Comparison of Neural Networks, Evolutionary Techniques and Thermodynamic Group Contribution Methods for the Prediction of Heats of Vaporization

Martin Mandischer[1] Hannes Geyer[2] Peter Ulbig[2]

[1] Department of Computer Science XI, University of Dortmund, Germany

[2] Department of Chemical Engineering, Chair of Thermodynamics, University of Dortmund, Germany

September 7, 1999

Abstract

In this paper we report results for the prediction of thermodynamic properties based on neural networks, evolutionary algorithms and a combination of them. We compare backpropagation trained networks and evolution strategy trained networks with two physical models. Experimental data for the enthalpy of vaporization were taken from the literature in our investigation. The input information for both neural network and physical models consists of parameters describing the molecular structure of the molecules and the temperature. The results show the good ability of the neural networks to correlate and to predict the thermodynamic property. We also conclude that backpropagation training outperforms evolutionary training as well as simple hybrid training.

Keywords: Neural Networks, Evolution Strategies, Hybrid-Learning, Chemical Engineering

Contents

1 Introduction

In chemical engineering the simulation of chemical plants is an important task. Millions of chemical compounds are known yet and experimental data are often not available. For this reason there is a need for calculation methods which are able to predict thermodynamic properties. Usually models are developed, which have a physical background and where the model parameters have to be fitted to experimental data. This leads usually to nonlinear regression models with a multi-modal objective function where encapsulated evolution strategies [GUS99, GUS98, GUSB99] are successfully used. In contrast to models with physical background simple so-called incremental methods are widely used, too. Each functional group of a molecule gives a contribution to the thermodynamic property and the sum of all contributions have to be calculated. A new way for the calculation and prediction of thermodynamic properties is the use of neural networks. Descriptors, which can be derived from the molecular structure, have to be defined for the input layer. Then experimental data for a specific thermodynamic property can be used for training. Predictions of this thermodynamic property are then possible by using the molecular structure for a chemical compound, where no experimental data are available. In this investigation the enthalpy of vaporization was taken. In section 2 we give a brief overview of the models used and continue in section 3 with an experimental comparison of physical models, networks trained with backpropagation, networks trained with evolutionary algorithms and a combination of the latter two.

2 Models for the enthalpy of vaporization

2.1 Physical Models

The physical background for the enthalpy of vaporization ΔH_v consists of electrostatic interactions forced by the atoms of the molecules. Equations can be derived from statistical thermodynamics in order to describe the interactions between molecules (first level) and between functional groups of these molecules (second level). Physical models, such as UNIFAC (UNIversal Functional Activity Coefficient) [FJP75] were developed in order to describe the real behavior of liquid mixtures. The part of the UNIFAC model, which summarizes the interactions between functional groups of the molecules within a pure liquid were taken as a basis for the development of the so-called UNIVAP model (UNIversal enthalpies of VAPorization) [KSU94, UKS96, Ulb96]. This model consists of sums of exponential terms, which include the interaction parameters and the temperature. The interactions are weighed by the surface fractions of functional groups of a molecule. The interaction parameters have to be fitted to experimental data of enthalpies of vaporization. This leads to a non-linear regression problem which objective function consists of the mean absolute error (MAE) over all experimental data points N between the calculated values (physical model) and the experimental data:

$$MAE = \frac{1}{N} \sum_{N} \left| \Delta H_v^{calc.} - \Delta H_v^{exp.} \right| \qquad (1)$$

Due to the complex structure of the physical model, especially the exponential terms, multimodality usually occurs. An encapsulated evolution strategy for solving this problem was developed [GUS99, GUS98, UFG+97, Ulb96].

In contrast to UNIFAC an extended temperature dependence was used in order to describe the behavior of the enthalpy of vaporization in principle. For the UNIVAP model it was difficult to reach the critical point, where the enthalpy of vaporization reaches null. Therefore a modified temperature dependence was used in this investigation (UNIVAP).

Another theoretical approach is the so-called EBGCM (Enthalpy Based Group Contribution Model) [KFUS99, Ulb96] in order to describe the enthalpy of mixing of binary liquid mixtures. This model is similar to UNIFAC, but has a slight different background. It was used to derive an equation for the enthalpy of vaporization, which is similar to the UNIVAP model. This so-called EBGVAP model (Enthalpy Based Group contribution model for enthalpies of VAPorization) was used in our investigation, too. For UNIVAP and EBGVAP three parameters for the interactions between functional groups of the same type have to be fitted by non-linear regression. For interactions between different kinds of functional groups six parameters have to be estimated. In principle the enthalpy of vaporization can be calculated as follows:

$$\Delta H_v = \sum_k \nu_k^{(i)} \varepsilon_k^{(i)} \quad / \, kJ/mol \tag{2}$$

R is defined as the universal gas constant of 8.314 J/(mol·K) and $\nu_k^{(i)}$ is the number of groups of kind k within the molecule i. The term $\varepsilon_k^{(i)}$ / J/mol is called group enthalpic factor of group k. This factor can be written for UNIVAP (Eq. 3):

$$\varepsilon_k^{(i)} = RT^2 Q_k \left\{ \frac{\sum_m \Theta_m^{(i)} \frac{\partial \Psi_{mk}}{\partial T}}{\sum_m \Theta_m^{(i)} \Psi_{mk}} + \sum_m \Theta_m^{(i)} \left[\frac{\frac{\partial \Psi_{km}}{\partial T}}{\sum_p \Theta_p^{(i)} \Psi_{pm}} - \Psi_{km} \frac{\sum_p \Theta_p^{(i)} \frac{\partial \Psi_{pm}}{\partial T}}{\left(\sum_p \Theta_p^{(i)} \Psi_{pm} \right)^2} \right] \right\} \tag{3}$$

and for EBGVAP (Eq. 4):

$$\begin{aligned} \varepsilon_k^{(i)} = RT^2 Q_k \Bigg\{ & \sum_m \Theta_m^{(i)} a_{mk} \frac{\sum_p \Theta_p^{(i)} \frac{\partial \Psi_{pk}}{\partial T}}{\sum_p \Theta_p^{(i)} \Psi_{pk}} \\ & + \sum_m \Theta_m^{(i)} \sum_j \Theta_j^{(i)} a_{jm} \left[\frac{\frac{\partial \Psi_{km}}{\partial T}}{\sum_p \Theta_p^{(i)} \Psi_{pm}} - \Psi_{km} \frac{\sum_p \Theta_p^{(i)} \frac{\partial \Psi_{pm}}{\partial T}}{\left(\sum_p \Theta_p^{(i)} \Psi_{pm} \right)^2} \right] \Bigg\} \end{aligned} \tag{4}$$

with Q_k is defined as the relative van der Waals surface of group k and the surface fraction of a group m within a molecule i can be calculated with:

$$\Theta_m^{(i)} = \frac{\nu_m^{(i)} Q_m}{\sum\limits_p \nu_p^{(i)} Q_p} \tag{5}$$

