Abstract: Self-organizing maps (SOM) have been successfully applied in many fields of research. In this paper, we demonstrate the use of a neural-network-based tool for a data analysis in fluidized bed energy plants. The software is based on self-organizing maps. Reference vectors of SOMs can be classified by K-means algorithm into clusters, which represented different states of processes. The differences between reference vectors could further be used for determination of reasons of non-optimal process states. The results show that this method can also be successfully applied to process state monitoring and optimisation in fluidized bed energy plants.

Keywords: neural networks, self-organizing maps, process optimization, analysis tool.
of the combustion chamber. Circulating fluidized bed boilers use high fluidizing velocities, so the particles are constantly held in the flue gases, and pass through the main combustion chamber into a separator, from which the larger particles are extracted and returned to the combustion chamber, while the finer particles are removed from the flue gases by an electrostatic precipitator or a baghouse located downstream of the boiler’s convection section. Due to the large heat capacity of the bed, the combustion is stable and supporting fuels such as oil or gas are needed only during the start-up. The intense turbulence of the circulating fluidized bed ensures good mixing and combustion of fuel. Combustion typically takes place at about 850 – 900 °C bed temperature.

The raw data are extracted monthly from databases of utility scale CFB boilers. The time resolution of the data set is typically 15 minutes. The size of an example data matrix is 2880x46 (2880 rows, 46 variables in columns).

3 COMPUTATIONAL METHODS

3.3 Self-organizing maps

SOMs can be used to map n-dimensional input vectors to the neurons in a two-dimensional array, whereby the input vectors sharing common features end up in the same or neighbouring neurons (Kohonen, 2001). This preserves the topological order of the original input data. The map reflects variations in the statistics of the data sets and selects common features, which approximate to the distribution of the data points. Each neuron is associated with an n-dimensional reference vector, which provides a link between the output and input spaces. This lattice type of array of neurons, i.e. the map, can be illustrated as a rectangular, hexagonal, or even irregular organisation. However, hexagonal organization is most often used, as it best presents the connections between neighbouring neurons. The size of the map as defined by the number of neurons can be varied depending on the application; the more neurons, the more details appear.

The SOM analysis includes an unsupervised learning process. First, random values for the initial reference vectors are sampled from an even distribution, whereby the limits are determined by the input data. During the learning, the input data vector is mapped onto a given neuron (best matching unit, BMU) based on a minimal n-dimensional distance between the input vector and the reference vectors of the neurons. The neighbours of the central activated neuron are also activated according to a network-topology-dependent neighbourhood function, a Gaussian distribution. The usual procedure is to utilize an initially wide function, which is subsequently reduced in width during learning to the level of individual neurons. Reference vectors of activated neurons will become updated after this procedure. This procedure features a local smoothing effect on the reference vectors of neighbouring neurons leading eventually to a global ordering (Kohonen, 2001). The analysis software has been made using the Matlab-software platform (Mathworks, Natick, MA, USA).

3.3 K-means method

The K-means algorithm was applied to the clustering of the map. The K-means method is a well-known non-hierarchical cluster algorithm (MacQueen, 1967). The basic version begins by randomly picking
K cluster centres, assigning each point to the cluster whose mean is closest in a Euclidean-distance-sense, then computing the mean vectors of the points assigned to each cluster, and using these as new centres in an iterative approach.

3.3 Optimization and subtraction analysis

For each neuron the reference vectors, which represent the common features of the data in each neuron, are defined during the training of the map. Therefore the components of the reference vectors vary in different parts of the map. An optimal neuron for the whole map or a cluster can be determined using one or more components of the reference vectors. For example, if we are interested in the control of the NOx concentration, thus the optimal neurons of each cluster were simply the neurons, where the NOx component of the reference vectors is the smallest one.

In the subtraction analysis, reference vectors of two neurons are subtracted from each other. This method can be used for identification of any differences in factors between corresponding subgroups of two neurons. So the difference between reference vectors of the optimal neuron and the winning neuron, when for example the NOx concentration is high, can indicate the reasons for high emissions.

