7

Construction of Optimal Artificial Neural Network Architectures for Application to Chemical Systems: Comparison of Generalized Pattern Search Method and Evolutionary Algorithm

Matthias Ihme
Department of Aerospace Engineering, University of Michigan
Ann Arbor, MI 48109,
USA

1. Introduction

Artificial neural networks (ANNs) are computational models of their biological counterparts. They consists of densely interconnected computing units that work together to solve a specific problem. The information, which is acquired during a learning process, is stored in the synaptic weights of the internodal connections. The main advantage of neural networks is their ability to represent complex functions and the efficient storage of information. ANNs are frequently employed in applications involving data classification, function approximation, and signal processing (Haykin, 1994).

The topology of ANNs consists of an arrangement of neurons, which are equipped with a transfer function and synaptic weights, and the nodal connections. Despite these simple topological elements, the flexible arrangement of neurons and connections allows the generation of ANNs with arbitrary complexity. The resulting topological complexity, however, directly affects the network performance. The performance, or fitness, is a measure of the accuracy of a network in representing an input-output relation. For instance, network topologies with only few neurons and synaptic weights provide only limited flexibility in representing complex functions. They have therefore typically only a poor fitness. On the other side, complex networks that provide large flexibility in representing new data, can lead to poor generalizability and extensive computational costs for training and data retrieval (Yao, 1999). Considering the integration of such networks in large-scale simulations, data retrieval from such large networks can lead to a significant increase in overall computing time. Because of the inherent topological complexity it is apparent that the *a priori* identification of a network topology with near-optimal performance is a challenging task, and is often guided by heuristics or trial-and-error.

The design of a specific network topology with optimal performance can be formulated as an optimization problem. The choice of the method to solve this problem is determined by the inherent properties of the ANN (Miller et al., 1989):

1. The dimension of the architecture space is infinite since the number of neurons and nodal connections is unbounded.

2. Assessing the performance of a particular network topology usually requires the prior training of the network, which in itself is an optimization problem and usually time consuming.
3. The design parameters determining the topology of the network can be both of continuous and categorical type.
4. The objective function that quantifies the network performance with respect to the design variables is noisy and non-differentiable.
5. Different ANN topologies can lead to similar performance, implying that the solution space is multimodal.

Over recent years considerable research has been conducted on the evolution of topological structures of networks using evolutionary strategies (Bornholdt & Graudenz 1992; Fogel & Fogel 1990; Husken et al. 2005; Koza & Rice 1991; Miller et al. 1989; Porto et al. 1995; Tang et al. 1995 and references in Yao 1999). Evolutionary strategies (ESs) are global search algorithms and have frequently been used to find optimal network topologies and nodal transfer functions. ESs can conveniently be implemented in an existing code and do not require gradient information. Despite their popularity, ESs are known to be expensive, typically characterized by slow convergence, and usually lack formal convergence theory. Other methods which have been employed for the automatic design of near-optimal networks are so-called construction and destruction algorithms (Frean, 1990; Mozer & Smolensky, 1989), in which neurons are systematically added or deleted with the objective to improve the network fitness. These methods, however, search only in a restricted subset of possible network architectures (Angeline et al., 1994).

As an alternative to these optimization techniques Ihme et al. (2008) proposed to use a generalized pattern search (GPS) method for the generation of optimal ANNs. They applied the GPS method to the optimization of multi-layer perceptrons (MLPs), and considered the number of neurons, transfer functions, and the nodal connectivity as free parameters in the optimization problem. The GPS algorithm is a derivative-free, mesh-based optimization method and provides robust convergence properties (Audet & Dennis, 2003). To increase the efficiency of the GPS for computationally expensive problems, this method can be complemented by a surrogate representation, which was developed by Serafini (1998) and Booker et al. (1999).

The objective of this work is to utilize the GPS method and an evolutionary strategy for the generation of optimal artificial neural networks (OANNs) to approximate non-linear functions, that are, for instance, encountered in representing chemical systems. These chemical system can be formulated as:

$$\mathcal{D}(\underline{\phi}) = \underline{w}(\underline{\phi}) \,, \tag{1}$$

where $\mathcal{D}$ is a linear operator acting on $\underline{\phi}$ (in the simplest case, $\mathcal{D}$ is the temporal derivative ∂_t), $\underline{\phi} \in \mathbb{R}^{N+1}$ denotes the vector of N chemical species and temperature, and $\underline{w} : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$ is a function, representing the chemical source terms and the heat release rate.

Chemical mechanisms often comprise thousands of reactions among hundreds of species. In numerical simulations of combustion systems the direct solution of transport equations for all of these species is usually not feasible. Alternative approaches, such as intrinsic low dimensional manifolds (Maas & Pope, 1992) or computational singular perturbation (Lam & Goussis, 1988) have been developed, in which a part of the chemical species are projected onto a lower dimensional manifold, resulting in a reduced chemical reaction mechanism. However,

the necessary introduction of certain assumptions, required for reducing the mechanism, can result in a degradation of the accuracy and generality.

On the other side, tabulation techniques, such as conventional structured look-up tables, are often employed for the parameterization of thermochemical quantities. Since, however, the memory requirement for the tabulation rapidly increases with the number of independent parameters, this method imposes drastic restrictions when more than three or four independent parameters are used. Other tabulation techniques include the in situ adaptive tabulation (ISAT) (Pope, 1997) or solution mapping using piecewise polynomial approximation (PRISM) (Tonse et al., 1999).

Over recent years, ANNs have successfully been employed for the approximation of chemical systems (Blasco, Fueyo, Dopazo & Ballester, 1999; Blasco et al., 2000; Blasco, Fueyo, Larroya, Dopazo & Chen, 1999; Chen et al., 2000; Christo, Masri & Nebot, 1996; Christo, Masri, Nebot & Pope, 1996; Flemming et al., 2005; Ihme et al., 2009; Sen & Menon, 2008; 2010). Important advantages of ANNs over tabulation methods are the modest memory requirement, and cost-effective and smooth function representation. However, in many if not all of these applications *ad hoc* network topologies were used, that were not fully optimized for the particular problem, so that the optimal performance could not be achieved.

Motivated by the chemistry application, the objective of this work is to demonstrate the potential of OANNs for application to chemical reacting flows. To this end, two different chemical systems of increasing complexity are considered. Specifically, the first problem considers an one-step chemical reaction in a homogeneous flow, representing decaying turbulence. The particular advantage of this problem is that it allows us to systematically evaluate different ANN-representations and compare the results against other predictions. The second problem considers the unsteady three-dimensional combustion of a methane/hydrogen-air mixture in a technical-relevant burner system. For this, large-scale simulations are employed and the accuracy of ANNs and conventional tabulation methods are assesses in the context of high-performance computations of turbulent reacting flows.

In order to assess the ANN performance, we consider two metrics, namely the ANN fitness evaluation under static and dynamic conditions. To explain both metrics, we consider Eq. (1), in which the source term is now approximated through an ANN:

$$\mathcal{D}(\underline{\varphi}) = \underline{w}_{\mathcal{A}}(\underline{\varphi}) \; . \tag{2}$$

where $\underline{\varphi} = \underline{\phi} + \underline{\varepsilon}$ and $\underline{w}_{\mathcal{A}}$ is the source-term representation by the ANN. After writing $\underline{w}_{\mathcal{A}}$ as

$$\underline{w}_{\mathcal{A}} = \underline{w} + \underline{\Omega} \; , \tag{3}$$

with $\underline{\Omega}$ denoting the ANN approximation error, Eqs. (1) and (2) can be combined to derive the following expression for the evolution of the error $\underline{\varepsilon}$:

$$\mathcal{D}(\underline{\varepsilon}) = \underline{\Omega}(\underline{\phi}) + \underline{\varepsilon}\,\underline{w}'_{\mathcal{A}}(\underline{\phi}) \; . \tag{4}$$

This expression shows that the ANN-approximation error acts as a spurious source term on the solution $\underline{\phi}$. Depending on the functional form of $\underline{\Omega}$ and $\underline{w}_{\mathcal{A}}$, the error $\underline{\varepsilon}$ can either grow, decay, or cancel.

The static ANN analysis characterizes the ability of the network to accurately represent the function $\underline{w}$. For this, a large set of sample data is used to evaluate the network fitness following the ANN training process. This metric is widely-used and is typically referred to as "testing." As such, the static analysis allows for the assessment of the ANN approximation error $\underline{\Omega}$.

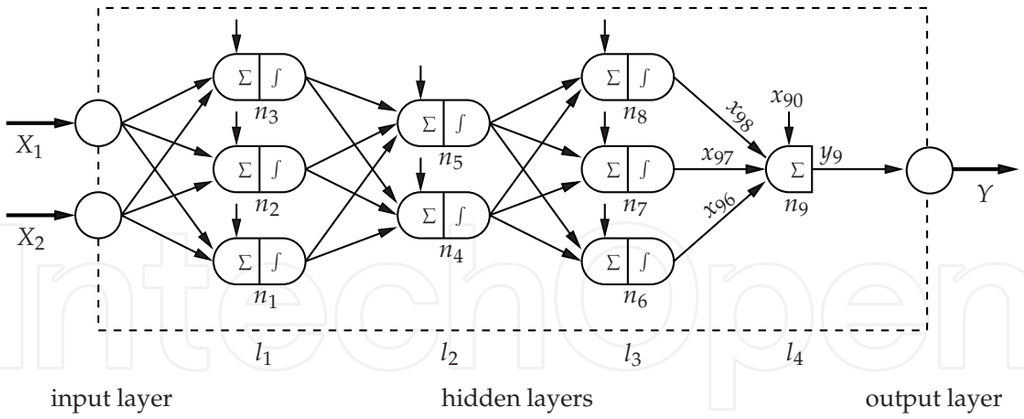


Fig. 1. Architecture of a multilayer perceptron, consisting of an input layer with two input channels X_1 and X_2, one output channel Y, and 4 hidden layers with, respectively, 3, 2, 3, and 1 neurons in each layer. The neurons in the hidden layers are denoted by n_i, with $i = 1, \ldots, 9$, and the number of neurons in this network is $|\underline{N}_\mathrm{N}| = 9$.

Since ANNs are integrated into dynamic systems of the form of Eq. (2), a main shortcoming of this static analysis is that it does not account for the feedback on the solution vector $\underline{\varphi}$. In particular, small errors arising from the network approximation could either cancel, or – worse – induce a drift in the solution vector leading to long-time instability issues of the governing equations.

To address this issue, we will also assess the ANN-performance under dynamic conditions. To this end, the temporal evolution of the error $\underline{\varepsilon}$ in Eq. (4) is evaluated in order to assess the dynamic stability of the system. This metric, which we refer to as dynamic ANN performance measure, allows us to directly characterize feedback-effects of ANN-approximation errors on the solution. It will be shown in the second part of this article that this dynamic ANN performance evaluation provides more realistic estimates of the ANN-fitness potential, whereas the static ANN analysis gives typically too optimistic estimates.

The remainder of this article is organized as follows. Section 2 discusses the ANN model and describes the training process. The GPS method and ES are presented in Sec. 3. The performance of OANNs are assesses by considering two combustion-chemical problem of increasing complexity. Specifically, Sec. 4 considers the evolution of a chemical species in decaying homogeneous isotropic turbulence. The combustion process is described by a reversible one-step chemical reaction, in which the mixing and combustion are described by a Lagrangian Fokker-Planck model. In the second problem, OANNs are integrated into a high-fidelity large-eddy simulation to predict the turbulent combustion in a swirl-stabilized burner system of practical relevance. Model formulation, experimental setup and comparisons with experimental data and conventional tabulation methods are presented in Sec. 5, and conclusions are drawn in Sec. 6.