The interaction parameter Ψ_{mk} between the groups of kind m and k is defined as:

$$\Psi_{mk} = \exp\left(\frac{-\Delta u_{mk}}{RT}\right) \tag{6}$$

The equations for the temperature dependence for UNIVAP (Eq. 7 and EBGVAP (Eq. 8) are:

$$\Delta u_{mk} = R \cdot (a_{mk} + b_{mk}T + \exp(c_{mk}T^2) \cdot \mathrm{K}). \tag{7}$$

$$\Delta u_{mk} = b_{mk}T + 1K \exp\left(c_{mk}T^2\right) \tag{8}$$

Here a_{mk}, b_{mk} and c_{mk} are the interaction parameters, which have to be fitted. Considering Eq. 3 and 4, the heat of vaporization in Eq. 2 should have the unit J/mol. Usually, heats of vaporization extend over a range between 0 J/mol and $> 10^5$ J/mol. This leads to difficulties in the optimization procedure, because exponential terms describing the temperature dependence as given in Eq. 7 and 8 cannot correlate data within this large range with satisfying results. A factor of about 1000 is introduced and therefore the output of Eq. 2 is set to kJ/mol.

2.2 Neural Networks

Neural networks are able to acquire an internal model of a process by learning from examples. After successful training the network will be a model for the process which led to the experimental data. Theoretical results show that feed-forward networks are capable of arbitrary exact function approximation, given an unlimited number of free parameters or infinite precision [HSW89].

In our experiments we used simple feed-forward networks with non-linear sigmoid activation functions. The network model can be written as Eq. 9:

$$o = \frac{1}{1 + e^{-\left(\left(\sum\limits_{i=1}^{n} w_{i,o} \cdot hid_i\right) + w_{bias,o}\right)}} \tag{9}$$

$$hid_i = \frac{1}{1 + e^{-\left(\left(\sum\limits_{j=1}^{m} I_j \cdot w_{j,i}\right) + w_{bias,i}\right)}} \tag{10}$$

with n as number of hidden units, m as number of input units and I_j as input value for unit j.

As training algorithms for the networks weights $w_{i,j}$ we employed the standard Backpropagation algorithm [RM86] and various (μ,λ) evolution strategies [Sch95, Bäc96] as well as a combination of both.

3 Experiments and Results

Comparing different methods or models is at least two-fold. On the one hand a fair comparison should allow all models the same number of free parameters to adjust to the problem. On the other hand, one can say that it is sufficient if a model performs good on formerly unseen data regardless of the number of parameters it needed. In both cases our main concern is generalization capability.

Some of our experiments were designed to find good neural models under the most similar conditions for the calculations as the physical models. Here the number of adjustable parameters was almost the same for all models. In other experiments we searched for good results independent of the number of free parameters (weights) used. One difficulty is to find the optimal structure of the neural network and the optimal structure of the temperature dependent equation of the physical model. Here we only investigated the structure of the network. Another important issue is to have the same input information for all methods, which can be derived from the structure of the molecules.

3.1 Generation and Description of the Data

3.1.1 Selection of data

The experimental data concerning the enthalpy of vaporization were taken from different data handbooks [MSK85, SVZ+87, SJO86]. Data for three different classes of chemical compounds were used: normal alkanes, 1-alcohols, and branched alcohols. These data were chosen for the investigation of three (3MG) and five (5MG) different functional groups, the so-called main groups: CH_3, CH_2 and CH_nOH. The group CH_nOH contains the functional groups CH_3OH, CH_2OH and CHOH. The experimental data for both data sets cover a temperature range from 92 K to 776 K. The number of carbon atoms in the n-alkanes ranges from 2 (Ethane) to 19 (Nonadecane), for the 1-alcohols from 1 (Methanol) to 14 (Tetradecanol) and for the branched alcohols from 4 (2-Methyl-2-propanol) to 6 (2-Methyl-2-pentanol). The preprocessing steps and experimental setting were the same for the 3MG and 5MG data sets.

3.1.2 Selection of descriptors

There are several possibilities for the definition of descriptors as input variables for a neural network: number of atoms, number of single bonds, molar mass, dipole moment and topological parameters concerning the connectivity between atoms [EJ93]. In our investigation the descriptors for the input layer are the surface fractions of the functional groups within a molecule and the temperature. Therefore a definition of functional groups is necessary. Here the definition of the UNIVAP model [KSU94, UKS96, Ulb96] shall be used.

3.1.3 Partitioning into subsets for cross validation

After generating a data set (either 3MG or 5MG) it was subdivided into 3 classes: training (50%), validation (25%) and test (25%) set. Only the training set was used to adapt the parameters for all our models. The validation set could be used for NN during the adaption process to evaluate the algorithms performance on unknown data and stop the adaption process if the error on the validation set increases. The validation set do not have any influence in the parameter fitting procedure of the NN or our physical models UNIVAP and EBGVAP. Validation and test set therefore measure the generalization ability of all our models. However, 50% of the data were used only for comparison, i. e. for a test of the prediction of the enthalpy of vaporization. The distribution of the data in the 3MG and 5MG data sets can be seen in Table 1 and Table 2.

3.1.4 Transformation

For the use with the neural network the data were normalized via separate linear transformations of main-groups, temperature and enthalpy to the interval [0.1 .. 0.9]. Network responses outside of this interval were mapped onto the boundaries and then re-transformed to the original scale.