4 RESULTS AND DISCUSSION

Here we shortly describe an analysis tool developed for process state monitoring and optimization of fluidized bed energy plants. Figure 2 shows the schematic diagram of the software. The system contains four parts: (1) pre-processing, (2) variable selection and process lag, (3) modelling and (4) post-processing parts. In the post-processing part, we have three applications namely state determination, process optimisation and emission reporting. The software has been made under a Matlab-software platform exploiting the SOM method and subtraction analysis of reference vectors.

The pre-processing part of the software can be used for plotting, scaling and filtering of large data sets. Users can also calculate easily new derivative variables using the older ones.

The second part is also important in the case of process data. The process lags can be long and so they should be taken into account before modelling of the data. The manual or automatic selection of the most important variables is also a useful operation, when a problem is new and its causes are unknown.

![Fig. 2: A schematic diagram of the SOM-based analysing software. The system contains four parts: (1) pre-processing, (2) variable selection and process lag, (3) modelling and (4) post-processing parts.](attachment:image1.png)

![Fig. 3: The modelling interface of the analysis software. A SOM is trained using the data of a fluidized bed energy plant.](attachment:image2.png)
Different sub-processes or the whole process can be modelled by the software using the process data and the SOM method. The interface of the modelling part of the software is shown in Figure 3, in which a SOM has been obtained by training a self-organizing network with the data of a fluidized bed energy plant.

SOMs can be clustered by using reference vectors and the K-means method. These clusters represent possible states of the process. An example map and its five clusters are shown in Figure 4. The brief descriptions of the clusters are also in Figure 4.

In the software, the optimal neurons for each cluster can be determined based on reference vectors, e.g. they can be neurons, where the NOx emissions are smallest. The difference between reference vectors of the optimal neurons and the neuron in each time step could be used for determination of possible reasons of non-optimal process states. Figure 5 illustrates the interface of this optimisation part, which contains three parts: (1) The rules for process optimisation, (2) the optimal neuron and the winning neuron of an input vector are visualized on the SOM and (3) the differences between reference vectors are shown by a bar graph representation.

Nowadays energy plants have to report monthly the exceeding of emissions. Our software can also be used for writing pre-reports on these exceedings and their possible reasons (Figure 6). The user can define the limits of emission, which can be several in the same analysis. The software finds all exceedings from the data set and calculates possible reasons using the subtraction analysis of reference vectors.

![Fig. 4: The clustering interface of the analysis software. A SOM is clustered using the K-means method and the reference vectors. The colours visualize the five clusters of the map, which represent the states of the process. Short descriptions for each cluster are also shown.](image)

![Fig. 5: The optimization interface of the analysis software, which contains three parts: (1) The rules for process optimisation, (2) the optimal neuron and the winning neuron of an input vector visualized on the SOM and (3) the differences between reference vectors by a bar graph representation.](image)
The software finds all exceedings and calculates possible reasons using the subtraction analysis of reference vectors, which can be saved into a file in a text format.

The exceedings can be gone through manually and checked that the reasons are correct. The exceedings and reasons can also be saved into a file in a text format, which can be used as a pre-report for emission exceedings.

The software has been made for processing large data sets, which are typically in the process industry. In this application a self-organising map forms the basis of the process model. However, building non-linear models requires inevitably more expertise from the users than building linear models, which also require non-trivial amount of statistical knowledge. Thus these generic modelling tools may normally be used by process experts, who are able to analyse and validate the results before taking them to operational use. On the other hand, laborious data processing stages can be automated using the software, which facilitates the routine and regular analysis work of process experts and so enables to do efficiently service business.

5 CONCLUSION

We have presented an intelligent tool, which can be used in service business of a manufacturer of energy plans. The results indicated that the SOM analysis provides an efficient method for data analysis in the process industry. Therefore this kind of intelligent data-driven approach is a fruitful way of developing tools for the process state monitoring and optimization in the process industry.

REFERENCES


ACKNOWLEDGEMENTS

This research was funded by National Technology Agency of Finland (Tekes), An-dritz Ltd., Foster Wheeler Energy Ltd., Honeywell Ltd. and Varenso Ltd.