2. Artificial neural networks

In the following, the class of multilayer perceptrons (MLPs) is considered. A MLP, schematically shown in Fig. 1, consists of an input layer with N_I input channels, an output layer having N_O channels, and N_L hidden layers. The number of neurons in each hidden layer is denoted by $\underline{N}_\mathrm{N} \in \mathbb{Z}^{N_\mathrm{L}}$. The connectivity of the network is denoted by $\underline{\underline{C}}$, corresponding to

a binary matrix, in which the element C_{ij} is unity if neurons i and j are connected and zero otherwise. The output y_i of neuron i is computed according to

$$y_i = \psi_i \left(\sum_{j=0}^{N_C} C_{ij} \omega_{ij} x_{ij} \right) , \tag{5}$$

where N_C is the number of connections in the network, ω_{ij} are the synaptic weights, and x_{ij} are the input signals. Note that $x_{i0} = -1$ is a threshold value. The transfer function of the neuron i is denoted by ψ_i. The synaptic weights in the network are adapted during the training process, which in itself represents an optimization problem and can be written as

$$\min_{\omega \in \mathbb{R}^{N_\omega}} E(\omega) , \tag{6}$$

where N_ω denotes the number of synaptic weights in the network, and $E : \mathbb{R}^{N_O \times N_t} \rightarrow \mathbb{R}$ is a measure of the error between the actual and desired output of the network. The number of training samples is denoted by N_t. In the following, E is defined as

$$E = \frac{1}{2N_t} \sum_{j=1}^{N_t} e^t(j) , \tag{7}$$

with

$$e^t(j) = \sum_{i=1}^{N_O} \left[Y_i(j) - Y_i^t(j) \right]^2 , \tag{8}$$

and $Y_i^t(j)$ represents the j^{th} training sample of the output signal i.
The topology of a particular network is defined by its neural arrangement, consisting of the number of layers N_L and neurons per layer $\underline{N}_N$, the connectivity $\underline{\underline{C}}$, and the neural transfer functions $\underline{\psi}$. In the following, this network topology is formally written as

$$\mathcal{A} = \mathcal{A}(N_L, \underline{N}_N, \underline{\underline{C}}, \underline{\psi} | N_I, N_O, Y, J) , \tag{9}$$

in which the last four arguments are constrained by the problem and the desired performance of the ANN. The performance – or generalization potential – of a trained network can be characterized by the cost function J, which is evaluated using test samples that typically differ from the training data. Testing – or static ANN performance analysis – is done after training to evaluate the ability of the network in representing untrained samples. In the present work, the following cost function is used:

$$J(\mathcal{A}) = \log_{10} \left(\sqrt{\frac{1}{N_s} \sum_{j=1}^{N_s} (e^s(j))^2} \right) , \tag{10}$$

where N_s refers to the number of test samples, and the decadic logarithm is introduced to enlarge the resolution of the cost function.
The objective is to identify a particular network topology $\mathcal{A}$ that minimizes the cost function, or in other words, maximizes the generalizability. In the following, we will restrict our discussion to a multidimensional optimization problem that only includes continuous parameters, i.e., real- or integer-valued variables. For this, a network will be considered, in

which a single neuron in the last hidden layer is used having a linear transfer function,

$$\psi(s) = s \ . \tag{11}$$

A sigmoidal function,

$$\psi(s) = \gamma_1 \tanh(\gamma_2 s) \ , \tag{12}$$

is used for all other neurons in the $N_L - 1$ hidden layers. The parameters γ_1 and γ_2, characterizing the saturation and slope of the transfer function, are considered as free parameters that will be adjusted during the optimization process. Furthermore, the network architecture will be constrained to a fully connected feed-forward network. With this, the resulting optimization problem may be formulated as

$$\begin{aligned} \min_{N_\mathrm{L}, \underline{N}_\mathrm{N}, \underline{\gamma}} \quad & J(\mathcal{A}) \\ \text{subject to} \quad & N_\mathrm{L} = N_\mathrm{L}^{\max} \ , \\ & N_{\mathrm{N},i} \in \{0, 1, \ldots, N_\mathrm{N}^{\max}\}, \qquad \text{for } i = 1, 2, \ldots, N_{\mathrm{L}-1} \ , \\ & N_{\mathrm{N},N_\mathrm{L}} = 1 \ , \\ & \underline{\gamma} \in \mathcal{G} \ , \end{aligned} \tag{13}$$

where $\underline{\gamma} \in \mathbb{R}^{2\times(|\underline{N}_\mathrm{N}|-1)}$, $\mathcal{G}$ denotes the bounded parameter space for the transfer function coefficients, and the number of neurons in all layers is denoted by $|\underline{N}_\mathrm{N}|$.
In this context it is important to point out that the herein employed optimization method is not restricted to continuous variables, and can also be applied to the optimization of categorical or discrete variables. This has been discussed by Ihme et al. (2008), in which the connectivity and transfer functions were included in the optimization of the network topology.

3. Topological optimization

3.1 Generalized pattern search method

In the following, a generalized pattern search method is used to solve the optimization problem (13). The GPS method is a derivative-free method, in which a sequence of iterates is generated whose cost function is non-increasing (Audet & Dennis, 2003). All points at which the cost function is evaluated are restricted to lie on a mesh, and the limited point of the sequence of iterates corresponds to a local optimal solution that is defined with respect to a user-specified neighborhood. GPS methods for unbounded problems have been discussed and analyzed by Torczon (1997); they were later extended by Lewis & Torczon (1999; 2000) to bounded and linearly constrained problems.
The GPS algorithm, schematically shown in Fig. 2, proceeds in two steps, namely a search and a poll step. In the search step a finite set of search points are evaluated to facilitate a global exploration of the parameter space. Since the search step is not required for convergence, different strategies can be employed to identify a promising region in the parameter space that potentially results in an improved cost function. For instance, *a priori* knowledge, random sampling using Latin hypercube sampling (LHS) (McKay et al., 1979), or a surrogate can be employed. In the case of a surrogate, the cost function is approximated by a lower-dimensional model, whose evaluation is typically less expensive. This surrogate approximation is continuously updated during the simulation, and is then used to identify a new point with a potentially lower cost function. Kriging is frequently employed as surrogate

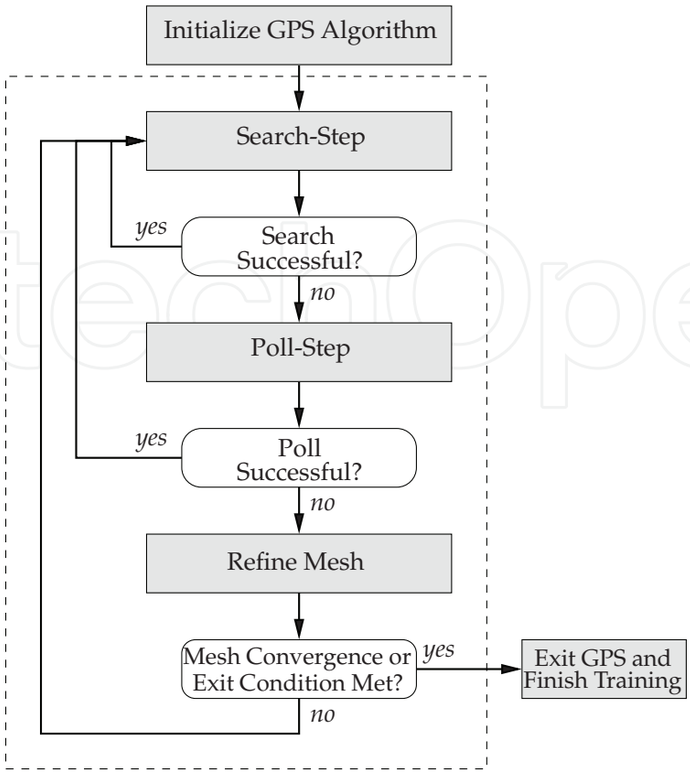


Fig. 2. Schematic diagram of the GPS algorithm.

approximation, and its multidimensional extension makes this method particularly attractive for application to network optimization.

In the case that the search step does not result in an improvement in the cost function, the algorithm continues with a poll step, which guarantees the convergence of the GPS algorithm. The poll step is restricted to a mesh that is constrained by the parameter space, and centered around the incumbent point $\underline{\eta}$. The vector $\underline{\eta}$ is of dimension $N_{\mathrm{L}} - 1 + 2(|\underline{N}_{\mathrm{N}}| - 1)$, and contains information about the number of non-linear neurons and coefficients in its respective non-linear transfer functions. At iteration k the mesh for the poll step is defined by

$$M_k = \left\{ \underline{\eta}_k + \underline{\Delta}_k \underline{\underline{D}} \right\} \tag{14}$$

where $\underline{\Delta}_k > 0$ is the mesh size parameter, which is obtained through a successive refinement of the mesh according to

$$\underline{\Delta}_k = \tau^k \underline{\Delta}_0 \,, \tag{15}$$

and $\tau = 1/2$ is used in the following application. The discrete directions in the parameter space that are evaluated during a poll step are specified by the matrix $\underline{\underline{D}}$ whose columns form a positive spanning set. That is, if $\underline{\underline{I}}$ denotes the identity matrix and $\underline{i}$ is the vector of ones, then $\underline{\underline{D}}$ is typically chosen as $\underline{\underline{D}} = [\underline{\underline{I}}, -\underline{i}]$ or $\underline{\underline{D}} = [\underline{\underline{I}}, -\underline{\underline{I}}]$ (Audet & Dennis, 2003). During the poll step the cost function is evaluated for these poll candidates. In the case that the poll step is unsuccessful the mesh is refined, and the algorithm continues with a new iteration, starting with a search step (see Fig. 2). This process continues until a convergence criterion is met or $\underline{\Delta}_k$ reaches a minimum mesh size. More details on the algorithm and convergence proofs

can be found in Audet & Dennis (2003) and extensions of the algorithm to include categorical parameters are discussed in Audet & Dennis (2000) and Abramson (2004).

3.2 Evolutionary strategy

In addition to the GPS method, an evolutionary strategy is used to solve Eq. (13). Evolutionary strategies have been proposed as simple mutation selection mechanisms by Rechenberg (1973), and were later generalized by Schwefel (1977; 1981). ES belongs to the general category of evolutionary algorithms, and is based on a collective evolution process of individuals in a population. The evolution of this population follows a biologically inspired process, consisting of a sequence of steps, involving mutation, recombination, and selection. In this work, a so-called (μ, λ)-ES is used, in which μ denotes the number of parents and λ corresponds to the number of offsprings in each generation. For completeness, the general form of the ES algorithm (Bäck & Schwefel, 1993) is briefly summarized.

In ES, the independent parameters for each individual network candidate are represented by a joint Gaussian distribution. This distribution is characterized by a zero expectation, a variance for each optimization parameter, and a rotation vector to ensure positive definiteness of the parameter covariance matrix. The mutation of each individual is sampled from the joint Gaussian distribution, which is modified by mutating the standard deviation and the rotation angle. More details on this mutation strategy and the self-adaptation, which are employed in the present work, can be found in Bäck & Schwefel (1993). For the recombination of individuals and strategy parameters, different mechanisms are used. Here, a discrete recombination of individuals is used, in which a new individual is produced from a random sampling of parameter components from two parents. For the recombination of the strategy parameters, i.e., mutation step size and rotation angle, pairwise intermediate recombination is used. Following the recombination and mutation, the ES continues with a deterministic selection step. In the (μ, λ)-ES, the μ-best candidates out of λ offsprings are selected. The best candidates are then used as parents in the subsequent generation. In the following application $\mu = 2$, $\lambda = 12$, and at most 80 iterations were used in the ES algorithm. It was found that the results for the optimal ANN topology showed some sensitivity to the initialization and the choice of the exogenous parameters in the ES algorithm. It is assumed that this sensitivity is mainly attributed to the heterogeneous search space, and the slow convergence of the ES.