3.2 Physical Model Experiments

Only the training set was used for the non-linear regression of the interaction parameters and for the training of the neural network. First the parameters were computed successive, i. e. first the 12 parameters for the interactions $CH_3 \leftrightarrow CH_3$, $CH_3 \leftrightarrow CH_2$ and $CH_2 \leftrightarrow CH_2$ were fitted to the training data set. After this optimization process (corresponding to Table 1), these 12 parameters are needed in the fitting procedure of the remaining parameters of the interactions $CH_nOH \leftrightarrow CH_nOH$, $CH_3 \leftrightarrow CH_nOH$ and $CH_2 \leftrightarrow CH_nOH$, because data points of substances are used, which contain the main groups CH_3 and CH_2, too. The advantage of an sequential fitting procedure is to keep the dimension space as small as possible. These sequential experiments for the physical models were done with the aid of an repeatedly started encapsulated evolution strategy

Group interaction	n_p	$n_{data,total}$	$n_{data,training}$
CH_3 CH_3/CH_3 CH_2/CH_2 CH_2	12	248	128 (51.61 %)
CH_nOH CH_nOH/CH_3 CH_nOH/CH_2 CH_nOH	15	181	86 (47.51 %)
total:	27	429	214 (49.88 %)

Table 1: Number of experimental data for the different group interactions (3MG)

[GUS98] by using a multidimensional but non-correlated step-length control and a panmictic generalized intermediate recombination of the objective and the strategic variables: `5*[GG 4+8(GG 7+19)`^{300}`]`^{30}. The used encapsulated evolution strategies were repeated four times with $6.8 \cdot 10^6$ function calls in all, in order to guarantee the determination of the global minimum or of a fairly good local minimum.

The determined results were optimized by a multi-start simplex-algorithm [NM65] with 50 different runs of 2500 iterations each. The best result for UNIVAP (seq) and EBGVAP (seq) can be found in Table 3 and 4. In contrast to this sequential regression of the model parameters a simultaneous regression (sim) of all 27 (3MG) respectively 75 (5MG) parameters was investigated by using the same encapsulated ES as for the sequential experiments. The determined mean absolute errors of these runs were improved by a multi-start simplex-method as well by using 50 different runs of 3000 iterations each. The results can be seen in Table 3 and 4, too.

3.3 Neural Networks Experiments (Backpropagation)

The learning rate η and the architecture of the network (number of hidden units and connections) have the biggest influence on the performance of the network [Man95]. In order to find good neural network solutions we did a primitive parameter study. We first varied the learning rate with a fixed architecture which had approximately the same number of free parameters (connections) as the UNIVAP respectively as the EBGVAP model. With the best learning rate found, we searched for a good number of hidden units. All runs were performed 10 times.

3.3.1 Variation of the learning rate

We fixed the architecture of the network at 4 input, 4 hidden and 1 output units (4-4-1) for the 3MG data at 6 input, 5 hidden and 1 output units (6-5-1) for the 5MG data. This was done to have approximately the same number of free parameters ($25 = 4 \times 4 + 4\,\text{bias} + 4 + 1\,\text{bias}$) as the UNIVAP method.

For both data sets (3MG and 5MG) we started with a very low learning rate $\eta = 0.001$ and ended with a far too high rate $\eta = 10.0$ (0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0). The momentum term α was fixed to 0.2. A training run was stopped after it reached the error limit ($tss \leq 5 \times 10^{-5}$) or

Group interaction	n_p	$n_{data,total}$	$n_{data,training}$
CH_3 CH_3/CH_3 CH_2/CH_2 CH_2	12	248	130 (52.42 %)
CH CH/CH CH_2/CH CH_3	15	133	58 (43.94 %)
C C/C CH/C CH_2/C CH_3	21	52	28 (53.85 %)
CH_nOH CH_nOH/CH_3 CH_nOH/CH_2 CH_nOH	15	181	89 (49.17 %)
CH CH_nOH/C CH_nOH	12	40	22 (55.00 %)
total:	75	654	327 (50.00 %)

Table 2: Number of experimental data for the different group interactions (5MG)

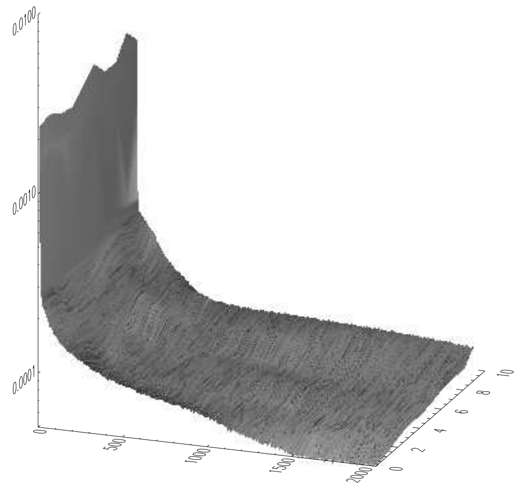

Figure 1: Training error 3MG (η=0.8)

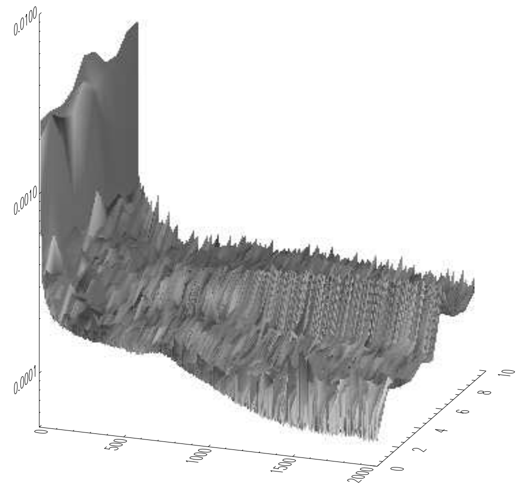

Figure 2: Validation error 3MG (η=0.8)

exceeded a maximum number of 100,000 pattern presentations (epochs). The error is defined as:

$$tss = \frac{1}{2} \times \sum_{i=1}^{n} (\theta_i - o_i)^2 \tag{11}$$

with θ as target vector and o as output activation of the network. Figure 1 and 2 show the curves for 10 different runs (3MG) with the best learning rate which was used throughout all other experiments. The left-hand side figure gives the error on the training set and on the right-hand side we see the validation error. If an error curve reaches the base of the graph it satisfied a specified error limit for the whole training set. Networks with very low learning rate never reached the specified error limit, due to the very slow learning progress. A too high rate resulted in oscillating error curves. The figures look the same for the 5MG data set.

