Figure 4.4. In addition, the evolutions of the pressure drop along the membrane channel for both stages, as well as for the overall process are represented in Figure 4.5.

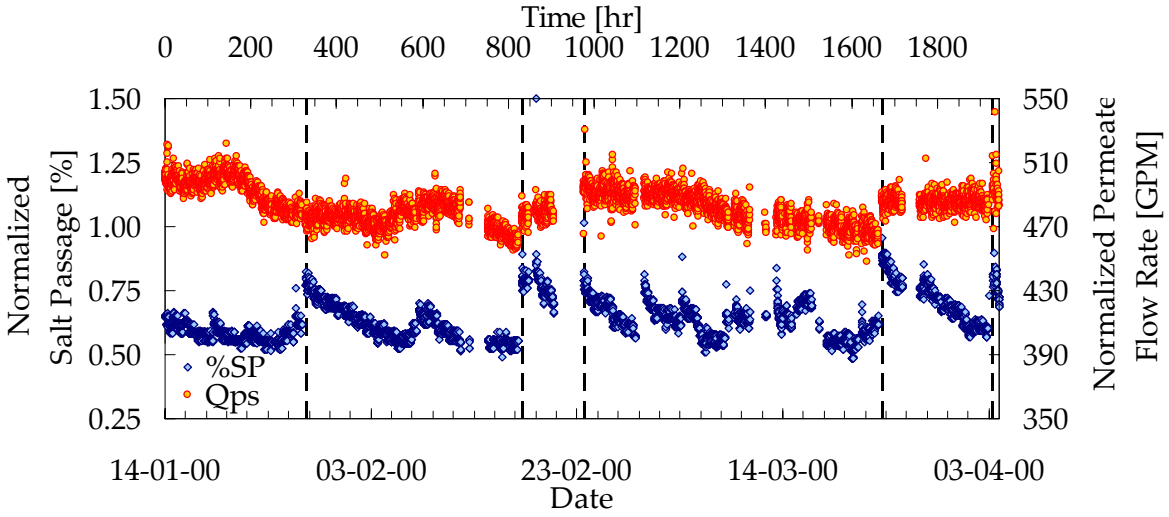


Figure 4.4. Time evolution of the normalized salt passage and normalized permeate flow. The vertical dotted lines represent startup moments after process interruptions for membrane cleaning and/or replacement.

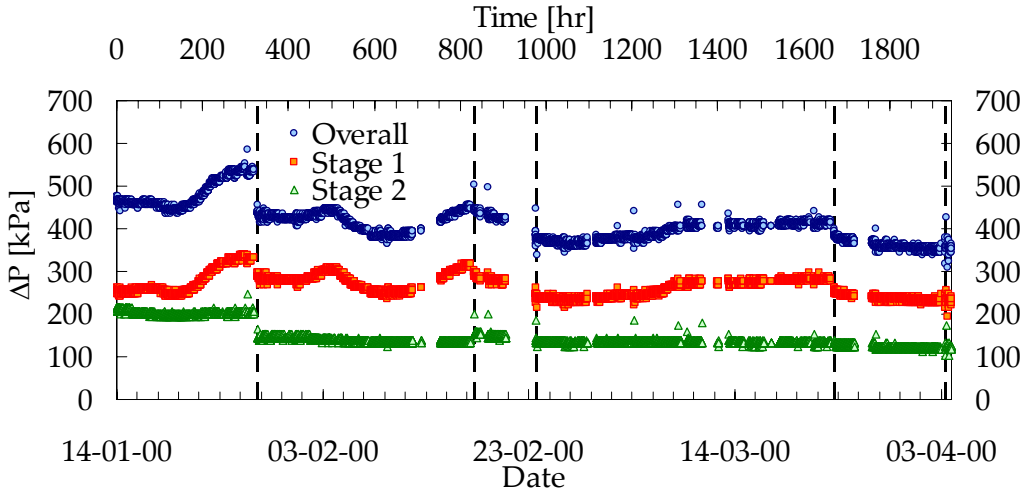


Figure 4.5. Time evolution of the pressure drop along the membrane channel, for each one of the two stages and for the overall process. The vertical dotted lines represent startup moments after process interruptions for membrane cleaning and/or replacement.

Discontinuities in the represented variables, especially when accompanied by large periods of missing data, since no measurements are performed during the cleaning procedure, are clear information of operational problems and process interruptions for membrane cleaning and/or replacement. The startup times after these interruptions are indicated by the dotted vertical lines (at t = 328, 832, 976, 1671, and 1928 hrs) as shown in Figures 4.4 and 4.5 for the percent salt passage, normalized permeate flow and pressure drop, respectively; the plant was shut down for various periods just prior to these times. There can also be noted other data gaps in the time evolution of represented variables. However, the missing of data cannot be used as unique criteria for the identification of operational problems, since it can be attributed to process shutdown or problems in the data acquisition system.

4.2. Back-propagation approach for modeling plant performance

Data preprocessing and analysis

The RO process performance was modeled by developing ANNs capable of describing temporal variations in normalized permeate flow and normalized salt passage. Accordingly, as illustrated in Figure 4.6, the output of the models was expressed as the normalized permeate flow, or the normalized salt passage, related to the first standardized value of each operating period identified in Figures 4.4 and 4.5. As input variables, process parameters physically meaningful, independent, easy and inexpensive to measure and capable of being monitored in real-time were selected. Therefore, the flow rate, conductivity, pressure, pH and temperature of the feed were chosen as inputs. Since the impact of membrane fouling occurrence on membrane performance is a cumulative effect over time, with flux decline becoming more severe with fouling progression, a time measure is needed as additional input parameter.

However, the time scales associated with plant readjustment to changes in operational conditions (e.g., pressure, flow rate) can be much shorter than the fouling time scale. These shorter time scales, termed here as “short term memory” (STM), were set by dividing the process operational period into equal time intervals, with the length of time varying from 7 to 125 hrs. To capture the STM, a time variable ranging for each time interval from t = 0 to the length of time considered was added as input.

ISBN:978-84-691-2701-8/DL:T.386-2008

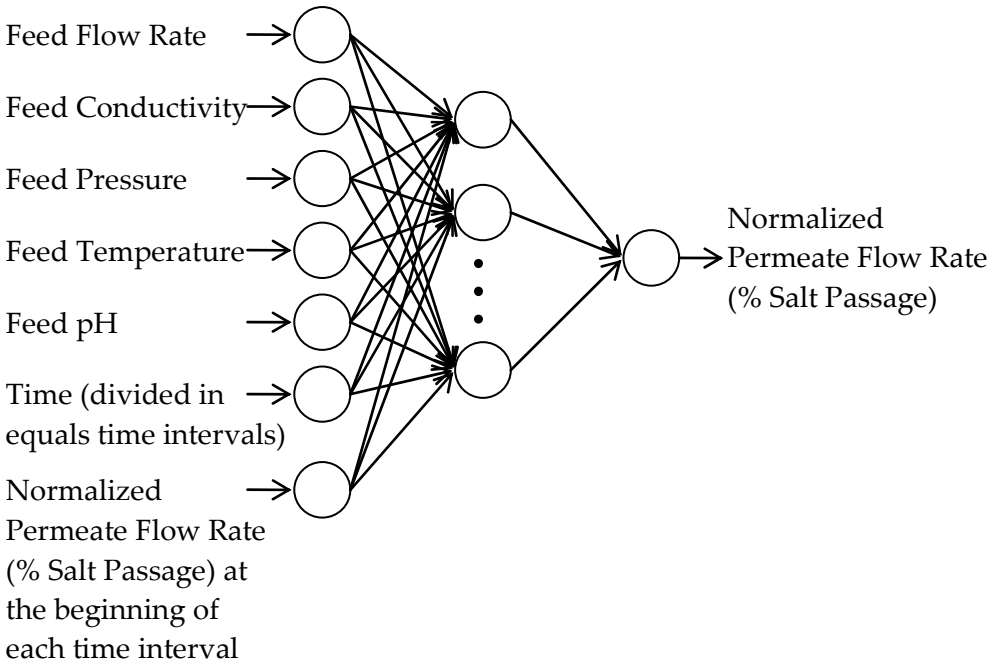


Figure 4.6. Identification of input and output variables used for modeling the RO plant performance.

In order to account for the large time-scale events, a "long term memory" that makes the link between two consecutive STM time intervals is needed. This was accomplished by considering the normalized permeate flow (or normalized salt passage, respectively) at the beginning of each time interval as additional input. Due to the fact that the monitored variables are measured every 10 minutes, dividing the operational time period in 7 hrs equal intervals, a maximum of 42 experimental measurements can correspond to each time interval. The use of a smaller number of points in each time interval would practically merge the two memory terms; therefore, time lengths smaller than 7 hrs were not considered.

ANN models for permeate flow and salt passage were built using the back-propagation algorithm, based on an architecture with one input layer, one hidden layer and one output layer, as presented in Figure 4.6. The linear transfer function was utilized for the input and output layers and a hyperbolic tangent transfer function was used for the hidden layer (see Figure 2.3) [71]. For model development, the input and output parameters were normalized in the range [0,1] using Eq. (2.1). A statistical analysis was performed for the selection of the optimal neural network architecture for each length of time considered for dividing the time space in equal intervals, by varying the number of hidden neurons from 2 to 11. The ANN models were trained using data from the first three periods of operation (0-906 hrs) which contained a total of 4569 data points. Subsequent to training, model validation was carried out with 20% of the training data selected randomly. The validation step served as a

stopping criterion for the ANN learning algorithm. Model testing was performed using the entire data set from the three operational periods starting at t = 976 hrs and forward in time (Figures 4.4 and 4.5). It is emphasized that the data from the last three operational periods were not used in the learning phase, neither for model training, nor for internal validation; hence, model testing was accomplished with an external data set. Overall, approximately 40% of the entire plant data set was used for ANN model training, 10% of the data were used for validation and 50% of the data (last three operational periods) were used for external model testing. The adequacy of model ANN architecture and length of time intervals was assessed using the model explained variance in prediction index, q^2, calculated separately for the training, validation, and test data sets [129], defined in (3.4). In addition, model performance was also evaluated based on the relative error (average and maximum), standard deviation, and coefficient of determination for the model fit (ANNEX VI).

Results

Various ANN architectures, with 2-11 hidden neurons, were evaluated ranging the STM time interval length form 7 to 125 hrs, for both normalized permeate flow and normalized salt passage models. The selection of the best architecture for each length of time considered was made based on the explained variance in prediction index calculated for the test data set, as presented in Figures 4.7 and 4.9 for the normalized permeate flow rate and normalized salt passage, respectively.

The accuracy of capturing the dynamic changes in normalized permeate flow generally increased as the STM time interval length decreased (Figure 4.7). For each one of the NN architectures considered, the highest q^2 index for the test data set was obtained when the time space was divided into 7 hrs of equal intervals. Also, the lowest q^2 index was generally obtained when using the 125 hrs time intervals, except the 7:7:1, 7:8:1 and 7:10:1 architectures when the lowest performance was obtained for 100 hrs time intervals, and the 7:2:1 architecture when the lowest performance was obtained for 50 hrs time intervals. In order to find the optimal NN architecture, the results were analyzed for each particular STM time interval length. When using time intervals of 15, 25 or 100 hrs, the q^2 index increased from 2 to 5 neurons in the hidden layer, and subsequently decreased slowly with increasing the

ISBN:978-84-691-2701-8/DL:T.386-2008

number of hidden neurons. In these cases, a bimodal graph was obtained with the highest values of q^2 obtained for 5 and 9 hidden neurons, respectively. For time intervals of 7, 50 or 75 hrs, the models performance followed a similar behavior of increasing q^2 index as the number of hidden neurons increased from 2 to 5 (or 6 in the case of 50 hrs time intervals), remaining nearly constant for higher values of neurons in the hidden layer. In the case of dividing the time space into equal intervals of 125 hrs, the maximum q^2 indices were obtained for 7:2:1 and 7:7:1 architectures. The ANN model that yielded the highest q^2 index for predicting the normalized permeate flow was obtained when using 5 neurons in the hidden layer, and dividing the time space into equal intervals of 7 hrs.

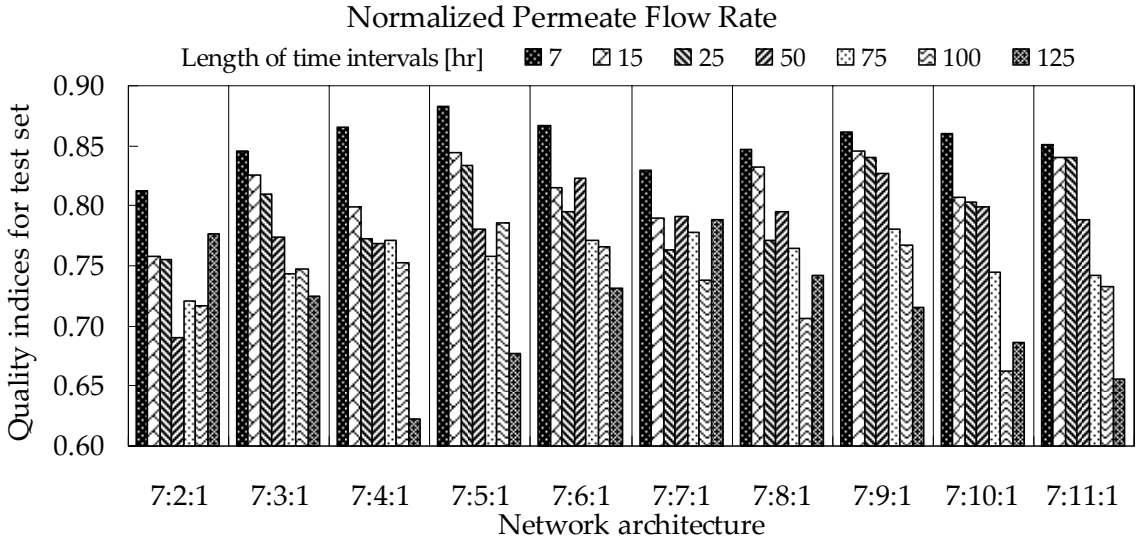


Figure 4.7. Statistical analysis for determination of the optimal network structure and optimal length of time intervals based on q^2 quality indices, for modeling the normalized permeate flow rate.

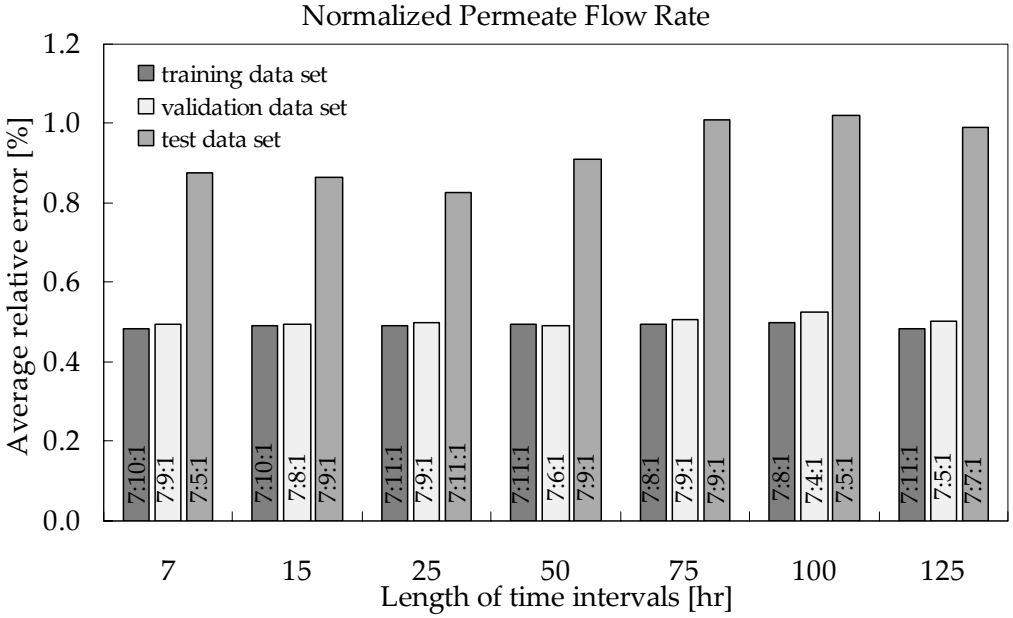


Figure 4.8. Average relative error for the best architecture (chosen based on q^2 index) for each length of time, for the normalized permeate flow rate models.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

The average relative errors for the best performing architecture for each length of time considered are presented in Figure 4.8. For the training and validation data sets the average relative errors were similar for all lengths of time intervals considered. In the case of the test data set, the average relative error decreased with increasing the length of time from 7 to 25 hrs, increasing thereafter with the increase of time interval length from 25 to 100 hrs. Accordingly, the lowest average relative error was obtained when using 25 hrs time intervals and 11 neurons in the hidden layer. However, as can be seen from Figure 4.8, the average relative errors of the best performing architecture for each one of the time interval lengths of 7, 15 and 25 hrs, presented very similar values around 0.85%. It should be noticed also that the complexity of the best architectures for these three cases increased as increasing the length of the time interval, presenting 5, 9 and 11 neurons in the hidden layer, respectively. In order to avoid the increase in models complexity, generally is preferable to use a smaller number of hidden neurons as long as the loss in model performance is small and acceptable. Based on the above analysis, the optimal NN architecture-length of time intervals combination for predicting the normalized permeate flow was selected to be 7:5:1 and 7 hrs.

Figure 4.9 summarizes the models ability to capture the changes in normalized salt passage during the operation of the RO process. High performance models were obtained when dividing the time space in equal intervals of 7-25 hrs. In these cases, the explained variance in prediction calculated for the test set were similar ($q^2 \approx 0.90$), irrespective of the length of time intervals and NN architecture used. Big differences were observed between the model performances obtained in the previously mentioned cases and the ones when higher time lengths were used for dividing the time space. When larger time intervals were used for dividing the time space, lower values were obtained for the test set explained variance in prediction index. The lowest value of $q^2 \approx 0.46$ was obtained when using 100 hrs time intervals.

The average relative errors for the best performing architecture for each length of time considered are in agreement with the former analysis based on explained variance in prediction index, as illustrated in Figure 4.10. A smooth increase of the average relative errors of the training and validation data sets was observed as the length of the time intervals increased from 7 to 75 hrs. The training and validation data sets average relative errors decreased slightly when the time space was divided into 100 hrs intervals, increasing

ISBN:978-84-691-2701-8/DL:T.386-2008

again when the length of the time intervals was 125 hr. Meanwhile, the test data set average relative errors were comparable when the length of the time intervals was 7, 15 or 25 hrs (2.4%, 2.5% and 2.8%, respectively), increasing drastically for larger lengths of time intervals, arriving to a maximum of 6.5% when the time length used was 100 hrs.

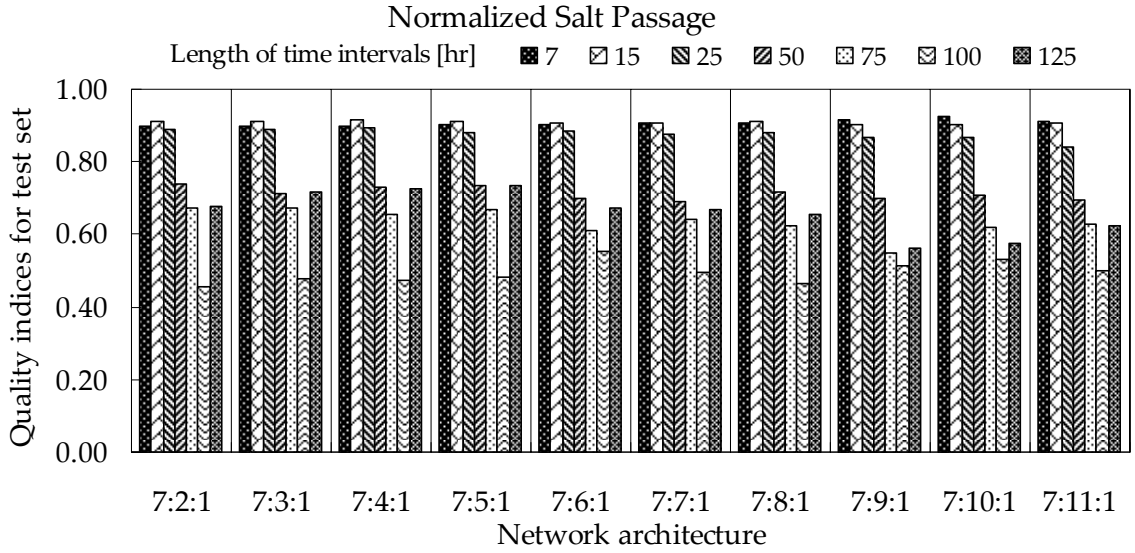


Figure 4.9. Statistical analysis for determination of the optimal network structure and optimal length of time interval based on q^2 quality indices, for modeling the normalized salt passage.

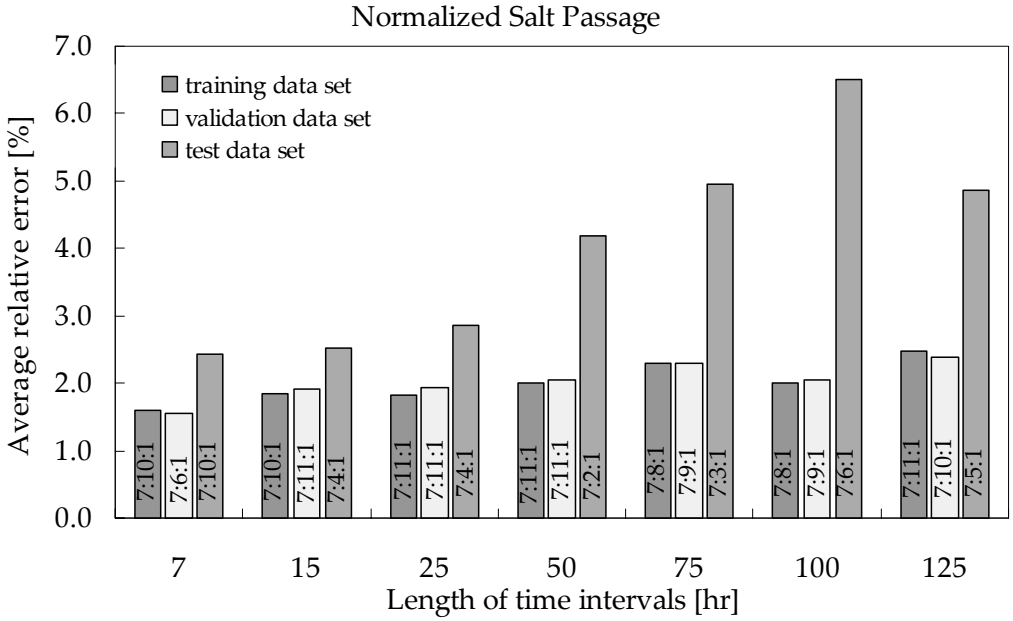


Figure 4.10. Average relative error for the best architecture (chosen based on q^2 index) for each length of time, for the normalized salt passage.

Therefore, without affecting significantly the model performance, the optimal characteristics (i.e., NN architecture and length of time intervals) for the normalized salt passage model were selected to be identical with the ones chosen for the normalized permeate flow model: 7:5:1 architecture, dividing the tine space into 7 hrs equal intervals.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

The selected optimal NN architecture and length of the time interval were used for modeling the RO process plant performance in terms of normalized permeate flow rate (Figure 4.11) and normalized salt passage (Figure 4.12). A wider scatter of model predictions for the test set relative to the test set is observed for both predicted parameters. This observation is in agreement with the higher average relative errors and maximum relative errors computed for the test data set compared with the ones calculated for the training data set. Accordingly, for the normalized permeate flow rate, the average relative errors were 0.50% and 0.88%, with maximum relative errors of 3.18% and 5.92% for the training and test data, respectively. When modeling the normalized salt passage, the average relative errors for the training and test data sets were 1.61% and 2.73% with the maximum relative errors of 15.83% and 35.52%, respectively.

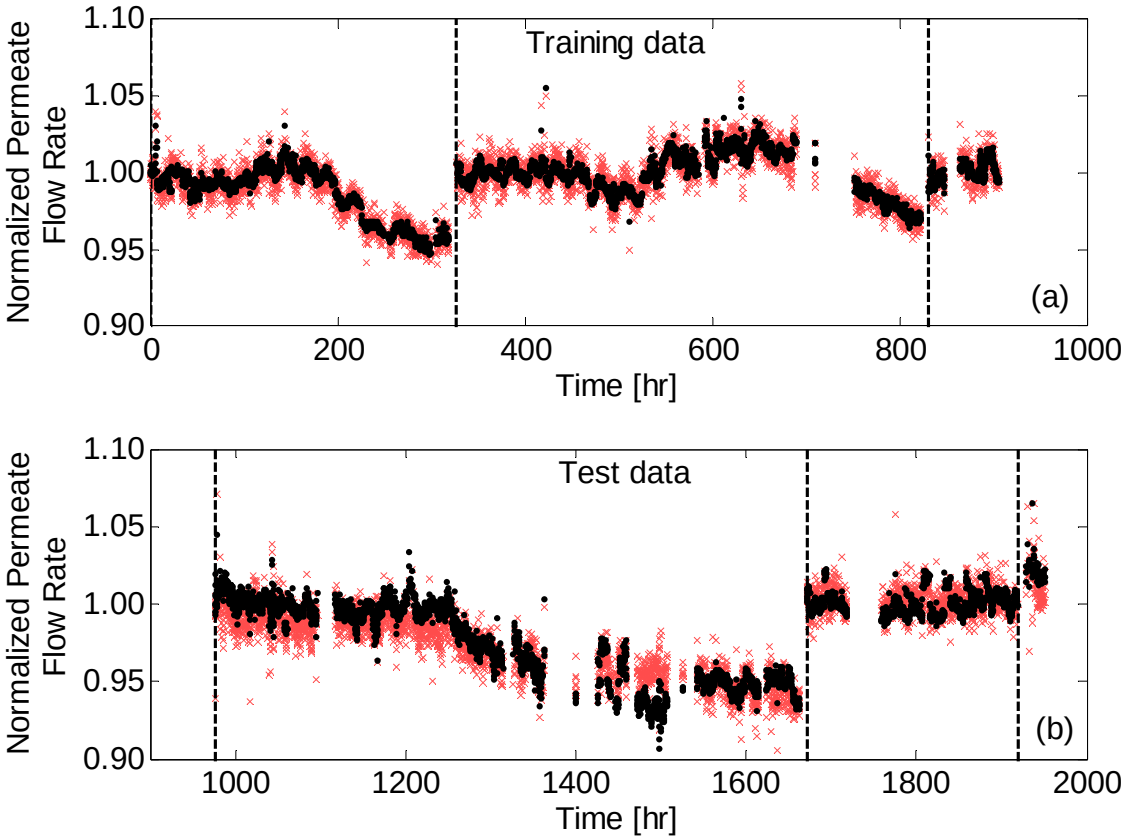


Figure 4.11. Normalized permeate flow rate training (a) and test (b) data predictions for the model built using 7 hrs time intervals and the neural network architecture 7:5:1. ×, experimental data; •, NN predictions. The vertical dotted lines represent startup moments after process interruptions for membrane cleaning and/or replacement.