4. Combustion in decaying homogeneous isotropic turbulence

4.1 Mathematical model

In the following problem, the mixing and reaction of fuel and oxidizer in decaying homogeneous isotropic turbulence are considered. The corresponding reaction equation can be written as

$$\mathrm{F} + r\mathrm{O} \rightleftarrows (1+r)\mathrm{P}\,. \tag{16}$$

The reaction rate w of this one-step reversible reaction follows an Arrhenius expression

$$\begin{aligned} w(Z,C) = (1+r)A \exp\{-\beta\} \exp\left\{-\frac{\beta(1-C)}{C}\right\} \\ \times \left[Z_{\mathrm{st}}(1-Z_{\mathrm{st}})\left(\frac{Z}{Z_{\mathrm{st}}} - C\right)\left(\frac{1-Z}{1-Z_{\mathrm{st}}} - C\right) - \frac{1}{K}C^2\right]\,, \end{aligned} \tag{17}$$

in which r is the stoichiometric coefficient, A is the frequency factor, β is the Zeldovich number, Z_{st} is the stoichiometric mixture fraction, and K is the equilibrium constant. Values for these

A	β	K	Z_{st}	r
80,000	4.0	100	0.5	1

Table 1. Parameters used for one-step chemical reaction.

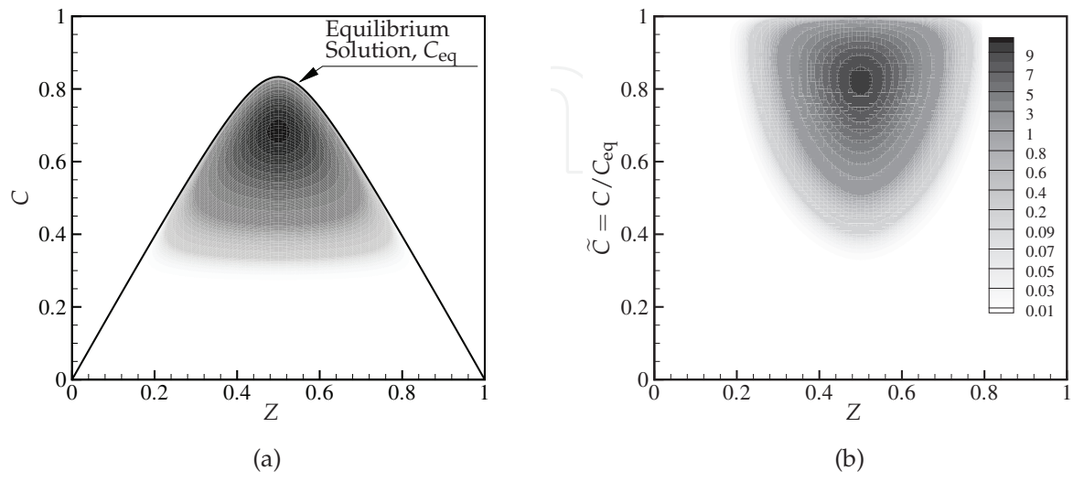


Fig. 3. Chemical reaction rate w as function of mixture fraction and reaction progress variable.

parameters are summarized in Tab. 1, and the interested reader is referred to Sripakagorn et al. (2004) for a more detailed discussion on the chemical reaction and parameter specifications.
The chemical reaction rate given in Eq. (17) is only a function of mixture fraction Z and reaction progress variable C, corresponding to the non-dimensional temperature. Here, $Z \in [0,1]$ and $C \in [0, C_{eq}]$, and the reaction rate as function of Z and C is shown in Fig. 3(a). The progress variable at equilibrium, C_{eq}, is a function of Z, and is obtained by solving Eq. (17) for $w = 0$, resulting in:

$$C_{eq} = \frac{25}{24}\left(1 - \sqrt{1 - \frac{96}{25}Z(1-Z)}\right) , \tag{18}$$

in which prescribed values for Z_{st} and K from Tab. 1 are used. Using C_{eq}, a normalized progress variable can be introduced, $\widetilde{C} = C/C_{eq}$, so that $\widetilde{C} \in [0,1]$, and w as function of $\widetilde{C}$ and Z is illustrated in Fig. 3(b). In the following, OANNs are generated for the approximation of the chemical reaction rate as function of Z and $\widetilde{C}$.
The temporal evolution of reactants and products is obtained from the solution of a Lagrangian Fokker-Planck (LFP) model. In this model, the trajectories of individual particles in composition space are described by the solution of a set of stochastic differential equations (SDEs). For the one-step chemical reaction, which is fully characterized by mixture fraction Z and reaction progress variable C, the Fokker-Planck model can be written as (Fox, 2003):

$$\mathrm{d}\begin{pmatrix} Z \\ C \end{pmatrix} = \begin{pmatrix} 0 \\ w(Z,C) \end{pmatrix}\mathrm{d}t - \underline{\underline{A}}\begin{pmatrix} Z - \langle Z \rangle \\ C - \langle C \rangle \end{pmatrix}\mathrm{d}t + \underline{\underline{B}}\,\mathrm{d}\underline{W}(t) . \tag{19}$$

where W is a Wiener process, and the angular brackets denote the mean value which is defined by $\langle \phi \rangle = \int \phi P(\phi)\mathrm{d}\phi$ for $\phi = \{Z, C\}$. The PDF is denoted by P, and the scalar fluctuation ϕ' is

computed as $\phi' = \phi - \langle\phi\rangle$. The drift matrix $\underline{\underline{A}}$ is given by

$$\underline{\underline{A}} = \begin{pmatrix} \langle\chi_Z\rangle/\langle Z'^2\rangle & 0 \\ 0 & \langle\chi_Z\rangle/\langle C'^2\rangle \end{pmatrix} , \tag{20}$$

and the diffusion matrix is

$$\underline{\underline{B}} = \begin{pmatrix} \langle\chi_Z\rangle \dfrac{(Z - Z_{\mathrm{eq}}^-)(Z_{\mathrm{eq}}^+ - Z)}{\langle(Z - Z_{\mathrm{eq}}^-)(Z_{\mathrm{eq}}^+ - Z)\rangle} & 0 \\ 0 & \langle\chi_C\rangle \dfrac{C(C_{\mathrm{eq}} - C)}{\langle C(C_{\mathrm{eq}} - C)\rangle} \end{pmatrix}^{1/2} , \tag{21}$$

with

$$Z_{\mathrm{eq}}^{\pm} = \frac{1}{2}\left(1 \pm \sqrt{1 - 2C\left(1 - \frac{12}{25}C\right)}\right) \tag{22}$$

from Eq. (18). $\langle\chi_Z\rangle$ and $\langle\chi_C\rangle$ are the mean scalar dissipation rates for mixture fraction and progress variable, and are modeled by an exponential decay

$$\langle\chi_\phi\rangle = \langle\chi_\phi\rangle_0 \exp\left\{-t/\tau_\phi\right\} . \tag{23}$$

The initial particle distribution is sampled from a beta distribution with $\langle Z\rangle(t=0) = \langle Z\rangle_0 = 0.5$ and $\langle Z'^2\rangle(t=0) = \langle Z'^2\rangle_0 = 0.2$, and the progress variable is set to the equilibrium condition $C_{\mathrm{eq}}(Z)$. The mean lifetime for the decay rates of $\langle\chi_Z\rangle$ and $\langle\chi_C\rangle$ correspond to the averaged decay constant from the direct numerical simulation by Sripakagorn et al. (2004) with $\tau_Z = 1.5$ and $\tau_C = 1.0$. Note that the time in Eq. (19) is non-dimensionalized by the initial large-eddy turnover time (Sripakagorn et al., 2004). Since the evolution of $\langle Z'^2\rangle$ obeys the ODE $\mathrm{d}_t\langle Z'^2\rangle = -\langle\chi_Z\rangle$ and $\lim_{t\to\infty}\langle Z'^2\rangle(t) = 0$, the initial condition for $\langle\chi_Z\rangle$ is $\langle\chi_Z\rangle_0 = \langle Z'^2\rangle/\tau_Z$, and $\langle\chi_C\rangle_0$ is set to 0.25. After the initialization of 2×10^6 particles, the LFP model is advanced over $T = 10$ non-dimensional time units with a step size of $\mathrm{d}t = 1\times10^{-3}$. In order to allow for a direct comparison of the different simulations, the increment in the Wiener process is kept identical for all runs.

4.2 Network optimization

The chemical source term, given in Eq. (17), is approximated by optimal ANNs. For this, GPS and ES are used to identify optimal network topologies that result in the lowest approximation error. For the network optimization the following constraints on the topological parameter space are imposed: The maximum number of hidden layers (including the last linear layer) is $N_{\mathrm{L}}^{\max} = 3$, and the maximum number of non-linear neurons per hidden layer is restricted to $N_{\mathrm{L}}^{\max} = 8$. Only fully connected networks are considered, and sigmoidal transfer functions are used in all non-linear neurons. The free parameters γ_1 and γ_2 in Eq. (12) are adjusted during the optimization process. Since the GPS algorithm is a mesh-based method these parameters are constrained to $0.4 \le \gamma_1 \le 1.2$ and $1 \le \gamma_2 \le 5$ having a maximum mesh size of $\Delta\gamma_1^{\max} = 0.4$ and $\Delta\gamma_2^{\max} = 1$, respectively, and the minimum mesh size is $\Delta\gamma_1^{\min} = \Delta\gamma_2^{\min} = 10^{-4}$.
To evaluate the fitness of a particular network candidate, the synaptic weights are first adjusted using a supervised learning strategy. In this technique, a set of training data is presented to the network and the weights are adjusted to reproduce the input-output relation. The training set consists of 50,000 randomly chosen samples, and a Levenberg-Marquardt

algorithm (Hagan et al., 1996) is used to iteratively adjust the synaptic weights. Since this training step is computationally expensive only a maximum of 100 iterations in the Levenberg-Marquardt algorithm are used to train each individual network. Although typically more iterations are necessary to ensure that the synaptic weights are fully converged, it was concluded that this number of iterations is sufficient to assess the fitness potential of a particular network architecture. After an optimal network architecture is identified by the optimization algorithm, this ANN is further trained until the synaptic weights are fully converged (see Fig. 2). It was found that the outcome of the training process shows some sensitivity to the initialization of the synaptic weights that are typically sampled from a uniform distribution. To allow for an objective comparison between different optimization methods, this sensitivity is here eliminated by initializing all synaptic weights with $\omega = 0.1$. For the evaluation of the static performance of each network candidate during the GPS, the cost function (10) is evaluated using $N_s = 50{,}000$ test samples.

4.3 Static ANN performance analysis

Results from the GPS and ES network optimization are presented in Tab. 2. The second column in this table lists the network architecture of the optimal ANN. The static network fitness, characterized by the cost-function, is shown in the third column, and the last column presents the memory requirement which is necessary to store the network architecture. In addition to these optimal network structures, results for conventional look-up tables are also presented. For this, the chemical source term, Eq. (17), is tabulated in terms of Z and $\widetilde{C}$ using an equidistant grid.