3.3.2 Variation of the number of hidden units

After variation of η we used the best rate as a constant for the hidden unit search[1]. The number of hidden units were varied between 1 and 40. Networks with less then 3 units failed to learn the task. Up to 40 units the results on training as well as validation data were almost independent of the number units employed. We therefore used our initial 4-4-1 network for the 3MG data and a 6-5-1 network for the 5MG data. This is an additional advantage because results can now be directly compared to other methods which use the same number of free parameters.

[1]This does not mean that both parameter are independent of each other. We consider this value to be a first estimate to start with.

3.4 Neural Networks Experiments (Evolution Strategy)

In this experiment we substituted the Backpropagation algorithm with an evolution strategy. Some authors [Wie93] reported good results when training a network with an ES. Again we systematically searched for a good parametrisation of the (15,100)–ES. Parameters under consideration were the number of mutation step-sizes σ_i and the recombination scheme used on the object variables x_i (the network weights). Each parameter setting was run for 10,000 generations (1,000,000 pattern presentations) and repeated 10 times to have some statistical validity. All of the following variations of the bisexual recombination scheme were done with 1 and 25 σ.

- no recombination of x_i and σ_i,
- discrete recombination of x_i and discrete of $\sigma_{v_i v_j}$,
- discrete recombination of x_i and intermediate of $\sigma_{v_i v_j}$,
- intermediate recombination of x_i and discrete of $\sigma_{v_i v_j}$,
- intermediate recombination of x_i and intermediate of σ_i.

For details on ES and recombination types see [Bäc96, Sch95].

None of the parameter settings lead to good and reliable results. Only one out of all ES trained network performed comparable to Backpropagation. All other networks give rather poor results. The quality of the average result did improve when using backpropagation as local search procedure (an additional training of 250,000 epochs) after ES optimization but was not as good as Backpropagation alone. Figure 3 shows the best run, which we regard as a very rare event, with a (15,100)–ES. We did not perform any ES experiments on the 5MG data set.

3.5 Results and Comparison

3.5.1 The 3 main groups data set (3MG)

For a comparison between the physical models and NN, we took two network architectures with learning rates gained by the previous experiments. Architecture A has 4 hidden units and nearly the same number of free parameters (25 weights) as the UNIVAP respectively the EBGVAP model (27). Architecture B performs alike and has 6 hidden units (37 weights).

1. Parameters for NN-A (4-4-1): η=0.8, epochs=250,000
2. Parameters for NN-B (4-6-1): η=0.8, epochs=250,000
3. Parameters for NN-ES (4-4-1): best (15,100)–ES $\#\sigma = n$, intermediate recombination of x_i and σ_i (100,000 generations)

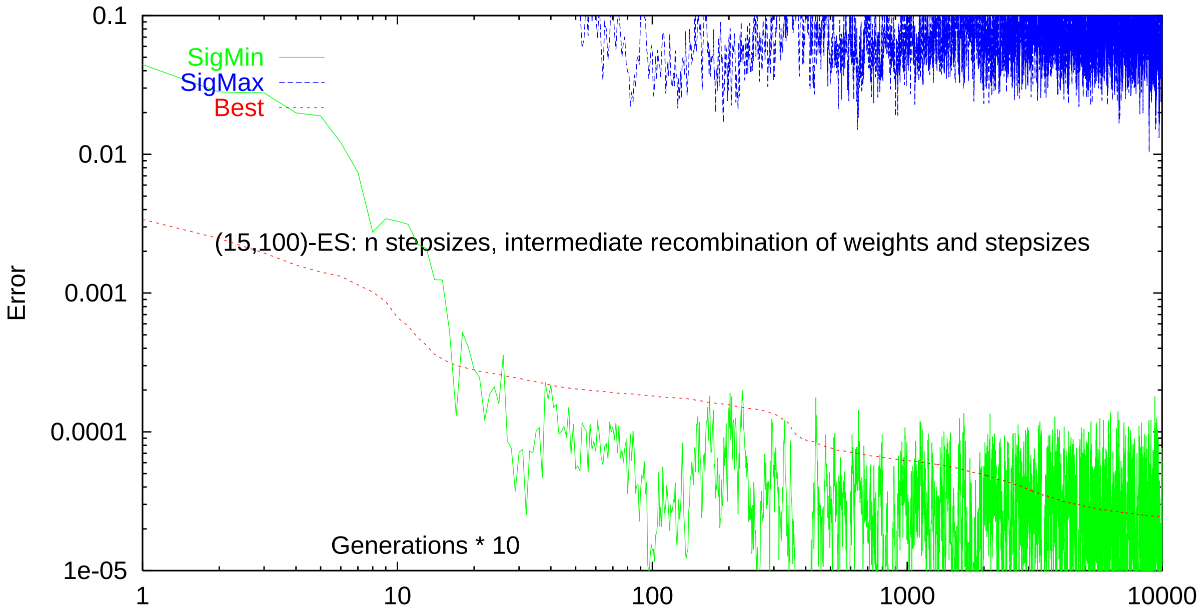

Figure 3: (15,100)–ES on 3MG (error during training)

	UNIVAP (seq)	EBGVAP (seq)	UNIVAP (sim)	EBGVAP (sim)
Train	0.681	0.617	0.881	0.720
Valid	0.941	0.750	0.966	0.829
Test	0.557	0.576	1.003	0.766
All	0.716	0.640	0.933	0.759
	NN-A	NN-B	NN-ES (best)	NN-ES (avrg)
Train	0.652	0.570	0.612	1.143
Valid	0.566	0.878	0.876	1.536
Test	0.686	0.703	0.747	1.357
All	0.638	0.679	0.711	1.292

Table 3: Mean absolut error per pattern for different data sets and models (3MG)

4. Parameters for NN-ES (4-4-1): average (15,100)–ES $\#\sigma = n$, intermediate recombination of x_i and σ_i (100,000 generations)

Table 3 gives an overview of all experiments. In the first place, the results determined by our physical models UNIVAP and EBGVAP show, that the newer group contribution model EBGVAP is more suitable than UNIVAP for the correlation and prediction of heats of vaporization because of its better physical background. On the other hand the results show a superiority of a sequential fitting procedure. The Neural Network performs even slightly better than the best physical model.

From the errors of the validation and test set we can derive the generalisation capabilities of the different models. The best generalisation is given by network A (backpropagation) very closely followed by EBGVAP (seq), whereas the worst generalisation is delivered by the same network trained with an Evolution Strategy and the UNIVAP (sim) model.