However, it is apparent that the ANN models captured the plant performance evolution not only for the period of time corresponding to the experimental data used in the models training phase, but also for the operational period for which the model was tested. It is

ISBN:978-84-691-2701-8/DL:T.386-2008

noted, however, that there are prediction outliers for both modeled parameters (Figures 4.11 and 4.12); this behavior is attributed to the scatter of the monitored variables during the considered operation period. The good agreement between the predicted and experimental values for both normalized permeate flow and salt passage demonstrate that the NN-based RO models can be successfully used for interpolation, as well as for reasonable forecasting of the plant performance evolution.

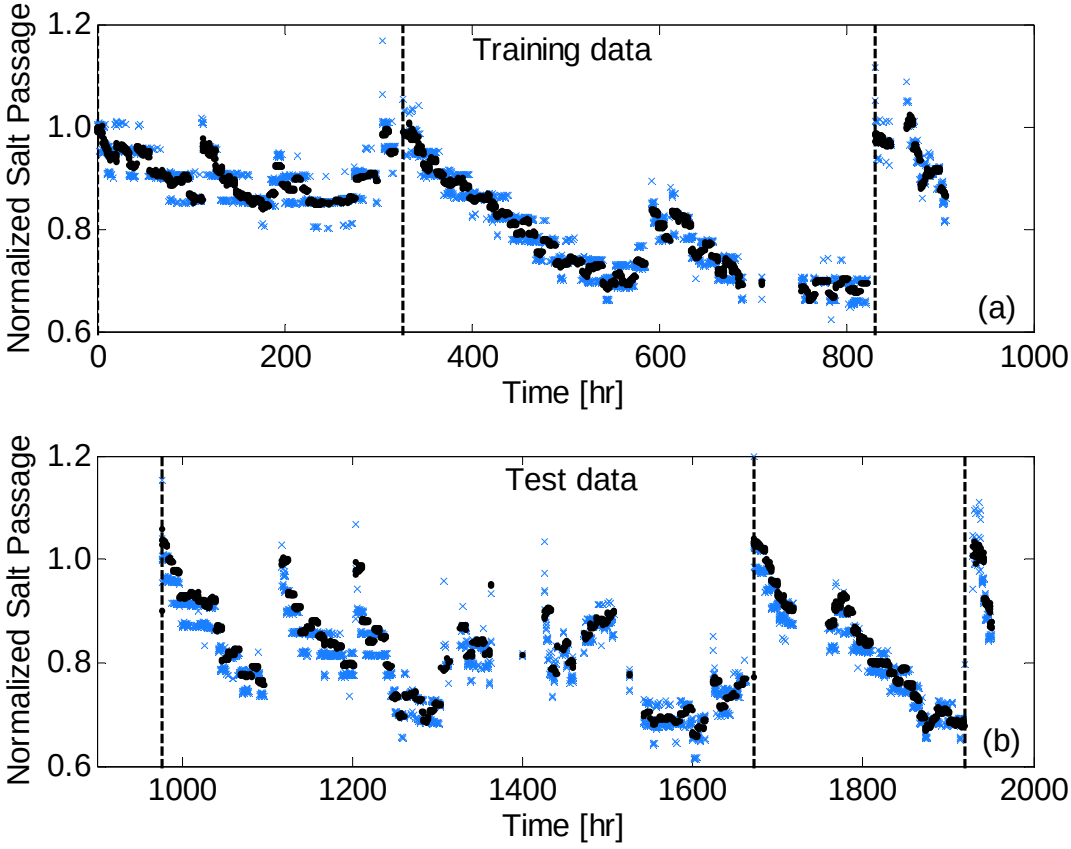


Figure 4.12. Normalized salt passage training (a) and test (b) data predictions for the model built using 7 hrs time intervals and the neural network architecture 7:5:1. ×, experimental data; •, NN predictions. The vertical dotted lines represent startup moments after process interruptions for membrane cleaning and/or replacement.

The normalized permeate flow (or normalized salt passage, depending on the parameter selected for modeling) at the beginning of each STM time interval is an important input variable to the present ANN RO plant process methodology. It assures the necessary information from the past, to make present time predictions. In order to illustrate the effectiveness of the approach, the operational time period 1750–1950 hrs was selected. During this period the normalized permeate flow rate (Figure 4.13) was stable, and therefore it is expected that an actual state model (i.e., correlate the present time plant performance with the present time process variables) would be sufficient to predict plant performance.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Figure 4.13 presents a comparison between the normalized permeate flow predicted by four different NN models, using different length of the time intervals for dividing the time space. It should be noted that the model developed using time intervals of 0 hrs (Figure 4.13a) is equivalent to predict based only on process variables (i.e., flow rate, conductivity, pressure, pH and temperature of the feed) without using any past information. Therefore, the network architecture used for the actual state model was 5:5:1. Similar to the other three models presented in Figure 4.13b,c,d, the actual state model was trained using data from the first three periods of operation (0-906 hrs) and tested with the remaining data (operational period > 976 hrs). The results presented in Figure 4.13 reveal that the actual state model (Figure 4.13a) is not able to predict correctly the normalized permeate flow for the considered operational time period (1750-1950 hrs), whereas when past information is taken into account for developing the model, the predictions get closer to the experimental values.

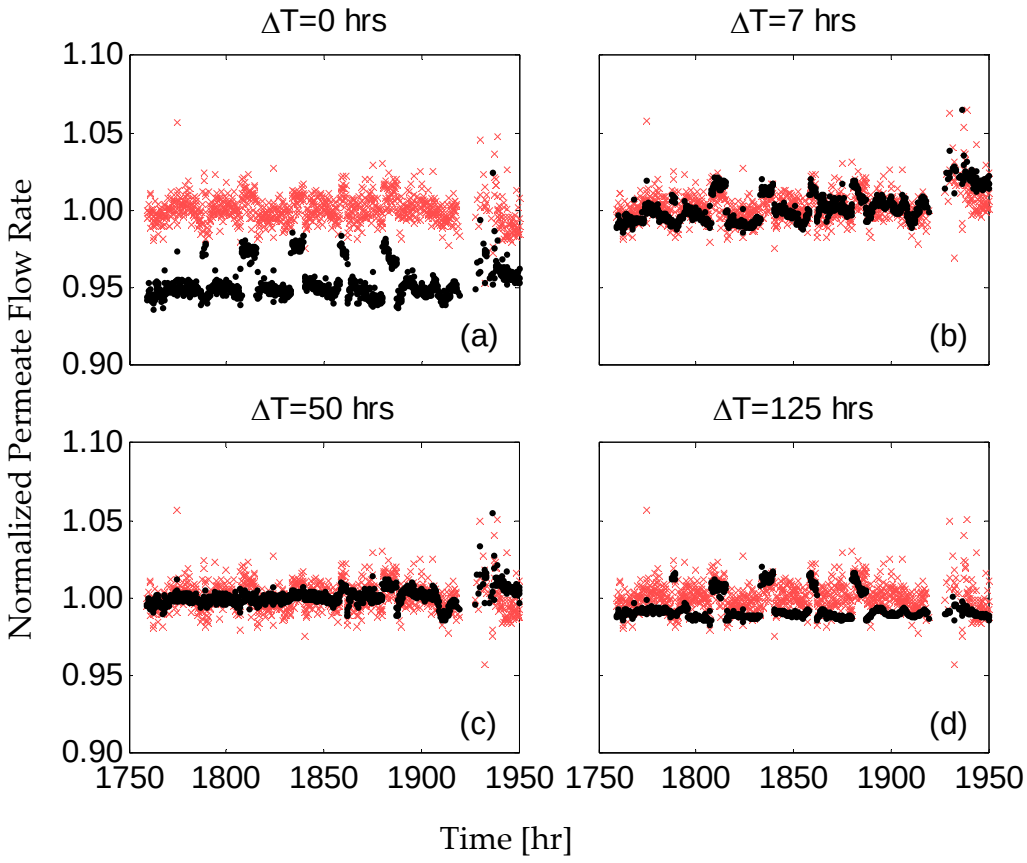


Figure 4.13. Comparison of normalized permeate flow predictions for different length of time intervals, for the operational period 1750-1950 hrs. (a) $\Delta t = 0$ hrs (actual state model); (b) $\Delta t = 7$ hrs; (c) $\Delta t = 50$ hrs; (d) $\Delta t = 125$ hrs. ×, experimental data; •, NN predictions.

The inability of the actual state ANN model to capture the evolution of the normalized permeate flow, and thus the need for past time information, can be best understood by

ISBN:978-84-691-2701-8/DL:T.386-2008

identifying major RO plant operational data clusters by means of Self-Organizing Map [77,78]. The operational patterns, represented by complete operational data set (five input variables and the normalized permeate flow) were classified using a SOM followed by a Davies-Bouldin index procedure [81] resulting in 12 separate clusters (Figure 4.14).

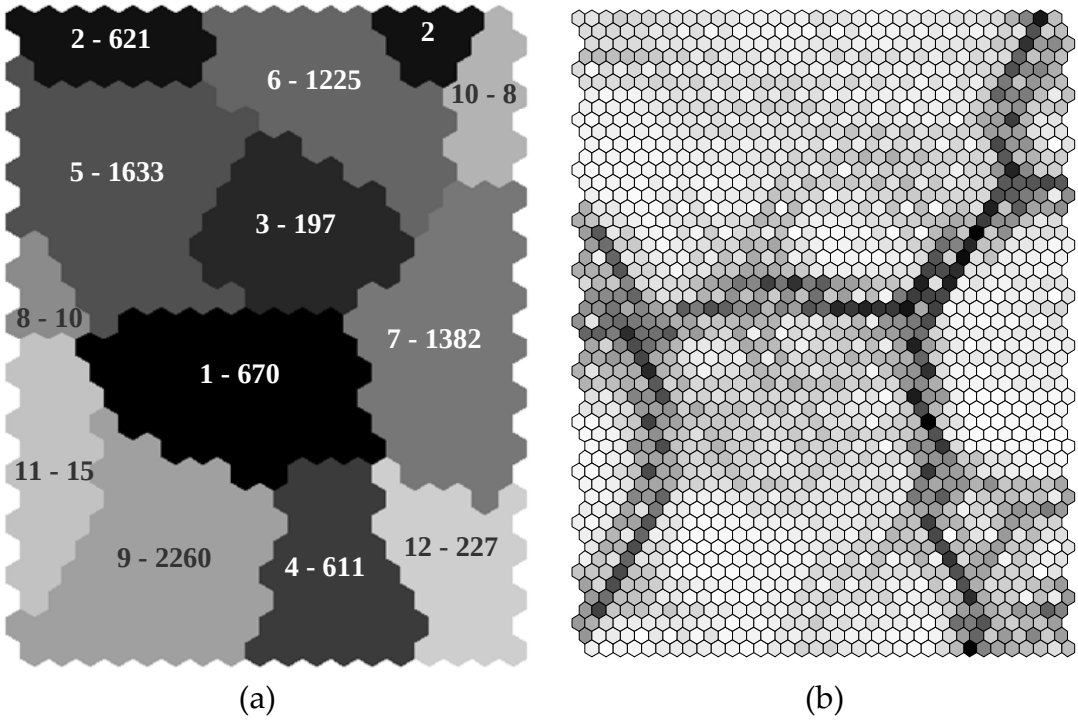


Figure 4.14. SOM clustering of normalized permeate flow operational patterns. (a) 12 separate clusters with the corresponding number of patterns allocated to each one; (b) Unified distance matrix.

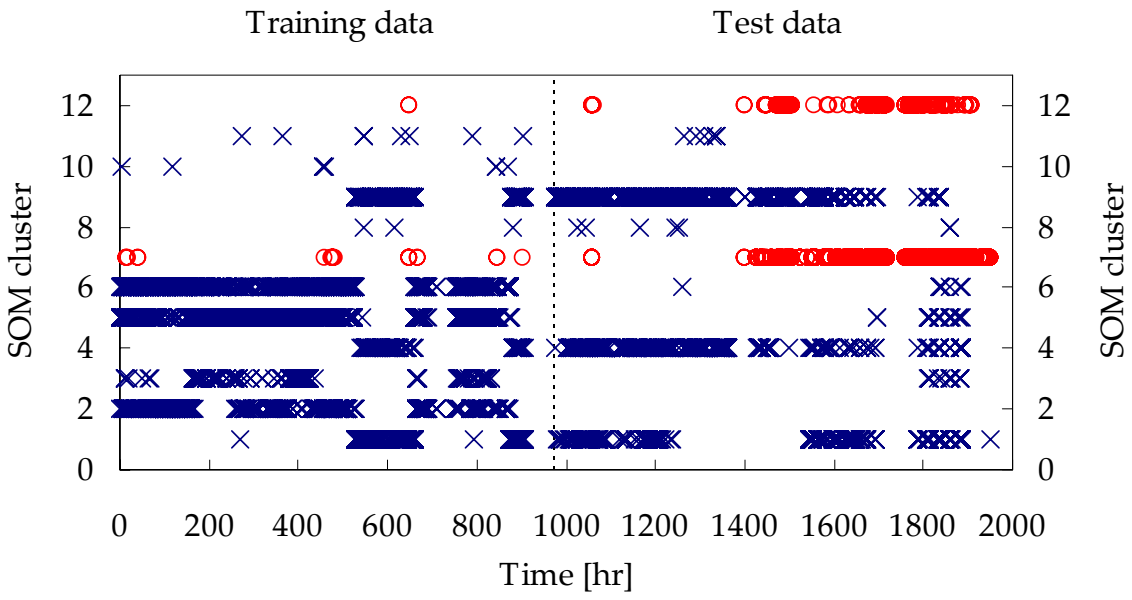


Figure 4.15. Representation of normalized permeate flow operational patterns in 12 clusters. The vertical dotted line represents the separation between the training data set and test data set.

ISBN:978-84-691-2701-8/DL:T.386-2008

The SOM classification together with the training-test distribution of experimental data is shown in Figure 4.15. The open circles represent operational patterns that were classified in clusters 7 and 12, while the crosses represent the other operational patterns. It is noted that for the period of time represented in Figure 4.13 (operational time higher than 1750 hrs), the patterns are mainly classified in clusters 7 and 12. Also, it can be observed that these clusters are scarcely represented in the training set, which is mainly formed by patterns belonging to clusters 2, 5 and 6. As illustrated in Figure 4.14, the later three clusters are located far from clusters 7 and 12 (in Figure 4.14b the light color represents map cells close to each other, while the dark color represents cells far from each other, suggesting also a separating border between different clusters).

Similar analysis was developed for the normalized salt passage models, with the comparison of four different model predictions presented in Figure 4.16. The same operational period of time (i.e., 1750-1950 hrs) like for the normalized permeate flow models was considered. In this case it was also noted that the actual state model (Figure 4.16a) did not predict correctly the plant performance for the considered period, while prediction improvement was observed when the past information is considered (Figure 4.16b,c,d). The SOM analysis was reprocessed this time for classifying the normalized salt passage patterns, followed by a Davies-Bouldin index procedure [81] to identify 11 distinct clusters (Figure 4.17). Figure 4.18 reveal that the operational patterns corresponding to the period considered in Figure 4.16 (1750-1950 hrs) are mainly classified in clusters 3, 6 and 8. Moreover, these clusters are scarcely represented in the training set, which is mainly formed by patterns classified to cluster 5. The unified distance matrix represented in Figure 4.17b denoted high distance between the map positions of the former three clusters and the later one.

These analyses indicate a high dissimilitude between the patterns corresponding to the period of time represented in Figures 4.13 and 4.16, and the ones used in the model training phase. Previous studies reveal that usually, the ANNs are not able to extrapolate beyond the range of data used for training, a solution to this problem being an optimal division of data into training and testing sets [134]. Thus, an actual state model cannot be used for the forecast of RO plant performance. Even a periodically retraining of the NN may not assure an acceptable forecast. In order to overcome this problem, time should be used as an input of the model.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

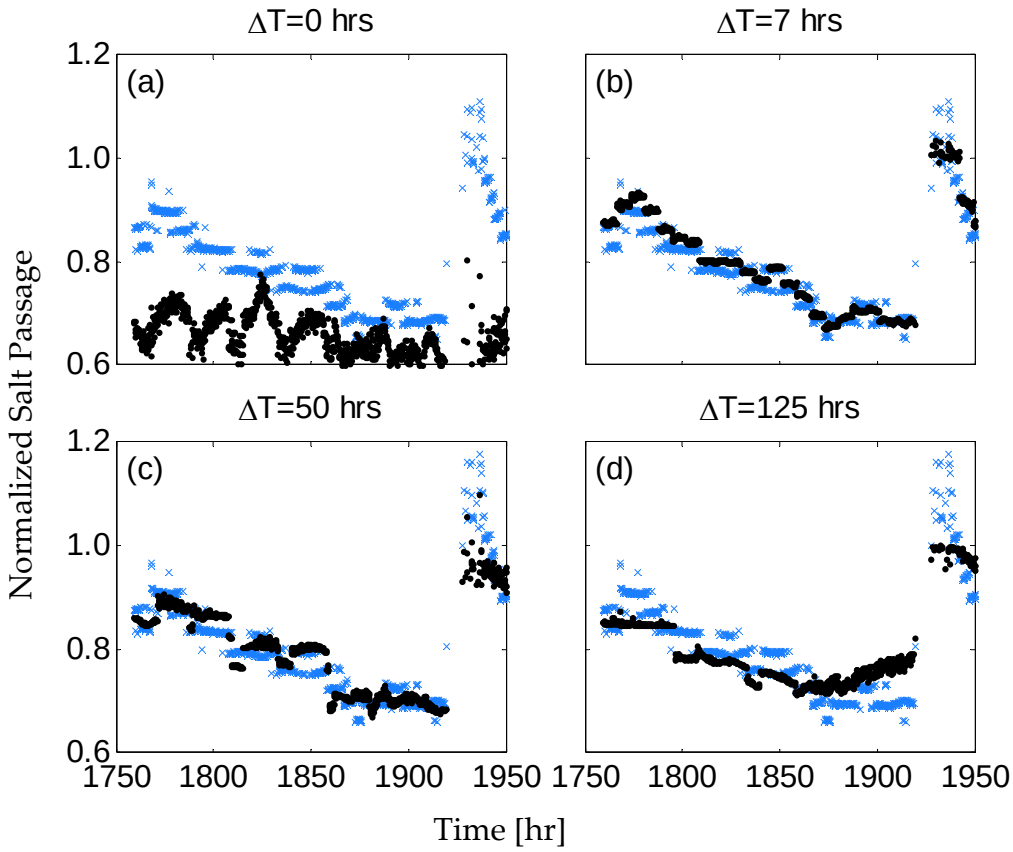


Figure 4.16. Comparison of normalized salt passage predictions for different length of time intervals, for the operational period 1750-1950 hrs. (a) Δt = 0 hrs (actual state model); (b) Δt = 7 hrs; (c) Δt = 50 hrs; (d) Δt = 125 hrs. ×, experimental data; •, NN predictions.

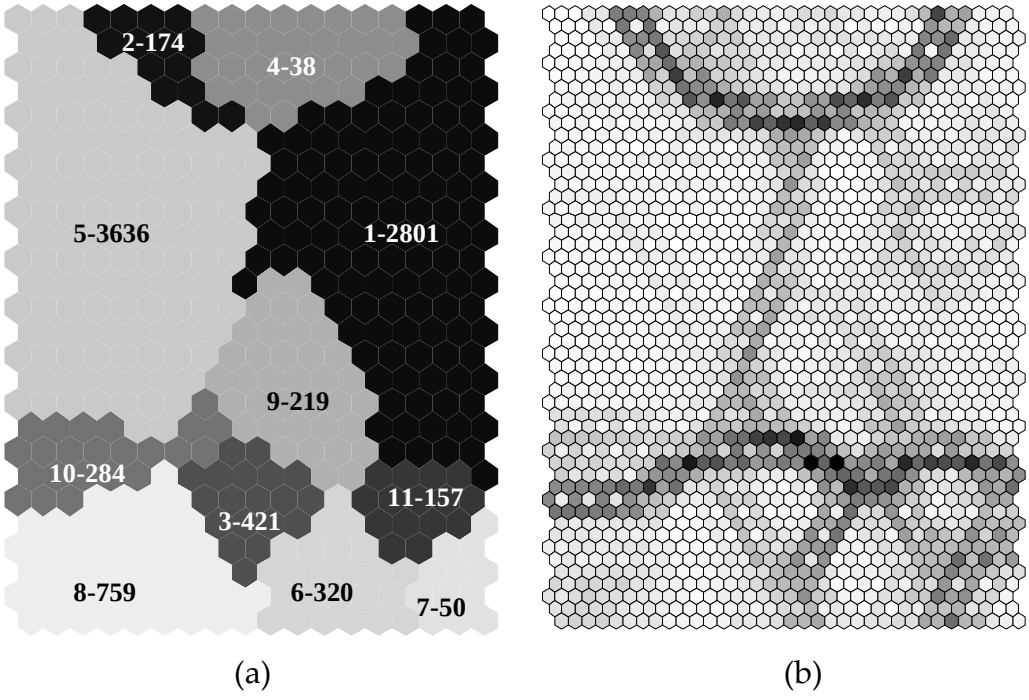


Figure 4.17. SOM clustering of normalized salt passage operational patterns. (a) 11 separate clusters with the corresponding number of patterns allocated to each one; (b) Unified distance matrix.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

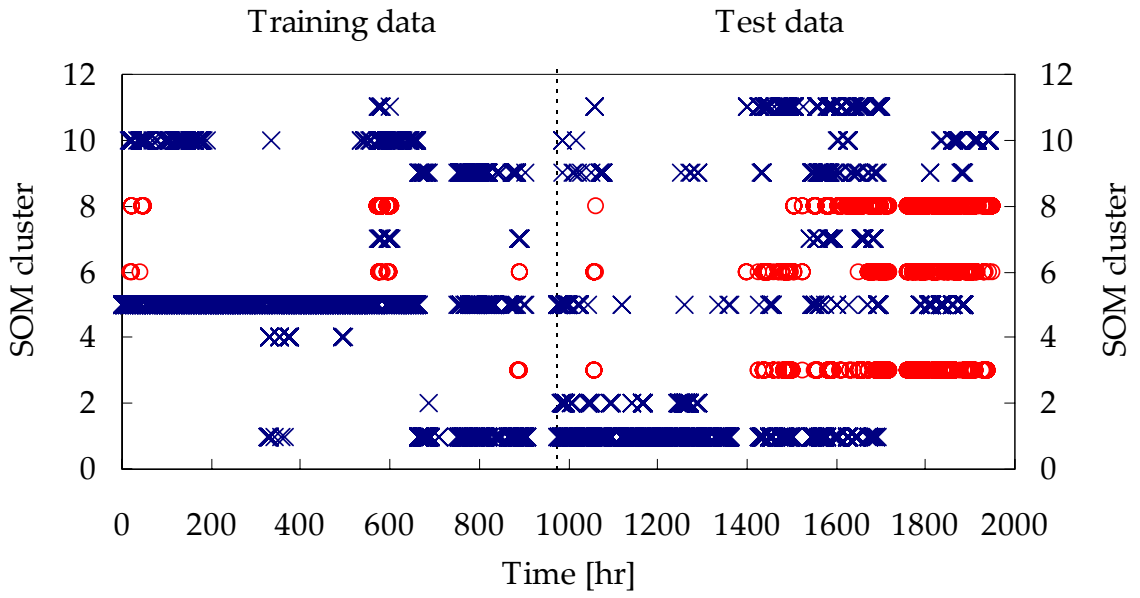


Figure 4.18. Representation of normalized salt passage operational patterns in 11 clusters. The vertical dotted line represents the separation between the training data set and test data set.

4.3. Fuzzy ARTMAP approach for modeling plant performance

An alternative approach based on Fuzzy ARTMAP classification has been developed for the forecasting of the reverse osmosis plant performance in terms of permeate flow rate and salt passage. The methodology uses the same experimental data presented in Section 4.1 and the predictive Fuzzy ARTMAP algorithm introduced by Giralt et al. [88] to anticipate the process performance evolution based on present and past information. Separate models were developed for normalized permeate flow rate and normalized salt passage with the aim of predicting these parameters several steps in the future based on their experimental variation.

Accordingly, for each parameter, three parallel Fuzzy ARTMAP models were implemented using the configuration of input (input pattern to ART_a module, called pattern A) and output (input pattern to ART_b module, called pattern B) data as presented in Figure 4.19. Each one of the three parallel Fuzzy ARTMAP models used the same set of input parameters, consisting in n successive experimental values of the parameter that is to be modeled. The present time (time t) together with the previous n-1 measurements, were used to predict the considered parameter (normalized permeate flow rate or normalized salt passage, respectively) for the subsequent three time moments. Hence, the first Fuzzy ARTMAP model was designed for predicting one time step ahead, the second model had the aim of predicting

ISBN:978-84-691-2701-8/DL:T.386-2008

two time steps ahead, while the third one was developed for predicting three time steps ahead.

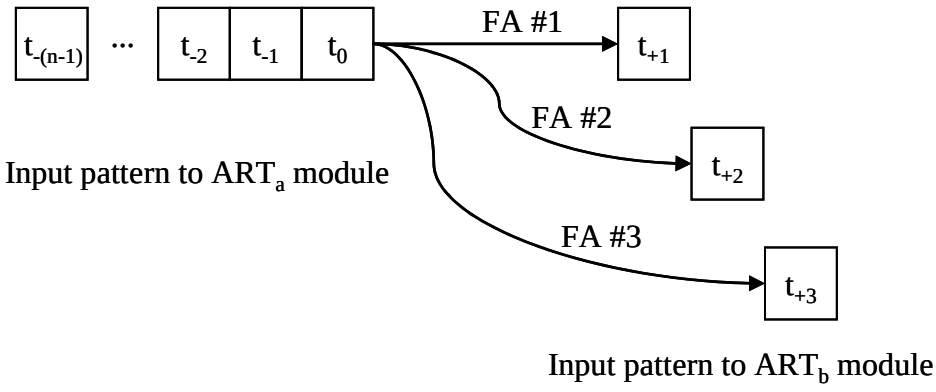


Figure 4.19. Input and output data configuration for the three parallel Fuzzy ARTMAP models developed for both normalized permeate flow rate and normalized salt passage.