OANN/Table	Architecture	Cost Function	Iterations	Memory [kB]
GPS-OANN	7-8-1	-3.289/-3.727	876	7
ES-OANN	7-4-1	-3.077/-3.144	960	4
ANN (fixed $\underline{\gamma}$)	7-8-1	-3.480	–	7
Table	50×50	-2.023	–	20
Table	100×100	-2.624	–	78
Table	200×200	-3.228	–	313
Table	300×300	-3.579	–	703
Table	400×400	-3.831	–	1,250
Table	500×500	-4.022	–	1,953

Table 2. Comparison of OANN architecture and performance obtained from GPS and ES with conventional tabulation. Comparisons with conventional tabulation techniques are also summarized, showing results for tables with increasing grid resolution.

The optimal network structure that was returned from the GPS algorithm consists of a 7-8-1 ANN. This architecture was found after 876 function evaluations in the 34-dimensional parameter space. The cost function of this GPS-OANN is $J(\mathcal{A}) = -3.289$, and was further reduced to $J(\mathcal{A}) = -3.727$ during the training following the GPS optimization. The evolution of $J(\mathcal{A})$ for the GPS-OANN is shown in Fig. 4(a). From this figure it can be seen that after a transition phase the cost function decays continuously, and the evaluation of J after 100 iterations is adequate to assess the fitness potential of a particular network structure.
The coefficients in the neural transfer functions for the GPS-OANN are shown in Fig. 4(b). Note that these parameters are typically kept constant for all neurons with values for the saturation $\gamma_1 = 1.075$ and the slope $\gamma_2 = 2.0$ (Haykin, 1994). It is interesting to point out that the transfer function coefficients in the GPS-OANN are considerably different from these

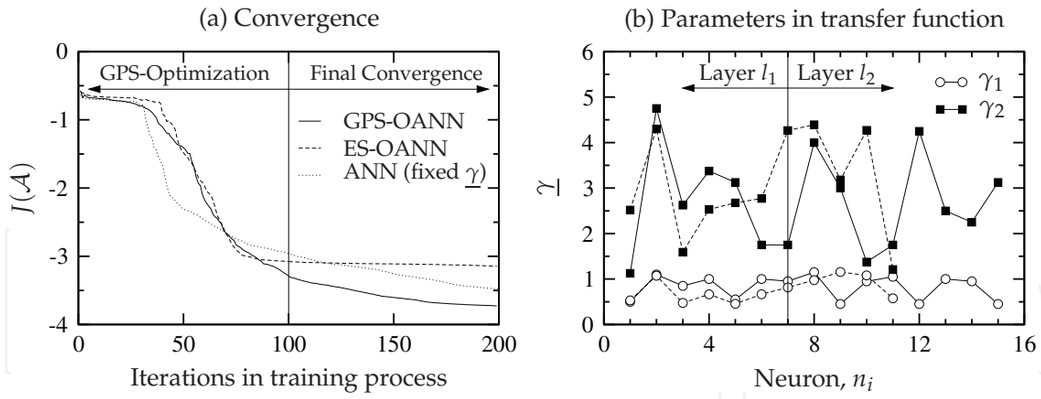


Fig. 4. Performance evolution of the 7-8-1 OANN obtained from the GPS method and optimal coefficients in the neural transfer functions. The dotted line in the left panel shows the fitness for a 7-8-1 ANN having constant transfer function coefficients with $\gamma_1 = 1.075$ and $\gamma_2 = 2.0$ for all neurons.

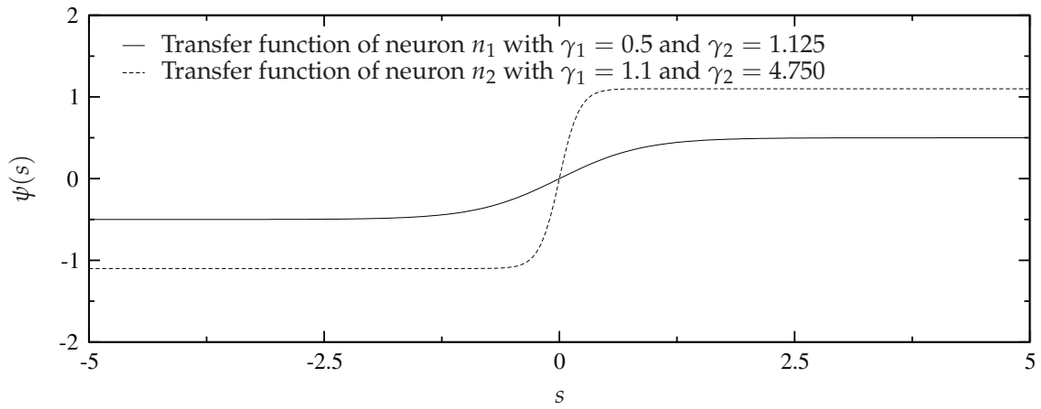


Fig. 5. Transfer function for the first two neurons in the OANN obtained from the GPS algorithm.

values, showing distinct spreading for all neurons. As an example, the transfer functions for the first two neurons are shown in Fig. 5. While γ_1 for the transfer function of the second neuron is generally in good agreement with the frequently employed value of 1.075, the saturation and the slope for the first neuron are considerably different, resulting in a distinctly different neural response characteristics between both neurons. To emphasize the effect of adaptive transfer function coefficients on the network fitness, a 7-8-1 network with constant γ_1 and γ_2 for all neurons is trained resulting in a cost function of $J(\mathcal{A}) = -3.480$. Note, however, that $J(\mathcal{A})$ corresponds to the decadic logarithm of the L_2 error norm. Therefore, this difference of seven percent in the cost function corresponds to an deviation of more than 75 percent in the L_2-norm.

The optimal topology identified by the evolutionary algorithm consists of a 7-4-1 ANN, and was obtained from a total ES population of 960 networks candidates, which have been evolved over 80 generations with two parents and 12 offsprings per generation. The termination of the ES optimization after 80 iterations was mainly motivated by the argument to have comparable computational cost for both GPS and ES optimization. The cost function of the ES-OANN is

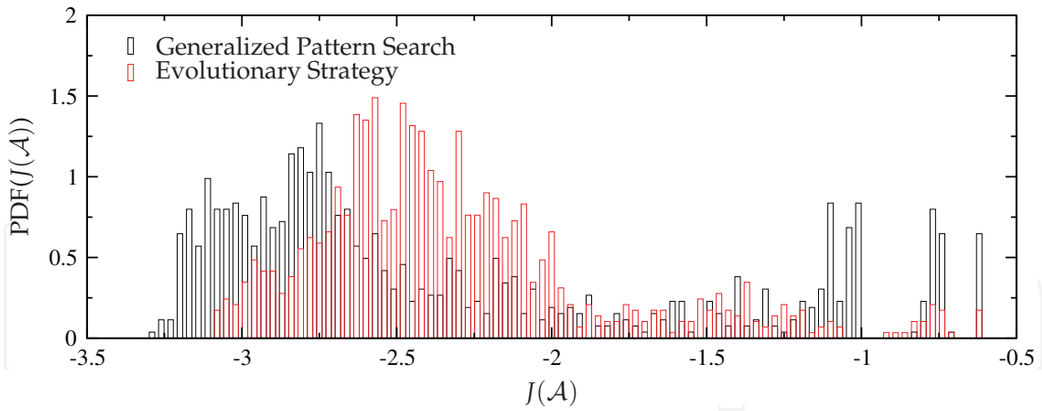


Fig. 6. Probability density function of the cost function evaluated from all network candidates that were evaluated during the GPS and ES optimization.

$J(\mathcal{A}) = -3.144$, and the evolution of $J(\mathcal{A})$ is shown by the dashed line in Fig. 4(a). The initial convergence of the cost function for the ES-OANN is similar to that of the GPS-OANN; however, $J(\mathcal{A})$ decreases only marginally after 80 iterations. This apparent saturation of the fitness can primarily be attributed to the small number of neurons in the second layer. The transfer function coefficients for this network are illustrated in Fig. 4(b). It is interesting to point out that both ES and GPS give similar values for slope and saturation of the neural transfer function in the first hidden layer.

A statistical comparison of the fitness of all network candidates that were evaluated during the GPS and ES optimization is shown in Fig. 6. The distributions of the network performance from both optimization methods is considerably different. For instance, the bimodal PDF from the GPS optimization is strongly skewed towards lower values of $J(\mathcal{A})$. Network candidates with poor fitness were mainly obtained from the random sampling of the parameter space during the search step. In comparison, the PDF from the ES is nearly unimodal and peaks around $J(\mathcal{A}) \approx -2.5$. In this context it is important to point out that the outcome of the ES method is sensitive to the initial conditions and prescribed step size. Therefore, it can be anticipated that a different set of parameters and initial conditions can potentially lead to a different optimal network topology.

In addition to the smooth function representation, a main advantage of ANNs over conventional tabulation techniques is the high knowledge density. This is reflected by the modest memory requirement necessary to store a network architecture. While the ANN-memory demand is nearly independent from the number of input parameters, the storage requirements for look-up tables grows exponentially with the dimensionality of the function. A comparison of the knowledge-density, which is here defined as the ratio between accuracy and memory requirement, is illustrated in Fig. 7. This figure illustrates that ANNs perform significantly better than conventional look-up tables, and for equivalent accuracy the memory savings can be in excess of two orders of magnitude.

4.4 Dynamic ANN Performance Analysis

In the previous section, the fitness of the network architectures obtained from GPS and ES were compared with the conventional tabulation method. It was found that the performance of the GPS-OANN is comparable to that of the tabulation method with a resolution of more than 300 grid points in both Z and $\widetilde{C}$ directions. In this static comparison, a homogeneously

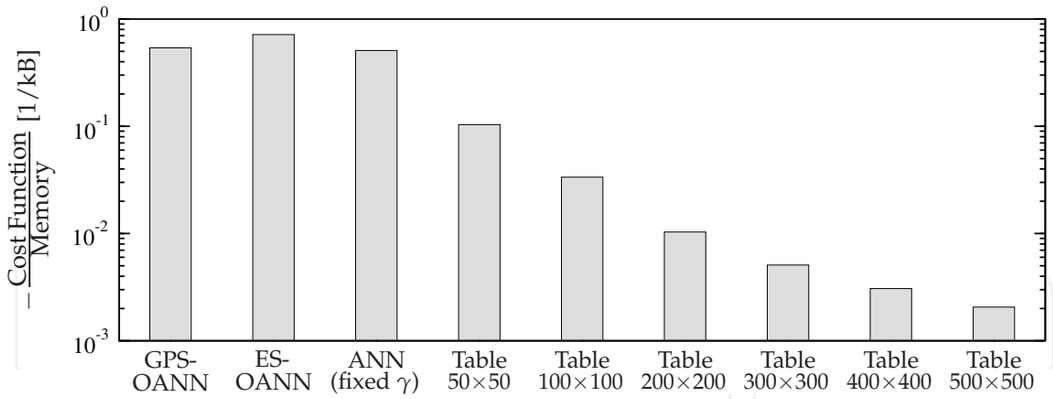


Fig. 7. Comparison of the knowledge-density between ANNs and conventional tabulation techniques. The knowledge density is defined as the ratio between static network fitness and memory requirement.