Figure 4 shows the errors of network A on all data sets. The errors are sorted by size and bars depict the target, whereas the dots are the networks predicted values. Except for the

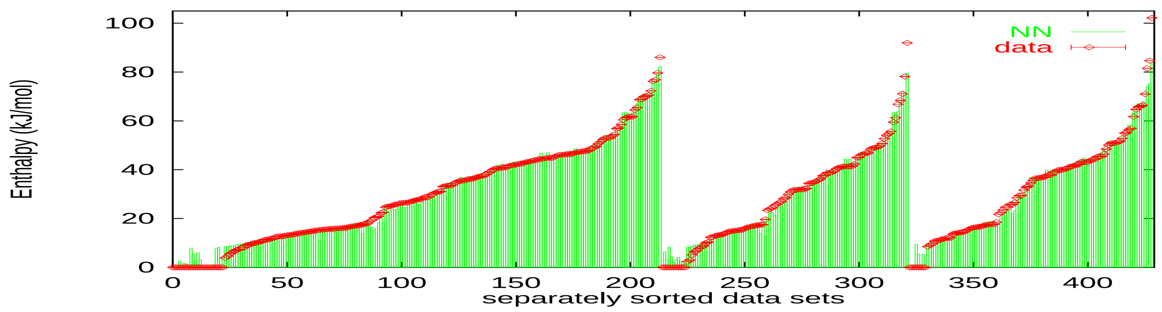

Figure 4: NN (backpropagation) errors on training, validation and test set (3MG)

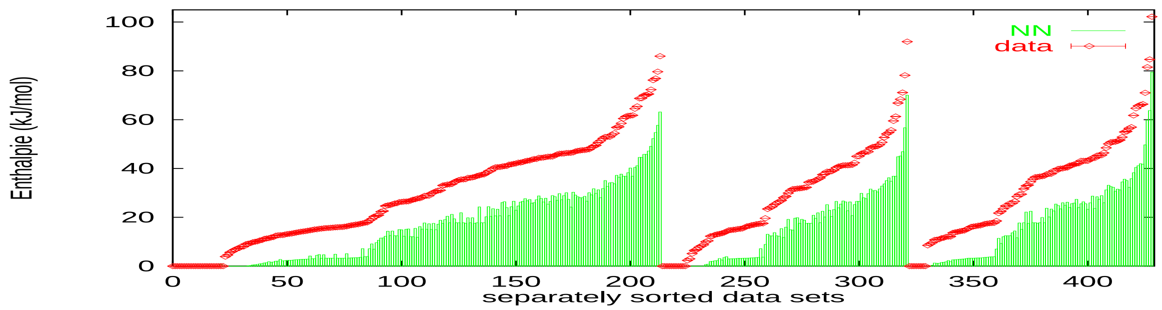

Figure 5: NNES (Evolution Strategy) errors on training, validation and test set (3MG)

critical regions close to $\Delta H_v(T_{cr}) = 0$ J/mol, the network comes very close to the desired values. The network performs equally well on training, validation and test data. In figure 5 we see the errors for the same network trained with an Evolution Strategy. The errors are rather high for all three data sets.

As an additional test for generalisation ability, we used all data of an ethane molecule in a range from 92 K to 305 K. In figures 6 and 7 we compare all models on the enthalpy prediction for ethane. It can be seen that the physical model EBGVAP and the neural network performs equally well on this task, except for the critical regions near $T \to T_{cr}$ and $\Delta H_v(T_{cr}) = 0$ J/mol, where the network outperforms all other models. The prediction by using the network is however characterized by a point of discontinuity near $T \approx 100$ K, which is not thermodynamically interpretable. A prediction by Neural Networks can be therefore only used partly over the whole temperature range.

Almost all networks trained with an ES give only a poor approximation of the enthalpy curve. In comparison to Table 3 the superiority of EBGVAP towards UNIVAP can be also seen in the figures 6 and 7 because of its smaller deviation at temperatures smaller than $T = 150$ K and at temperatures near the critical temperature of T_{cr}(Ethane)= 305.4 K.

3.5.2 The 5 main groups data set (5MG)

As in the 3MG experiments, we took two network architectures with learning rates gained by the previous experiments. Architecture A has 5 hidden units (41 weights), architecture B has 9 hidden units (73 weights) that is nearly the same number of free parameters of the physical models UNIVAP and EBGVAP (75 parameters).

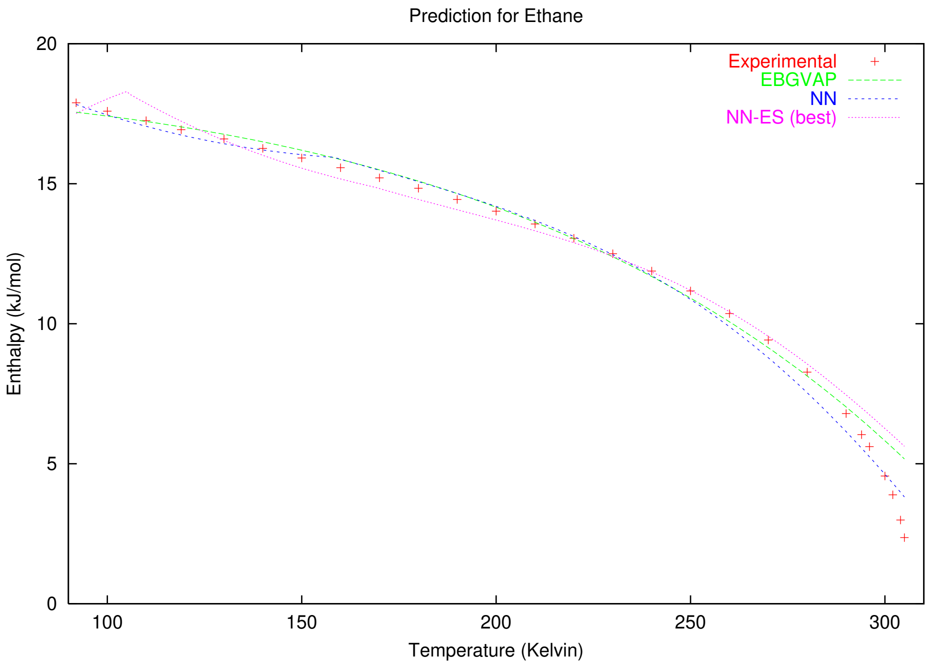

Figure 6: 3MG - Performance on ethane (good)