The number of successive experimental points used to form the input patterns of ART_a module of the three parallel Fuzzy ARTMAP (FA) models was established in order to avoid the multi-evaluation. This occurs when two identical input patterns presented to ART_a module (pattern A) have two different corresponding patterns that are submitted to ART_b module (pattern B), as presented in Eq. (4.8):

$$\begin{aligned} A_1 &= [a_1, a_2, ..., a_n] \rightarrow B_1 = [b_1] \\ A_2 &= [a_1, a_2, ..., a_n] \rightarrow B_2 = [b_2] \end{aligned} \tag{4.8}$$

Multi-evaluation can be avoided by increasing the length of patterns A, and therefore increasing n. The smallest number of successive experimental points to avoid multi-evaluation was found to be 6 for modeling the normalized permeate flow rate and 7 for modeling the normalized salt passage.

Similar with the approach based on back-propagation algorithm presented in Section 4.2, the models were trained using the experimental data corresponding to the first three periods of operation (0-906 hrs). The rest of experimental data, corresponding to the operation periods starting at t = 976 hrs and forward in time were used for testing the models. In order to use at a maximum level the information presented in the training data set, the formation of the highest possible number of classes in ART_b module was allowed by setting the ρ_b parameter value very closed to unity. Accordingly, the training data set is perfectly modeled, and therefore only the test data set prediction will be further presented.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

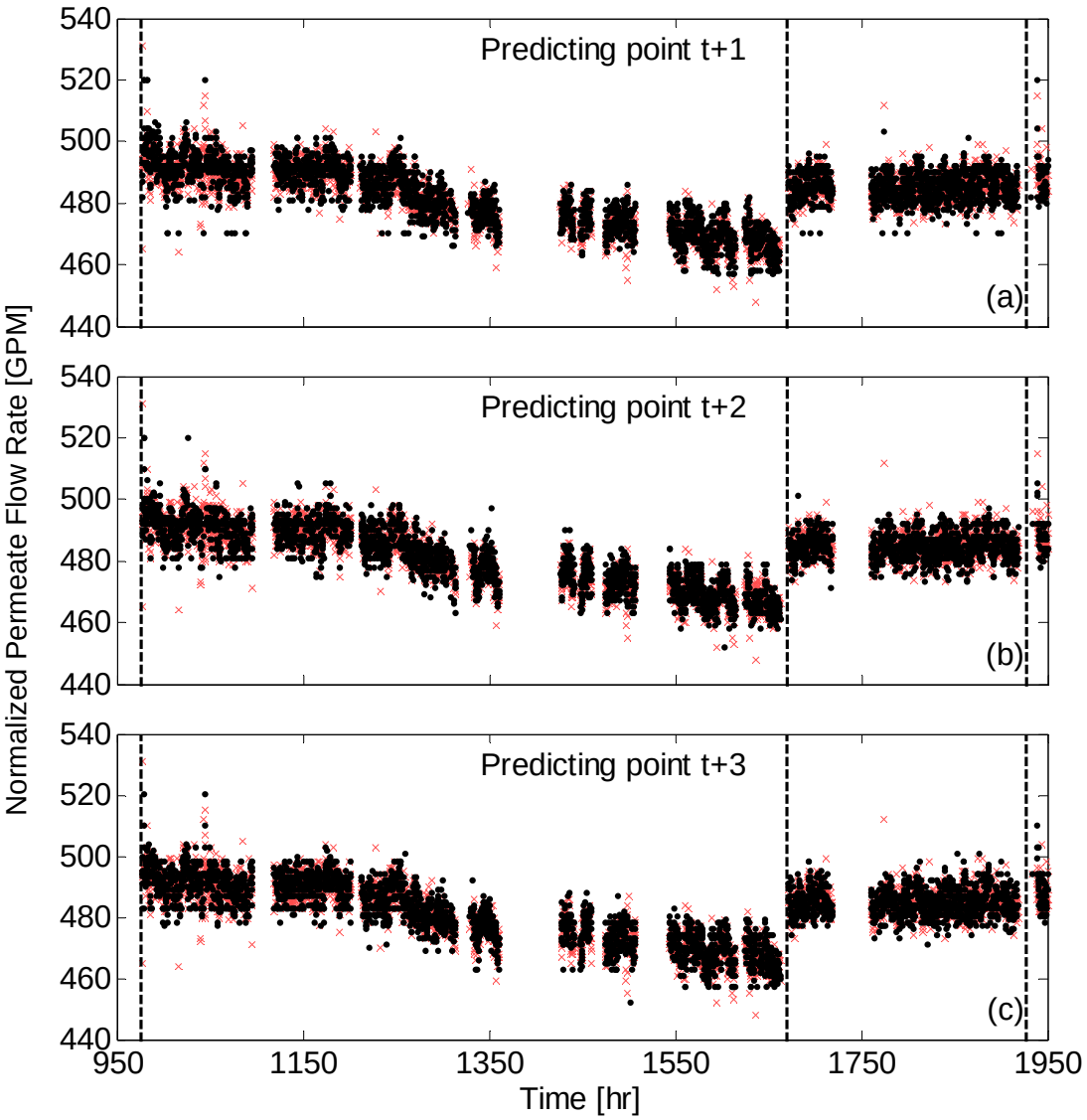


Figure 4.20. Test data set predictions for the normalized permeate flow rate using the Fuzzy ARTMAP approach. ×, experimental data; •, NN predictions. Normalized permeate flow rate predicted for (a) one time step ahead; (b) two time steps ahead; (c) three time steps ahead. The vertical dotted lines represent startup moments after process interruptions for membrane cleaning and/or replacement.

As illustrated in Figures 4.20 and 4.21, the developed Fuzzy ARTMAP models captured very well the dynamics of the process performance parameters. The predicted normalized permeate flow rates were in very good agreement with the experimental values, the average relative error being in all cases lower than 1%. Accordingly, the Fuzzy ARTMAP model designed to predict the normalized permeate flow rate one time step ahead (Figure 4.20a) presented an average relative error of 0.89%, with the corresponding standard deviation of 0.76%. The models developed for predicting the normalized permeate flow rate for two and three time steps ahead presented the average relative errors of 0.84% with standard deviation of 0.71% (Figure 4.20b) and 0.91% with the corresponding standard deviations of 0.75% (Figure 4.20), respectively.

ISBN:978-84-691-2701-8/DL:T.386-2008

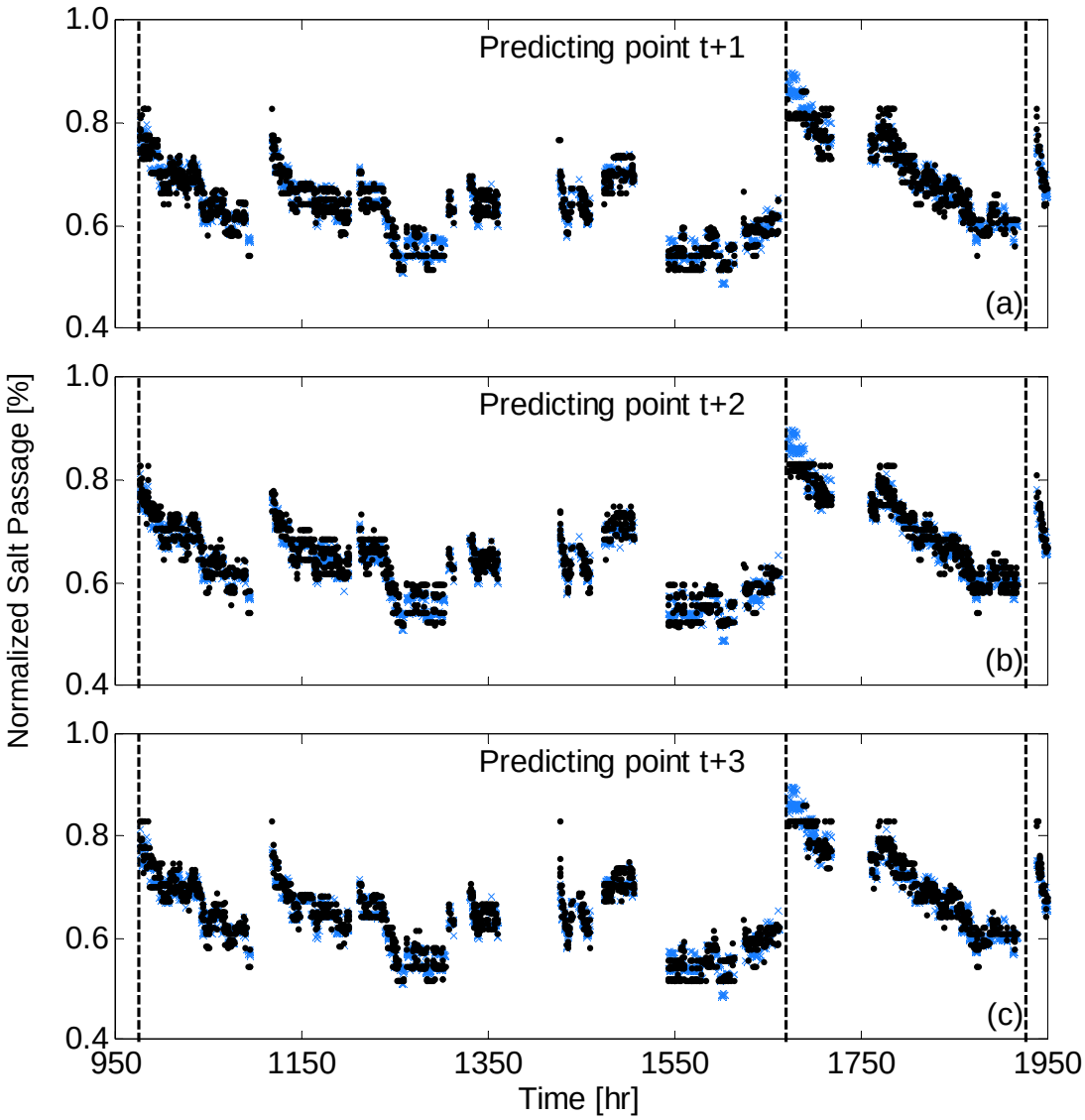


Figure 4.21. Test data set predictions for the normalized salt passage using the Fuzzy ARTMAP approach. ×, experimental data; •, NN predictions. Normalized salt passage predicted for (a) one time step ahead; (b) two time steps ahead; (c) three time steps ahead. The vertical dotted lines represent startup moments after process interruptions for membrane cleaning and/or replacement.

Slightly larger average absolute errors were obtained in the case of the three parallel Fuzzy ARTMAP models developed for predicting the normalized salt passage for one, two and three time steps ahead. However, the models predicted the normalized salt passage with a very good accuracy, since the average relative errors did not exceed 3% in none of the three cases. The model design for predicting the normalized salt passage for one time step ahead presented an average relative error of 2.79% with the corresponding standard deviation of 2.29% (Figure 4.21a). The normalized salt passage for two and three time steps ahead was predicted with similar average relative errors of 2.66% with standard deviation of 2.24% (Figure 4.21b) and 2.61% with standard deviation of 2.29% (Figure 4.21c), respectively.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

The results of this approach based on Fuzzy ARTMAP encourage in considering the integration of NN-based RO models in development of optimization and control strategies. The prediction of process performance parameters was achieved for three time steps ahead based on the last experimental measurements. Taking into account that the time sampling period for monitoring the process parameters of the considered RO brackish water desalination plant was 10 minutes, a 30 minutes forecast could be attained. Hence, the current approach could be useful for control as it will give sufficient lead time for decision making.

ISBN:978-84-691-2701-8/DL:T.386-2008

5. Conclusions

The RO membrane process operation performance was successfully modeled by means of artificial neural networks-based on direct analysis of experimental data. Two approaches of process operation performance modeling were considered: one based on characterizing the organic compounds passage through RO membranes, and a second one based on modeling the dynamics of permeate flow and separation performances for a full-scale RO plant.

The passage, sorption and rejection of organics in RO filtration were studied using quantitative chemical structure-property analysis based on available experimental data for 50 compounds that included specific chemicals of public heath concern in addition to amino acids and selected antibiotics. It was demonstrated that organic solute passage and sorption in RO membranes can be qualitatively and quantitatively related to chemical structure. Three feature selections methods, CFS, SOM-DA and ANNIGMA, were effectively used to discriminate the most relevant set of molecular descriptors to account for organic solute sorption by RO membranes and passage through these membranes. Very good agreement between the three feature selection methods was observed. The most significant molecular descriptors to characterize the sorbed fraction included size of the smallest ring, dipole moment, dipole hybridization, LUMO energy and heat of formation, with the dipole vector Y as additional parameter specific for the polyamide membranes. For the passage fraction the most relevant molecular descriptors were the size of the smallest ring, molecular weight, shape index kappa 2 and LUMO energy, with the dipole hybridization as additional descriptor specific for the cellulose acetate membrane. The chemical space of the 50 organic compounds and the applicability domain of the models developed were analyzed by means of PCA and Self-Organizing Maps. Families that included chemicals of public health concern, amino acids and antibiotics were identified and successfully discriminated by functional group counts.

ISBN:978-84-691-2701-8/DL:T.386-2008

Leave-one-out cross-validation and externally validated quantitative structure-property relationship models for organic solute sorption and passage for polyamide and cellulose acetate membranes were developed using artificial neural networks. Three kinds of ANN-based QSPRs were developed: Independent, Membrane-Composite and MP-Composite ANQ models. Predictions of organic solute rejection were made based on an overall mass balance using the predicted solute sorption and passage. Highly performing Independent ANQ models were built with the explained variance in prediction indices (q^2) exceeding 0.975 for the LOO internal validation and 0.900 in most cases of the external validation, i.e., with a good correlation between the predicted and experimental values. The absolute average errors for predicted organic passage, sorption and rejection fractions were generally lower than 0.077 (70.9%) for all LOO cross-validation and externally validated models. For the Membrane-Composite ANQ models lower explained variance in predictions were obtained, i.e. q^2 as low as 0.928 for the internal validation, and 0.793 for the external validation, respectively. The average absolute errors of these models were as high as 0.067 (34.9%) for the LOO cross-validation, and 0.088 (52.4%) for the external test set validation, respectively. Worst results were obtained for the MP-Composite ANQ models built for the simultaneous predictions of the two experimental fractions. Accordingly, for the internal validation, the explained variance in prediction index was as low as 0.842, while the average absolute errors were up to 0.054 (34.7%). For the external test set validation, the q^2 index decreased down to 0.692 while the average absolute errors increased up to 0.145 (109.8%). This was mainly attributed to the reduced number of organic compounds with available experimental data, and the increase in the number of input parameters for these models. Therefore, a reduced number of hidden neurons had to be used, which directly affected the models performance.

Predictions were consistent with the fact that higher organic solute rejection and lower organic solute passage occur in the polyamide membranes compared to the cellulose acetate membrane. The results are encouraging and suggest the potential application of the applied methods for developing comprehensive and predictive ANN-based QSPR models, using expanded databases, which will provide the analysis and forecasting capability necessary for public health protection that is afforded by RO water treatment processes.

ISBN:978-84-691-2701-8/DL:T.386-2008

The applicability of the ANN-based QSPR models was assessed by means of a mass balance validating test. The models were tested for the 50 organic compounds with available experimental data, as well as on a number of 143 organic micro pollutants of health concern, that were not experimentally characterized. The mass balance test was applied without a priori prescreening the new compounds with respect to the applicability domain for which the models were developed. The test results demonstrate the applicability of the QSPR approach as the mass balance based on predicted fractions was fulfilled in most cases. A reduced number of compounds out of the total of 193 considered, presented a mass balance relative error higher than 25%, i.e. 15 compounds in the case of BW30 membrane, 14 in the case of TFCHR membrane and 15 in the case of CA membrane, respectively. A SOM analysis revealed that these compounds are not well represented by the set of 50 chemicals used to train the models.

The performance dynamics of a full-scale brackish water RO desalination plant were successfully modeled by two different neural networks approaches. A back-propagation neural network was trained to integrate the effect of operating parameters, feed water quality and fouling occurrence on the time evolution of permeate flow rate and salt passage. The experimental data collected from a RO pilot plant in Port Hueneme, California were normalized following the ASTM 4516-00 method. It was showed that an actual state model (i.e., which correlate the present time plant performance with the present time process variables) cannot be used for predictions of operational patterns different from the ones considered in model training. Therefore, the process past information necessary to make present time predictions was incorporated into the model by dividing the time space into equals intervals and selecting the normalized permeate flow rate (or normalized salt passage, respectively) at the beginning of each time interval as additional model input. The spectrum of possible network architectures and input variables configuration was scanned to arrive at the optimal model. The best results were obtained when dividing the time space into equal intervals of 7 hrs, and a network architecture with 5 neurons in the hidden layer. Using this approach, reasonable process performance parameters forecasting can be attained and thus provide the capability of inferring the occurrence of membrane fouling. Accordingly, the test data set normalized permeate flow rate and normalized salt passage were predicted with average relative errors of 0.88% and 2.73%, respectively.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

An alternative approach based on the use of Fuzzy ARTMAP was developed for the forecasting of the two process performance parameters considered. The prediction of normalized permeate flow rate and normalized salt passage for three time steps ahead was achieved based on the last 6 and 7 experimental measurements, respectively. The forecasting average relative error did not exceed 0.89% in the case of normalized permeate flow rate, and 2.79% for the normalized salt passage. These errors are comparable with the ones obtained using the back-propagation models.

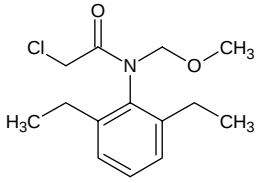
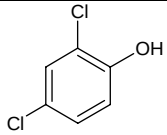
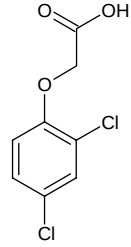
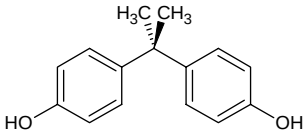
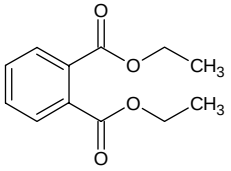
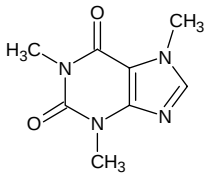
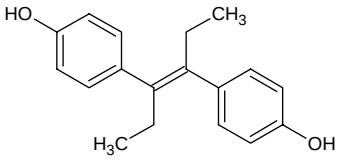
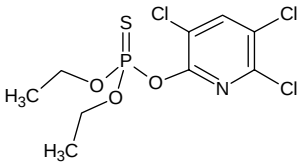
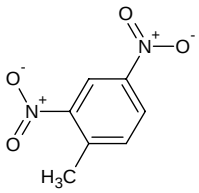
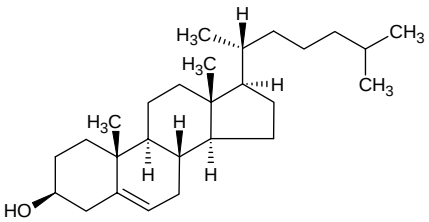
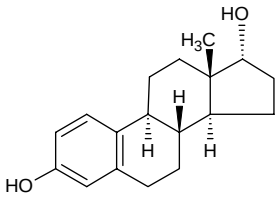
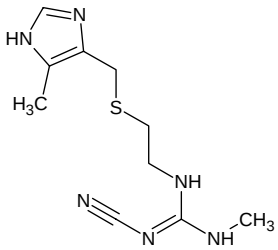
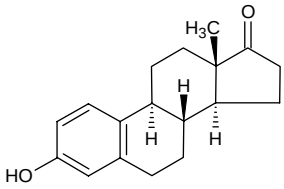
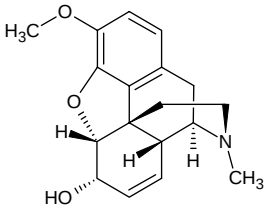
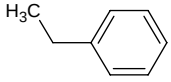
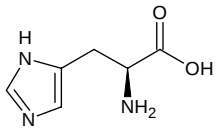
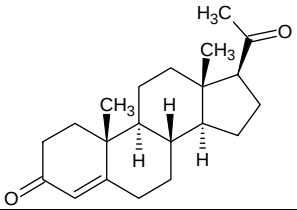
UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

ANNEXES

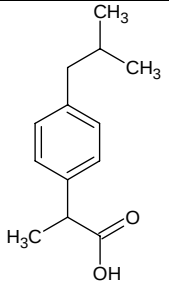
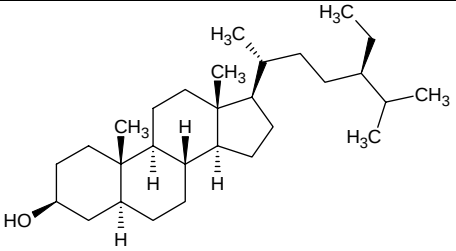
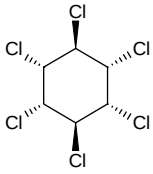
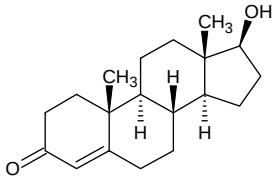
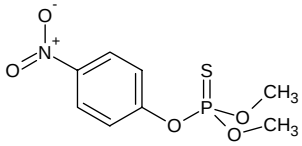
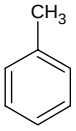
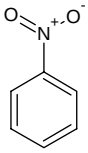
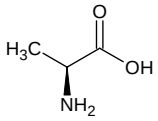
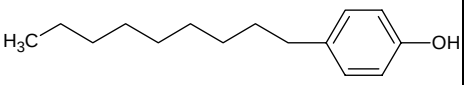
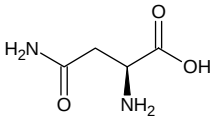
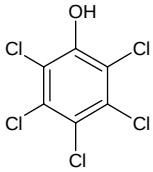
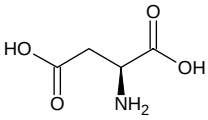
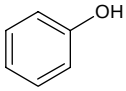
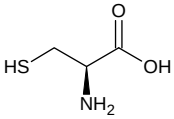
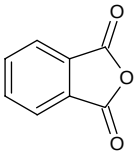
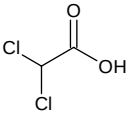
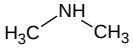
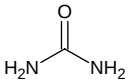
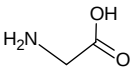
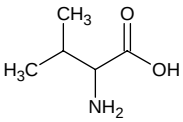
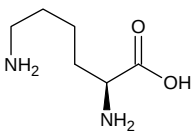
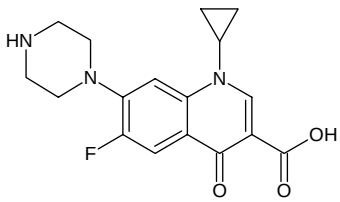
ANNEX I. Chemical structures of the 50 organic compounds used for developing QSPR models.

Chemical structures of the 50 organic compounds with available experimental data presented in Table 3.1, drawn using ACD/ChemSketch 8.00.

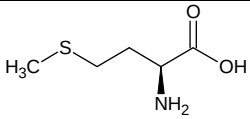
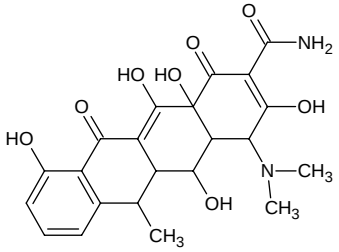
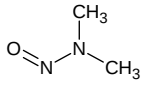
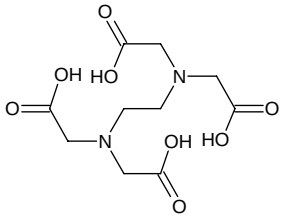
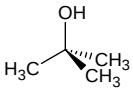
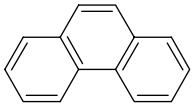
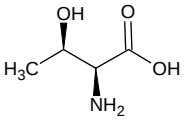
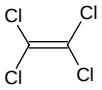
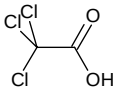
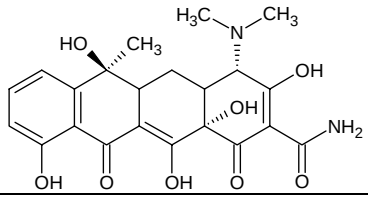
UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Chemical structure	CAS	Chemical structure
15972-60-8		120-83-2	
71-43-2		94-75-7	
80-05-7		84-66-2	
58-08-2		56-53-1	
2921-88-2		121-14-2	
57-88-5		57-91-0	
51481-61-9		53-16-7	
76-57-3		100-41-4	
71-00-1		57-83-0	

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Chemical structure	CAS	Chemical structure
15687-27-1		19466-47-8	
58-89-9		58-22-0	
298-00-0		108-88-3	
98-95-3		56-41-7	
104-40-5		70-47-3	
87-86-5		56-84-8	
108-95-2		52-90-4	
85-44-9		79-43-6	
124-40-3		57-13-6	
56-40-6		72-18-4	
56-87-1		85721-33-1	

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Chemical structure	CAS	Chemical structure
63-68-3		564-25-0	
62-75-9		60-00-4	
75-65-0		85-01-8	
72-19-5		127-18-4	
76-03-9		60-54-8	

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

ANNEX II. Molecular descriptors and reverse osmosis experimental data for the 50 organic compounds used to develop QSPR.

Database containing the molecular descriptors calculated using CAChe Pro Version 6.1 (Oxford Molecular Ltd.) for the 50 organic compounds, together with the average experimental values and experimental standard deviations for the three fractions measured for each molecule and each membrane.

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	a_count	b_count	b_single	b_double	chi0	chi1	chi2	dipole [Debye]	dipole X [Debye]
15972-60-8	38	38	34	4	13.54	8.69	6.68	6.12	-2.23
71-43-2	12	12	9	3	4.24	3.00	2.12	0.00	0.00
80-05-7	33	34	28	6	12.47	8.00	7.74	2.08	-0.45
58-08-2	24	25	21	4	10.46	6.54	6.23	5.13	4.95
2921-88-2	29	29	25	4	13.76	8.42	7.68	6.93	-0.69
57-88-5	74	77	76	1	20.10	13.25	13.07	2.26	0.64
51481-61-9	33	33	29	3	12.51	8.27	6.37	6.62	-6.06
76-57-3	43	47	43	4	15.06	10.66	10.23	5.34	-0.49
120-83-2	13	13	10	3	6.85	4.20	3.87	0.57	0.47
94-75-7	19	19	15	4	9.85	6.09	5.63	4.61	4.01
84-66-2	30	30	25	5	11.97	7.70	6.14	1.81	-1.41
56-53-1	40	41	34	7	14.54	9.65	8.18	0.09	-0.08
121-14-2	19	19	14	5	10.01	6.02	5.70	7.27	2.89
57-91-0	44	47	44	3	13.91	9.59	9.38	0.54	0.37
53-16-7	42	45	41	4	13.91	9.59	9.38	2.89	0.35
100-41-4	18	18	15	3	5.82	3.93	2.91	0.38	-0.38
71-00-1	20	20	17	3	8.27	5.20	4.61	3.96	2.25
15687-27-1	33	33	29	4	11.42	7.00	6.51	2.75	-1.76
58-89-9	18	18	18	0	9.46	5.46	5.16	2.97	0.72
298-00-0	26	26	21	5	12.18	7.50	6.75	5.95	-0.70
98-95-3	14	14	10	4	6.69	4.31	3.64	6.93	0.03
104-40-5	40	40	37	3	11.64	7.83	6.04	1.81	-1.37
87-86-5	13	13	10	3	9.46	5.46	5.16	1.91	-1.64
108-95-2	13	13	10	3	5.11	3.39	2.74	1.70	-1.32
85-44-9	15	16	11	5	7.85	5.29	4.80	8.64	-8.64
57-83-0	53	56	53	3	16.41	10.86	11.01	6.08	1.16
19466-47-8	82	85	85	0	21.68	14.20	13.73	2.02	-0.25
58-22-0	49	52	50	2	14.83	9.95	10.06	5.36	5.24
108-88-3	15	15	12	3	5.11	3.39	2.74	0.31	0.00
56-41-7	13	12	11	1	5.16	2.64	2.49	2.61	-1.84
70-47-3	17	16	14	2	7.44	4.04	3.85	2.36	-1.48
56-84-8	16	15	13	2	7.44	4.04	3.85	5.56	3.57
52-90-4	14	13	12	1	5.86	3.18	2.63	4.05	0.28
79-43-6	8	7	6	1	5.16	2.64	2.49	2.94	1.13
124-40-3	10	9	9	0	2.71	1.41	0.71	1.50	-0.01
56-40-6	10	9	8	1	4.28	2.27	1.80	2.07	-1.19
56-87-1	24	23	22	1	7.98	4.68	3.72	5.42	-0.22
63-68-3	20	19	18	1	7.28	4.18	3.36	4.09	-2.14
62-75-9	11	10	9	1	4.28	2.27	1.80	4.87	4.05
75-65-0	15	14	14	0	4.50	2.00	3.00	2.05	1.62
72-19-5	17	16	15	1	6.73	3.55	3.35	2.29	1.83
76-03-9	8	7	6	1	6.08	2.94	3.52	1.64	0.21
57-13-6	8	7	6	1	3.58	1.73	1.73	5.71	0.00
72-18-4	19	18	17	1	6.73	3.55	3.35	3.31	-2.87
85721-33-1	42	45	39	6	16.85	11.56	10.84	13.87	-1.92
564-25-0	56	59	51	8	23.86	14.87	14.85	15.80	10.91
60-00-4	36	35	31	4	15.71	9.20	8.83	15.60	-0.20
85-01-8	24	26	19	7	9.38	6.95	5.99	0.11	0.00
127-18-4	6	5	4	1	5.16	2.64	2.49	0.00	0.00
60-54-8	56	59	51	8	23.91	14.77	15.24	4.73	-1.65

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	dipole Y [Debye]	dipole Z [Debye]	dielEn [kcal/mol]	sterEn [kcal/mol]	heatform [kcal/mol]	mr	weight [Da]	Po [Å³]	r_count
15972-60-8	-1.41	-5.52	-0.37	-76.89	-76.88	73.93	269.77	29.90	1
71-43-2	0.00	0.00	-0.09	-8.08	20.16	26.06	78.11	10.39	1
80-05-7	2.02	-0.16	-0.46	-9.94	-57.06	68.21	228.29	27.60	2
58-08-2	1.36	-0.06	-0.67	8.24	8.24	48.28	194.19	20.65	2
2921-88-2	-3.40	6.00	-0.58	-144.94	-144.86	79.00	350.58	27.93	1
57-88-5	1.90	-1.06	-0.21	63.20	-130.52	120.62	386.66	48.84	4
51481-61-9	2.10	1.64	-1.41	103.49	103.47	72.46	252.34	27.32	1
76-57-3	5.26	0.81	-0.48	73.61	-58.55	84.53	299.37	34.30	5
120-83-2	0.32	-0.01	-0.17	-7.69	-38.72	37.36	163.00	15.32	1
94-75-7	0.03	-2.28	-0.66	-10.62	-119.78	48.22	221.04	19.97	1
84-66-2	0.92	-0.66	-0.44	-78.72	-149.07	58.61	222.24	23.63	1
56-53-1	-0.01	-0.05	-0.44	-8.94	-55.98	82.67	268.36	33.64	2
121-14-2	-6.67	-0.23	-1.08	26.02	26.02	45.75	182.14	17.41	1
57-91-0	-0.32	-0.22	-0.41	20.54	-116.44	79.62	272.39	32.71	4
53-16-7	-1.09	-2.65	-0.47	16.20	-100.53	78.80	270.37	32.30	4
100-41-4	-0.03	-0.01	-0.10	-6.77	7.10	35.70	106.17	14.24	1
71-00-1	3.20	-0.65	-0.88	-49.20	-49.21	37.00	155.16	15.56	1
15687-27-1	1.23	-1.71	-0.37	-17.56	-109.89	60.73	206.28	24.36	1
58-89-9	-1.26	-2.59	-0.38	29.40	-62.10	54.08	290.83	23.06	1
298-00-0	1.63	-5.68	-1.04	-126.28	-126.34	63.19	263.21	20.88	1
98-95-3	-6.93	0.00	-0.41	-7.56	17.57	33.38	123.11	12.87	1
104-40-5	1.19	-0.01	-0.26	-2.09	-88.10	69.60	220.35	28.03	1
87-86-5	0.99	-0.02	-0.17	3.86	-48.17	51.78	266.34	21.70	1
108-95-2	1.06	0.01	-0.24	-7.90	-27.06	27.75	94.11	11.13	1
85-44-9	0.00	0.00	-0.54	-15.25	-78.58	34.63	148.12	14.98	2
57-83-0	5.92	0.78	-0.56	39.93	-120.32	92.80	314.47	37.78	4
19466-47-8	2.00	0.01	-0.16	-161.30	-161.30	128.92	416.73	52.58	4
58-22-0	1.14	0.08	-0.47	43.89	-123.66	84.52	288.43	34.56	4
108-88-3	0.31	0.03	-0.10	-8.60	12.42	31.10	92.14	12.36	1
56-41-7	-1.65	-0.84	-0.49	-4.54	-110.98	20.50	89.09	8.30	0
70-47-3	1.83	-0.17	-0.76	-34.60	-157.44	28.36	132.12	11.62	0
56-84-8	2.58	-3.39	-0.96	-19.97	-205.95	26.53	133.10	10.91	0
52-90-4	-2.09	-3.46	-0.57	-2.57	-105.23	28.17	121.15	9.24	0
79-43-6	-0.82	2.59	-0.35	8.01	-114.55	22.62	128.94	9.08	0
124-40-3	-0.37	1.46	-0.11	-5.66	-5.66	14.69	45.08	5.83	0
56-40-6	1.70	-0.04	-0.44	-5.74	-110.18	16.00	75.07	6.52	0
56-87-1	1.54	-5.19	-0.67	0.86	-127.11	37.81	146.19	15.10	0
63-68-3	-3.44	-0.55	-0.62	-5.89	-117.37	37.83	149.21	13.10	0
62-75-9	2.63	0.59	-0.29	-1.97	3.19	20.05	74.08	7.54	0
75-65-0	-1.25	0.05	-0.17	3.74	-75.10	22.07	74.12	8.64	0
72-19-5	-1.09	-0.85	-0.46	-1.80	-165.19	26.46	119.12	10.76	0
76-03-9	-1.40	-0.82	-0.28	17.69	-109.16	28.16	163.39	11.17	0
57-13-6	-5.71	0.00	-0.68	-61.97	-56.14	13.14	60.06	5.31	0
72-18-4	0.29	1.62	-0.44	-1.14	-122.99	29.49	117.15	11.95	0
85721-33-1	13.74	0.01	-1.20	293.82	-93.10	87.12	331.35	36.69	4
564-25-0	-11.40	-0.67	-1.40	7.23	-273.95	113.13	444.44	45.63	4
60-00-4	-0.06	-15.59	-1.96	62.83	-368.92	62.35	292.25	25.21	0
85-01-8	-0.11	0.00	-0.22	-24.64	53.02	58.96	178.23	24.56	3
127-18-4	0.00	0.00	-0.04	8.91	-13.27	30.95	165.83	12.21	0
60-54-8	4.42	0.33	-0.73	-268.98	-268.99	113.42	444.44	45.97	4

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	small_ring	large_ring	kier1	kier2	kier3	SASA [Å²]	chi0v	chi1v	chi2v	HOMO [eV]
15972-60-8	6	6	16.06	8.23	4.00	281.99	11.96	6.69	4.27	-9.62
71-43-2	6	6	4.17	2.22	1.33	119.17	3.46	2.00	1.16	-9.75
80-05-7	6	6	13.43	5.33	3.06	256.48	10.01	5.59	4.72	-8.95
58-08-2	5	6	10.52	3.54	1.46	211.43	8.18	4.11	3.23	-8.96
2921-88-2	6	6	16.06	6.96	4.57	315.41	13.63	8.71	7.03	-9.74
57-88-5	5	6	21.24	7.92	3.66	420.39	19.34	12.63	12.19	-9.45
51481-61-9	5	5	15.06	9.00	5.93	303.75	10.77	6.38	4.37	-8.58
76-57-3	5	6	14.35	4.76	1.64	286.85	12.95	8.10	7.11	-8.70
120-83-2	6	6	7.11	2.72	1.70	166.08	5.95	3.10	2.44	-9.09
94-75-7	6	6	11.08	5.02	3.70	216.52	7.97	4.15	3.10	-9.46
84-66-2	6	6	14.06	7.35	4.08	247.48	9.36	5.14	2.99	-10.29
56-53-1	6	6	16.37	7.85	4.25	297.20	11.93	6.96	4.76	-8.91
121-14-2	6	6	11.08	4.48	2.72	192.38	6.76	3.42	2.49	-11.03
57-91-0	5	6	13.65	4.75	1.96	285.30	12.18	8.09	7.44	-8.90
53-16-7	5	6	13.65	4.75	1.96	284.70	12.06	7.95	7.22	-8.96
100-41-4	6	6	6.13	3.11	1.80	155.66	5.09	2.97	1.84	-9.41
71-00-1	5	5	9.09	4.13	2.84	183.92	5.82	3.16	2.23	-9.03
15687-27-1	6	6	13.07	5.92	4.17	254.81	9.53	5.32	4.40	-9.57
58-89-9	6	6	10.08	3.40	1.56	220.87	10.27	5.93	5.69	-10.33
298-00-0	6	6	14.06	6.07	4.08	261.77	10.36	6.72	5.80	-10.37
98-95-3	6	6	7.11	3.24	2.00	148.16	4.65	2.50	1.59	-10.60
104-40-5	6	6	14.06	9.07	7.06	305.86	10.41	6.61	4.54	-8.94
87-86-5	6	6	10.08	3.40	1.56	211.00	9.12	4.56	3.81	-9.14
108-95-2	6	6	5.14	2.34	1.50	129.42	3.83	2.13	1.34	-9.18
85-44-9	5	6	7.64	2.80	1.21	163.60	5.53	3.14	2.22	-10.82
57-83-0	5	6	16.47	5.50	2.29	322.12	14.86	9.60	9.25	-10.16
19466-47-8	5	6	23.17	8.74	3.97	438.95	21.13	13.88	13.28	-10.24
58-22-0	5	6	14.58	4.75	1.93	296.84	13.40	8.87	8.61	-10.08
108-88-3	6	6	5.14	2.34	1.50	138.20	4.39	2.41	1.66	-9.44
56-41-7	0	0	6.00	2.22	3.00	123.33	3.51	1.63	1.13	-9.96
70-47-3	0	0	9.00	3.92	4.50	155.58	4.70	2.30	1.62	-9.88
56-84-8	0	0	9.00	3.92	4.50	155.62	4.57	2.24	1.54	-10.37
52-90-4	0	0	7.00	3.06	2.67	146.11	4.56	2.41	1.49	-9.64
79-43-6	0	0	6.00	2.22	3.00	129.34	4.20	2.03	1.74	-10.88
124-40-3	0	0	3.00	2.00	2.00	92.15	2.50	1.00	0.50	-9.38
56-40-6	0	0	5.00	2.25	4.00	105.88	2.64	1.19	0.60	-9.93
56-87-1	0	0	10.00	5.76	5.53	190.41	5.92	3.37	2.23	-9.38
63-68-3	0	0	9.00	4.84	4.50	184.87	6.15	4.05	2.71	-9.17
62-75-9	0	0	5.00	2.25	4.00	111.91	3.30	1.28	0.93	-9.82
75-65-0	0	0	5.00	1.00	0.00	122.70	3.95	1.72	2.17	-11.28
72-19-5	0	0	8.00	3.11	2.81	147.83	4.54	2.22	1.61	-9.82
76-03-9	0	0	7.00	1.85	2.67	144.57	5.26	2.38	3.08	-10.94
57-13-6	0	0	4.00	1.33	0.00	90.67	2.06	0.78	0.40	-10.10
72-18-4	0	0	8.00	3.11	2.81	156.13	5.09	2.54	2.11	-9.88
85721-33-1	3	6	17.42	6.96	3.13	329.09	13.09	8.13	6.37	-8.83
564-25-0	6	6	25.10	8.59	3.37	384.28	17.60	10.02	8.79	-9.23
60-00-4	0	0	20.00	10.69	12.49	285.10	10.56	5.52	4.00	-9.75
85-01-8	6	6	9.24	3.87	1.65	209.22	7.77	4.82	3.51	-8.74
127-18-4	0	0	6.00	2.22	3.00	145.11	5.54	2.52	2.42	-9.22
60-54-8	6	6	25.10	8.29	3.37	387.91	17.66	9.97	8.96	-9.32

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	LUMO [eV]	dipole_P [Debye]	dipole_H [Debye]	E1_e-n [eV]	E1_e-e [eV]	E1_total [eV]	E2_res [eV]	E2_ex [eV]	E2_e-e [eV]
15972-60-8	-0.03	5.95	0.21	-4682.6	1919.2	-2763.4	-493.8	-225.3	18581.4
71-43-2	0.40	0.00	0.00	-1134.7	471.3	-663.4	-169.9	-80.1	2299.9
80-05-7	0.29	1.90	0.25	-3883.7	1607.0	-2276.7	-473.3	-214.9	14339.6
58-08-2	-0.35	5.02	0.76	-3834.2	1582.5	-2251.6	-372.7	-157.1	11395.4
2921-88-2	-1.50	7.37	0.74	-5963.3	2473.2	-3490.1	-378.0	-162.9	17787.2
57-88-5	1.03	1.84	0.61	-5845.5	2398.9	-3446.6	-882.7	-422.1	39378.3
51481-61-9	0.18	5.31	1.31	-4181.6	1728.1	-2453.5	-462.6	-203.7	14120.1
76-57-3	0.28	4.36	1.59	-5297.9	2178.5	-3119.4	-609.5	-274.4	25297.7
120-83-2	-0.24	0.75	0.75	-2837.5	1147.6	-1689.9	-193.5	-84.4	4946.0
94-75-7	-0.48	3.80	0.76	-4257.0	1732.5	-2524.5	-286.1	-121.8	9490.8
84-66-2	-0.58	2.00	0.23	-4370.3	1802.2	-2568.1	-424.3	-187.9	14338.1
56-53-1	0.16	0.07	0.04	-4458.4	1842.4	-2616.0	-564.2	-258.2	18562.4
121-14-2	-1.84	7.63	0.36	-3994.4	1639.1	-2355.3	-315.6	-130.9	9476.1
57-91-0	0.32	0.84	0.42	-4510.7	1860.3	-2650.3	-584.7	-270.0	21096.9
53-16-7	0.26	3.08	0.60	-4484.6	1849.6	-2635.0	-572.2	-263.9	20331.1
100-41-4	0.39	0.42	0.06	-1535.2	634.7	-900.5	-237.2	-112.7	4080.9
71-00-1	0.59	1.21	2.85	-3179.3	1322.8	-1856.5	-296.0	-124.3	7511.0
15687-27-1	0.04	3.28	0.63	-3561.9	1471.9	-2090.0	-432.4	-199.3	12420.6
58-89-9	0.06	3.01	0.09	-4795.2	1914.2	-2881.1	-200.9	-91.5	10272.8
298-00-0	-2.06	6.00	0.05	-5001.1	2119.3	-2881.8	-369.3	-154.8	13961.5
98-95-3	-1.14	7.17	0.30	-2466.3	1015.9	-1450.5	-225.6	-97.2	4861.1
104-40-5	0.32	1.47	0.62	-3462.9	1426.4	-2036.5	-492.0	-232.4	12895.6
87-86-5	-0.79	1.37	0.64	-4595.7	1825.6	-2770.1	-198.5	-83.1	8565.1
108-95-2	0.29	1.21	0.64	-1656.0	687.4	-968.6	-190.7	-85.8	3150.2
85-44-9	-1.41	8.58	0.00	-3009.9	1239.3	-1770.7	-262.8	-112.3	6450.0
57-83-0	-0.10	5.90	0.21	-5116.7	2107.3	-3009.4	-676.9	-317.7	27439.1
19466-47-8	3.32	1.65	0.58	-6270.5	2570.0	-3700.5	-959.4	-460.4	44978.8
58-22-0	-0.09	4.87	0.74	-4739.7	1952.2	-2787.5	-622.6	-291.3	24237.6
108-88-3	0.38	0.33	0.04	-1334.4	552.6	-781.9	-203.8	-96.5	3146.6
56-41-7	0.95	2.71	1.09	-1989.3	827.5	-1161.8	-166.7	-70.4	3153.8
70-47-3	0.68	2.42	0.76	-3028.7	1267.1	-1761.6	-240.2	-97.7	5905.6
56-84-8	0.36	3.54	2.29	-3213.7	1337.4	-1876.3	-228.4	-91.9	5882.4
52-90-4	-0.07	2.87	2.16	-2299.3	952.4	-1347.0	-174.3	-74.5	3995.4
79-43-6	-0.09	3.09	0.51	-2639.1	1067.4	-1571.7	-104.7	-41.7	2903.9
124-40-3	3.48	0.56	0.99	-759.9	312.6	-447.3	-105.1	-49.1	1088.3
56-40-6	0.75	2.58	1.72	-1789.4	746.2	-1043.2	-133.8	-54.1	2229.5
56-87-1	1.11	4.06	1.49	-2926.0	1214.7	-1711.3	-297.2	-130.4	7470.0
63-68-3	0.04	3.20	0.92	-2704.5	1120.7	-1583.8	-241.0	-107.0	5961.1
62-75-9	0.32	4.72	0.84	-1577.0	644.5	-932.5	-139.8	-60.0	2311.7
75-65-0	3.25	1.65	0.64	-1357.9	560.8	-797.2	-161.2	-75.9	2624.4
72-19-5	0.52	2.75	2.32	-2715.8	1128.8	-1587.0	-219.5	-92.0	5337.9
76-03-9	-0.46	1.63	0.41	-3223.7	1291.4	-1932.3	-105.1	-41.2	3885.5
57-13-6	0.99	5.14	0.59	-1406.5	600.3	-806.2	-115.1	-44.2	1477.1
72-18-4	0.98	3.27	0.95	-2391.1	991.9	-1399.2	-233.6	-103.0	5291.2
85721-33-1	-0.76	14.63	1.57	-6532.7	2699.5	-3833.2	-636.6	-274.9	26156.3
564-25-0	-1.02	15.18	1.49	-8979.7	3716.3	-5263.3	-841.7	-360.9	45631.9
60-00-4	0.65	13.54	2.03	-6789.4	2813.9	-3975.4	-510.7	-210.1	22893.2
85-01-8	-0.54	0.10	0.01	-2585.9	1071.4	-1541.5	-381.2	-176.3	8982.0
127-18-4	-0.32	0.00	0.00	-2762.5	1083.4	-1679.1	-70.2	-29.9	2688.0
60-54-8	-0.84	6.81	2.11	-8982.8	3719.2	-5263.6	-841.4	-360.6	45492.9

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	E2_e-n [eV]	E2_n-n [eV]	E2_el [eV]	E2_total [eV]	E_total [eV]	pmiX [10^{-40} g cm^2]	pmiY [10^{-40} g cm^2]	pmiZ [10^{-40} g cm^2]
15972-60-8	-37313.3	18924.9	193.0	-526.1	-3289.5	2303.1	2533.9	3160.1
71-43-2	-4639.2	2402.4	63.1	-186.9	-850.3	147.8	147.8	295.6
80-05-7	-28809.5	14653.5	183.6	-504.6	-2781.3	981.3	3181.7	3537.0
58-08-2	-22949.6	11707.2	152.9	-376.9	-2628.5	811.5	1216.9	2011.3
2921-88-2	-35927.3	18254.9	114.8	-426.1	-3916.2	1999.1	5346.9	6632.1
57-88-5	-78924.2	39882.9	336.9	-967.8	-4414.4	2768.9	10504.2	11382.6
51481-61-9	-28419.6	14485.0	185.6	-480.8	-2934.3	822.5	5801.9	6564.6
76-57-3	-50746.2	25691.5	243.0	-640.9	-3760.3	1698.5	3051.4	4000.4
120-83-2	-9969.3	5100.0	76.7	-201.2	-1891.1	395.1	1131.6	1526.7
94-75-7	-19115.2	9741.6	117.2	-290.8	-2815.2	849.4	2520.4	3282.9
84-66-2	-28829.2	14660.4	169.3	-442.9	-3011.0	1345.4	1722.7	2699.4
56-53-1	-37272.0	18927.5	217.9	-604.6	-3220.5	965.9	5418.1	5702.2
121-14-2	-19146.0	9804.1	134.2	-312.3	-2667.6	556.6	1558.0	2108.9
57-91-0	-42335.4	21465.6	227.1	-627.5	-3277.8	891.4	4627.4	5086.1
53-16-7	-40813.7	20704.0	221.4	-614.8	-3249.8	1008.9	4392.1	4870.1
100-41-4	-8211.9	4219.5	88.5	-261.4	-1162.0	176.2	559.6	723.9
71-00-1	-15163.8	7776.1	123.3	-297.0	-2153.5	294.0	1516.4	1739.9
15687-27-1	-24972.0	12718.4	167.1	-464.6	-2554.5	631.4	3245.0	3345.3
58-89-9	-20667.0	10473.2	79.0	-213.3	-3094.4	1673.5	1964.1	3244.7
298-00-0	-28329.5	14483.3	115.2	-408.8	-3290.6	787.3	4257.2	4465.3
98-95-3	-9827.6	5058.5	92.1	-230.7	-1681.1	208.1	668.5	876.6
104-40-5	-25906.8	13198.8	187.6	-536.8	-2573.4	299.1	8400.4	8650.0
87-86-5	-17218.2	8733.9	80.9	-200.8	-2970.9	1443.1	1818.4	3261.5
108-95-2	-6360.5	3284.5	74.1	-202.3	-1170.9	150.6	320.2	470.8
85-44-9	-13027.0	6683.4	106.4	-268.7	-2039.4	460.1	733.6	1193.7
57-83-0	-55068.9	27889.4	259.7	-735.0	-3744.4	1164.1	6309.5	6634.2
19466-47-8	-90127.3	45514.7	366.2	-1053.7	-4754.2	2051.0	14280.7	15077.1
58-22-0	-48643.2	24646.1	240.4	-673.4	-3460.9	1011.3	4811.4	5315.3
108-88-3	-6337.2	3266.6	75.9	-224.3	-1006.2	153.2	328.6	476.3
56-41-7	-6395.1	3310.9	69.7	-167.5	-1329.3	165.9	259.9	399.5
70-47-3	-11950.9	6146.0	100.7	-237.3	-1998.9	243.6	943.6	1120.3
56-84-8	-11915.2	6130.9	98.1	-222.2	-2098.5	247.1	971.2	1112.3
52-90-4	-8095.3	4172.2	72.3	-176.6	-1523.6	180.1	733.1	860.7
79-43-6	-5888.6	3030.2	45.5	-100.9	-1672.7	319.6	431.5	607.2
124-40-3	-2205.0	1158.7	42.0	-112.2	-559.6	24.2	89.8	101.3
56-40-6	-4543.8	2371.8	57.5	-130.5	-1173.7	83.1	215.6	288.4
56-87-1	-15055.6	7707.2	121.5	-306.1	-2017.4	394.1	1344.8	1425.8
63-68-3	-12046.9	6182.5	96.6	-251.3	-1835.1	240.4	1583.2	1775.9
62-75-9	-4681.7	2431.0	61.0	-138.7	-1071.3	91.2	189.4	269.5
75-65-0	-5297.5	2736.2	63.0	-174.1	-971.2	177.6	179.3	184.6
72-19-5	-10794.7	5549.5	92.6	-218.9	-1805.9	253.5	564.6	695.3
76-03-9	-7853.5	4014.2	46.3	-100.0	-2032.3	549.5	616.1	675.5
57-13-6	-3029.7	1600.2	47.6	-111.7	-917.9	73.7	86.3	160.0
72-18-4	-10685.9	5489.6	94.9	-241.7	-1640.9	295.2	547.5	686.7
85721-33-1	-52577.1	26675.9	255.2	-656.3	-4489.6	1839.5	6459.3	8188.2
564-25-0	-91632.5	46343.3	342.7	-859.9	-6123.3	3161.9	9327.4	11807.6
60-00-4	-46117.1	23441.2	217.4	-503.4	-4478.9	2481.2	3062.8	5054.7
85-01-8	-18042.5	9203.2	142.7	-414.7	-1929.2	521.2	1510.4	2031.6
127-18-4	-5406.4	2746.8	28.4	-71.7	-1750.8	482.9	602.0	1085.0
60-54-8	-91347.7	46196.2	341.4	-860.7	-6124.3	3147.9	9616.8	11652.3

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	BW30					
	M fraction		P fraction		R fraction	
	average	st. dev.	average	st. dev.	average	st. dev.
15972-60-8	0.052	0.019	0.004	0.001	0.944	0.020
71-43-2	0.740	0.054	0.260	0.054	0.000	0.000
80-05-7	0.283	0.089	0.031	0.020	0.685	0.105
58-08-2	0.141	0.065	0.179	0.029	0.681	0.093
2921-88-2	0.257	0.135	0.008	0.000	0.735	0.135
57-88-5	0.134	0.013	0.001	0.000	0.865	0.013
51481-61-9	0.134	0.029	0.078	0.033	0.789	0.058
76-57-3	0.131	0.041	0.094	0.047	0.775	0.085
120-83-2	0.926	0.013	0.074	0.013	0.000	0.000
94-75-7	0.060	0.015	0.133	0.071	0.808	0.071
84-66-2	0.370	0.178	0.068	0.040	0.562	0.212
56-53-1	0.373	0.081	0.001	0.001	0.626	0.081
121-14-2	0.949	0.015	0.051	0.015	0.000	0.000
57-91-0	0.776	0.128	0.002	0.001	0.222	0.128
53-16-7	0.696	0.233	0.006	0.002	0.298	0.234
100-41-4	0.964	0.004	0.036	0.004	0.000	0.000
71-00-1	0.062	0.018	0.162	0.010	0.776	0.022
15687-27-1	0.184	0.007	0.162	0.025	0.655	0.028
58-89-9	0.663	0.056	0.024	0.009	0.314	0.059
298-00-0	0.120	0.019	0.010	0.001	0.870	0.019
98-95-3	0.996	0.001	0.004	0.001	0.000	0.000
104-40-5	0.366	0.058	0.003	0.001	0.631	0.059
87-86-5	0.534	0.053	0.004	0.002	0.462	0.053
108-95-2	0.600	0.051	0.354	0.057	0.046	0.047
85-44-9	0.017	0.010	0.068	0.027	0.915	0.036
57-83-0	0.253	0.021	0.000	0.000	0.746	0.022
19466-47-8	0.289	0.094	0.005	0.001	0.706	0.094
58-22-0	0.117	0.034	0.009	0.005	0.874	0.037
108-88-3	0.985	0.000	0.015	0.000	0.000	0.000
56-41-7	0.048	0.007	0.136	0.035	0.815	0.038
70-47-3	0.024	0.007	0.069	0.026	0.907	0.033
56-84-8	0.055	0.026	0.126	0.025	0.819	0.048
52-90-4	0.126	0.013	0.178	0.064	0.696	0.064
79-43-6	0.078	0.011	0.165	0.044	0.757	0.054
124-40-3	0.069	0.018	0.347	0.044	0.584	0.037
56-40-6	0.034	0.008	0.149	0.044	0.817	0.051
56-87-1	0.031	0.008	0.140	0.032	0.829	0.039
63-68-3	0.079	0.027	0.241	0.055	0.679	0.072
62-75-9	0.186	0.022	0.868	0.068	0.000	0.000
75-65-0	0.071	0.018	0.184	0.038	0.745	0.048
72-19-5	0.040	0.006	0.092	0.025	0.867	0.026
76-03-9	0.088	0.012	0.244	0.054	0.668	0.058
57-13-6	0.014	0.001	0.894	0.021	0.092	0.022
72-18-4	0.046	0.014	0.230	0.064	0.725	0.061
85721-33-1	0.027	0.014	0.021	0.020	0.952	0.032
564-25-0	0.105	0.025	0.033	0.008	0.862	0.030
60-00-4	0.030	0.009	0.053	0.015	0.917	0.021
85-01-8	0.997	0.000	0.003	0.000	0.000	0.000
127-18-4	1.000	0.000	0.000	0.000	0.000	0.000
60-54-8	0.077	0.009	0.034	0.015	0.889	0.016

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	ESPA2					
	M fraction		P fraction		R fraction	
	average	st. dev.	average	st. dev.	average	st. dev.
15972-60-8	0.200	0.039	0.023	0.006	0.777	0.039
71-43-2	0.767	0.041	0.233	0.041	0.000	0.000
80-05-7	0.255	0.075	0.019	0.007	0.726	0.081
58-08-2	0.191	0.044	0.206	0.042	0.603	0.059
2921-88-2	0.596	0.111	0.007	0.001	0.397	0.110
57-88-5	0.179	0.034	0.001	0.000	0.821	0.034
51481-61-9	0.341	0.021	0.196	0.064	0.463	0.057
76-57-3	0.477	0.066	0.154	0.052	0.369	0.115
120-83-2	0.826	0.361	0.070	0.106	0.104	0.256
94-75-7	0.173	0.021	0.158	0.063	0.669	0.081
84-66-2	0.315	0.089	0.049	0.016	0.636	0.102
56-53-1	0.217	0.057	0.001	0.001	0.782	0.057
121-14-2	0.965	0.008	0.035	0.008	0.000	0.000
57-91-0	0.859	0.085	0.017	0.006	0.124	0.084
53-16-7	0.998	0.001	0.002	0.001	0.000	0.000
100-41-4	0.968	0.002	0.032	0.002	0.000	0.000
71-00-1	0.079	0.007	0.160	0.025	0.761	0.026
15687-27-1	0.089	0.029	0.045	0.016	0.867	0.041
58-89-9	0.583	0.102	0.021	0.005	0.396	0.100
298-00-0	0.282	0.068	0.015	0.002	0.703	0.068
98-95-3	0.995	0.001	0.005	0.001	0.000	0.000
104-40-5	0.210	0.031	0.003	0.001	0.786	0.032
87-86-5	0.447	0.064	0.029	0.016	0.525	0.068
108-95-2	0.633	0.054	0.304	0.074	0.063	0.069
85-44-9	0.031	0.012	0.080	0.048	0.889	0.056
57-83-0	0.342	0.063	0.003	0.001	0.655	0.064
19466-47-8	0.486	0.043	0.005	0.000	0.509	0.043
58-22-0	0.279	0.101	0.023	0.012	0.697	0.109
108-88-3	0.916	0.009	0.084	0.009	0.000	0.000
56-41-7	0.056	0.024	0.185	0.065	0.758	0.073
70-47-3	0.067	0.010	0.220	0.030	0.713	0.029
56-84-8	0.035	0.006	0.157	0.019	0.808	0.021
52-90-4	0.050	0.009	0.101	0.030	0.850	0.038
79-43-6	0.080	0.013	0.306	0.071	0.613	0.070
124-40-3	0.206	0.123	0.337	0.081	0.456	0.090
56-40-6	0.046	0.016	0.269	0.027	0.685	0.031
56-87-1	0.069	0.021	0.142	0.033	0.789	0.051
63-68-3	0.251	0.098	0.167	0.070	0.582	0.088
62-75-9	0.141	0.018	0.808	0.086	0.052	0.102
75-65-0	0.052	0.012	0.170	0.026	0.779	0.038
72-19-5	0.036	0.005	0.118	0.023	0.846	0.024
76-03-9	0.069	0.023	0.233	0.067	0.698	0.061
57-13-6	0.083	0.137	0.851	0.102	0.066	0.042
72-18-4	0.088	0.024	0.215	0.088	0.697	0.107
85721-33-1	0.187	0.038	0.106	0.057	0.708	0.080
564-25-0	0.145	0.015	0.044	0.020	0.812	0.023
60-00-4	0.091	0.038	0.143	0.053	0.766	0.069
85-01-8	0.853	0.087	0.005	0.001	0.143	0.087
127-18-4	0.997	0.002	0.003	0.002	0.000	0.000
60-54-8	0.189	0.017	0.071	0.026	0.740	0.040

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	LFC1					
	M fraction		P fraction		R fraction	
	average	st. dev.	average	st. dev.	average	st. dev.
15972-60-8	–	–	–	–	–	–
71-43-2	0.640	0.119	0.195	0.056	0.165	0.172
80-05-7	0.161	0.036	0.011	0.005	0.828	0.039
58-08-2	0.218	0.049	0.146	0.041	0.636	0.075
2921-88-2	0.212	0.033	0.011	0.001	0.777	0.033
57-88-5	0.126	0.028	0.003	0.001	0.872	0.027
51481-61-9	0.290	0.027	0.052	0.017	0.658	0.040
76-57-3	0.389	0.028	0.121	0.026	0.491	0.027
120-83-2	0.977	0.010	0.023	0.010	0.000	0.000
94-75-7	0.039	0.004	0.048	0.010	0.913	0.013
84-66-2	0.299	0.037	0.055	0.012	0.646	0.042
56-53-1	0.184	0.062	0.002	0.001	0.814	0.061
121-14-2	0.983	0.003	0.017	0.003	0.000	0.000
57-91-0	0.673	0.065	0.007	0.001	0.320	0.066
53-16-7	0.837	0.046	0.009	0.001	0.154	0.046
100-41-4	0.981	0.001	0.019	0.001	0.000	0.000
71-00-1	0.080	0.014	0.173	0.023	0.747	0.035
15687-27-1	0.086	0.018	0.052	0.022	0.862	0.034
58-89-9	0.373	0.058	0.011	0.003	0.616	0.059
298-00-0	0.239	0.067	0.013	0.002	0.748	0.069
98-95-3	0.997	0.001	0.003	0.001	0.000	0.000
104-40-5	0.234	0.007	0.003	0.000	0.763	0.007
87-86-5	0.607	0.036	0.007	0.001	0.386	0.035
108-95-2	0.653	0.022	0.347	0.022	0.000	0.000
85-44-9	0.031	0.004	0.061	0.013	0.909	0.016
57-83-0	0.232	0.078	0.000	0.000	0.767	0.078
19466-47-8	0.143	0.057	0.005	0.002	0.852	0.059
58-22-0	0.412	0.161	0.017	0.009	0.571	0.167
108-88-3	0.810	0.048	0.190	0.048	0.000	0.000
56-41-7	0.058	0.011	0.155	0.030	0.788	0.035
70-47-3	0.074	0.033	0.120	0.027	0.806	0.051
56-84-8	0.028	0.012	0.097	0.022	0.875	0.031
52-90-4	0.058	0.011	0.155	0.030	0.788	0.035
79-43-6	0.071	0.007	0.234	0.015	0.695	0.019
124-40-3	0.288	0.030	0.313	0.034	0.399	0.031
56-40-6	0.064	0.006	0.228	0.049	0.707	0.046
56-87-1	0.024	0.003	0.062	0.018	0.914	0.018
63-68-3	0.041	0.008	0.103	0.016	0.856	0.019
62-75-9	0.005	0.007	0.842	0.018	0.152	0.022
75-65-0	0.062	0.006	0.259	0.042	0.679	0.047
72-19-5	0.039	0.004	0.120	0.018	0.841	0.021
76-03-9	0.020	0.005	0.128	0.027	0.852	0.029
57-13-6	0.017	0.003	0.955	0.028	0.029	0.026
72-18-4	0.049	0.017	0.122	0.020	0.829	0.032
85721-33-1	0.305	0.039	0.076	0.023	0.619	0.050
564-25-0	0.160	0.014	0.053	0.004	0.787	0.015
60-00-4	0.021	0.002	0.067	0.023	0.912	0.023
85-01-8	0.993	0.002	0.007	0.002	0.000	0.000
127-18-4	0.999	0.000	0.001	0.000	0.000	0.000
60-54-8	0.176	0.038	0.035	0.015	0.789	0.046

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	TFCHR					
	M fraction		P fraction		R fraction	
	average	st. dev.	average	st. dev.	average	st. dev.
15972-60-8	0.212	0.066	0.024	0.008	0.764	0.073
71-43-2	0.786	0.034	0.214	0.034	0.000	0.000
80-05-7	0.240	0.050	0.006	0.002	0.754	0.051
58-08-2	0.174	0.045	0.148	0.035	0.678	0.028
2921-88-2	0.526	0.068	0.007	0.002	0.467	0.069
57-88-5	0.133	0.046	0.004	0.001	0.863	0.046
51481-61-9	0.261	0.040	0.140	0.092	0.600	0.115
76-57-3	0.168	0.058	0.077	0.029	0.756	0.072
120-83-2	0.980	0.006	0.020	0.006	0.000	0.000
94-75-7	0.096	0.030	0.060	0.029	0.844	0.056
84-66-2	0.412	0.123	0.015	0.003	0.574	0.124
56-53-1	0.477	0.106	0.001	0.000	0.521	0.106
121-14-2	0.981	0.005	0.019	0.005	0.000	0.000
57-91-0	0.849	0.127	0.005	0.002	0.146	0.127
53-16-7	0.998	0.000	0.002	0.000	0.000	0.000
100-41-4	0.984	0.002	0.016	0.002	0.000	0.000
71-00-1	0.046	0.012	0.117	0.034	0.838	0.028
15687-27-1	0.104	0.015	0.040	0.011	0.857	0.015
58-89-9	0.663	0.100	0.009	0.002	0.327	0.101
298-00-0	0.256	0.064	0.035	0.047	0.709	0.064
98-95-3	0.996	0.001	0.004	0.001	0.000	0.000
104-40-5	0.699	0.170	0.003	0.001	0.298	0.169
87-86-5	0.687	0.079	0.051	0.018	0.262	0.088
108-95-2	0.647	0.028	0.351	0.030	0.002	0.005
85-44-9	0.034	0.009	0.081	0.022	0.884	0.021
57-83-0	0.339	0.218	0.000	0.000	0.661	0.218
19466-47-8	0.248	0.040	0.004	0.000	0.748	0.040
58-22-0	0.145	0.044	0.005	0.002	0.849	0.045
108-88-3	0.881	0.006	0.119	0.006	0.000	0.000
56-41-7	0.042	0.006	0.102	0.044	0.856	0.047
70-47-3	0.066	0.014	0.319	0.044	0.615	0.050
56-84-8	0.055	0.005	0.147	0.037	0.798	0.034
52-90-4	0.057	0.012	0.070	0.018	0.873	0.024
79-43-6	0.088	0.011	0.258	0.049	0.654	0.055
124-40-3	0.079	0.012	0.286	0.042	0.635	0.037
56-40-6	0.054	0.005	0.180	0.039	0.766	0.040
56-87-1	0.038	0.010	0.107	0.028	0.855	0.029
63-68-3	0.065	0.007	0.200	0.056	0.735	0.057
62-75-9	0.213	0.021	0.787	0.021	0.000	0.000
75-65-0	0.101	0.016	0.239	0.033	0.660	0.039
72-19-5	0.049	0.010	0.106	0.010	0.845	0.018
76-03-9	0.065	0.025	0.290	0.087	0.645	0.098
57-13-6	0.017	0.001	0.900	0.018	0.082	0.019
72-18-4	0.051	0.011	0.111	0.050	0.838	0.059
85721-33-1	0.121	0.045	0.066	0.026	0.813	0.068
564-25-0	0.167	0.047	0.102	0.078	0.730	0.080
60-00-4	0.020	0.008	0.111	0.037	0.869	0.041
85-01-8	0.983	0.021	0.005	0.000	0.013	0.021
127-18-4	1.000	0.000	0.000	0.000	0.000	0.000
60-54-8	0.145	0.043	0.029	0.009	0.826	0.051

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	CA					
	M fraction		P fraction		R fraction	
	average	st. dev.	average	st. dev.	average	st. dev.
15972-60-8	–	–	–	–	–	–
71-43-2	0.434	0.032	0.566	0.032	0.000	0.000
80-05-7	0.991	0.001	0.009	0.001	0.000	0.000
58-08-2	0.101	0.006	0.755	0.035	0.144	0.032
2921-88-2	0.971	0.005	0.029	0.005	0.000	0.000
57-88-5	0.165	0.149	0.003	0.001	0.833	0.148
51481-61-9	0.217	0.024	0.599	0.068	0.184	0.064
76-57-3	0.262	0.030	0.573	0.050	0.165	0.070
120-83-2	0.976	0.004	0.024	0.004	0.000	0.000
94-75-7	0.053	0.011	0.437	0.099	0.510	0.103
84-66-2	0.835	0.006	0.165	0.006	0.000	0.000
56-53-1	0.997	0.001	0.003	0.001	0.000	0.000
121-14-2	0.929	0.008	0.071	0.008	0.000	0.000
57-91-0	0.975	0.005	0.025	0.005	0.000	0.000
53-16-7	0.973	0.003	0.027	0.003	0.000	0.000
100-41-4	0.666	0.075	0.256	0.057	0.078	0.099
71-00-1	0.088	0.009	0.453	0.035	0.459	0.032
15687-27-1	0.205	0.022	0.580	0.056	0.216	0.077
58-89-9	0.985	0.004	0.015	0.004	0.000	0.000
298-00-0	0.978	0.002	0.022	0.002	0.000	0.000
98-95-3	0.655	0.030	0.345	0.030	0.000	0.000
104-40-5	0.960	0.036	0.007	0.001	0.033	0.037
87-86-5	0.978	0.003	0.022	0.003	0.000	0.000
108-95-2	0.283	0.018	0.717	0.018	0.000	0.000
85-44-9	0.062	0.012	0.291	0.085	0.647	0.088
57-83-0	0.985	0.002	0.015	0.002	0.000	0.000
19466-47-8	0.282	0.090	0.007	0.001	0.711	0.090
58-22-0	0.744	0.059	0.210	0.043	0.046	0.055
108-88-3	0.522	0.032	0.478	0.032	0.000	0.000
56-41-7	0.068	0.014	0.539	0.082	0.393	0.078
70-47-3	0.007	0.002	0.650	0.045	0.343	0.044
56-84-8	0.085	0.023	0.342	0.045	0.573	0.049
52-90-4	0.094	0.024	0.439	0.063	0.468	0.049
79-43-6	0.062	0.010	0.413	0.041	0.525	0.048
124-40-3	0.133	0.013	0.549	0.039	0.318	0.029
56-40-6	0.067	0.017	0.564	0.044	0.369	0.037
56-87-1	0.093	0.010	0.518	0.042	0.389	0.042
63-68-3	0.086	0.020	0.469	0.044	0.445	0.049
62-75-9	0.035	0.001	0.941	0.046	0.024	0.047
75-65-0	0.040	0.004	0.874	0.045	0.085	0.044
72-19-5	0.075	0.025	0.457	0.040	0.467	0.020
76-03-9	0.042	0.004	0.601	0.050	0.357	0.053
57-13-6	0.031	0.004	0.906	0.025	0.063	0.024
72-18-4	0.056	0.017	0.627	0.061	0.316	0.067
85721-33-1	0.270	0.053	0.351	0.115	0.378	0.077
564-25-0	0.306	0.056	0.180	0.038	0.514	0.069
60-00-4	0.075	0.019	0.481	0.053	0.444	0.065
85-01-8	0.996	0.001	0.004	0.001	0.000	0.000
127-18-4	0.678	0.055	0.308	0.052	0.014	0.020
60-54-8	0.144	0.028	0.323	0.098	0.533	0.081

ISBN:978-84-691-2701-8/DL:T.386-2008

CAS – chemical abstracts service; **a_count** – atom count (all atoms); **b_count** – bond count (all bonds); **b_single** – bond count (single bonds); **b_double** – bond count (double bonds); **chi0** – connectivity index (order 0, standard); **chi1** – connectivity index (order 1, standard); **chi2** – connectivity index (order 2, standard); **dipole** – dipole moment [Debye]; **dipoleX** – dipole vector X [Debye]; **dipoleY** – dipole vector Y [Debye]; **dipoleZ** – dipole vector Z [Debye]; **dielEn** – dielectric energy [kcal/mole]; **sterEn** – steric energy [kcal/mole]; **heatform** – heat of formation [kcal/mole]; **mr** – molar refractivity; **weight** – molecular weight [Da]; **Po** – polarizability [$Å^3$]; **r_count** – ring count (all rings); **small_ring** – size of smallest ring; **large_ring** – size of largest ring; **kier1** – shape index (basic kappa, order 1); **kier2** – shape index (basic kappa, order 2); **kier3** – shape index (basic kappa, order 3); **SASA** – solvent accessibility surface area [$Å^2$]; **chi0v** – valence connectivity index (order 0, standard); **chi1v** – valence connectivity index (order 1, standard); **chi2v** – valence connectivity index (order 2, standard); **HOMO** – HOMO energy [eV]; **LUMO** – LUMO energy [eV]; **dipole_P** – dipole point-charge [Debye]; **dipole_H** – dipole hybridization [Debye]; **E1_e-n** – one term electron-nuclear [eV]; **E1_e-e** – one term electron-electron [eV]; **E1_total** – one term total [eV]; **E2_res** – two center resonance [eV]; **E2_ex** – two center exchange [eV]; **E2_e-e** – two center electron-electron [eV]; **E2_e-n** – two center electron-nuclear [eV]; **E2_n-n** – two center nuclear-nuclear [eV]; **E2_el** – two center total electrostatic [eV]; **E2_total** – two center total [eV]; **E_total** – total energy [eV]; **pmiX** – moments of inertia A [10^{-40} g cm^2]; **pmiY** – moments of inertia B [10^{-40} g cm^2]; **pmiZ** – moments of inertia C [10^{-40} g cm^2]; **M fraction** – solute fraction sorbed by the membrane; **P fraction** – permeating fraction of the solute; **R fraction** – rejected solute fraction; **average** – average of several replicate measurements; **st. dev.** – experimental standard deviation of several replicate measurements for a membrane-solute combination.

ISBN:978-84-691-2701-8/DL:T.386-2008

ANNEX III. Internal validation performance of the ANN-based QSPRs built.

For each model are presented the explained variance in prediction index (q^2), the average absolute error ($\overline{\varepsilon}$), the standard deviation of the absolute error (σ_ε) and the maximum absolute error ($\varepsilon_{\max}$). In each case, in parenthesis are indicated the corresponding values for the relative error (average relative error, standard deviation of the relative error and maximum relative error). The presented external validation results refer to the performance achieved for the test set compounds.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Internal validation (LOO) for the Independent ANQ models built with parameters selected by the CFS method.

CFS	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.992	0.018	0.030	0.143	0.993	0.012	0.020	0.098	0.978	0.020	0.030	0.119
		(12.4%)	(40.4%)	(270.2%)		(3.3%)	(4.4%)	(18.0%)		(11.1%)	(19.0%)	(66.8%)
BW30	0.999	0.006	0.009	0.042	0.996	0.005	0.010	0.059	0.998	0.007	0.011	0.060
		(5.1%)	(11.8%)	(73.1%)		(5.2%)	(8.5%)	(39.8%)		(1.2%)	(1.6%)	(7.3%)
ESPA2	0.989	0.013	0.031	0.213	0.998	0.004	0.006	0.036	0.987	0.015	0.031	0.215
		(6.4%)	(9.4%)	(40.4%)		(2.1%)	(2.0%)	(8.3%)		(3.2%)	(5.8%)	(28.7%)
LFC1	0.998	0.010	0.013	0.055	0.995	0.007	0.012	0.055	0.996	0.012	0.018	0.110
		(5.1%)	(8.5%)	(38.2%)		(7.7%)	(17.3%)	(100.0%)		(3.9%)	(11.6%)	(72.2%)
TFCHR	0.997	0.010	0.018	0.082	0.998	0.004	0.006	0.028	0.996	0.012	0.018	0.081
		(6.3%)	(15.0%)	(74.7%)		(3.3%)	(3.6%)	(11.6%)		(2.4%)	(3.0%)	(9.4%)

Internal validation (LOO) for the Independent ANQ and Membrane-Composite ANQ models built with parameters selected by the SOM-DA method.

SOM-DA	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.996	0.014	0.021	0.124	0.981	0.014	0.028	0.137	0.978	0.020	0.030	0.126
		(8.2%)	(13.5%)	(73.7%)		(3.0%)	(3.9%)	(16.7%)		(11.6%)	(18.5%)	(69.2%)
BW30	0.998	0.008	0.014	0.078	0.992	0.008	0.014	0.084	0.997	0.010	0.016	0.078
		(4.2%)	(5.3%)	(24.3%)		(7.0%)	(12.8%)	(68.7%)		(1.8%)	(4.2%)	(26.2%)
ESPA2	0.998	0.008	0.011	0.064	0.995	0.005	0.019	0.133	0.998	0.008	0.011	0.061
		(3.7%)	(5.9%)	(25.2%)		(1.9%)	(3.9%)	(16.4%)		(2.4%)	(6.7%)	(42.7%)
LFC1	0.992	0.015	0.025	0.117	0.990	0.006	0.017	0.113	0.990	0.018	0.029	0.121
		(10.0%)	(21.7%)	(115.3%)		(5.7%)	(23.6%)	(139.2%)		(4.9%)	(15.3%)	(73.8%)
TFCHR	0.999	0.006	0.007	0.028	0.976	0.007	0.027	0.186	0.994	0.010	0.025	0.174
		(3.1%)	(3.4%)	(12.0%)		(3.2%)	(4.8%)	(23.6%)		(1.2%)	(1.3%)	(5.4%)
PA	0.964	0.044	0.046	0.212	0.953	0.024	0.030	0.230	0.952	0.052	0.050	0.212
		(26.0%)	(29.0%)	(149.6%)		(24.4%)	(19.9%)	(111.9%)		(14.5%)	(33.2%)	(370.2%)
PACA	0.939	0.060	0.061	0.322	0.933	0.038	0.043	0.240	0.928	0.067	0.061	0.288
		(34.9%)	(45.6%)	(274.9%)		(33.5%)	(39.0%)	(306.9%)		(20.1%)	(36.7%)	(295.7%)

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Internal validation (LOO) for the Independent ANQ models built with parameters selected by the ANNIGMA method.

ANNIGMA	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.991	0.014	0.035	0.217	0.994	0.009	0.020	0.124	0.975	0.014	0.035	0.216
		(7.4%)	(17.9%)	(105.8%)		(2.1%)	(3.8%)	(19.2%)		(9.8%)	(20.6%)	(100.0%)
BW30	0.988	0.013	0.035	0.230	0.996	0.006	0.010	0.049	0.986	0.017	0.036	0.230
		(4.2%)	(5.9%)	(27.9%)		(6.2%)	(8.3%)	(30.4%)		(1.8%)	(2.4%)	(12.0%)
ESPA2	0.993	0.012	0.025	0.159	0.997	0.005	0.009	0.047	0.991	0.014	0.025	0.161
		(5.2%)	(12.3%)	(75.8%)		(2.5%)	(3.6%)	(17.9%)		(4.5%)	(11.6%)	(60.0%)
LFC1	0.995	0.009	0.021	0.140	0.998	0.004	0.007	0.035	0.995	0.011	0.021	0.127
		(3.8%)	(6.3%)	(28.9%)		(7.0%)	(15.8%)	(68.0%)		(2.5%)	(6.9%)	(43.0%)
TFCHR	0.993	0.012	0.027	0.140	0.997	0.004	0.008	0.043	0.992	0.015	0.027	0.140
		(6.6%)	(14.8%)	(83.7%)		(3.8%)	(4.6%)	(17.9%)		(1.8%)	(3.0%)	(13.6%)

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Internal validation (LOO) for the MP-Composite ANQ models built with parameters selected by the CFS method.

CFS	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.977	0.033	0.049	0.238	0.947	0.041	0.048	0.180	0.888	0.051	0.062	0.236
		(26.7%)	(54.3%)	(331.5%)		(10.7%)	(10.1%)	(37.0%)		(30.9%)	(43.3%)	(141.1%)
BW30	0.985	0.025	0.032	0.151	0.968	0.018	0.027	0.159	0.975	0.029	0.045	0.256
		(14.8%)	(19.4%)	(81.2%)		(16.3%)	(22.5%)	(100.0%)		(2.9%)	(3.1%)	(12.5%)
ESPA2	0.971	0.035	0.042	0.236	0.976	0.020	0.018	0.082	0.956	0.040	0.051	0.275
		(21.1%)	(39.6%)	(219.4%)		(17.5%)	(24.1%)	(122.8%)		(11.4%)	(31.2%)	(203.1%)
LFC1	0.971	0.033	0.045	0.250	0.948	0.027	0.033	0.134	0.969	0.040	0.042	0.186
		(24.1%)	(79.5%)	(522.7%)		(31.5%)	(49.2%)	(263.8%)		(9.1%)	(15.1%)	(83.8%)
TFCHR	0.954	0.035	0.067	0.413	0.842	0.029	0.064	0.438	0.967	0.033	0.051	0.312
		(13.6%)	(21.4%)	(118.7%)		(17.5%)	(18.9%)	(72.9%)		(7.3%)	(16.8%)	(100.0%)

Internal validation (LOO) for the MP-Composite ANQ models built with parameters selected by the ANNIGMA method.

ANNIGMA	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.987	0.030	0.032	0.126	0.941	0.045	0.050	0.220	0.889	0.054	0.059	0.228
		(28.2%)	(54.5%)	(237.0%)		(12.3%)	(11.3%)	(36.4%)		(34.7%)	(50.5%)	(231.2%)
BW30	0.961	0.030	0.059	0.339	0.940	0.022	0.038	0.200	0.981	0.028	0.038	0.209
		(13.2%)	(19.2%)	(92.5%)		(13.7%)	(15.4%)	(72.9%)		(5.4%)	(15.6%)	(100.0%)
ESPA2	0.959	0.037	0.054	0.294	0.932	0.027	0.036	0.198	0.933	0.052	0.061	0.307
		(21.4%)	(39.7%)	(229.6%)		(23.9%)	(31.0%)	(132.3%)		(12.9%)	(30.6%)	(176.2%)
LFC1	0.986	0.025	0.030	0.166	0.958	0.024	0.029	0.161	0.982	0.032	0.033	0.112
		(27.2%)	(82.4%)	(534.0%)		(32.8%)	(37.3%)	(151.8%)		(7.4%)	(14.7%)	(73.4%)
TFCHR	0.980	0.027	0.041	0.196	0.957	0.023	0.029	0.145	0.982	0.029	0.035	0.176
		(15.1%)	(21.3%)	(92.0%)		(28.6%)	(44.2%)	(232.0%)		(4.5%)	(4.1%)	(18.6%)

ISBN:978-84-691-2701-8/DL:T.386-2008

ANNEX IV. External validation performance of the ANN-based QSPRs built.

For each model are presented the explained variance in prediction index (q^2), the average absolute error ($\overline{\varepsilon}$), the standard deviation of the absolute error (σ_ε) and the maximum absolute error ($\varepsilon_{\max}$). In each case, in parenthesis are indicated the corresponding values for the relative error (average relative error, standard deviation of the relative error and maximum relative error). The presented external validation results refer to the performance achieved for the test set compounds.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

External validation for the Independent ANQ models built with parameters selected by the CFS method.

CFS	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.828	0.112	0.135	0.362	0.982	0.041	0.025	0.069	0.331	0.101	0.113	0.394
		(44.3%)	(67.3%)	(176.8%)		(10.1%)	(4.5%)	(13.4%)		(33.9%)	(29.6%)	(95.7%)
BW30	0.981	0.034	0.040	0.113	0.932	0.024	0.015	0.039	0.959	0.048	0.040	0.108
		(17.6%)	(14.1%)	(42.3%)		(42.6%)	(49.8%)	(100.0%)		(7.5%)	(4.7%)	(13.4%)
ESPA2	0.905	0.057	0.071	0.190	0.969	0.020	0.008	0.030	0.939	0.046	0.046	0.170
		(26.1%)	(29.5%)	(75.6%)		(17.0%)	(14.3%)	(30.9%)		(10.3%)	(9.7%)	(29.3%)
LFC1	0.987	0.031	0.018	0.054	0.823	0.033	0.027	0.070	0.988	0.034	0.021	0.073
		(28.3%)	(25.1%)	(63.3%)		(31.6%)	(10.0%)	(40.7%)		(8.1%)	(9.2%)	(32.0%)
TFCHR	0.995	0.015	0.017	0.046	0.926	0.025	0.015	0.045	0.981	0.027	0.035	0.117
		(8.7%)	(13.8%)	(36.5%)		(38.5%)	(25.2%)	(56.4%)		(5.9%)	(8.0%)	(23.4%)

External validation for the Independent ANQ and Membrane-Composite ANQ models built with parameters selected by the SOM-DA method.

SOM-DA	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.999	0.012	0.008	0.024	0.963	0.043	0.030	0.072	0.904	0.061	0.056	0.203
		(8.5%)	(10.6%)	(27.0%)		(10.4%)	(5.3%)	(14.6%)		(30.1%)	(33.0%)	(100.0%)
BW30	0.929	0.066	0.064	0.171	0.922	0.018	0.021	0.053	0.982	0.044	0.054	0.171
		(70.9%)	(88.2%)	(225.4%)		(44.5%)	(70.2%)	(148.5%)		(7.8%)	(9.4%)	(30.4%)
ESPA2	0.988	0.034	0.019	0.068	0.839	0.036	0.028	0.074	0.934	0.060	0.070	0.230
		(28.5%)	(19.5%)	(51.2%)		(43.4%)	(31.9%)	(88.4%)		(12.6%)	(10.1%)	(28.0%)
LFC1	0.976	0.041	0.044	0.118	0.961	0.019	0.012	0.039	0.972	0.037	0.051	0.153
		(23.5%)	(28.6%)	(81.3%)		(8.6%)	(3.5%)	(12.6%)		(5.0%)	(6.4%)	(20.0%)
TFCHR	0.997	0.017	0.012	0.032	0.943	0.021	0.003	0.024	0.992	0.025	0.023	0.072
		(20.2%)	(27.2%)	(66.6%)		(15.9%)	(3.3%)	(19.5%)		(3.7%)	(2.8%)	(8.4%)
PA	0.940	0.063	0.046	0.184	0.904	0.027	0.025	0.106	0.942	0.069	0.060	0.192
		(43.3%)	(37.0%)	(93.8%)		(52.4%)	(41.5%)	(100.0%)		(12.9%)	(15.2%)	(54.5%)
PACA	0.873	0.088	0.081	0.345	0.793	0.052	0.057	0.226	0.936	0.065	0.062	0.271
		(43.4%)	(51.6%)	(238.4%)		(45.2%)	(39.5%)	(135.1%)		(13.4%)	(15.5%)	(79.0%)

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

External validation for the Independent ANQ models built with parameters selected by the ANNIGMA method.

ANNIGMA	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.993	0.025	0.019	0.057	0.885	0.060	0.039	0.127	0.889	0.045	0.050	0.155
		(15.6%)	(18.4%)	(45.0%)		(18.7%)	(8.5%)	(30.7%)		(35.9%)	(54.0%)	(144.8%)
BW30	0.966	0.053	0.035	0.093	0.987	0.013	0.007	0.022	0.942	0.059	0.039	0.132
		(16.5%)	(12.1%)	(29.1%)		(8.8%)	(4.9%)	(14.8%)		(7.1%)	(5.2%)	(17.7%)
ESPA2	0.972	0.041	0.024	0.076	0.972	0.017	0.009	0.027	0.954	0.049	0.021	0.085
		(24.5%)	(15.1%)	(44.1%)		(17.9%)	(17.4%)	(42.1%)		(16.5%)	(26.6%)	(81.9%)
LFC1	0.944	0.060	0.055	0.138	0.968	0.020	0.009	0.036	0.941	0.049	0.046	0.144
		(29.8%)	(34.0%)	(86.0%)		(45.0%)	(73.3%)	(154.7%)		(10.7%)	(9.1%)	(24.4%)
TFCHR	0.927	0.077	0.063	0.176	0.874	0.030	0.019	0.065	0.946	0.057	0.056	0.176
		(55.1%)	(48.3%)	(125.7%)		(28.3%)	(25.1%)	(60.3%)		(10.1%)	(9.1%)	(26.7%)

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

External validation for the MP-Composite ANQ models built with parameters selected by the CFS method.

CFS	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.989	0.041	0.029	0.088	0.963	0.057	0.038	0.105	0.925	0.042	0.046	0.095
		(16.3%)	(23.7%)	(64.2%)		(9.2%)	(8.2%)	(16.8%)		(86.6%)	(91.6%)	(151.3%)
BW30	0.916	0.076	0.053	0.157	0.919	0.018	0.017	0.033	0.880	0.085	0.052	0.149
		(51.0%)	(31.2%)	(85.7%)		(18.2%)	(14.5%)	(43.2%)		(12.5%)	(9.7%)	(25.0%)
ESPA2	0.747	0.134	0.074	0.267	0.815	0.032	0.021	0.061	0.692	0.145	0.086	0.287
		(109.8%)	(70.8%)	(177.0%)		(21.9%)	(14.5%)	(35.1%)		(44.6%)	(67.8%)	(181.4%)
LFC1	0.875	0.109	0.051	0.175	0.908	0.022	0.018	0.054	0.831	0.113	0.066	0.188
		(77.9%)	(65.1%)	(161.2%)		(24.9%)	(18.6%)	(56.6%)		(13.5%)	(10.0%)	(27.4%)
TFCHR	0.837	0.117	0.082	0.246	0.827	0.031	0.024	0.076	0.797	0.120	0.091	0.280
		(85.8%)	(86.9%)	(184.7%)		(21.3%)	(9.3%)	(31.9%)		(13.9%)	(10.7%)	(32.4%)

External validation for the MP-Composite ANQ models built with parameters selected by the ANNIGMA method.

ANNIGMA	**Predicted M fraction**				**Predicted P fraction**				**Calculated R fraction**			
	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}	q^2	$\overline{\varepsilon}$	σ_ε	ε_{max}
CA	0.980	0.048	0.037	0.099	0.908	0.080	0.068	0.209	0.843	0.052	0.048	0.122
		(65.1%)	(107.3%)	(279.7%)		(15.2%)	(6.4%)	(23.1%)		(58.6%)	(90.7%)	(194.0%)
BW30	0.967	0.052	0.042	0.119	0.764	0.039	0.019	0.062	0.976	0.035	0.031	0.081
		(36.9%)	(56.1%)	(135.4%)		(53.8%)	(71.1%)	(197.5%)		(4.9%)	(4.7%)	(12.1%)
ESPA2	0.956	0.044	0.042	0.114	0.962	0.018	0.012	0.032	0.954	0.039	0.028	0.081
		(22.3%)	(22.9%)	(61.3%)		(9.8%)	(6.4%)	(14.0%)		(14.2%)	(20.7%)	(55.1%)
LFC1	0.840	0.081	0.071	0.192	0.855	0.027	0.025	0.072	0.843	0.090	0.068	0.178
		(39.8%)	(21.6%)	(68.7%)		(34.3%)	(38.5%)	(83.5%)		(16.3%)	(10.6%)	(31.1%)
TFCHR	0.813	0.109	0.131	0.361	0.871	0.026	0.019	0.047	0.757	0.116	0.129	0.362
		(99.3%)	(144.2%)	(347.5%)		(22.4%)	(19.0%)	(46.3%)		(23.0%)	(23.8%)	(54.4%)

ISBN:978-84-691-2701-8/DL:T.386-2008

ANNEX V. List of 143 new organic compounds used for validating the QSPR models

Organic compounds of public health concern including endocrine disruptors, pharmaceutically active compounds, antibiotics and antimicrobial agents, neuroactive drugs, insecticides, herbicides, pesticides, disinfect byproducts, solvents, industrial pollutants, fuel hydrocarbons and amino acids, with the identification of application and/or effects. These compounds were not experimentally characterized in terms of membrane sorption, passage or rejection, nor used for models development.

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Name	Compounds class, known use and/or toxicity endpoint
100-75-4	1-Nitrosopiperidine	Carcinogen
1031-07-8	6,7,8,9,10,10-Hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepin 3,3-dioxide (Endosulfan sulfate)	Pesticide
103-23-1	Adipic acid bis (2-ethylhexyl) ester	Plasticizer-Industrial/household waste water product
103-90-2	4-Acetamidophenol	Pharmaceutical-Analgesic-Human drug
106-44-5	1-Hydroxy-4-methylbenzene	Wood preservative-Industrial/household waste water product
106-46-7	1,4-Dichlorobenzene	Fumigant-Carcinogen-Industrial/household waste water product
108-67-8	1,3,5-Trimethyl benzene	Fuel hydrocarbon
108-86-1	Bromobenzene	Solvent
1141-38-4	2,6-Naphthalenedicarboxylic acid	Manufacture polyethylenenaphthalate and polyethylenepthalate polymers
115-29-7	1,2,3,4,7,7-Hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepin-3-oxide (Thiosulfan)	Endocrine disruptor
115-32-2	1,1-Bis-(p-chlorophenyl)-2,2,2-trichloroethanol (Dicofol)	Endocrine disruptor
115-86-6	Triphenyl phosphate (TPP)	Plasticizer-Industrial/household waste water product
115-96-8	Tris(2-chloroethyl)phosphate (TCEP)	Plasticizer-Industrial/household waste water product
117-81-7	1,2-Benzenedicarboxylic acid bis(2-ethylhexyl) ester	Carcinogen
117-84-0	1,2-Benzenedicarboxylic acid, dioctyl ester	Plasticizer
118-74-1	Hexachlorobenzene	Endocrine disruptor
120-12-7	Anthracene	Polycyclic aromatic hydrocarbon
121-82-4	1,3,5-Triaza-1,3,5-trinitrocyclohexane	Carcinogen
122-11-2	6-Sulfanilamido-2,4-dimethoxypyrimidine (Sulfadimethoxine)	Pharmaceutical human/veterinary antibiotic
122-34-9	1-Chloro-3,5-bisethylamino-2,4,6-triazine (Simazine)	Carcinogen
124-48-1	Dibromochloromethane	Disinfection byproduct
127-79-7	Sulfamerazine	Pharmaceutical human/veterinary antibiotic
12789-03-6	1,2,4,5,6,7,10,10-octachloro-4,7,8,9-tetrahydro-4,7-methyleneindane (Chlordane)	Endocrine disruptor
128-37-0	2,6-bis(1,1-Dimethylethyl)-4-methylphenol	Antioxidant/antiskimming agent
128-39-2	2,6-bis(1,1-Dimethylethyl)phenol	Intermediate for preparation of antioxidants and UV stabilizer
129-00-0	Pyrene	Polycyclic aromatic hydrocarbon
13071-79-9	O,O-diethyl S-(((1,1-dimethylethyl)thio)methyl) phosphorodithoic acid (Terbufos)	Insecticide
134-62-3	N,N-diethyl-3-methylbenzamide	Insecticide

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Name	Compounds class, known use and/or toxicity endpoint
136-85-6	5-Methyl-1H-benzotriazole	Antioxidant-Industrial/household waste water product
139-13-9	Nitrilo-2,2',2''-triacetic acid	Carcinogen
1401-69-0	Tylosin	Pharmaceutical human/veterinary antibiotic
14345-90-8	Cylindrospermopsin	Algal toxin
144-82-1	2-(p-Aminobenzenesulfonamido)-5-methyl-1,3,4-thiadiazole (Sulfamethizole)	Pharmaceutical human/veterinary antibiotic
154-21-2	Lincomycin	Pharmaceutical human/veterinary antibiotic
1610-18-0	2-Methoxy-4,6-bis(isopropylamino)-1,3,5-triazine (Pramitol)	Herbicide
1634-04-4	Methyl tert-butyl ether (MTBE)	Fuel hydrocarbon-Carcinogen
1646-88-4	2-methyl-2-(methylsulfonyl)propanal O-((methylamino)carbonyl)oxime (Aldoxycarb)	Agricultural product residue
16655-82-6	2,3-Dihydro-2,2-dimethyl-3,7-benzofurandiol, 7-(methylcarbamate)	Pesticide
1672-46-4	Digoxigenin	Pharmaceutical human drug
16752-77-5	Acetamidic acid, thio-, N-[(methyl-carbamoyl)oxy]-, methyl ester (Methomyl)	Endocrine disruptor
1836-75-5	2,4-Dichloro-1-(4-nitrophenoxy)benzene (Nitrofen)	Endocrine disruptor
18559-94-9	(alpha1-((tert-Butylamino)methyl)-4-hydroxy-m-xylene-alpha,alpha-diol) (Salbutamol)	Pharmaceutical human drug
1912-24-9	1-Chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine (Atrazine)	Endocrine disruptor
206-44-0	1,2-(1,8-Naphthalenediyl)benzene (Fluoranthene)	Polycyclic aromatic hydrocarbon
20830-75-5	(3beta,5beta,12beta)-3-[(O-2,6-dideoxy-beta-D-ribo-hexopyranosyl-(1->4)-O-2,6-dideoxy-beta-D-ribo-hexapyranosyl-(1->4)-2,6-dideoxy-beta-D-ribo-hexopyranosyl)oxy]-12,14-dihydroxycard-20(22)-enolide (Digoxin)	Pharmaceutical human drug
21087-64-9	4-Amino-6-(1,1-dimethylethyl)-3-(methylthio)-1,2,4-triazin-5(4H)-one (Metribuzin)	Endocrine disruptor
2136-79-0	2,3,5,6-Tetrachloro-1,4-benzenedicarboxylic acid (Chlorthal)	Herbicide
2169-87-1	2,3-Naphthalenedicarboxylic acid	Plasticizer
2212-67-1	1H-Azepine-1 carbothioic acid, hexahydro-S-ethyl ester (Molinate)	Herbicide
2385-85-5	1,2,3,4,5,5-Hexachloro-1,3-cyclopentadiene dimer (Mirex)	Endocrine disruptor
25013-16-5	2(3)-tert-Butyl-4-hydroxyanisole	Antioxidant-Industrial/household waste water product
25812-30-0	2,2-Dimethyl-5-(2,5-xylyloxy)valeric acid (Gemfibrozil)	Pharmaceutical human drug
26638-19-7	Dichloropropane	Chemical intermediate of perchloroethylene and other chlorinated chemicals
27304-13-8	2,3,4,5,6,6a,7,7-Octachloro-1a,1b,5,5a,6,6a-hexahydro-2,5-methano-2H-indeno[1,2-b]oxirene, (1aalpha,1bbeta 2alpha,5alpha,5abeta,6beta,6aalpha) (Oxychlordane)	Endocrine disruptor
298-04-4	O,O-diethyl S-(2-(ethylthio)ethyl) phosphorodithioate (Disulfoton)	Insecticide

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Name	Compounds class, known use and/or toxicity endpoint
3018-12-0	Dichloroacetonitrile	Disinfection byproduct
302-17-0	1,1,1-Trichloro-2,2-ethanediol	Disinfection byproduct
309-00-2	1,2,3,4,10,10-Hexachloro-1,4,4a,5,8,8a-hexahydro-endo-1,4-exo-5,8-dimethanonaphthalene (Aldrin)	Insecticide
3252-43-5	Dibromoacetonitrile	Disinfection byproduct
330-54-1	1,1-Dimethyl-3-(3,4-dichlorophenyl)urea (Diuron)	Herbicide
330-55-2	1-Methoxy-1-methyl-3-(3,4-dichlorophenyl)urea (Linuron)	Herbicide
333-41-5	O,O-Diethyl O-(2-isopropyl-4-methyl-6-pyrimidinyl) thiophosphoric acid (Diazinon)	Insecticide
3380-34-5	2,4,4'-Trichloro-2-hydroxydiphenyl ether (Triclosan)	Antimicrobial-Industrial/household waste water product
34256-82-1	2-Chloro-2'-methyl-6'-ethyl-N-ethoxymethyl-acetanilide (Acetochlor)	Herbicide
35523-89-8	Saxitoxin	Algal toxin
42399-41-7	Diltiazem	Pharmaceutical human drug
474-86-2	1,3,5(10),7-Estratetraen-3-ol-17-one (Equilin)	Pharmaceutical-Sex/steroid hormone
486-56-6	(S)-1-Methyl-5-(3-pyridinyl)-2-pyrrolidinone (Cotinine)	Nicotine metabolite
50-27-1	1,3,5(10)-Estratriene-3,16a,17b-triol (Estriol)	Pharmaceutical-Sex/steroid hormone
50-29-3	1,1,1-Trichloro-2,2-bis(p-chlorophenyl)ethane (DDT)	Endocrine disruptor
50-32-8	3,4-Benzopyrene	Polycyclic aromatic hydrocarbon
5103-71-9	(1alpha,2alpha,3aalpha,4beta,7beta,7aalpha)-1,2,4,5,6,7,8,8-Octachloro-2,3,3a,4,7,7a-hexahydro-4,7-methano-1H-indene (cis-Chlordane)	Insecticide
51218-45-2	2-Chloro-6'-ethyl-N-(2-methoxy-1-methylethyl)-o-acetoluidide (Metolachlor)	Pesticide
51-28-5	2,4-Dinitrophenol	Released from mines, metals, petroleum and dye plants
513-88-2	1,1-Dichloroacetone	Disinfection byproduct
517-04-4	beta-Estradiol	Pharmaceutical-Estrogen-Sex/Steroid hormone
517-09-9	1,3,5-10,6,8-Estrapentaen-3-ol-17-one (Equilenin)	Pharmaceutical-Sex/steroid hormone
53-41-8	3alpha-Hydroxy-17-androstanone (Androsterone)	Pharmaceutical-Sex/steroid hormone
54910-89-3	Fluoxetine	Pharmaceutical-Human drug
55-18-5	N-nitrosodiethylamine	Carcinogen
5589-96-8	Bromochloroacetic acid	Disinfection byproduct
56-45-1	2-Amino-3-hydroxypropionic acid (Serine)	Amino acid
57-62-5	7-Chloro-4-(dimethylamino)-1,4,4a,5,5a,6,11,12a-octahydro-3,6,10,12,12a-pentahydroxy-6-methyl-1,11-dioxo-2-naphthacenecarboxamide (Chlorotetracycline)	Pharmaceutical human/veterinary antibiotic
57-68-1	2-(p-Aminobenzenesulfonamido)-4,6-dimethylpyrimidine (Sulfamethazine)	Pharmaceutical human/veterinary antibiotic

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Name	Compounds class, known use and/or toxicity endpoint
5902-51-2	3-tert-Butyl-5-chloro-6-methyluracil (Terbacil)	Herbicide
59-89-2	4-Nitrosomorpholine	Carcinogen
60-57-1	1,2,3,4,10,10-Hexachloro-6,7-epoxy-1,4,4a,5,6,7,8,8a-octahydro-1,4-endo-exo-5,8-dimethanonaphthalene (Dieldrin)	Insecticide-Industrial/household waste water product
606-20-2	1,3-Dinitro 2-methyl benzene	Production of polyurethane foams, ammunition and explosives
608-73-1	1,2,3,4,5,6-Hexachlorocyclohexane	Carcinogen
611-59-6	3,7-Dihydro-1,7-dimethyl-1H-purine-2,6-dione (Paraxanthine)	Caffeine metabolite
61-82-5	Triazol-3-amine (Diurol)	Endocrine disruptor
61869-08-7	Paroxetine	Pharmaceutical human drug
61-90-5	2-Amino-4-methylvaleric acid (Leucine)	Amino acid
621-64-7	N-nitroso di-n-propylamine	Carcinogen
631-64-1	Dibromoacetic acid	Disinfection byproduct
63-25-2	1-Naphthalenol methylcarbamate	Endocrine disruptor
637-92-3	2-Ethoxy-2-methylpropane (ETBE)	Fuel oxygenate-Carcinogen
63-91-2	2-Amino-3-phenylpropanoic acid	Amino acid
64285-06-9	1-[(1R,6R)-9-Azabicyclo[4.2.1]non-4-en-5-yl]ethanone (Anatoxin-a)	Algal toxin
657-24-9	1,1-Dimethylbiguanide (Metformin)	Pharmaceutical human drug
66357-35-5	Ranitidine	Pharmaceutical human drug
67-66-3	Trichloromethane	Disinfection byproduct-Carcinogen
67708-83-2	Dibromochloropropane	Carcinogen
6804-07-5	2-(2-Quinoxalinylmethylene)hydrazine-carboxylic acid methyl ester N,N'-dioxide (Carbadox)	Pharmaceutical human/veterinary antibiotic
68-22-4	17a-Ethynyl-19-nor-delta4-androstan-17b-ol-3-one (Norethindrone)	Form of progesterone
70458-96-7	Norfloxacin	Pharmaceutical human/veterinary antibiotic
719-22-2	2,6-bis(1,1-Dimethylethyl)-2,5-cyclohexadiene-1,4-dione	Insecticide
72-14-0	2-(p-Aminobenzenesulfonamido)thiazole (Sulfathiazole)	Pharmaceutical human/veterinary antibiotic
72-33-3	17a Ethynyl-estradiol-3-methyl ether (Mestranol)	Pharmaceutical-Sex/steroid hormone
723-46-6	4-Amino-N-(5-methyl-3-isoxazolyl)benzenesulfonamide (Sulfamethoxazole)	Pharmaceutical human/veterinary antibiotic
72-43-5	1,1,1-Trichloro-2,2-bis(p-anisyl)ethane (Methoxychlor)	Endocrine disruptor
72-54-8	1,1-bis(p-Chlorophenyl)-2,2-dichloroethane (DDD)	Endocrine disruptor
72-55-9	1,1-Dichloro-2,2-bis(p-chlorophenyl)ethylene (DDE)	Pesticide-Carcinogen

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

CAS	Name	Compounds class, known use and/or toxicity endpoint
738-70-5	2,4-Diamino-5-(3,4,5-trimethoxybenzyl)pyrimidine (Trimethoprim)	Pharmaceutical human/veterinary antibiotic
74-83-9	Bromomethane	Fumigant-Solvent
74-95-3	Dibromomethane	Solvent-Intermediate in production of herbicides
74-97-5	Bromochloromethane	Disinfect byproduct
75-09-2	Dichloromethane	Solvent-Found in aerosol and pesticide products, photographic film
75-25-2	Tribromomethane	Disinfection byproduct-Carcinogen
75-27-4	Dichlorobromomethane	Disinfect byproduct
75-71-8	Dichlorodifluoromethane	Refrigerant gas
759-94-4	Dipropylthiocarbamic acid S-ethyl ester	Herbicide
7601-90-3	Perchloric acid	Used to produce perchlorate, oxidant-Carcinogen
76420-72-9	Enalaprilat	Pharmaceutical human drug
76-44-8	1(3a),4,5,6,7,8,8-Heptachloro-3a(1),4,7,7a-tetrahydro-4,7-methanoindene (Heptachlor)	Endocrine disruptor
79-01-6	Trichloroethylene (TCE)	Solvent-Carcinogen
79-34-5	1,1,2,2-Tetrachloroethane	Solvent
79-57-2	4-(Dimethylamino)-1,4,4a,5,5a,6,11,12a-octahydro-3,5,6,10,12,12a-hexahydroxy-6-methyl-1,11-dioxo-2-naphthacenecarboxamide (Terramycin)	Antibiotic
80-32-0	Sulfachlorpyridazine	Pharmaceutical human/veterinary antibiotic
83463-62-1	Bromochloroacetonitrile	Disinfect byproduct
84-74-2	1,2-Benzenedicarboxylic acid dibutyl ester	Plasticizer
87-68-3	1,1,2,3,4,4-Hexachloro-1,3-butadiene	Used to make rubber compounds-Solvent
924-16-3	N-nitrosodibutylamine	Carcinogen
930-55-2	1-Nitroso-pyrrolidine	Carcinogen
93106-60-6	Enrofloxacin	Antibiotic-Industrial/household waste water product
93-76-5	2,4,5-Trichlorophenoxyacetic acid	Endocrine disruptor
944-22-9	O-ethyl S-phenyl ethylphosphonodithioate (Fonofos)	Insecticide
95-47-6	1,2-Dimethylbenzene	Fuel hydrocarbon-Carcinogen
95-48-7	1-Hydroxy-2-methylbenzene	Intermediate for production of pesticides, pharmaceuticals
95-50-1	1,2-Dichlorobenzene	Fumigant
95-63-6	1,2,4-Trimethylbenzene	Fuel hydrocarbon
994-05-8	2-Methyl-2-methoxybutane	Solvent
99-87-6	1-Methyl-4-isopropylbenzene (Cymene)	Manufacture of synthetic resins

ISBN:978-84-691-2701-8/DL:T.386-2008

ANNEX VI. Performance of training, validation and test data sets for the normalized permeate flow rate and normalize salt passage models based on back-propagation

Performance of all developed models is presented. For each length of time (7 to 125 hrs) considered for dividing the time space in equal intervals, several ANN models were built by varying the number of hidden neurons from 2 to 11. The model performance is characterized by the explained variance in prediction index (q^2), the correlation of determination (R^2), the average relative error ($\overline{\varepsilon}$), the standard deviation of the relative error (σ_ε) and the maximum relative error ($\varepsilon_{\max}$).

ISBN:978-84-691-2701-8/DL:T.386-2008

Performance of normalized permeate flow rate models.

NN archeology	TRAINING SET					VALIDATION SET					TEST SET				
	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]
Equal time intervals of 7 hrs															
7:2:1	0.86	0.86	0.52	0.40	2.87	0.86	0.87	0.52	0.41	3.23	0.81	0.75	1.10	0.86	8.44
7:3:1	0.87	0.87	0.50	0.39	2.94	0.87	0.87	0.50	0.39	2.42	0.85	0.78	0.98	0.81	5.38
7:4:1	0.87	0.87	0.50	0.39	2.90	0.87	0.87	0.51	0.39	3.03	0.87	0.80	0.93	0.73	6.00
7:5:1	0.86	0.86	0.51	0.40	3.46	0.86	0.86	0.51	0.39	2.49	0.88	0.81	0.88	0.70	5.92
7:6:1	0.87	0.87	0.50	0.38	2.85	0.86	0.86	0.51	0.42	3.19	0.87	0.82	0.94	0.72	7.66
7:7:1	0.87	0.87	0.50	0.39	2.83	0.85	0.85	0.49	0.41	3.07	0.83	0.75	1.02	0.86	6.58
7:8:1	0.87	0.87	0.49	0.38	3.06	0.87	0.87	0.50	0.39	2.43	0.85	0.77	1.02	0.77	6.37
7:9:1	0.87	0.87	0.50	0.38	3.64	0.88	0.88	0.49	0.39	2.90	0.86	0.77	0.96	0.73	8.94
7:10:1	0.88	0.88	0.48	0.38	2.78	0.86	0.86	0.51	0.39	3.11	0.86	0.80	0.97	0.72	7.79
7:11:1	0.87	0.87	0.49	0.38	3.34	0.87	0.87	0.50	0.38	2.61	0.85	0.76	1.00	0.78	7.01
Equal time intervals of 15 hrs															
7:2:1	0.84	0.84	0.54	0.42	3.14	0.85	0.85	0.52	0.41	3.05	0.76	0.70	1.08	0.81	7.99
7:3:1	0.85	0.85	0.51	0.40	3.17	0.85	0.85	0.54	0.41	2.27	0.83	0.76	0.91	0.70	5.81
7:4:1	0.85	0.85	0.52	0.41	2.64	0.84	0.84	0.54	0.42	3.12	0.80	0.73	0.97	0.77	6.39
7:5:1	0.86	0.86	0.51	0.39	3.45	0.85	0.85	0.53	0.41	2.56	0.84	0.78	0.87	0.65	5.17
7:6:1	0.86	0.86	0.51	0.39	2.71	0.84	0.84	0.52	0.41	3.14	0.82	0.75	0.94	0.72	7.75
7:7:1	0.86	0.86	0.51	0.40	3.11	0.86	0.86	0.51	0.39	2.14	0.79	0.72	1.02	0.73	6.62
7:8:1	0.86	0.87	0.50	0.39	3.12	0.86	0.86	0.49	0.37	2.10	0.83	0.81	0.90	0.68	6.68
7:9:1	0.86	0.86	0.50	0.39	3.26	0.85	0.85	0.53	0.42	3.50	0.85	0.79	0.87	0.66	5.63
7:10:1	0.87	0.87	0.49	0.38	2.95	0.85	0.85	0.52	0.41	2.69	0.81	0.74	0.95	0.77	8.42
7:11:1	0.86	0.86	0.50	0.39	3.32	0.85	0.85	0.52	0.40	2.62	0.84	0.84	0.85	0.70	6.92
Equal time intervals of 25 hrs															
7:2:1	0.84	0.84	0.54	0.41	3.42	0.85	0.85	0.51	0.41	2.21	0.76	0.73	1.08	0.76	6.64
7:3:1	0.85	0.85	0.52	0.40	3.23	0.86	0.86	0.53	0.40	2.45	0.81	0.76	0.93	0.68	6.88
7:4:1	0.86	0.86	0.51	0.40	3.25	0.84	0.84	0.53	0.41	2.48	0.77	0.69	1.03	0.75	7.00
7:5:1	0.85	0.85	0.53	0.41	3.29	0.84	0.84	0.53	0.41	1.99	0.83	0.79	0.86	0.67	8.61
7:6:1	0.84	0.84	0.54	0.41	3.43	0.85	0.85	0.53	0.40	2.00	0.80	0.73	0.98	0.71	7.68
7:7:1	0.86	0.87	0.50	0.39	3.19	0.86	0.86	0.51	0.40	2.59	0.76	0.79	1.03	0.81	7.21
7:8:1	0.85	0.85	0.53	0.41	3.32	0.85	0.85	0.52	0.40	2.89	0.77	0.69	0.98	0.81	6.92
7:9:1	0.87	0.87	0.49	0.38	3.09	0.88	0.88	0.50	0.38	2.11	0.84	0.81	0.83	0.68	4.84
7:10:1	0.86	0.86	0.50	0.39	3.37	0.85	0.85	0.52	0.39	2.39	0.80	0.83	0.94	0.74	7.89
7:11:1	0.87	0.87	0.49	0.39	2.40	0.86	0.86	0.50	0.40	3.21	0.84	0.81	0.83	0.68	7.15
Equal time intervals of 50 hrs															
7:2:1	0.84	0.84	0.54	0.42	3.50	0.85	0.85	0.52	0.39	2.44	0.69	0.65	1.25	0.90	8.32
7:3:1	0.84	0.84	0.53	0.41	3.38	0.84	0.84	0.56	0.44	3.03	0.77	0.71	1.00	0.84	7.17
7:4:1	0.84	0.84	0.53	0.42	3.47	0.86	0.86	0.52	0.41	2.24	0.77	0.74	1.00	0.87	7.83
7:5:1	0.85	0.85	0.53	0.41	3.42	0.85	0.85	0.53	0.42	3.18	0.78	0.74	1.00	0.84	7.95
7:6:1	0.86	0.86	0.51	0.40	3.37	0.87	0.87	0.49	0.38	2.79	0.82	0.78	0.91	0.73	6.59
7:7:1	0.84	0.84	0.53	0.41	2.83	0.84	0.84	0.56	0.43	3.55	0.79	0.71	0.98	0.79	7.46
7:8:1	0.86	0.86	0.51	0.40	2.74	0.85	0.85	0.51	0.41	3.59	0.79	0.77	1.00	0.73	7.94
7:9:1	0.86	0.86	0.51	0.40	3.18	0.85	0.85	0.53	0.39	2.62	0.83	0.74	0.91	0.70	9.31
7:10:1	0.86	0.86	0.50	0.39	3.35	0.86	0.86	0.50	0.39	2.97	0.80	0.72	0.94	0.80	6.84
7:11:1	0.87	0.87	0.49	0.39	2.92	0.87	0.87	0.49	0.37	2.55	0.79	0.71	1.00	0.78	8.35

ISBN:978-84-691-2701-8/DL:T.386-2008

NN archeology	TRAINING SET					VALIDATION SET					TEST SET				
	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]
Equal time intervals of 75 hrs															
7:2:1	0.84	0.84	0.54	0.42	3.00	0.83	0.83	0.55	0.42	2.85	0.72	0.69	1.13	0.93	7.99
7:3:1	0.82	0.82	0.57	0.43	3.76	0.84	0.84	0.55	0.43	2.23	0.74	0.64	1.12	0.84	7.67
7:4:1	0.85	0.85	0.52	0.41	3.94	0.83	0.83	0.55	0.41	2.69	0.77	0.70	1.04	0.80	7.41
7:5:1	0.85	0.85	0.52	0.40	3.16	0.85	0.85	0.51	0.40	3.08	0.76	0.67	1.09	0.80	7.87
7:6:1	0.84	0.84	0.53	0.42	3.90	0.85	0.85	0.55	0.43	2.73	0.77	0.74	1.07	0.80	7.34
7:7:1	0.85	0.85	0.52	0.42	4.16	0.85	0.85	0.52	0.41	2.59	0.78	0.71	1.01	0.85	5.95
7:8:1	0.87	0.87	0.50	0.39	3.07	0.84	0.84	0.53	0.40	2.41	0.76	0.72	1.00	0.88	8.30
7:9:1	0.86	0.86	0.50	0.39	3.03	0.86	0.87	0.51	0.39	2.83	0.78	0.75	1.01	0.81	7.98
7:10:1	0.86	0.86	0.52	0.40	3.21	0.84	0.84	0.53	0.42	3.15	0.74	0.76	1.10	0.86	7.92
7:11:1	0.86	0.86	0.51	0.40	3.47	0.85	0.85	0.52	0.42	2.50	0.74	0.72	1.10	0.91	7.75
Equal time intervals of 100 hrs															
7:2:1	0.83	0.83	0.57	0.43	3.83	0.83	0.83	0.56	0.43	2.69	0.72	0.61	1.16	0.87	6.51
7:3:1	0.85	0.85	0.53	0.42	2.97	0.84	0.84	0.53	0.40	2.68	0.75	0.67	1.08	0.87	7.18
7:4:1	0.85	0.85	0.52	0.40	2.98	0.86	0.86	0.53	0.42	3.40	0.75	0.66	1.06	0.87	7.90
7:5:1	0.83	0.83	0.56	0.43	3.53	0.84	0.84	0.54	0.42	2.68	0.79	0.73	1.02	0.77	6.50
7:6:1	0.86	0.86	0.50	0.39	3.13	0.85	0.85	0.52	0.41	3.17	0.77	0.80	1.08	0.81	8.57
7:7:1	0.85	0.85	0.52	0.41	3.74	0.84	0.84	0.54	0.42	2.80	0.74	0.69	1.15	0.83	7.83
7:8:1	0.87	0.87	0.50	0.39	3.30	0.85	0.85	0.52	0.41	3.17	0.71	0.59	1.19	0.95	6.46
7:9:1	0.86	0.86	0.50	0.39	3.42	0.85	0.85	0.52	0.41	2.90	0.77	0.68	1.03	0.85	5.83
7:10:1	0.85	0.85	0.53	0.41	3.58	0.85	0.85	0.52	0.40	3.10	0.66	0.63	1.30	0.93	7.09
7:11:1	0.85	0.85	0.52	0.41	3.73	0.84	0.84	0.53	0.42	3.22	0.73	0.66	1.17	0.81	9.34
Equal time intervals of 125 hrs															
7:2:1	0.83	0.83	0.55	0.42	3.25	0.84	0.84	0.55	0.46	4.14	0.78	0.71	1.03	0.80	7.37
7:3:1	0.84	0.84	0.54	0.42	4.01	0.84	0.84	0.53	0.41	2.70	0.72	0.61	1.12	0.90	7.09
7:4:1	0.86	0.86	0.50	0.40	2.63	0.86	0.86	0.51	0.41	3.44	0.62	0.56	1.35	1.01	8.15
7:5:1	0.86	0.86	0.51	0.40	3.33	0.87	0.87	0.50	0.38	2.89	0.68	0.69	1.30	0.88	9.72
7:6:1	0.84	0.84	0.53	0.42	4.33	0.84	0.84	0.52	0.43	3.37	0.73	0.65	1.15	0.87	6.14
7:7:1	0.85	0.85	0.53	0.41	3.69	0.83	0.83	0.53	0.41	2.19	0.79	0.73	0.99	0.81	11.74
7:8:1	0.86	0.86	0.49	0.39	3.34	0.87	0.87	0.50	0.39	2.76	0.74	0.68	1.16	0.80	7.61
7:9:1	0.87	0.87	0.49	0.38	2.70	0.87	0.87	0.50	0.40	3.24	0.72	0.71	1.15	0.95	8.27
7:10:1	0.86	0.86	0.50	0.39	3.68	0.86	0.86	0.50	0.39	2.04	0.69	0.56	1.26	0.90	6.64
7:11:1	0.87	0.87	0.48	0.38	2.80	0.86	0.86	0.50	0.39	2.09	0.66	0.70	1.34	0.92	11.07

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

Performance of normalized salt passage models.

NN archeology	TRAINING SET					VALIDATION SET					TEST SET				
	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]
Equal time intervals of 7 hrs															
7:2:1	0.95	0.96	1.80	1.56	16.01	0.95	0.95	1.85	1.58	8.65	0.90	0.93	2.92	2.32	35.37
7:3:1	0.96	0.96	1.60	1.48	16.02	0.95	0.95	1.69	1.61	15.26	0.90	0.93	2.85	2.30	35.47
7:4:1	0.96	0.96	1.63	1.51	15.80	0.96	0.96	1.67	1.51	14.94	0.90	0.93	2.97	2.30	35.27
7:5:1	0.96	0.96	1.61	1.46	15.12	0.95	0.95	1.65	1.67	15.83	0.90	0.93	2.73	2.29	35.52
7:6:1	0.96	0.96	1.60	1.49	15.35	0.96	0.96	1.56	1.50	15.94	0.90	0.92	2.81	2.31	35.64
7:7:1	0.96	0.96	1.61	1.45	15.99	0.96	0.96	1.57	1.59	15.48	0.91	0.93	2.89	2.27	35.35
7:8:1	0.96	0.96	1.58	1.46	15.67	0.96	0.96	1.65	1.58	14.67	0.91	0.93	2.79	2.24	35.45
7:9:1	0.96	0.96	1.59	1.45	15.57	0.96	0.96	1.62	1.56	16.10	0.92	0.91	2.51	2.29	35.57
7:10:1	0.96	0.96	1.59	1.47	15.50	0.96	0.96	1.67	1.53	16.42	0.93	0.91	2.44	2.12	36.34
7:11:1	0.96	0.96	1.61	1.46	15.86	0.96	0.96	1.58	1.52	16.69	0.91	0.91	2.72	2.21	35.52
Equal time intervals of 15 hrs															
7:2:1	0.94	0.94	2.11	1.79	15.71	0.95	0.95	2.08	1.74	12.87	0.91	0.90	2.54	2.27	37.03
7:3:1	0.94	0.94	2.10	1.78	15.74	0.95	0.95	2.01	1.76	12.64	0.91	0.90	2.52	2.28	37.41
7:4:1	0.95	0.95	1.99	1.67	12.05	0.94	0.94	2.04	1.75	17.74	0.91	0.90	2.53	2.22	37.18
7:5:1	0.95	0.95	2.03	1.69	17.68	0.95	0.95	1.95	1.69	12.02	0.91	0.90	2.54	2.37	35.92
7:6:1	0.95	0.95	1.95	1.69	17.22	0.95	0.95	1.97	1.71	10.29	0.91	0.90	2.62	2.35	36.02
7:7:1	0.95	0.95	1.88	1.67	17.13	0.95	0.95	1.88	1.62	11.64	0.91	0.90	2.62	2.33	37.66
7:8:1	0.95	0.95	1.92	1.64	16.76	0.95	0.95	1.89	1.64	9.61	0.91	0.90	2.58	2.35	37.20
7:9:1	0.95	0.95	1.90	1.62	17.10	0.95	0.95	1.89	1.63	15.66	0.90	0.90	2.70	2.42	36.10
7:10:1	0.96	0.96	1.85	1.56	13.00	0.95	0.95	1.98	1.79	18.02	0.90	0.89	2.75	2.42	36.91
7:11:1	0.95	0.95	1.87	1.62	16.38	0.96	0.96	1.92	1.55	7.69	0.91	0.89	2.68	2.38	36.00
Equal time intervals of 25 hrs															
7:2:1	0.93	0.93	2.36	2.02	23.07	0.92	0.92	2.47	2.06	13.53	0.89	0.87	2.88	2.58	39.24
7:3:1	0.92	0.92	2.50	2.12	23.37	0.93	0.93	2.44	2.06	12.55	0.89	0.87	2.85	2.59	38.73
7:4:1	0.94	0.94	2.24	1.80	23.38	0.94	0.94	2.23	1.81	11.96	0.90	0.87	2.86	2.46	38.25
7:5:1	0.93	0.93	2.28	1.93	22.75	0.93	0.93	2.37	2.00	12.76	0.88	0.86	3.05	2.65	38.51
7:6:1	0.94	0.94	2.20	1.90	22.77	0.94	0.94	2.22	1.90	13.55	0.89	0.86	2.97	2.59	38.45
7:7:1	0.95	0.95	2.07	1.71	13.82	0.93	0.93	2.07	1.99	23.49	0.88	0.87	3.12	2.69	40.59
7:8:1	0.94	0.94	2.17	1.82	21.84	0.94	0.94	2.17	1.75	13.74	0.88	0.85	3.07	2.64	39.09
7:9:1	0.94	0.94	2.09	1.78	22.97	0.95	0.95	2.06	1.74	11.22	0.87	0.85	3.35	2.87	37.55
7:10:1	0.95	0.95	1.90	1.68	23.37	0.95	0.95	1.94	1.71	13.20	0.87	0.84	3.42	2.67	40.88
7:11:1	0.95	0.95	1.83	1.65	22.80	0.95	0.95	1.93	1.64	11.06	0.84	0.82	3.60	3.04	38.65
Equal time intervals of 50 hrs															
7:2:1	0.90	0.90	2.78	2.27	31.38	0.92	0.92	2.74	2.12	13.05	0.74	0.73	4.19	3.82	42.32
7:3:1	0.91	0.91	2.74	2.23	28.96	0.91	0.91	2.70	2.13	12.12	0.71	0.75	4.45	3.92	42.33
7:4:1	0.91	0.91	2.63	2.23	33.54	0.91	0.91	2.71	2.12	13.77	0.73	0.73	4.20	3.97	41.22
7:5:1	0.91	0.91	2.63	2.18	18.26	0.91	0.91	2.64	2.23	27.26	0.74	0.72	4.49	3.94	37.13
7:6:1	0.92	0.92	2.59	2.16	28.07	0.92	0.92	2.54	2.07	11.21	0.70	0.70	4.74	4.00	39.17
7:7:1	0.93	0.93	2.32	1.98	29.56	0.93	0.93	2.37	1.91	13.46	0.69	0.73	4.94	4.07	42.86
7:8:1	0.89	0.89	2.94	2.41	33.15	0.89	0.89	2.98	2.45	14.41	0.72	0.71	4.47	4.00	39.91
7:9:1	0.94	0.94	2.18	1.84	28.31	0.94	0.94	2.14	1.78	12.83	0.70	0.73	4.61	4.10	43.50
7:10:1	0.94	0.94	2.11	1.72	28.32	0.95	0.95	2.11	1.69	10.53	0.71	0.71	4.53	4.02	40.67
7:11:1	0.95	0.95	2.01	1.73	28.15	0.95	0.95	2.04	1.71	10.53	0.70	0.73	4.77	4.22	39.46

ISBN:978-84-691-2701-8/DL:T.386-2008

NN archeology	TRAINING SET					VALIDATION SET					TEST SET				
	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]	q^2	R^2	$\overline{\varepsilon}$ [%]	σ_ε [%]	ε_{max} [%]
Equal time intervals of 75 hrs															
7:2:1	0.86	0.86	3.37	3.01	18.76	0.85	0.85	3.48	3.10	17.47	0.67	0.65	5.39	4.18	39.72
7:3:1	0.90	0.90	2.82	2.30	18.79	0.89	0.89	2.88	2.36	14.53	0.67	0.66	4.96	4.47	41.37
7:4:1	0.90	0.90	2.90	2.28	17.91	0.91	0.91	2.80	2.11	13.70	0.66	0.62	4.86	4.88	41.57
7:5:1	0.91	0.91	2.72	2.28	18.92	0.91	0.91	2.69	2.30	13.46	0.67	0.65	5.46	4.28	41.30
7:6:1	0.89	0.89	2.98	2.69	17.88	0.88	0.88	3.03	2.75	15.49	0.61	0.59	6.00	4.57	41.48
7:7:1	0.92	0.92	2.54	2.19	15.02	0.90	0.91	2.63	2.34	14.65	0.64	0.67	5.47	4.71	41.34
7:8:1	0.92	0.92	2.46	2.22	16.50	0.92	0.92	2.41	2.21	17.13	0.62	0.60	5.40	5.01	42.24
7:9:1	0.94	0.94	2.29	1.87	17.90	0.94	0.94	2.29	1.81	10.96	0.55	0.63	6.10	4.91	43.28
7:10:1	0.93	0.93	2.25	2.01	15.99	0.94	0.94	2.32	2.09	15.98	0.62	0.65	5.71	4.41	38.11
7:11:1	0.88	0.88	3.10	2.76	20.18	0.88	0.88	3.11	2.60	14.77	0.63	0.57	5.54	4.84	42.82
Equal time intervals of 100 hrs															
7:2:1	0.87	0.87	3.24	2.77	21.55	0.86	0.86	3.21	2.81	20.81	0.46	0.47	7.19	5.25	43.21
7:3:1	0.89	0.89	2.83	2.70	19.86	0.88	0.88	2.90	2.85	18.46	0.48	0.52	6.85	4.96	42.39
7:4:1	0.90	0.90	2.96	2.31	15.41	0.90	0.90	2.97	2.35	16.85	0.47	0.49	7.11	5.16	44.19
7:5:1	0.92	0.92	2.56	2.18	19.69	0.92	0.92	2.54	2.13	11.89	0.48	0.50	6.99	5.28	44.44
7:6:1	0.94	0.94	2.15	1.75	18.35	0.95	0.95	2.08	1.69	10.22	0.55	0.55	6.50	4.69	44.21
7:7:1	0.94	0.94	2.20	1.79	17.46	0.93	0.93	2.35	2.01	11.73	0.49	0.54	7.15	5.12	46.99
7:8:1	0.95	0.95	1.99	1.65	18.66	0.95	0.95	2.05	1.68	8.97	0.47	0.47	7.26	5.26	44.85
7:9:1	0.95	0.95	2.00	1.59	17.07	0.95	0.95	2.04	1.63	10.87	0.51	0.53	7.00	5.07	43.89
7:10:1	0.95	0.95	2.00	1.65	16.92	0.95	0.95	2.05	1.68	15.38	0.53	0.56	6.82	4.56	44.43
7:11:1	0.94	0.94	2.33	1.94	16.85	0.92	0.92	2.52	2.12	11.77	0.50	0.52	7.15	4.55	41.70
Equal time intervals of 125 hrs															
7:2:1	0.88	0.88	3.08	2.57	17.75	0.86	0.86	3.15	2.73	26.30	0.68	0.72	5.76	4.11	37.07
7:3:1	0.91	0.91	2.69	2.29	27.63	0.89	0.89	2.84	2.40	19.56	0.71	0.69	5.16	4.31	36.06
7:4:1	0.89	0.89	2.96	2.59	33.60	0.88	0.88	2.91	2.53	16.97	0.73	0.69	4.81	4.08	39.75
7:5:1	0.90	0.90	2.79	2.34	28.59	0.90	0.90	2.76	2.35	15.37	0.73	0.69	4.86	4.17	42.34
7:6:1	0.89	0.89	2.91	2.42	16.11	0.89	0.89	2.90	2.50	26.17	0.67	0.65	5.66	4.24	44.23
7:7:1	0.92	0.92	2.38	2.14	26.91	0.91	0.91	2.51	2.24	19.29	0.67	0.68	5.55	3.98	39.96
7:8:1	0.88	0.88	2.98	2.59	17.79	0.86	0.86	3.04	2.71	26.19	0.66	0.65	5.43	4.53	41.42
7:9:1	0.91	0.91	2.69	2.15	17.86	0.90	0.90	2.73	2.22	25.79	0.56	0.60	6.56	4.66	42.91
7:10:1	0.93	0.93	2.47	1.97	27.25	0.94	0.94	2.38	1.86	21.19	0.58	0.56	6.58	5.22	40.87
7:11:1	0.89	0.89	2.88	2.45	30.67	0.89	0.89	2.91	2.44	17.62	0.63	0.64	5.68	4.63	39.91

UNIVERSITAT ROVIRA I VIRGILI
MODELING THE RESERVE OSMOSIS PROCESSES PERFORMANCE USING ARTIFICIAL NEURAL NETWORKS
Dan Mihai Libotean
ISBN:978-84-691-2701-8/DL:T.386-2008

ISBN:978-84-691-2701-8/DL:T.386-2008

References

[1] M. Abou Rayan and I. Khaled, Seawater desalination by reverse osmosis (case study), Desalination 153(1-3) (2003) 245-251.

[2] M. Abou Rayan, B. Djebedjian and I. Khaled, Water supply and demand and a desalination option for Sinai, Egypt, Desalination 136(1-3) (2001) 73-81.

[3] M. Mulder, Basic Principles of Membrane Technology, 2nd ed, Kluwer Academic Publishers, 1997.

[4] G. Al-Enezi and N. Fawzi, Design consideration of RO units: case studies, Desalination 153(1-3) (2003) 281-286.

[5] I.M. El-Azizi and A.A.M. Omran, Design criteria of 10,000 m(3)/d SWRO desalination plant of Tajura, Libya, Desalination 153(1-3) (2003) 273-279.

[6] D.P. Rico and M.F.C. Arias, A reverse osmosis potable water plant at Alicante University: first years of operation, Desalination 137(1-3) (2001) 91-102.

[7] V. Romero-Ternero, L. Garcia-Rodriguez and C. Gomez-Camacho, Thermoeconomic analysis of a seawater reverse osmosis plant, Desalination 181(1-3) (2005) 43-59.

[8] K.T. Chua, M.N.A. Hawlader and A. Malek, Pretreatment of seawater: Results of pilot trials in Singapore, Desalination 159(3) (2003) 225-243.

[9] K. Karakulski, M. Gryta and M. Sasim, Production of process water using integrated membrane processes, Chemical Papers-Chemicke Zvesti 60(6) (2006) 416-421.

[10] T. Heberer, Occurrence, fate, and assessment of polycyclic musk residues in the aquatic environment of urban areas - A review, Acta Hydrochimica Et Hydrobiologica 30(5-6) (2003) 227-243.

[11] D.W. Kolpin, E.T. Furlong, M.T. Meyer, E.M. Thurman, S.D. Zaugg, L.B. Barber and H.T. Buxton, Pharmaceuticals, hormones, and other organic wastewater

ISBN:978-84-691-2701-8/DL:T.386-2008

contaminants in US streams, 1999-2000: A national reconnaissance, Environmental Science & Technology 36(6) (2002) 1202-1211.

[12] C. Baronti, R. Curini, G. D'Ascenzo, A. Di Corcia, A. Gentili and R. Samperi, Monitoring natural and synthetic estrogens at activated sludge sewage treatment plants and in a receiving river water, Environmental Science & Technology 34(24) (2000) 5059-5066.

[13] Y. Zhao and J.S. Taylor, Incorporation of osmotic pressure in an integrated incremental model for predicting RO or NF permeate concentration, Desalination 174(2) (2005) 145-159.

[14] J.G. Wijmans and R.W. Baker, The Solution-Diffusion Model - a Review, Journal of Membrane Science 107(1-2) (1995) 1-21.

[15] K. Jamal, M.A. Khan and M. Kamil, Mathematical modeling of reverse osmosis systems, Desalination 160(1) (2004) 29-42.

[16] B. Nicolaisen, Developments in membrane technology for water treatment, Desalination 153(1-3) (2003) 355-360.

[17] S.S. Sablani, M.F.A. Goosen, R. Al-Belushi and M. Wilf, Concentration polarization in ultrafiltration and reverse osmosis: a critical review, Desalination 141(3) (2001) 269-289.

[18] W.Y. Shih, A. Rahardianto, R.W. Lee and Y. Cohen, Morphometric characterization of calcium sulfate dihydrate (gypsum) scale on reverse osmosis membranes, Journal of Membrane Science 252(1-2) (2005) 253-263.

[19] K. Kosutic and B. Kunst, RO and NF membrane fouling and cleaning and pore size distribution variations, Desalination 150(2) (2002) 113-120.

[20] S. Otles and S. Otles, Desalination Techniques, Electronic Journal of Environmental, Agricultural and Food Chemistry 4(4) (2004) 963-969.

[21] B. Pilat, Practice of water desalination by electrodialysis, Desalination 139(1-3) (2001) 385-392.

[22] Asociación Española de Desalación y Reutilización, www.aedyr.com.

ISBN:978-84-691-2701-8/DL:T.386-2008

[23] A. Valero, J. Uche and L. Serra, La desalación como alternativa al Plan Hidrológico Nacional, 2001: documento de trabajo Centro de Investigación de Recursos y Consumos Energéticos, Universidad de Zaragoza.

[24] V. Geraldes, N.E. Pereira and M.N. de Pinho, Simulation and optimization of medium-sized seawater reverse osmosis processes with spiral-wound modules, Industrial & Engineering Chemistry Research 44(6) (2005) 1897-1905.

[25] S. Senthilmurugan, A. Ahluwalia and S.K. Gupta, Modeling of a spiral-wound module and estimation of model parameters using numerical techniques, Desalination 173(3) (2005) 269-286.

[26] M.G. Marcovecchio, P.A. Aguirre and N.J. Scenna, Global optimal design of reverse osmosis networks for seawater desalination: modeling and algorithm, Desalination 184(1-3) (2005) 259-271.

[27] A.J. Dababneh and M.A. Al-Nimr, A reverse osmosis desalination unit, Desalination 153(1-3) (2003) 265-272.

[28] S. El-Manharawy and A. Hafez, Dehydration model for RO-membrane fouling: a preliminary approach, Desalination 153(1-3) (2003) 95-107.

[29] S. Jain and S.K. Gupta, Analysis of modified surface force pore flow model with concentration polarization and comparison with Spiegler-Kedem model in reverse osmosis systems, Journal of Membrane Science 232(1-2) (2004) 45-61.

[30] K.L. Chen, L.F. Song, S.L. Ong and W.J. Ng, The development of membrane fouling in full-scale RO processes, Journal of Membrane Science 232(1-2) (2004) 63-72.

[31] T. Matsuura and S. Sourirajan, Physicochemical Criteria for Reverse Osmosis Separation of Alcohols, Phenols, and Monocarboxylic Acids in Aqueous Solutions Using Porous Cellulose Acetate Membranes, Journal of Applied Polymer Science 15(12) (1971) 2905-2927.

[32] H. Ozaki and H.F. Li, Rejection of organic compounds by ultra-low pressure reverse osmosis membrane, Water Research 36(1) (2002) 123-130.

[33] L. Kastelan-Kunst, K. Kosutic, V. Dananic and B. Kunst, FT30 membranes of characterized porosities in the reverse osmosis organics removal from aqueous solutions, Water Research 31(11) (1997) 2878-2884.

ISBN:978-84-691-2701-8/DL:T.386-2008

[34] B. Van der Bruggen, J. Schaep, W. Maes, D. Wilms and C. Vandecasteele, Nanofiltration as a treatment method for the removal of pesticides from ground waters, Desalination 117(1-3) (1998) 139-147.

[35] B. Van der Bruggen, J. Schaep, D. Wilms and C. Vandecasteele, Influence of molecular size, polarity and charge on the retention of organic molecules by nanofiltration, Journal of Membrane Science 156(1) (1999) 29-41.

[36] Y. Kiso, T. Kon, T. Kitao and K. Nishimura, Rejection properties of alkyl phthalates with nanofiltration membranes, Journal of Membrane Science 182(1-2) (2001) 205-214.

[37] Y. Kiso, Y. Nishimura, T. Kitao and K. Nishimura, Rejection properties of non-phenylic pesticides with nanofiltration membranes, Journal of Membrane Science 171(2) (2000) 229-237.

[38] Y. Kiso, Y. Sugiura, T. Kitao and K. Nishimura, Effects of hydrophobicity and molecular size on rejection of aromatic pesticides with nanofiltration membranes, Journal of Membrane Science 192(1-2) (2001) 1-10.

[39] K. Kimura, S. Toshima, G. Amy and Y. Watanabe, Rejection of neutral endocrine disrupting compounds (EDCs) and pharmaceutical active compounds (PhACs) by RO membranes, Journal of Membrane Science 245(1-2) (2004) 71-78.

[40] A. Abbas and N. Al-Bastaki, Modeling of an Reverse Osmosis water desalination unit using neural networks, Chemical Engineering Journal 114 (2005) 139-143.

[41] W.R. Bowen, M.G. Jones and H.N.S. Yousef, Prediction of the rate of crossflow membrane ultrafiltration of colloids: A neural network approach, Chemical Engineering Science 53(22) (1998) 3793-3802.

[42] S. Chellam, Artificial neural network model for transient crossflow microfiltration of polydispersed suspensions, Journal of Membrane Science 258(1-2) (2005) 35-42.

[43] H.Q. Chen and A.S. Kim, Prediction of permeate flux decline in crossflow membrane filtration of colloidal suspension: a radial basis function neural network approach, Desalination 192(1-3) (2006) 415-428.

[44] N. Delgrange, C. Cabassud, M. Cabassud, L. Durand-Bourlier and J.M. Laine, Modelling of ultrafiltration fouling by neural network, Desalination 118(1-3) (1998) 213-227.

ISBN:978-84-691-2701-8/DL:T.386-2008

[45] N. Delgrange, C. Cabassud, M. Cabassud, L. Durand-Bourlier and J.M. Laine, Neural networks for prediction of ultrafiltration transmembrane pressure - application to drinking water production, Journal of Membrane Science 150 (1998) 111-123.

[46] N. Delgrange-Vincent, C. Cabassud, M. Cabassud, L. Durand-Bourlier and J.M. Laine, Neural networks for long term prediction of fouling and backwash efficiency in ultrafiltration for drinking water production, Desalination 131 (2000) 353-362.

[47] M. Dornier, M. Decloux, G. Trystram and A. Lebert, Dynamic Modeling of Cross-Flow Microfiltration Using Neural Networks, Journal of Membrane Science 98(3) (1995) 263-273.

[48] H. Niemi, A. Bulsari and S. Palosaari, Simulation of membrane separation by neural networks, Journal of Membrane Science 120 (1995) 185-191.

[49] M.A. Razavi, A. Mortazavi and M. Mousavi, Application of neural networks for crossflow milk ultrafiltration simulation, International Dairy Journal 14(1) (2004) 69-80.

[50] G.B. Sahoo and C. Ray, Predicting flux decline in crossflow membranes using artificial neural networks and genetic algorithms, Journal of Membrane Science 283(1-2) (2006) 147-157.

[51] G.R. Shetty and S. Chellam, Predicting membrane fouling during municipal drinking water nanofiltration using artificial neural networks, Journal of Membrane Science 217(1-2) (2003) 69-86.

[52] Z. Zhang, V.M. Bright and A.R. Greenberg, Use of capacitive microsensors and ultrasonic time-domain reflectometry for in-situ quantification of concentration polarization and membrane fouling in pressure-driven membrane filtration, Sensors and Actuators, B: Chemical 117(2) (2006) 323-331.

[53] J.S. Vrouwenvelder, J.A.M. van Paassen, L.P. Wessels, A.F. van Dama and S.M. Bakker, The membrane fouling simulator: A practical tool for fouling prediction and control, Journal of Membrane Science 281(1-2) (2006) 316-324.

[54] X.Y. Li and X.M. Wang, Modelling of membrane fouling in a submerged membrane bioreactor, Journal of Membrane Science 278(1-2) (2006) 151-161.

ISBN:978-84-691-2701-8/DL:T.386-2008

[55] C.C. Ho and A.L. Zydney, Overview of fouling phenomena and modeling approaches for membrane bioreactors, Separation Science and Technology 41(7) (2006) 1231-1251.

[56] S. Lalot, On-line detection of fouling in a water circulating temperature controller (WCTC) used in injection moulding: Part 1: Principles, Applied Thermal Engineering 26(11-12) (2006) 1087-1094.

[57] S. Lalot, On-line detection of fouling in a water circulating temperature controller (WCTC) used in injection moulding: Part 2: Application, Applied Thermal Engineering 26(11-12) (2006) 1095-1105.

[58] K. Kimura, G. Amy, J.E. Drewes, T. Heberer, T.U. Kim and Y. Watanabe, Rejection of organic micropollutants (disinfection by-products, endocrine disrupting compounds, and pharmaceutically active compounds) by NF/RO membranes, Journal of Membrane Science 227(1-2) (2003) 113-121.

[59] C.N. Laabs, G.L. Amy and M. Jekel, Understanding the size and character of fouling-causing substances from effluent organic matter (EfOM) in low-pressure membrane filtration, Environmental Science & Technology 40(14) (2006) 4495-4499.

[60] C. Bellona, J.E.J.E. Drewes, P. Xu and G. Amy, Factors affecting the rejection of organic solutes during NF/RO treatment-a literature review, Water Research 38(12) (2004) 2795-2809.

[61] B. Van der Bruggen, A. Verliefde, L. Braeken, E.R. Cornelissen, K. Moons, J. Verberk, H.J.C. van Dijk and G. Amy, Assessment of a semi-quantitative method for estimation of the rejection of organic compounds in aqueous solution in nanofiltration, Journal of Chemical Technology and Biotechnology 81(7) (2006) 1166-1176.

[62] C.F. Schutte, The rejection of specific organic compounds by reverse osmosis membranes, Desalination 158(1-3) (2003) 285-294.

[63] J.W. Cho, G. Amy and J. Pellegrino, Membrane filtration of natural organic matter: comparison of flux decline, NOM rejection, and foulants during filtration with three UF membranes, Desalination 127(3) (2000) 283-298.

ISBN:978-84-691-2701-8/DL:T.386-2008

[64] J.W. Cho, G. Amy, Y.M. Yoon and J. Sohn, Predictive models and factors affecting natural organic matter (NOM) rejection and flux decline in ultrafiltration (UF) membranes, Desalination 142(3) (2002) 245-255.

[65] H.Y. Ng and M. Elimelech, Influence of colloidal fouling on rejection of trace organic contaminants by reverse osmosis, Journal of Membrane Science 244(1-2) (2004) 215-226.

[66] R. Boussahel, A. Montiel and M. Baudu, Effects of organic and inorganic matter on pesticide rejection by nanofiltration, Desalination 145(1-3) (2002) 109-114.

[67] T. Matsuura and S. Sourirajan, Reverse Osmosis Separation of Some Organic Solutes in Aqueous Solution Using Porous Cellulose Acetate Membranes, Industrial & Engineering Chemistry Process Design and Development 10(1) (1971) 102-108.

[68] H. Niemi and S. Palosaari, Calculation of permeate flux and rejection in simulation of ultrafiltration and reverse osmosis processes, Journal of Membrane Science 84 (1993) 123-137.

[69] R.S. Faibish, M. Elimelech and Y. Cohen, Effect of interparticle electrostatic double layer interactions on permeate flux decline in crossflow membrane filtration of colloidal suspensions: An experimental investigation, Journal of Colloid and Interface Science 204(1) (1998) 77-86.

[70] P. Bhagat, An Introduction to Neural Nets, Chemical Engineering Progress 86(8) (1990) 55-60.

[71] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 2002.

[72] S.P. Chitra, Use Neural Networks for Problem-Solving, Chemical Engineering Progress 89(4) (1993) 44-52.

[73] G.E. Hinton, How Neural Networks Learn from Experience, Scientific American 267(3) (1992) 145-151.

[74] K. Levenberg, A Method for the Solution of Certain Non-linear Problems in Least Squares, Quarterly of Applied Mathematics 2(2) (1944) 164-168.

ISBN:978-84-691-2701-8/DL:T.386-2008

[75] D.W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics 11(2) (1963) 431-441.

[76] M.T. Hagan and M.B. Menhaj, Training Feedforward Networks with the Marquardt Algorithm, Ieee Transactions on Neural Networks 5(6) (1994) 989-993.

[77] T. Kohonen, The self-organizing map, Neurocomputing 21(1-3) (1998) 1-6.

[78] T. Kohonen, The Self-Organizing Map, Proceedings of the IEEE 78(9) (1990) 1464-1480.

[79] R. Xu and D. Wunsch, Survey of clustering algorithms, Ieee Transactions on Neural Networks 16(3) (2005) 645-678.

[80] J. Vesanto and E. Alhoniemi, Clustering of the self-organizing map, Ieee Transactions on Neural Networks 11(3) (2000) 586-600.

[81] D.L. Davies and D.W. Bouldin, Cluster Separation Measure, Ieee Transactions on Pattern Analysis and Machine Intelligence 1(2) (1979) 224-227.

[82] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds and D.B. Rosen, Fuzzy Artmap - a Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps, Ieee Transactions on Neural Networks 3(5) (1992) 698-713.

[83] R. Rallo, J. Ferre-Gine, A. Arenas and F. Giralt, Neural virtual sensor for the inferential prediction of product quality from process variables, Computers & Chemical Engineering 26(12) (2002) 1735-1754.

[84] L.A. Zadeh, Fuzzy Sets, Information and Control 8(3) (1965) 338.

[85] G.A. Carpenter, S. Grossberg and D.B. Rosen, Fuzzy Art - Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System, Neural Networks 4(6) (1991) 759-771.

[86] G.A. Carpenter, Default ARTMAP, Proceedings of the International Joint Conference on Neural Networks, Portland (2003).

ISBN:978-84-691-2701-8/DL:T.386-2008

[87] G.A. Carpenter, S. Grossberg and J.H. Reynolds, Artmap - Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network, Neural Networks 4(5) (1991) 565-588.

[88] F. Giralt, A. Arenas, J. Ferre-Gine, R. Rallo and G.A. Kopp, The simulation and interpretation of free turbulence with a cognitive neural system, Physics of Fluids 12(7) (2000) 1826-1835.

[89] E. Frank, M. Hall, L. Trigg, G. Holmes and I.H. Witten, Data mining in bioinformatics using Weka, Bioinformatics 20(15) (2004) 2479-2481.

[90] M.A. Hall and L.A. Smith, Practical feature subset selection for machine learning, Australian Computer Science Conference, 1998, Springer.

[91] M.A. Hall, Correlation-based feature selection for discrete and numeric class machine learning, International Conference on Machine Learning, 2000, Stanford University, Morgan Kaufmann Publishers.

[92] R. Rallo, G. Espinosa and F. Giralt, Using an ensemble of neural based QSARs for the prediction of toxicological properties of chemical contaminants, Process Safety and Environmental Protection 83(B4) (2005) 387-392.

[93] C.N. Hsu, H.J. Huang and D. Schuschel, The ANNIGMA-wrapper approach to fast feature selection for neural nets, Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics 32(2) (2002) 207-212.

[94] I.H. Witten and E. Frank, Data Mining - Practical Machine Learning Tools and Techniques with Java Implementation, Morgan Kaufmann Publishers, 2000.

[95] M.A. Hall and S. L.A., Practical feature subset selection for machine learning, 21st Australasian Computer Science Conference, 1998, Perth, Australia.

[96] M.A. Hall and L.A. Smith, Feature subset selection: a correlation based filter approach, International Conference on Neural Information Processing and Intelligent Information Systems, 1997, Springer.

[97] S. Kaski and K. Lagus, Comparing self-organizing maps, ICANN'96, International Conference of Neural Networks, 1997, Springer, Berlin.

ISBN:978-84-691-2701-8/DL:T.386-2008

[98] G. Espinosa, D. Yaffe, Y. Cohen, A. Arenas and F. Giralt, Neural network based quantitative structural property relations (QSPRs) for predicting boiling points of aliphatic hydrocarbons, Journal of Chemical Information and Computer Sciences 40(3) (2000) 859-879.

[99] G. Espinosa, D. Yaffe, A. Arenas, Y. Cohen and F. Giralt, A fuzzy ARTMAP-based quantitative structure-property relationship (QSPR) for predicting physical properties of organic compounds, Industrial & Engineering Chemistry Research 40(12) (2001) 2757-2766.

[100] D. Yaffe, Y. Cohen, G. Espinosa, A. Arenas and F. Giralt, A fuzzy ARTMAP based on quantitative structure-property relationships (QSPRs) for predicting aqueous solubility of organic compounds, Journal of Chemical Information and Computer Sciences 41(5) (2001) 1177-1207.

[101] D. Yaffe and Y. Cohen, Neural network based temperature-dependent quantitative structure property relations (QSPRs) for predicting vapor pressure of hydrocarbons, Journal of Chemical Information and Computer Sciences 41(2) (2001) 463-477.

[102] D. Yaffe, Y. Cohen, G. Espinosa, A. Arenas and F. Giralt, Fuzzy ARTMAP and back-propagation neural networks based quantitative structure-property relationships (QSPRs) for octanol-water partition coefficient of organic compounds, Journal of Chemical Information and Computer Sciences 42(2) (2002) 162-183.

[103] D. Yaffe, Y. Cohen, G. Espinosa, A. Arenas and F. Giralt, A fuzzy ARTMAP-based quantitative structure-property relationship (QSPR) for the Henry's law constant of organic compounds, Journal of Chemical Information and Computer Sciences 43(1) (2003) 85-112.

[104] F. Giralt, G. Espinosa, A. Arenas, J. Ferre-Gine, L. Amat, X. Girones, R. Carbo-Dorca and Y. Cohen, Estimation of infinite dilution activity coefficients of organic compounds in water with neural classifiers, Aiche Journal 50(6) (2004) 1315-1343.

[105] G. Espinosa, A. Arenas and F. Giralt, An integrated SOM-fuzzy ARTMAP neural system for the evaluation of toxicity, Journal of Chemical Information and Computer Sciences 42(2) (2002) 343-359.

ISBN:978-84-691-2701-8/DL:T.386-2008

[106] P. Mazzatorta, M. Vracko, A. Jezierska and E. Benfenati, Modeling toxicity by using supervised Kohonen Neural Networks, Journal of Chemical Information and Computer Sciences 43(2) (2003) 485-492.

[107] G. Gini, M. Lorenzini, E. Benfenati, P. Grasso and M. Bruschi, Predictive carcinogenicity: A model for aromatic compounds, with nitrogen-containing substituents, based on molecular descriptors using an artificial neural network, Journal of Chemical Information and Computer Sciences 39(6) (1999) 1076-1080.

[108] R. Vendrame, R.S. Braga, Y. Takahata and D.S. Galvao, Structure-activity relationship studies of carcinogenic activity of polycyclic aromatic hydrocarbons using calculated molecular descriptors with principal component analysis and neural network methods, Journal of Chemical Information and Computer Sciences 39(6) (1999) 1094-1104.

[109] C.W. Yap and Y.Z. Chen, Quantitative structure-pharmacokinetic relationships for drug distribution properties by using general regression neural network, Journal of Pharmaceutical Sciences 94(1) (2005) 153-168.

[110] G. Rodriguez, S. Buonora, T. Knoell, D. Phipps and H. Ridgway, Rejection of pharmaceuticals by reverse osmosis (RO) membranes: quantitative structure activity relationship (QSAR) analysis, 2004: National Water Research Institute, NWRI Project No. 01-EC-002.

[111] H.T. Buxton, U.S. Geological Survey Fact Sheet FS-062-00, 2000: U.S. Geological Survey Toxic Substances Hydrology Program, p. 4.

[112] U.S. Environmental Protection Agency Unregulated Contaminant Monitoring Rule, 1999: U.S. Environmental Protection Agency, Federal Register: Volume 64, Number 180.

[113] U.S. Environmental Protection Agency Announcement of the Drinking Water Contaminant List, 1998: U.S. Environmental Protection Agency, Federal Register: Volume 63, Number 40.

[114] Unregulated Chemicals Requiring Monitoring, Title 22 of the California Code of Regulations, No. 64450, 2001: California Division of Drinking Water and Environmental Management.

[115] R.T. Riley, B.W. Kemppainen and W.P. Norred, Quantitative Tritium Exchange of H-3 Aflatoxin-B1 During Penetration through Isolated Human-Skin, Biochemical and Biophysical Research Communications 153(1) (1988) 395-401.

[116] ChemSketch 8.00, Advanced Chemistry Development Inc.

[117] CAChe Worksystem Pro 6.1, Oxford Molecular Ltd.

[118] CAChe for Windows User Guide, Fujitsu Limited, 2001.

[119] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart, The Development and Use of Quantum-Mechanical Molecular-Models. 76. AM1 - a New General-Purpose Quantum-Mechanical Molecular-Model, Journal of the American Chemical Society 107(13) (1985) 3902-3909.

[120] D.C. Young, Computational Chemistry - A Practical Guide for Applying Techniques to Real-World Problems, Wiley-Interscience, 2001.

[121] L.B. Kier and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.

[122] L.B. Kier and L.H. Hall, Molecular Connectivity in Structure-Activity Analysis, John Wiley & Sons Inc., New York, 1985.

[123] L.B. Kier, A Shape Index from Molecular Graphs, Quantitative Structure-Activity Relationships 4(3) (1985) 109-116.

[124] B. Lee and F.M. Richards, Interpretation of Protein Structures - Estimation of Static Accessibility, Journal of Molecular Biology 55(3) (1971) 379-400.

[125] J.J.P. Stewart, Optimization of Parameters for Semiempirical Methods. 1. Method, Journal of Computational Chemistry 10(2) (1989) 209-220.

[126] S.A. Wildman and G.M. Crippen, Prediction of physicochemical parameters by atomic contributions, Journal of Chemical Information and Computer Sciences 39(5) (1999) 868-873.

[127] J. Jaworska, T. Aldenberg and N. Nikolova, Review of Methods for QSAR applicability domain estimation by the training test, European Commission, Joint Research Centre, Institute of Health & Consumer Protection (2005).

ISBN:978-84-691-2701-8/DL:T.386-2008

[128] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. Palyulin, E. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk and V.V. Prokopenko, Virtual computational chemistry laboratory - design and description, Journal of Computer-Aided Molecular Design 19(6) (2005) 453-463.

[129] A. Tropsha, P. Gramatica and V.K. Gombar, The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models, Qsar & Combinatorial Science 22(1) (2003) 69-77.

[130] A. Golbraikh and A. Tropsha, Beware of q(2)!, Journal of Molecular Graphics & Modelling 20(4) (2002) 269-276.

[131] T.D. Wolfe, Membrane Process Optimization Technology, 2003: Bureau of Reclamation, Desalination and Water Purification Research and Development Report No. 100.

[132] OLI Analyzer 2.0, OLI Systems, Morris Plains, NJ.

[133] ASTM D 4516-00, Standard Practice for Standardizing Reverse Osmosis Performance Data, in American Society of Testing Materials, 2000.

[134] G.J. Bowden, H.R. Maier and G.C. Dandy, Optimal division of data for neural network models in water resources applications, Water Resources Research 38(2) (2002).