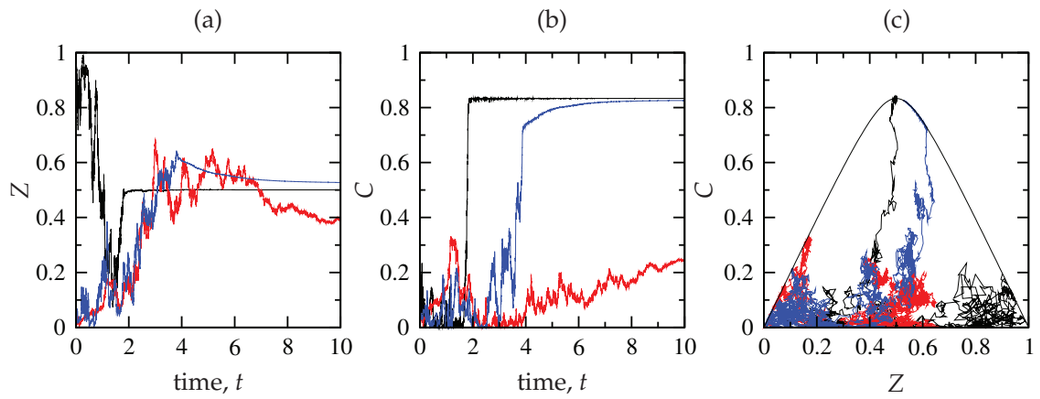


Fig. 8. Evolution of the chemical composition of three representative particles: (a) mixture fraction; (b) reaction progress variable; and (c) corresponding trajectories in composition space.

distributed set of random test samples was used for the performance evaluation. However, the reaction rate which is represented by ANNs and look-up tables corresponds to a source term in a partial differential equation describing the evolution of a chemical system. The primary variable which is of interest in the characterization of this system is the species composition, whose prediction is directly affected by the accurate representation of the chemical source term. This chemical reaction evolves along trajectories that typically occupy only a small region in composition space. This suggests that the static analysis as discussed in the previous section could have only limited relevance for the present application. Therefore, a dynamic OANN performance analysis is conducted in order to assess feedback-effects of ANN approximation errors on the evolution of the dynamic system.

Before analyzing the performance of OANNs and tabulation methods, the evolution of the chemical system as described by the LFP model, Eq. (19), is briefly discussed. The temporal evolutions of the mixture fraction and reaction progress variable for three representative particles are shown in Figs. 8(a) and (b), and the corresponding trajectories in Z-C-composition space are illustrated in Fig. 8(c). Following an initial phase of intense

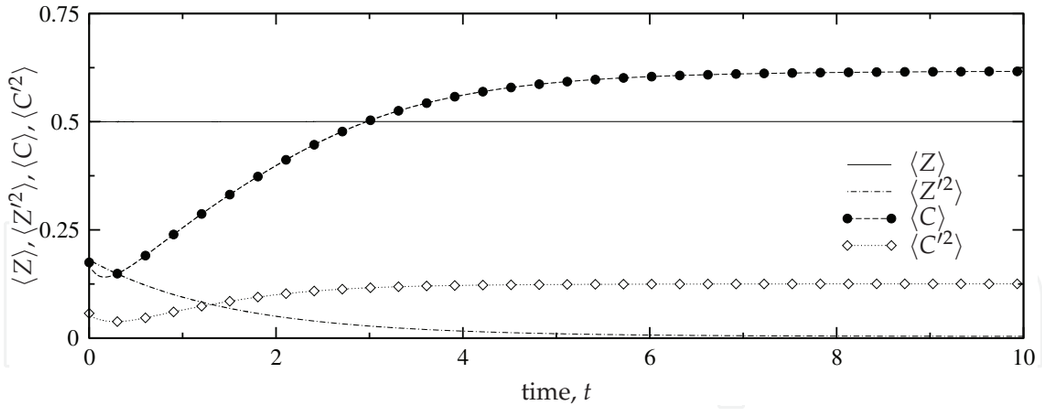


Fig. 9. Statistical evolution of the chemical system.

mixing, the trajectories of the particles, shown by the black and blue lines, converge and reach after approximately two respectively four eddy turn-over times the equilibrium condition. The trajectory of the particle shown by the red line indicates that the particle fails to reach a stably burning state. This can mainly be attributed to the initially large fluctuations in mixture fraction and progress variable.

The statistical evolution of the LFP model is illustrated in Fig. 9, in which the first two moments of mixture fraction and reaction progress variable are shown. From the governing equation for the mean mixture fraction, viz. $\mathrm{d}_t\langle Z\rangle = 0$, follows that $\langle Z\rangle$ is constant and equal to the initial condition, and the mixture fraction variance decays as $\langle Z'^2\rangle(t) = \langle Z'^2\rangle_0 \exp\{-t/\tau_Z\}$. The evolution of the reaction progress variable, which is equal to the normalized temperature, is shown by the dashed line in Fig. 9. Starting from the initial condition, the mixture slowly ignites and with increasing time $\langle C\rangle$ approaches a steady condition. Note that this final state corresponds to the equilibrium condition; however, the maximum temperature $\langle C\rangle = C_{\mathrm{eq}}(\langle Z\rangle) = 5/6$, is not reached, which is also evidenced by the non-vanishing progress variable variance, shown by the dotted line in Fig. 9.

Instead of performing the costly evaluation of the chemical source term from Eq. (17), in the following $w(Z, C)$ is obtained from the GPS-OANN and look-up tables. This comparison allows us to critically assess effects of approximation errors in the source term representation on the evolution of the reaction progress variable. To quantify this error, the following norm is used:

$$L_2^2(\langle C\rangle) = \frac{1}{T}\int_0^T [\langle C\rangle - \langle C\rangle_{\mathrm{ref}}]^2 \mathrm{d}t \,, \tag{24}$$

where the subscript "ref" denotes the solution obtained from the analytical evaluation of the chemical source term, and T corresponds to the simulation time.

Comparisons of the L_2-norm, obtained from the simulations with GPS-OANN and look-up tables, are summarized in Tab. 3. From these results the following observations can be made. First, the solution from the tabulation exhibits a monotonic convergence with quadratic convergence rate. Second, the overall accuracy of the results from the GPS-OANN simulation for this dynamic application is comparable to a tabulation having approximately 80×80 grid point resolution. This is different to the findings from the static analysis of Sec. 4.3. The main reasons for this are the non-linearity in the diffusion matrix $\underline{\underline{B}}$ and the rather strict constraints for $N_{\mathrm{L}}^{\max}$ and $N_{\mathrm{N}}^{\max}$ in Eq. (13). In this ANN optimization, the GPS method was restricted to include only two non-linear hidden layers with a maximum of eight neurons per layer. By

Architecture	$L_2(\langle C \rangle)$
GPS-OANN	4.820×10^{-4}
Table 50×50	1.038×10^{-3}
Table 100×100	2.712×10^{-4}
Table 200×200	6.294×10^{-5}
Table 300×300	2.549×10^{-5}
Table 400×400	1.336×10^{-5}
Table 500×500	5.321×10^{-6}

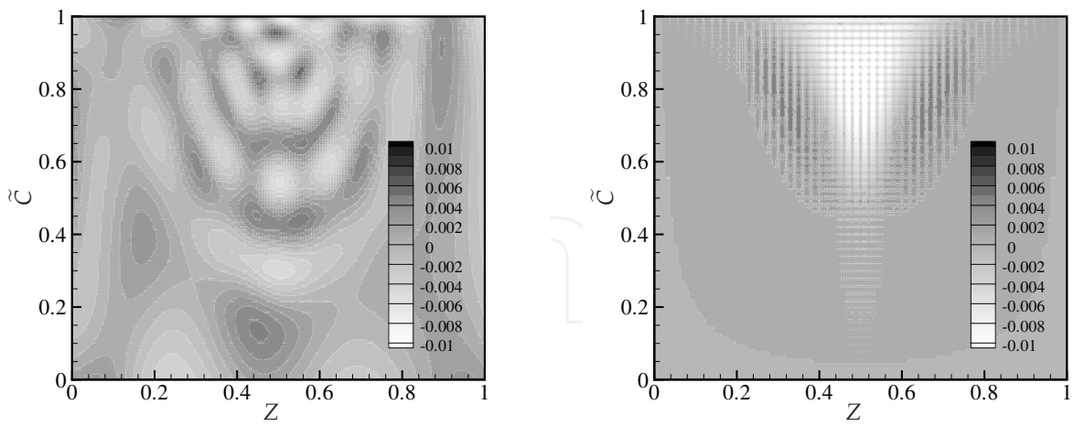
Table 3. Comparison of the convergence error for the solution of the LFP model.

relaxing these constraints and extending the search space to include a larger number of layers and neurons, the ANN topology becomes more flexible, and will lead to improvements in the ANN performance characteristics. To demonstrate this, an additional GPS optimization was conducted in which four hidden layers with a maximum of eight neurons per non-linear layer were used for the network optimization. For training and performance evaluation of the GPS network candidates respectively 100,000 samples were used, and all synaptic weights were initialized with random numbers. The GPS algorithm returned as optimal topology a 8-8-8-1 network having a cost function of $J(\mathcal{A}) = -4.242$. The application of this larger OANN in the LFP model resulted in an improvement of the L_2 norm by more than 20 % compared to the 7-8-1 GPS-OANN.

To quantify the approximation quality of the ANN and the look-up table, the error surface is compared in Fig. 10. For this, the error Ω is defined as the difference between ANN approximation $w_{\mathcal{A}}(Z, \widetilde{C})$ and the corresponding analytical value (see Eq. (3)). The apparent oscillations in the error surface are an indication for underfitting, and the small differences in the chemical source term representation can lead to incremental deviations in the particle trajectories. The analysis of the ANN function representation suggests that the training of networks with a larger number of samples or a biased distribution of sample points, using for instance an acception-rejection algorithm (Ihme et al., 2008), can lead to a further reduction of this error.

5. Turbulent combustion in a swirl-stabilized burner system

In the previous section, the advantages of optimal ANNs in application to a zero-dimensional combustion problem were discussed. The present section extends this analysis by considering the unsteady turbulent combustion in a technical-relevant burner system. In this application, the large-eddy simulation (LES) technique in combination with a flamelet-based combustion model is employed for predicting the turbulent reacting flow field, heat release, and pollutant formation. These high-fidelity LES computations of turbulent reacting flows are typically performed on massively parallel computing architectures. As such, the utilization of ANNs for chemistry representation and function approximation can provide significant benefits over conventional look-up tables in at least the following three aspects. First, the reduction in storage requirements allows to perform these large-scale combustion simulations on computing architectures with restricted memory. Second, the information retrieval from ANNs amounts to a direct function evaluation which is typically more efficient than table-interpolation. In addition, the smooth function representation of ANNs can result in improved model accuracy and faster convergence.



(a) GPS-OANN (b) Table: 100×100

Fig. 10. Error surface $\Omega(Z, \widetilde{C})$, showing the difference between the source term approximation from ANN (a) and tabulation (b) and the analytical expression from Eq. (17).

In the following, the potential of ANNs in application to turbulent reacting flows will be assessed, and qualitative comparisons with conventional look-up tables will be presented. The mathematical formulation and governing equations that are solved in the large-eddy simulation are summarized in the next section, and results of the OANN-technique in static and dynamic applications are presented in Secs. 5.3 and 5.4.