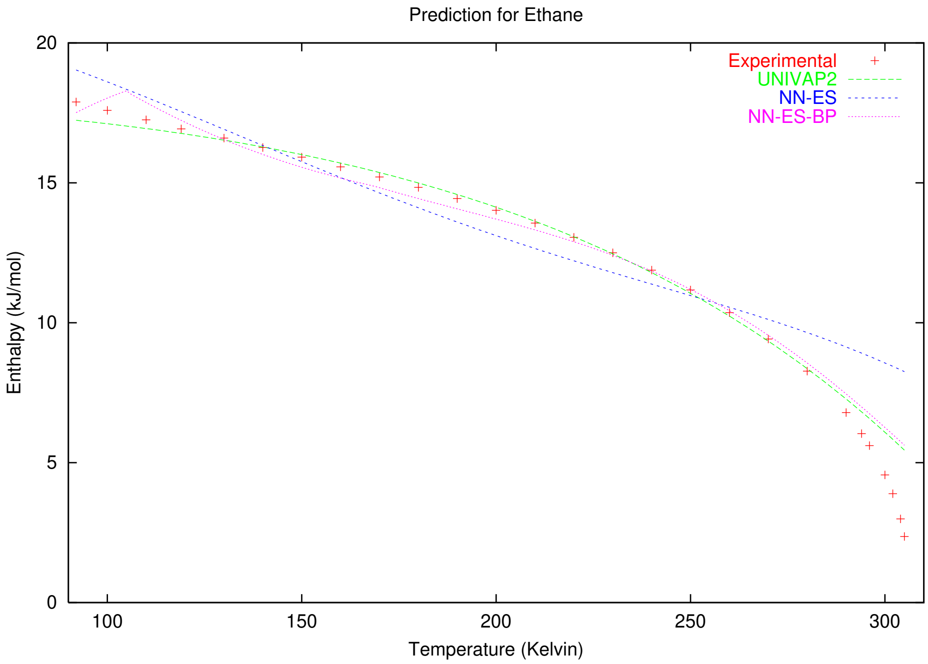

Figure 7: 3MG - Performance on ethane (poor)

	UNIVAP (seq)	EBGVAP (seq)	UNIVAP (sim)	EBGVAP (sim)
Train	0.667	0.609	1.764	1.275
Valid	1.050	0.805	2.077	1.579
Test	1.180	0.939	2.013	1.707
All	0.891	0.740	1.904	1.459
	NN-A	NN-B		
Train	0.702	0.564		
Valid	0.649	0.589		
Test	0.904	0.744		
All	0.737	0.614		

Table 4: Mean absolut error per pattern for different data sets and models (5MG)

1. Parameters for NN-A (6-5-1): η=0.9, epochs=250,000

2. Parameters for NN-B (6-9-1): η=0.9, epochs=250,000

Table 4 gives on overview of all 5MG experiments. Again we see that the EBGVAP model is superior to the UNIVAP model but both neural networks perform better than the best physical model. Network A has only 55 % of the free parameters of the models UNIVAP and EBGVAP but gives slightly better results, wheras network B with nearly the same number of parameters is significantly better. With the increased problem size the simultaneously adaption method of parameters loses even more ground compared to the sequential method. The results concering the physical models show the need of decreasing the dimension of the variable space. It is obvious, that an simultaneous optimization of 75 parameters in all did not lead to satisfying results. To split the optimization procedure of all 75 parameters into several sequential optimizations by using already fitted constant parameters lead to the best results which could be seen in the figures 10 and 11.

Figure 8 shows the errors of the network A on all data sets. The errors are sorted by size and bars depict the target, whereas the dots are the networks predicted values. Except for the critical regions close to $\Delta H_v(T_{cr}) = 0$ J/mol, the network comes very close to the desired values.

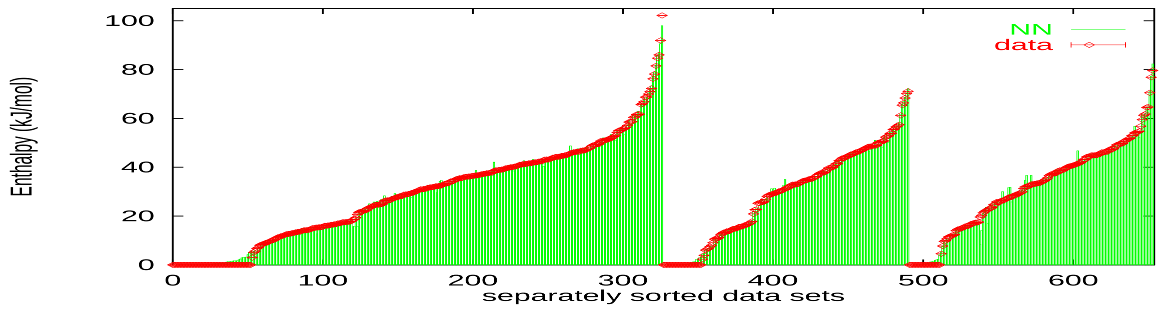

Figure 8: NN errors on training, validation and test set (5MG)

In figures 9 we again compare all models on the enthalpy prediction for ethane.

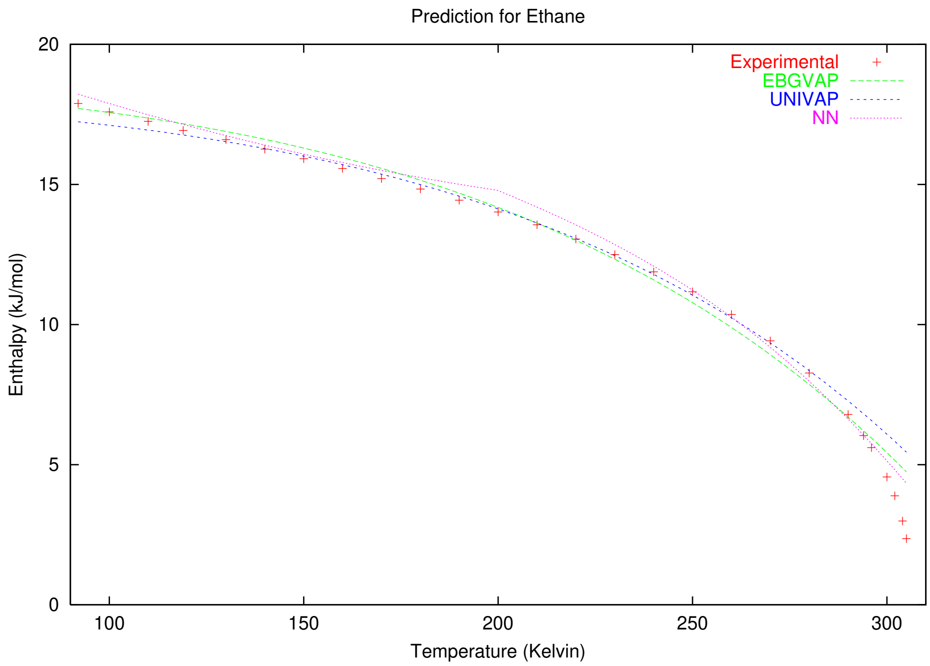

Figure 9: 5MG - Performance on ethane

When we take a closer look at the performance of the sequential and simulataneous adaption of the physical model (figures 10 and 11), we clearly see that the sequential method outperforms the other on both problem sizes. The superiority of the sequential fitting procedure can be explained by the negative influence of increasing numbers of parameters for the optimization process. The dimension of 27 simultaneously fitted parameters in all makes the adaption of the strategic variables used by (encapsulated) evolution strategies more difficult in contrast to sequential fitting procedures, which result in smaller dimensions. With the increased problem size from 3MG to 5MG the simultaneous adaption methods is 2-3 times worse than the sequential one and the generalisation performance is even worse. The NN instead is not sensitive to an increase in the number of free parameters.

4 Discussion

The most important result of this investigation is the good ability to correlate as well as to predict the enthalpy of vaporization with neural and physical methods. Neural networks with simple Backpropagation training are as good as the physical based group contribution methods UNIVAP and EBGVAP and especially at critical temperatures even slightly better, but their computational effort is much lower. Precitions by using Neural networks however are often characterized by points of discontinuity over the thermodynamic significant temperature range, which is shown in the figures 6,7 and 9. These points are not thermodynamically interpretable, so that a prediction by Neural Networks should only be done carefully and should only be used partly over the whole temperature range.

The results concerning the sequential versus simultaneous parameter optimization of the 3MG and the 5MG data sets show the need of a relative relative small dimension of variable space by carrying out a sequential optimization, where thermodynamic information is

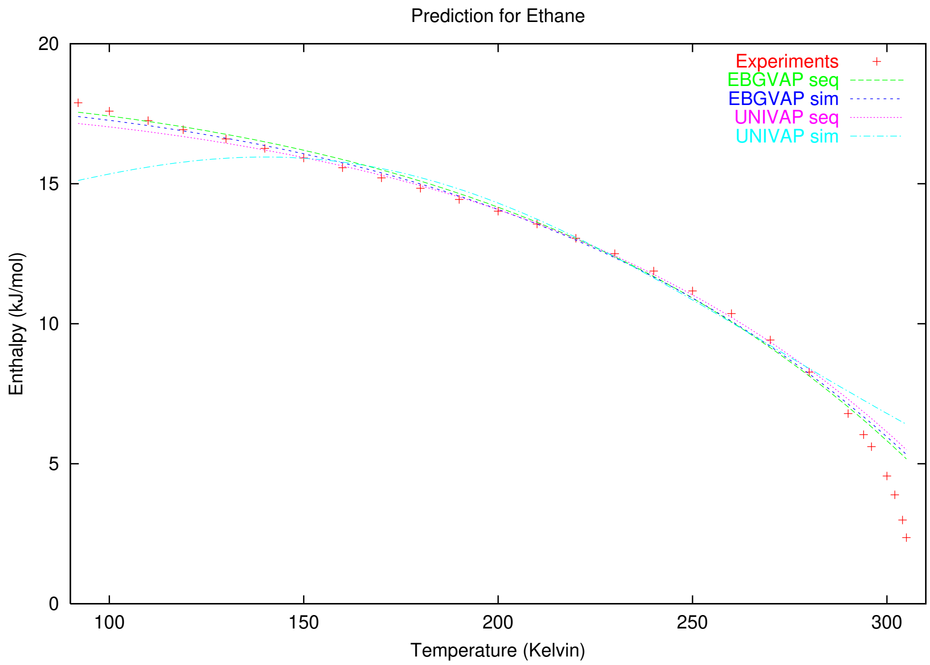

Figure 10: Physical models: sequential vs. simultaneous (3MG)

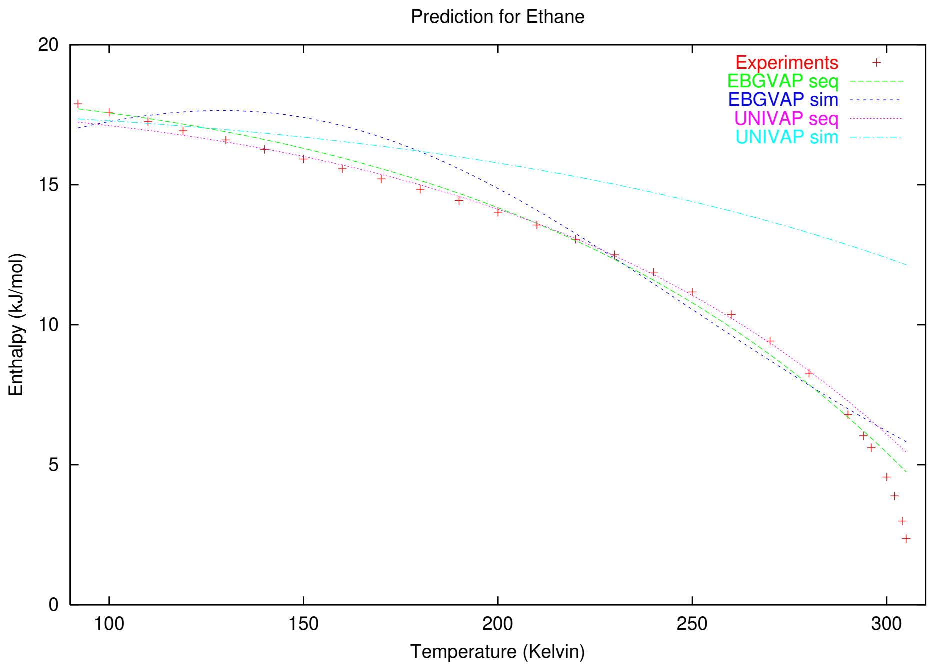

Figure 11: Physical models: sequential vs. simultaneous (5MG)

included as much as possible. Optimization of interaction parameters by using already fitted constant parameters could lead to incompatibilties during the fitting procedure [KFUS99] but results obviously in better determined mean absolute errors, which is shown in the figures 10 and 11.

The comparison of the results for UNIVAP and EBGVAP shows the influence of the structure of the model itself. Further investigations could use evolutionary algorithms to optimize the structure of the models with regard to the temperature dependence. For the neural networks it can be stated that the use of surface fractions of functional groups as descriptors for a neural network leads to good results for both correlation and prediction. The big advantage of this new procedure is, that the molecules can easily be divided into functional groups, which makes it easy to use in engineering applications and allows the direct comparison of neural networks and physical models, due to the same input information. The investigations concerning the architecture of the neural networks show, that a simple network structure is sufficient and a more complicated network does not give better results. In this context evolution strategies as training algorithms and combinations of ES with backpropagation failed to deliver useful models in almost all experiments.

From a thermodynamic point of view, it is interesting that a simple method like a neural network can give similar results in comparison with much more complicated physical motivated models. If a physical model gives results with a quality less than a neural model, the physical model should be improved. However, in chemical engineering there are many thermophysical properties, which are usually not described by physical methods, but by incremental methods. These methods, for example, for critical data, normal boiling points and so on, could be replaced by neural networks. However, these results are first steps in developping efficient network structures for our purpose and especially investigations with more functional groups will give a better comparison between physical models and neural networks.

5 Acknowledgements

The work presented is a result of the *Collaborative Research Center SFB 531* sponsored by the Deutsche Forschungsgemeinschaft (DFG).

References

[Bäc96] T. Bäck. *Evolutionary Algorithms in Theory and Practice.* Oxford Univ. Press, New York, 1996.

[EJ93] L. M. Egolf and P. S. Jurs. Prediction of boiling points of organic heterocyclic compounds using regression and neural network techniques. In *J. Chem. Inf. Comput. Sci. 33*, pages 616–625. 1993.

[FJP75] A. Fredenslund, R. L. Jones, and J. M. Prausnitz. In *AIChE Journal 21*, page 1086. 1975.

[GUS98] H. Geyer, P. Ulbig, and S. Schulz. Encapsulated evolution strategies for the determination of group contribution parameters in order to predict thermodynamic properties. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature 5*, pages 978–987. Springer, Amsterdam, 1998.

[GUS99] H. Geyer, P. Ulbig, and S. Schulz. Use of evolutionary algorithms for the calculation of group contribution parameters in order to predict thermodynamic properties. Part 2: Encapsulated evolution strategies. In *Computers and Chem. Eng. 27*, volume 7, pages 955–973. 1999.

[GUSB99] H. Geyer, P. Ulbig, S. Schulz, and P. Bräuer. Vergleich zwischen klassischen und verschachtelten Evolutionsstrategien am Beispiel einer nichtlinearen Regression an Oberflächenspannungen in $\mathbb{R}^2$. Reihe CI 66/99, Sonderforschungsbereich 531, Universität Dortmund, März 1999.

[HSW89] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[KFUS99] C. Kracht, T. Friese, P. Ulbig, and S. Schulz. Development of an Enthalpy Based Group Contribution g^E-Model (EBGCM). In *The Journal of Chemical Thermodynamics 31*, volume 5, pages 587–614. 1999.

[KSU94] M. Klüppel, S. Schulz, and P. Ulbig. UNIVAP - A Group Contribution Method for the Prediction of Enthalpies of Vaporization of Pure Substances. In *Fluid Phase Equilibria 102*, pages 1–15. 1994.

[Man95] M. Mandischer. Evolving recurrent neural networks with non-binary encoding. In *Proc. Second IEEE Int'l Conf. Evolutionary Computation (ICEC '95), vol. 2*, pages 584–589, Perth, Australia, 1995. IEEE Press, Piscataway NJ.

[MSK85] V. Majer, V. Svoboda, and H. V. Kehiaian. Enthalpies of vaporization of organic compounds. In *Blackwell Scientific Publications, Oxford*. 1985.

[NM65] J. A. Nelder and R. Mead. A simplex method for function minimization. In *Computer Journal 7*. 1965.

[RM86] D. E. Rummelhart and J. L. McClelland. *PDP: Explorations in the Microstructure of Cognition*, volume 1. MIT Press, Cambridge, MA, USA, 1986.

[Sch95] H.-P. Schwefel. *Evolution and Optimum Seeking*. Sixth-Generation Computer Technology. Wiley, New York, 1995.

[SJO86] K. A. Simmrock, R. Janowsky, and A. Ohnsorge. In *Critical Data of Pure Substances, Vol. I and II*. Dechema, Frankfurt, 1986.

[SVZ+87] V. Sychev, A. A. Vasserman, V. A. Zagoruchenko, G. A. Spiridonov, and V. A. Tsymarny. Thermodynamic properties of ethane (vol. 4). In *National Standard Reference Data Service of the USSR, A series of property tables*. 1987.

[UFG+97] P. Ulbig, T. Friese, H. Geyer, C. Kracht, and S. Schulz. Prediction of thermodynamic properties for chemical engineering with the aid of Computational Intelligence. In *Progress in Connectionist-Based Information Systems*, pages 1259–1262. Springer, 1997.

[UKS96] P. Ulbig, M. Klüppel, and S. Schulz. Extension of the UNIVAP group contribution method: enthalpies of vaporization of special alcohols in the temperature range vom 313 to 358 K. In *thermochimica acta 271*, pages 9–21. 1996.

[Ulb96] P. Ulbig. *Gruppenbeitragsmodelle UNIVAP & EBGCM*. Dr.-Ing. Thesis, University of Dortmund, Institute for Thermodynamics, 1996.

[Wie93] W. Wienholt. Minimizing the system error in feedforward neural networks with evolution strategy. In Stan Gielen and Bert Kappen, editors, *Proc. of the Int'l. Conf. on Artificial Neural Networks*, pages 490–493, London, 1993. Springer-Verlag.