5.1 Mathematical formulation and combustion model

The LES technique is based on the separation of large and small scales in a turbulent flow. The decomposition into the different scales is achieved by applying a filtering operator to the field quantities (Sagaut, 1998). Specifically, a Favre-filtered quantity of a scalar ϕ is defined as

$$\widetilde{\phi}(t, \underline{x}) = \frac{1}{\overline{\rho}} \int \rho(t, \underline{y}) \phi(t, \underline{y}) G(\underline{y}, \underline{x}; \Delta) \mathrm{d}\underline{y} \,, \tag{25}$$

where ρ is the density, Δ is the filter width and G is the filter-kernel. The time is denoted by t and $\underline{x}$ corresponds to the spatial coordinate. The residual field is then defined as $\phi''(t, \underline{x}) = \phi(t, \underline{x}) - \widetilde{\phi}(t, \underline{x})$, and Favre-filtered quantities are related to Reynolds-filtered quantities by $\overline{\rho}\widetilde{\phi} = \overline{\rho\phi}$.

In the following, the Favre-filtered form of the Navier-Stokes equations are solved in the low-Mach number limit. These equations, describing the conservation of mass and momentum, can be written as

$$\widetilde{\mathcal{D}}_t \overline{\rho} = -\overline{\rho} \nabla \cdot \widetilde{\underline{u}} \,, \tag{26a}$$

$$\overline{\rho} \widetilde{\mathcal{D}}_t \widetilde{\underline{u}} = -\nabla \overline{p} + \nabla \cdot \underline{\underline{\overline{\sigma}}} + \nabla \cdot \underline{\underline{\overline{\sigma}}}^{\mathrm{res}} + \overline{\rho} \underline{g} \,, \tag{26b}$$

where

$$\underline{\underline{\overline{\sigma}}} = 2\overline{\rho}\widetilde{\nu} \left(\underline{\underline{\widetilde{S}}} - \frac{1}{3} (\nabla \cdot \widetilde{\underline{u}}) \underline{\underline{I}} \right) \qquad \text{and} \qquad \underline{\underline{\overline{\sigma}}}^{\mathrm{res}} = \overline{\rho} \widetilde{\underline{u}\underline{u}} - \overline{\rho} \underline{\widetilde{u}} \underline{\widetilde{u}} \tag{27}$$

are the filtered viscous stress tensor and the residual stress tensor, respectively. In these equations, $\widetilde{\underline{u}}$ is the filtered velocity vector, $\overline{p}$ is the filtered pressure, $\underline{g}$ is the body force, $\widetilde{\nu}$ is the filtered kinematic viscosity, $\widetilde{\underline{\underline{S}}}$ is the filtered strain rate tensor, and $\widetilde{\mathcal{D}}_t \equiv \partial_t + \widetilde{\underline{u}} \cdot \nabla$ is the substantial derivative. The residual stress tensor, appearing as unclosed term in the momentum equations, is modeled by a dynamic Smagorinsky model (Germano et al., 1991; Lilly, 1992).

In addition to the residual stresses, the filtered density and kinematic viscosity are also unknown and are dependent on the temperature and the species distribution in the flame. To provide information about these quantities, a combustion model is employed which relates the density and all other thermochemical quantities to a reduced set of reaction coordinates.

In the present work, the flamelet/progress variable (FPV) approach (Pierce & Moin, 2004) is used for the prediction of the turbulent reactive flow field. This model is based on the steady laminar flamelet equations (Peters, 1983; 1984), in which all thermochemical quantities are obtained from the solution of the steady flamelet equations. These quantities are parameterized in terms of mixture fraction Z and scalar dissipation rate χ_Z. However, for reasons which are outlined in Pierce & Moin (2004) and Ihme et al. (2005), in the FPV model the scalar dissipation rate is replaced by a reaction progress variable C. This reactive scalar corresponds to a linear combination of major product mass fractions, and is here defined as $C = Y_{CO_2} + Y_{CO} + Y_{H_2O} + Y_{H_2}$. With this, all thermochemical quantities, denoted by $\underline{\phi}$, can then be written as $\underline{\phi} = \underline{\mathcal{G}}_{\phi}(Z, C)$, where $\underline{\mathcal{G}}$ denotes the FPV chemistry library, and $\widetilde{\underline{\phi}}$ is obtained by employing a presumed PDF approach, in which the joint PDF of Z and C is modeled by a beta distribution for the mixture fraction and a mixture fraction-conditioned Dirac function for the progress variable (Pierce & Moin, 2004). The Favre-filtered form of all thermochemical quantities can then be written as $\widetilde{\underline{\phi}} = \widetilde{\underline{\mathcal{G}}}_{\phi}(\widetilde{Z}, \widetilde{Z''^2}, \widetilde{C})$. In addition to the solution of the Favre-filtered Navier-Stokes equations, the FPV model requires also the solution of the Favre-filtered conservation equations for mixture fraction, progress variable, and residual mixture fraction variance, which is denoted by $\widetilde{Z''^2}$. These equations can be written as

$$\overline{\rho}\widetilde{\mathcal{D}}_t\widetilde{Z} = \nabla \cdot (\overline{\rho}\widetilde{\alpha}\nabla\widetilde{Z}) + \nabla \cdot \underline{\overline{\tau}}_Z^{\text{res}} \,, \tag{28a}$$

$$\overline{\rho}\widetilde{\mathcal{D}}_t\widetilde{C} = \nabla \cdot (\overline{\rho}\widetilde{\alpha}\nabla\widetilde{C}) + \nabla \cdot \underline{\overline{\tau}}_C^{\text{res}} + \overline{\rho}\widetilde{w}_C \,, \tag{28b}$$

$$\overline{\rho}\widetilde{\mathcal{D}}_t\widetilde{Z''^2} = \nabla \cdot (\overline{\rho}\widetilde{\alpha}\nabla\widetilde{Z''^2}) + \nabla \cdot \underline{\overline{\tau}}_{\widetilde{Z''^2}}^{\text{res}} - 2\overline{\rho}\widetilde{\underline{u}''Z''} \cdot \nabla\widetilde{Z} - \overline{\rho}\widetilde{\chi}_Z^{\text{res}} \,, \tag{28c}$$

where $\widetilde{\alpha}$ is the filtered diffusivity, $\widetilde{w}_C$ is the chemical source term of the progress variable. The residual scalar fluxes $\underline{\overline{\tau}}_{\psi}^{\text{res.}} = \overline{\rho}\widetilde{\underline{u}}\widetilde{\psi} - \overline{\rho}\widetilde{\underline{u}\psi}$, turbulent scalar transport $\widetilde{\underline{u}''Z''}$, and scalar dissipation rate $\widetilde{\chi}_Z^{\text{res}}$ are modeled using turbulence closure formulations.

In the LES computation, information about the filtered quantities of density $\overline{\rho}$, chemical source term for the progress variable $\widetilde{w}_C$, kinematic viscosity $\widetilde{\nu}$, and molecular diffusivity $\widetilde{\alpha}$ are required. In addition, for the computation of statistical properties of the flame structure, the filtered temperature $\widetilde{T}$, and mass fractions of CO_2, H_2O and other species are also obtained from the state relation. In order to increase the resolution in the direction of the mixture fraction variance, the unmixedness $\widetilde{S} = \widetilde{Z''^2}/(\widetilde{Z}(1-\widetilde{Z}))$ is introduced and all thermochemical quantities, which are of importance for the numerical simulation, are then parameterized as

$$\widetilde{\underline{\phi}} = \widetilde{\underline{\mathcal{G}}}_{\phi}(\widetilde{Z}, \widetilde{S}, \widetilde{C}) \,, \tag{29}$$

and $\widetilde{\phi} = (\overline{\rho}, \widetilde{w}_C, \widetilde{\nu}, \widetilde{\alpha}, \widetilde{T}, \widetilde{Y}_{CO_2}, \widetilde{Y}_{H_2O})^T$. This information is made accessible during the simulation through a structured look-up table or OANNs.

5.2 Experimental configuration

The SMH1-flame from a series of bluff-body/swirl-stabilized flame experiments is used to analyze the effects due to errors in the table interpolation and in the ANN approximation on the statistical flow field quantities in the context of an LES application.

The burner configuration (see Fig. 11) of this well-characterized flame consists of a central fuel nozzle of 3.6 mm diameter which is surrounded by a bluff body with $D_{\text{ref}} = 50$ mm diameter. Swirling air at an axial bulk velocity of $U_{\text{s}} = 42.8$ m/s is supplied through an annulus of 10 mm width. The burner is surrounded by a coflowing air stream with an axial velocity of $U_{\text{e}} = 20$ m/s. The fuel consists of a methane/hydrogen mixture in a volumetric ratio of 1:1. The bulk exit velocity of the fuel stream is $U_{\text{J}} = 140.8$ m/s. The geometric swirl number for this configuration is $S_{\text{g}} = 0.32$. The turbulent flow field of this flame series was measured by Kalt et al. (2002) and Al-Abdeli & Masri (2003), species measurements were performed by Kalt et al. (2002) and Masri et al. (2004), and all experimental results are available from Masri (2006).

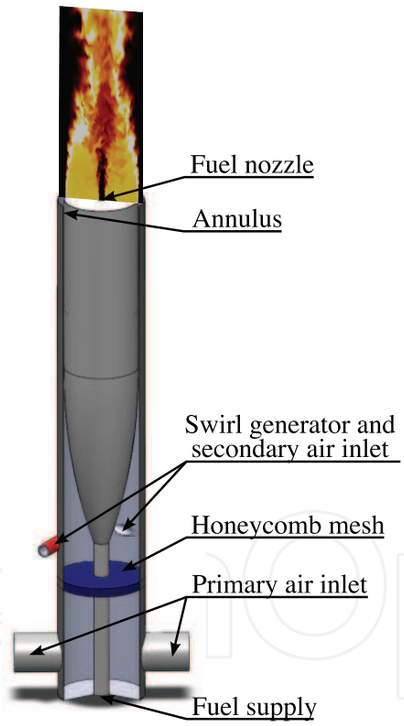


Fig. 11. Schematic of the swirl-stabilized burner (Masri, 2006).

The Favre-filtered transport equations for mass, momentum, mixture fraction, progress variable, and residual mixture fraction variance are solved in cylindrical coordinates (Pierce & Moin, 2004). The computational domain is $5\,D_{\text{ref}} \times 3\,D_{\text{ref}} \times 2\pi$ in axial, radial, and circumferential directions, respectively. The radial direction is discretized by 230 unevenly spaced grid points concentrated in the shear layer region surrounding the fuel jet and swirling annulus. The grid in axial direction uses 192 points and is stretched in downstream direction while the circumferential direction is equally spaced and uses 64 points.

The turbulent inflow profiles for the fuel nozzle and annulus are computed from a separate pipe flow simulation by enforcing the bulk axial and azimuthal velocity reported in the experiment. The GRI 2.11 mechanism (Bowman et al., 1997) is used for the description of the chemistry, consisting of 279 chemical reactions among 49 species.

5.3 Static ANN performance analysis

The accuracy, memory requirements, and computational cost associated with the table interpolation and the network evaluation are addressed in this section.

In the following, each thermochemical quantity is represented by one network, whose architecture is obtained from the GPS optimization. In this optimization process, the maximum size of the network is restricted to $N_{\mathrm{L}} = 5$ hidden layers, in which the last layer contains only one linear neuron. The maximum number of neurons in the remaining four layers is set to $N_{\mathrm{N}}^{\max} = 8$, having a sigmoidal transfer function. The coefficients in the transfer function, Eq. (12), are kept equal and constant for all neurons with $\gamma_1 = 1.075$ and $\gamma_2 = 2.0$. The training set used for adjusting the synaptic weights consists of 50,000 samples, and 1.2×10^6 data samples are used to evaluate the network fitness, defined in Eq. (10). The optimization of the networks for each thermochemical quantity is performed in parallel and takes approximately three days.

Quantity	OANN Architecture
$\overline{\rho}$	4-8-4-4-1
$\widetilde{w}_C$	6-6-8-4-1
$\widetilde{\nu}$	7-8-8-1
$\widetilde{\alpha}$	8-8-8-1
$\widetilde{T}$	8-4-7-8-1
$\widetilde{Y}_{\mathrm{CO_2}}$	4-8-8-4-1
$\widetilde{Y}_{\mathrm{H_2O}}$	3-8-8-8-1

Table 4. Architecture of the OANNs identified from the GPS algorithm for all thermochemical species in the LES calculation. The connectivity corresponds to a fully-connected feed-forward network with sigmoidal transfer function for all non-linear neurons.

The optimal architectures for the seven networks are summarized in Tab. 4. This table shows that all OANNs except those for $\widetilde{\nu}$ and $\widetilde{\alpha}$ have four non-linear layers and considerably different neural arrangements. These results support the assumption that a separate ANN for each thermochemical quantity provides greater flexibility in obtaining an optimal fitness characteristic.

Table 5 compares the memory requirements and the accuracy between four structured tables and GPS-OANNs. The accuracy of the table increases with increasing resolution. Note, however, that the finest table with 64 million grid points requires 3.4 GB of memory which is typically not available on massively parallel systems. The fitness of all OANNs is comparable to that of a tabulation with a $400 \times 50 \times 400$ resolution. However, the memory requirement to store the network is approximately 5,000 times smaller.

Figure 12 shows regression plots for two structured tables and the optimal ANNs for the temperature and the chemical source term. It is interesting to point out that the scattering for the chemical source term increases with decreasing values in the ANN representation, which might affect the solution of the flame structure in equilibrium-near regions in an LES application.

		Cost function						
Table size	Mem. [MB]	$\overline{\rho}$	$\widetilde{w}_C$	$\widetilde{\nu}$	$\widetilde{\alpha}$	$\widetilde{T}$	$\widetilde{Y}_{CO_2}$	$\widetilde{Y}_{H_2O}$
GPS-OANN	0.1	-2.48	-2.67	-2.52	-2.65	-2.49	-1.98	-2.62
100×13×100	7	-1.54	-1.81	-1.84	-1.83	-1.81	-1.61	-1.79
200×25×200	53	-1.86	-2.15	-2.16	-2.17	-2.12	-2.12	-2.10
400×50×400	427	-2.55	-2.78	-2.72	-2.83	-2.66	-2.07	-2.91
800×100×800	3,417	-2.55	-2.98	-2.79	-2.92	-2.74	-2.13	-2.86

Table 5. Comparison of memory requirements and network fitness between tabulation method with increasing resolution and OANNs. Note that a reduction of the table size by a factor of two can be achieved by storing the data in single precision format.

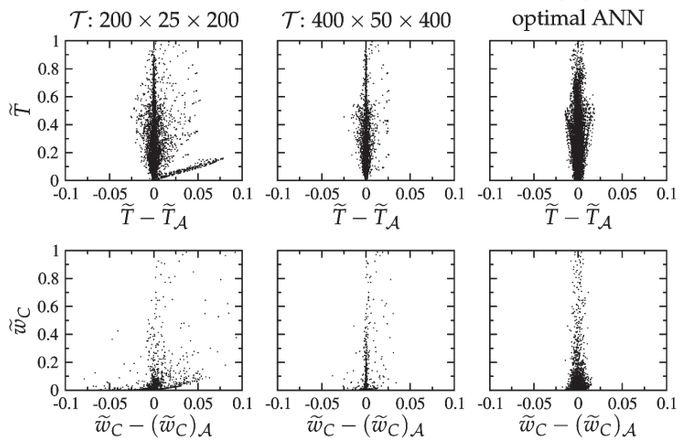


Fig. 12. Regression analysis for the table representation and ANN approximation for the normalized temperature (top) and chemical source term (bottom). The subscript "$\mathcal{A}$" refers to data evaluated from OANNs, and $\widetilde{T}$ and $\widetilde{w}_C$ are the sample data.

5.4 Dynamic ANN performance analysis

In this section, results from LES computations employing look-up tables and OANNs for the chemistry representation are presented. Statistical data are collected over four flow-through-times. For reference, the azimuthal and temporal averaged quantity of a scalar ϕ is denoted by $\langle\widetilde{\phi}\rangle$, and the resolved scalar variance is indicated by $\langle\widetilde{\phi}''^2_<\rangle$.

Figure 13 compares the averaged axial velocity field between experiments (left) and ANN simulation (right). The solid line in both figures separates the regions of positive and negative axial velocities. From both figures the decay of the central fuel jet and the contraction of the swirling oxidizer stream after the recirculation bubble are evident.

Radial profiles of the mean and rms mixture fraction are shown in Fig. 14. Both, simulations and experiments predict a region of homogeneous mixture directly behind the bluff body. The location of stoichiometry is aligned with the inner shear layer of the swirling stream. Since the value of the stoichiometric mixture fraction is $Z_{st} = 0.042$, accurate prediction of the shear layer is crucial for the determination of species and temperature distributions. Even though both simulations slightly over-predict the spreading rate of $\langle\widetilde{Z}\rangle$, the computed axial decay rates are in excellent agreement with the experiments.

Radial profiles of mean temperature and CO_2 mass fraction are shown in Fig. 15. The numerical results for $\langle\widetilde{Y}_{CO_2}\rangle$ from the simulations employing the tabulation method are in

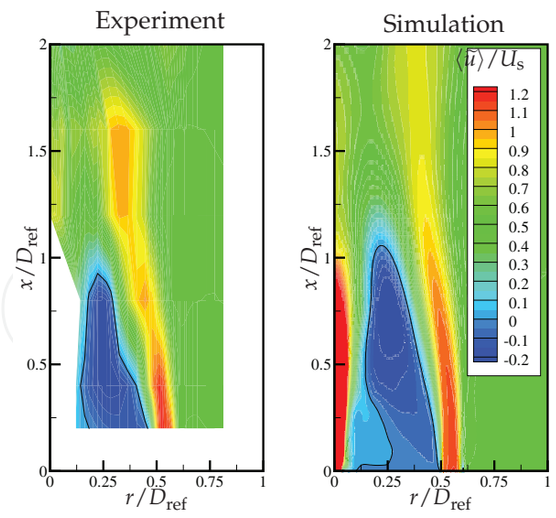


Fig. 13. Comparison of averaged axial velocity fields between experiment (left) and LES calculation employing an OANN chemistry representation (right).

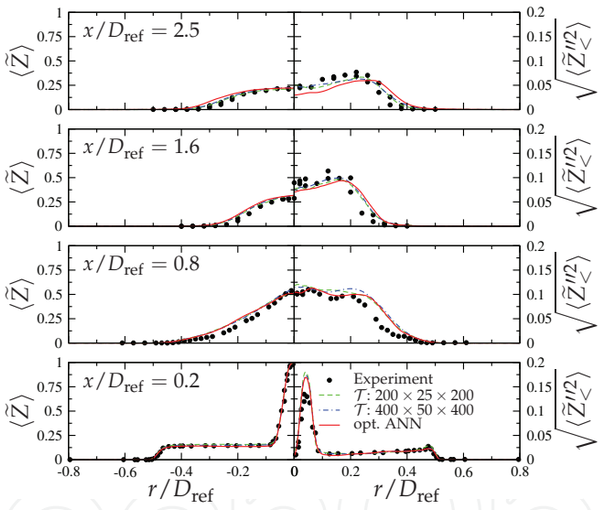


Fig. 14. Comparison of measured and calculated mean and resolved rms statistics of mixture fraction at different axial locations in the Sydney SMH1 flame.

reasonable agreement with experimental data; however, the simulation with ANN chemistry representation under-predicts $\widetilde{Y}_{CO_2}$ on the fuel-rich side of the flame. This is primarily attributed to the poor fitness of the CO_2-OANN. The shift of the profiles for $\langle\widetilde{T}\rangle$ and $\langle\widetilde{Y}_{CO_2}\rangle$ towards the lean side of the flame is due to the slight over-prediction of the mixture fraction in the shear layer surrounding the swirling oxidizer stream. The mean temperature in the region of the bluff body at $x/D_{ref} = 0.2$ is correctly predicted.

In summary, the comparisons of statistical flow field results shows that both methods for the chemistry representation yield similar results. Discrepancies for CO_2-species predictions are primarily attributed to the poor fitness of the ANN, which can be further improved

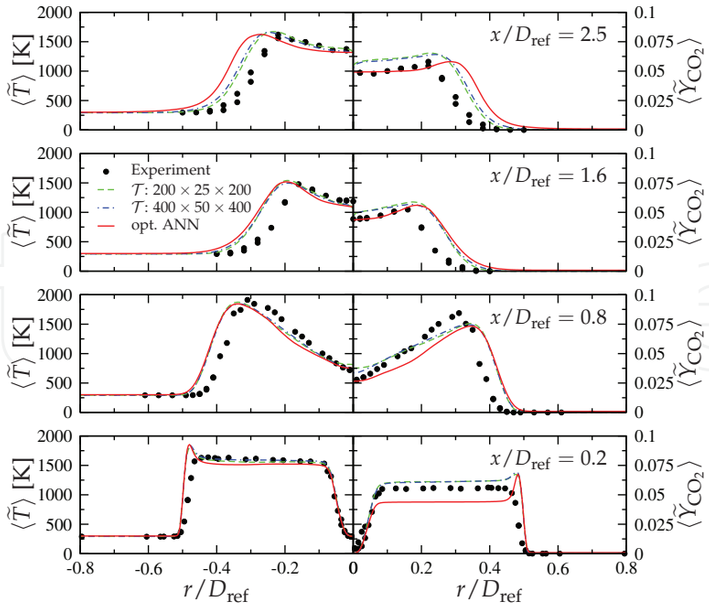


Fig. 15. Comparison of measured and calculated mean temperature and CO_2 mass fraction at different axial locations in the Sydney SMH1 flame.

by considering larger network architectures. These discrepancies emphasize the sensitivity of the flow field structure to the ANN-representation. Furthermore, they demonstrate the significance of evaluating the network performance in dynamic applications, in which the integrated effect of ANN-approximation errors on the system evolution can be assessed.

6. Conclusions

Topological optimization of ANNs for application to chemically reacting flows was conducted. To this end, a generalized pattern search method was presented as attractive alternative to previously employed ANN optimization strategies. This GPS method offers extensive flexibility in the implementation and allows for the consideration of optimization parameters that are of categorical and continuous type. In addition, an evolutionary strategy (ES) was also utilized in order to establish a direct comparison with the GPS method. For the present application, it was shown that, compared to ES, the GPS method is more robust, converges faster, and returns ANN topologies with better performance.

The GPS method was applied to the generation of optimal ANNs to approximate chemical source terms, species mass fractions, and other thermochemical quantities in chemical reacting systems. For this, two different combustion problems with increasing complexity were considered. In the first problem, the combustion of a fuel/air mixture in decaying homogeneous isotropic turbulence was modeled, in which the chemical source term, describing the fuel-conversion rate, was approximated by an optimal ANN. The particularly interesting aspect of this problem is that it allows for a systematic evaluation of ANN approximation errors on the transient evolution of the reacting system. The second problem focused on the unsteady three-dimensional combustion of a methane/hydrogen-air mixture in a technical-relevant burner system. For this, large-scale simulations were employed, and the accuracy of ANNs and conventional tabulation methods were assesses in the context of high-performance computations of turbulent reacting flows.

To enable a comprehensive assessment of the network fitness, two complementary metrics were examined. The first metric considers the static ANN performance analysis, and evaluates the ability of the network to represent untrained samples. A main shortcoming of this metric, which is also referred to as "testing," is that it does not account for feedback-effects of ANN approximation errors on the evolution of the dynamic system. To address this issue, a dynamic ANN performance measure is developed. This metric provides additional diagnostics, and is useful for applications in which ANNs are used as substitutes for complex function-evaluations in dynamic systems. It was shown that this dynamic performance metric provides a practical-relevant assessment of the network fitness, whereas the static ANN analysis gives typically too optimistic estimates.

7. Acknowledgments

The author gratefully acknowledges financial support through ONR under Grant No. N00014-10-1-0717. Helpful discussions with Heinz Pitsch, Christoph Schmitt, and Rodney Fox on the ANN optimization and the LFP model formulation are much appreciated.

8. References

Abramson, M. A. (2004). Mixed variable optimization of a load-bearing thermal insulation system using a filter pattern search algorithm, *Optimization and Engineering* 5(2): 157–177.

Al-Abdeli, Y. A. & Masri, A. R. (2003). Stability characteristics and flowfields of turbulent non-premixed swirling flames, *Comb. Theory Modelling* 7: 731–766.

Angeline, P. J., Saunders, G. M. & Pollack, J. B. (1994). Evolutionary algorithm that constructs recurrent neural networks, *IEEE Trans. Neural Networks* 5: 54–65.

Audet, C. & Dennis, Jr., J. E. (2000). Pattern search algorithms for mixed variable programming, *SIAM J. Optim.* 11(3): 573–594.

Audet, C. & Dennis, Jr., J. E. (2003). Analysis of generalized pattern searches, *SIAM J. Optim.* 13(3): 889–903.

Bäck, T. & Schwefel, H.-P. (1993). An overview of evolutionary algorithms for parameter optimization, *Evolutionary Computation* 1 (1): 1–23.

Blasco, J. A., Fueyo, N., Dopazo, C. & Ballester, J. (1999). Modelling the temporal evolution of a reduced combustion chemical system with an artificial neural network, *Combust. Flame* 113(1-2): 38–52.

Blasco, J. A., Fueyo, N., Dopazo, C. & Chen, J. Y. (2000). A self-organizing-map approach to chemistry representation in combustion applications, *Combust. Theory Modelling* 4(1): 61–76.

Blasco, J. A., Fueyo, N., Larroya, J. C., Dopazo, C. & Chen, J. Y. (1999). Single-step time-integrator of a methane-air chemical system using artificial neural networks, *Computers and Chemical Engineering* 23(9): 1127–1133.

Booker, A. J., Dennis, Jr., J. E., Frank, P. D., Serafini, D. B., Torczon, V. & Trosset, M. W. (1999). A rigorous framework for optimization of expensive functions by surrogates, *Structural Optimization* 17(1): 1–13.

Bornholdt, S. & Graudenz, D. (1992). General asymmetric neural networks and structure design by genetic algorithms, *Neural Networks* 5(2): 327–334.

Bowman, C. T., Hanson, R. K., Davidson, D. F., Gardiner, W. C., Lissianski, V., Smith, G. P., Golden, D. M., Frenklach, M. & Goldenberg, M. (1997). GRI-Mech 2.11. available from `http://www.me.berkeley.edu/gri-mech/`.

Chen, J. Y., Blasco, J. A., Fueyo, N. & Dopazo, C. (2000). An economical strategy for storage of chemical kinetics: Fitting in situ adaptive tabulation with artificial neural networks, *Proc. Comb. Institute* 28: 115–121.

Christo, F. C., Masri, A. R. & Nebot, E. M. (1996). Artificial neural network implementation of chemistry with PDF simulation of H_2/CO_2 flames, *Combust. Flame* 106(4): 406–427.

Christo, F. C., Masri, A. R., Nebot, E. M. & Pope, S. B. (1996). An integrated PDF/neural network approach for simulating turbulent reacting systems, *Proc. Comb. Institute* 26: 43–48.

Flemming, F., Sadiki, A. & Janicka, J. (2005). LES using artificial neural networks for chemistry representation, *Progress in Computational Fluid Dynamics* 5(7): 375–385.

Fogel, D. B. & Fogel, L. J. (1990). Evolving neural networks, *Biological Cybern.* 63(6): 487–493.

Fox, R. O. (2003). *Computational Models for Turbulent Reacting Flows*, Cambridge University Press, Cambridge.

Frean, M. (1990). The upstart algorithm: A method for constructing and training feedforward neural networks, *Neural Computation* 2(2): 198–209.

Germano, M., Piomelli, U., Moin, P. & Cabot, W. H. (1991). A dynamic subgrid-scale eddy viscosity model, *Phys. Fluids A* 3(7): 1760–1765.

Hagan, M. T., Demuth, H. B. & Beale, M. (1996). *Neural Network Design*, PWS Publishing Company, Boston, Massachusetts, USA.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Upper Saddle River, N.J.

Husken, M., Jin, Y. & Sendhoff, B. (2005). Structure optimization of neural networks for evolutionary design optimization, *Soft Computing* 9(1): 21–28.

Ihme, M., Cha, C. M. & Pitsch, H. (2005). Prediction of local extinction and re-ignition effects in non-premixed turbulent combustion using a flamelet/progress variable approach, *Proc. Combust. Inst.* 30: 793–800.

Ihme, M., Marsden, A. L. & Pitsch, H. (2008). Generation of optimal artificial neural networks using a pattern search algorithm: Application to approximation of chemical systems, *Neural Computation* 20(2): 573–601.

Ihme, M., Schmitt, C. & Pitsch, H. (2009). Optimal artificial neural networks and tabulation methods for chemistry representation in LES of a bluff-body swirl-stabilized flame, *Proc. Combust. Inst.* 32: 1527–1535.

Kalt, P. A. M., Al-Abdeli, Y. A., Masri, A. R. & Barlow, R. S. (2002). Swirling turbulent non-premixed flames of methane: Flow field and compositional structure, *Proc. Combust. Inst.* 29: 1913–1919.

Koza, J. R. & Rice, J. P. (1991). Genetic generation of both the weights and architecture for a neural network, *Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN'91 Seattle)*, pp. 397–404.

Lam, S. H. & Goussis, D. A. (1988). Understanding complex chemical kinetics with computational singular perturbation, *Proc. Comb. Inst.* 22: 931–941.

Lewis, R. M. & Torczon, V. (1999). Pattern search algorithms for bound constrained minimization, *SIAM J. Optim.* 9(4): 1082–1099.

Lewis, R. M. & Torczon, V. (2000). Pattern search methods or linearly constrained minimization, *SIAM J. Optim.* 10(3): 917–941.

Lilly, D. K. (1992). A proposed modification of the Germano subgrid-scale closure method, *Phys. Fluids A* 4(3): 633–635.

Maas, U. & Pope, S. B. (1992). Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space, *Combust. Flame* 88(3-4): 239–264.

Masri, A. R. (2006). Web site for the Sydney swirl flame series http://www.aeromech.usyd.edu.au/thermofluids.

Masri, A. R., Kalt, P. A. M. & Barlow, R. S. (2004). The compositional structure of swirl-stabilised turbulent nonpremixed flames, *Combust. Flame* 137: 1–37.

McKay, M. D., Beckman, R. J. & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21(2): 239–245.

Miller, G. F., Todd, P. M. & Hegde, S. U. (1989). Designing neural networks using genetic algorithms, *in* J. D. Schaffer (ed.), *Proc. 3rd Int. Conf. Genetic Algorithms and Their Applications*, pp. 379–384.

Mozer, M. C. & Smolensky, P. (1989). Skeletonization: A technique for trimming the fat from a network via relevance assessment, *Connection Sci.* 1(1): 3–26.

Peters, N. (1983). Local quenching due to flame stretch and non-premixed turbulent combustion, *Combust. Sci. Tech.* 30: 1–17.

Peters, N. (1984). Laminar diffusion flamelet models in non-premixed turbulent combustion, *Prog. Energy Combust. Sci.* 10(3): 319–339.

Pierce, C. D. & Moin, P. (2004). Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion, *J. Fluid Mech.* 504: 73–97.

Pope, S. B. (1997). Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation, *Combust. Theory Modelling* 1(1): 41–63.

Porto, V. W., Fogel, D. B. & Fogel, L. J. (1995). Alternative neural network training methods, *IEEE Expert* 10(3): 16–22.

Rechenberg, I. (1973). *Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, PhD thesis, Technical University Berlin, Germany.

Sagaut, P. (1998). *Large Eddy Simulation for Incompressible Flows: An Introduction*, Springer, Berlin.

Schwefel, H.-P. (1977). *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*, Birkhäuser, Basle, Switzerland.

Schwefel, H.-P. (1981). *Numerical Optimization of Computer Models*, Wiley, Chichester.

Sen, B. A. & Menon, S. (2008). Turbulent premixed flame modeling using artificial neural networks based chemical kinetics, *Proc. Combust. Inst.* 32: 1605–1611.

Sen, B. A. & Menon, S. (2010). Linear eddy mixing based tabulation and artificial neural networks for large eddy simulations of turbulent flames, *Combust. Flame* 157: 62–74.

Serafini, D. B. (1998). *A framework for managing models in nonlinear optimization of computationally expensive functions*, PhD thesis, Rice University, Houston, TX.

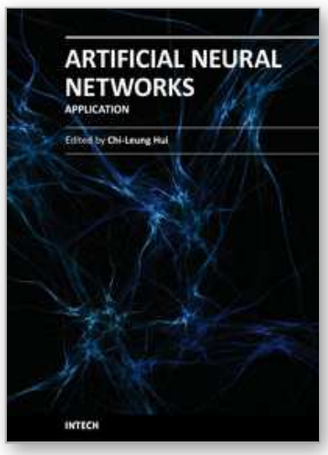
Sripakagorn, P., Mitarai, S., Kosály, G. & Pitsch, H. (2004). Extinction and reignition in a diffusion flame: A direct numerical simulation study, *J. Fluid Mech.* 518: 231–259.

Tang, K. S., Chan, C. Y., Man, K. F. & Kwong, S. (1995). Genetic structure for NN topology and weights optimization, *Proc. 1st IEE/IEEE International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications (GALESIA'95)*, pp. 250–255.

Tonse, S. R., Moriarty, N. W., Brown, N. J. & Frencklach, M. (1999). PRISM: piecewise reusable implementation of solution mapping. An economical strategy for chemical kinetics, *Israel Journal of Chemistry* 39(1): 97–106.

Torczon, V. (1997). On the convergence of pattern search algorithms, *SIAM J. Optim.* 7(1): 1–25.

Yao, X. (1999). Evolving artificial neural networks, *Proc. IEEE* 87(9): 1423–1447.



Artificial Neural Networks - Application

Edited by Dr. Chi Leung Patrick Hui

ISBN 978-953-307-188-6
Hard cover, 586 pages
Publisher InTech
Published online 11, April, 2011
Published in print edition April, 2011

This book covers 27 articles in the applications of artificial neural networks (ANN) in various disciplines which includes business, chemical technology, computing, engineering, environmental science, science and nanotechnology. They modeled the ANN with verification in different areas. They demonstrated that the ANN is very useful model and the ANN could be applied in problem solving and machine learning. This book is suitable for all professionals and scientists in understanding how ANN is applied in various areas.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen