

Simone Sieg · Bernhard Stutz · Timm Schmidt ·  
Fred Hamprecht · Wilhelm F. Maier

## A QCAR-approach to materials modeling

Received: 25 April 2005 / Accepted: 8 September 2005 / Published online: 25 April 2006  
© Springer-Verlag 2006

**Abstract** Little is known about the relationship between the function and structure of materials. Materials (solids with a function) are complex entities and a better knowledge of the parameters that contribute to function is desirable. Here, we present modeling approaches that correlate chemical composition with function of heterogeneous catalysts. The complete composition space of the mixed oxides of Ni–Cr–Mn and of Ni–Co–Mo–Mn (10% spacing) have been measured for the oxidation of propene to acroleine. The data have been collected, visualized and modeled. Different mathematical approaches such as Support Vector Machines, multilevel B-splines approximation and Kriging have been applied to model this relationship. High-throughput screening data of ternary and quaternary composition spreads are approximated to locate catalysts of high activity within the search space. For quaternary systems, slice plots offer a good tool for visualization of the results. Using these approximation techniques, the composition of the most active catalysts can be predicted. The study documents that distinct relationships between chemical composition and catalytic function exist and can be described by mathematical models.

**Keywords** Selective oxidation · Acroleine materials modeling · Kriging · Heterogeneous catalysts · Composition–activity relationship · QCAR

### Introduction

In the late seventies, quantum mechanical and force-field calculations had a low reputation and experimentalists did

not accept these calculations as serious and reliable data. During the last three decades through the work of Paul Schleyer and his colleagues around the globe this has changed completely and calculations and molecular modeling now complement experiments as well as predict new molecules. Today, nobody is surprised any longer if experiments confirm predictions from calculations or if calculations prove experiments to be wrong. Calculations have become an accepted tool in molecular science. One of the important acronyms here is QSAR (quantitative structure–activity relationships), where correlations between molecular structure and activity are established and used for prediction and understanding. Molecular structures are a well defined parameter space in organic and inorganic chemistry. In organometallic chemistry the parameter space becomes even larger, due to the many central atoms and their larger variation in coordination and electronic states.

However, drugs and molecular structures represent only a small portion of the needs of our society for better or new substances. Clearly, the greatest needs are in materials. Materials, solids with a function, are strongly affected by chemistry, but can mostly not be described by a structure. Instead, materials are often poorly defined composites containing amorphous, polymeric, and crystalline components. Materials development in the past has been a time-consuming process of sequential materials synthesis, property measurements and analysis in the “one at a time” approach. During the last 10 years high-throughput technologies have spread from pure drug-related applications in the pharmaceutical industry to materials laboratories in industry and academia. A large number of technological applications have been developed and materials can now be optimized rapidly and new materials are discovered [1–6].

There has been significant resistance to the use of High Throughput Experimentation (HTE) in many materials applications. Especially complex materials, such as catalysts, polymers or materials based on formulations have been predicted not to lend themselves to HTE. In the meantime HTE has been applied successfully to an ever-increasing

S. Sieg · B. Stutz · T. Schmidt · W. F. Maier (✉)  
Lehrstuhl für Technische Chemie,  
Universität des Saarlandes,  
66123 Saarbrücken, Germany  
e-mail: w.f.maier@mx.uni-saarland.de

F. Hamprecht  
Interdisciplinary Center for Scientific Computing (IWR),  
University of Heidelberg,  
D-69120 Heidelberg, Germany

number of complex materials-development problems including polymers, electronic materials and catalysts.

Some of the promises not delivered yet by HTE are discovery of knowledge and correlation of the function of a solid with its structure. Here things become difficult. While most drugs are clearly identified by their molecular structure, materials are complex entities that do not lend themselves to the idea of structure–activity relationships. Heterogeneous catalysts are typical complex materials and were chosen as models in this study. Today's state of theory does not allow us to predict the catalytic activity of new compositions. Catalyst optimization is therefore a tedious process of materials preparation followed by individual testing. Correlating activity with useful descriptors, such as composition, would open the complex field of materials to theoretical modeling and predictions. In an interesting study, the groups of Schüth and Mirodatos investigated 467 catalysts and their catalytic performance for the oxidation of propene with oxygen. The catalysts were of different origin and prepared by various methods [7]. The authors searched for general descriptors that correlate with the catalytic activity described by the dataset [8]. Because of the difficulty in assigning appropriate structure parameters to materials, a complex strategy was developed. In the dataset, parameters such as mean electron affinity, mean electronegativity or mean molar mass of all elements in the catalysts were identified as descriptors, while chemical composition does not correlate with the function of the materials [7]. This was to be expected, since catalytic activity is very sensitive to the preparation method of the catalyst. The function of a catalyst is not defined by composition, but by the exact synthesis protocol and the synthesis procedure, such as co-precipitation, impregnation, sol–gel-synthesis or powder synthesis and the precursors used. Changing the sequence of additions and allowing different drying times and calcination temperatures will result in a variety of different materials of identical composition, but different catalytic activity and selectivity. Such a dataset cannot be used in the search for the optimal composition of a catalyst. When optimization of chemical composition becomes the issue of HTE, great care has to be taken that the synthesis of the materials of varying composition is carried out under identical synthetic conditions. This implies synthesis recipes that tolerate a broad variation in chemical composition. We have found that this can be achieved very generally with specially optimized sol–gel-procedures. With such a set of materials, catalytic activity, measured under identical reaction conditions, should correlate with chemical composition and can be called QCAR (quantitative composition–activity relationship) [9]. If such correlations can be found, modeling catalytic activity becomes possible, as would be the prediction of optimal compositions. One essential aspect for such an approach is that the dependence of the material's function (catalytic activity or other desired properties) on chemical composition is continuous and not erratic.

---

## QCAR modeling applied to heterogeneous catalysts

The main goal of this work is the modeling of the function of materials dependent on their composition. As model materials a discrete set of catalysts has been prepared. The complete composition space of the mixed oxides of Ni, Cr, Mo, Co and Mn has been synthesized (spacing steps of 10%, 1,001 catalyst samples) based on a compositionally tolerant sol–gel-recipe developed for this purpose. The materials have been prepared with the help of a pipetting robot in arrays of HPLC-vials through the control of the software Plattenbau [10]. The thus prepared sols were allowed to gel, then calcined at 300°C and manually transferred into the wells of a 207-well slate plate. Prior to the measurement, the reactor was heated to 400°C in synthetic air for 3 h. Catalytic activity for the oxidation of propene to acroleine was monitored at 350°C (28.6 vol.% propene in synthetic air at ambient pressure) on a high-throughput reactor by sequential rapid analysis with a micro GC [11]. This dataset not only contains the activity and selectivity of the mixed oxides composed of five metals, but also all lower mixed oxides (pure oxides, binary, ternary and quaternary mixed oxides).

---

### Parameter space

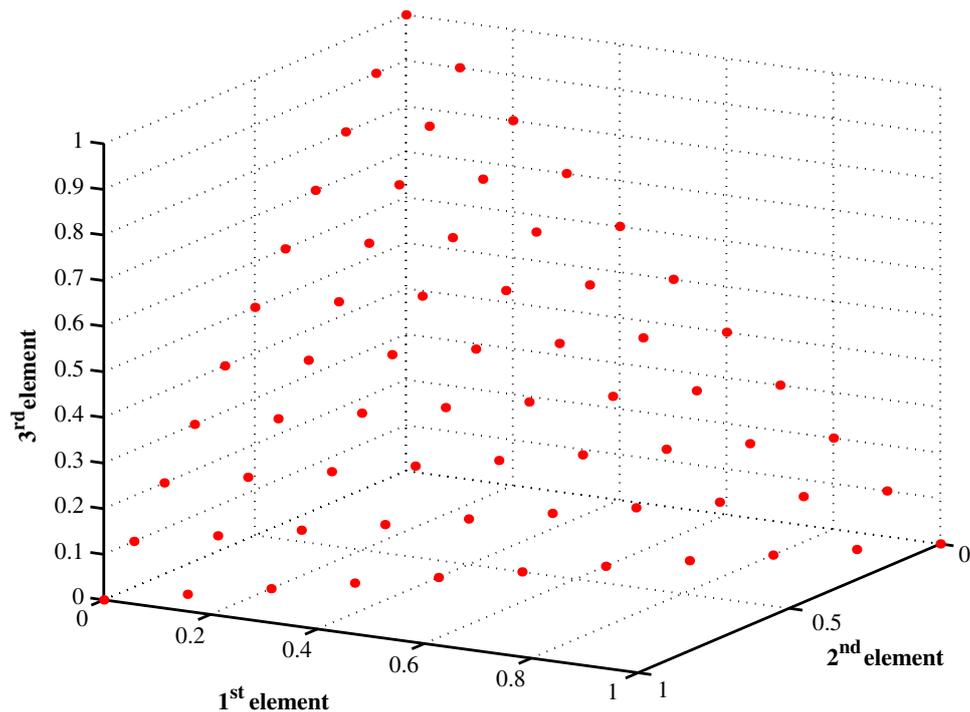
Mixed metal oxides are potential heterogeneous catalysts, composed of  $n$  different metals in various contents and oxidation states. Since the oxidation states, and therefore the oxygen content of the materials, may vary with preparation conditions, pretreatment or during catalysis, it is not specified. The total metal content (defined as 100%) is uniquely determined if the contents of  $n-1$  metals are specified. From a mathematical point of view, a catalyst of  $n$  metals can be regarded as a point from an  $n$ -dimensional space where each coordinate axis corresponds to one element. Figure 1 illustrates this for ternary catalysts. Ternary mixtures can be regarded as points lying on the diagonal of the unit cube in the Cartesian coordinate system if the composition is interpreted as Cartesian coordinates within the range from 0 to 1 (0–100%).

Because of the redundancy of the  $n$ th dimension, the points can be transformed by an adequate coordinate transform into a  $(n-1)$ th dimensional representation. Figure 2 shows the three-dimensional case.

Thus, a ternary composition spread can be illustrated in two dimensions. Analogously, catalysts consisting of four different metals can be illustrated in three dimensions: they form a tetrahedron as illustrated in Fig. 3. Clearly visible is the rather steady catalytic activity. A clear region of higher activity is in the tetrahedron towards higher Co and Cr contents. This also verifies that the catalytic activity changes smoothly with composition and may provide a good basis for modeling.

From Figs. 1 and 2 it becomes clear what the mathematical model should achieve. Modeling the three dimen-

**Fig. 1** Ternary composition spread interpreted as Cartesian coordinates in three dimensions



sional case, the result is a triangular surface providing catalytic activity for any composition within the parameter space. In more than two dimensions, the result will be a hyperplane describing the data. The goal of interpolation or approximation is to reconstruct an underlying function (e.g. surface) that may be evaluated at any desired set of positions. This serves to propagate the information associated with the discrete data smoothly onto all positions in the domain. The field of mathematical approximation techniques is very large and the approximation in itself can be done in a variety of ways such that no unique solution exists. Using different approaches, the approximating surfaces differ in shape and smoothness according to the model chosen. As many real-world applications, this problem ranks among the so-called *ill-posed* problems. The following figure stresses the spectrum of possible interpolations and gives an idea what ill-posed means in this context.

Figure 4 illustrates in a simple way what the difficulties of interpolation and approximation problems are. All four curves interpolate or approximate the data points but possess different shapes. As with catalyst data the underlying composition–activity relationship is an unknown function, the goodness of fit cannot be calculated directly (e.g. least-squares distance, mean squared error etc.) but one has to think about a criterion that handles the problem structure.

For the interpolation of catalyst data, we assume that samples of the same composition should show nearly the same catalytic behavior within the range of synthesis and measurement errors. Samples of similar compositions are assumed to cause a similar catalytic activity such that the

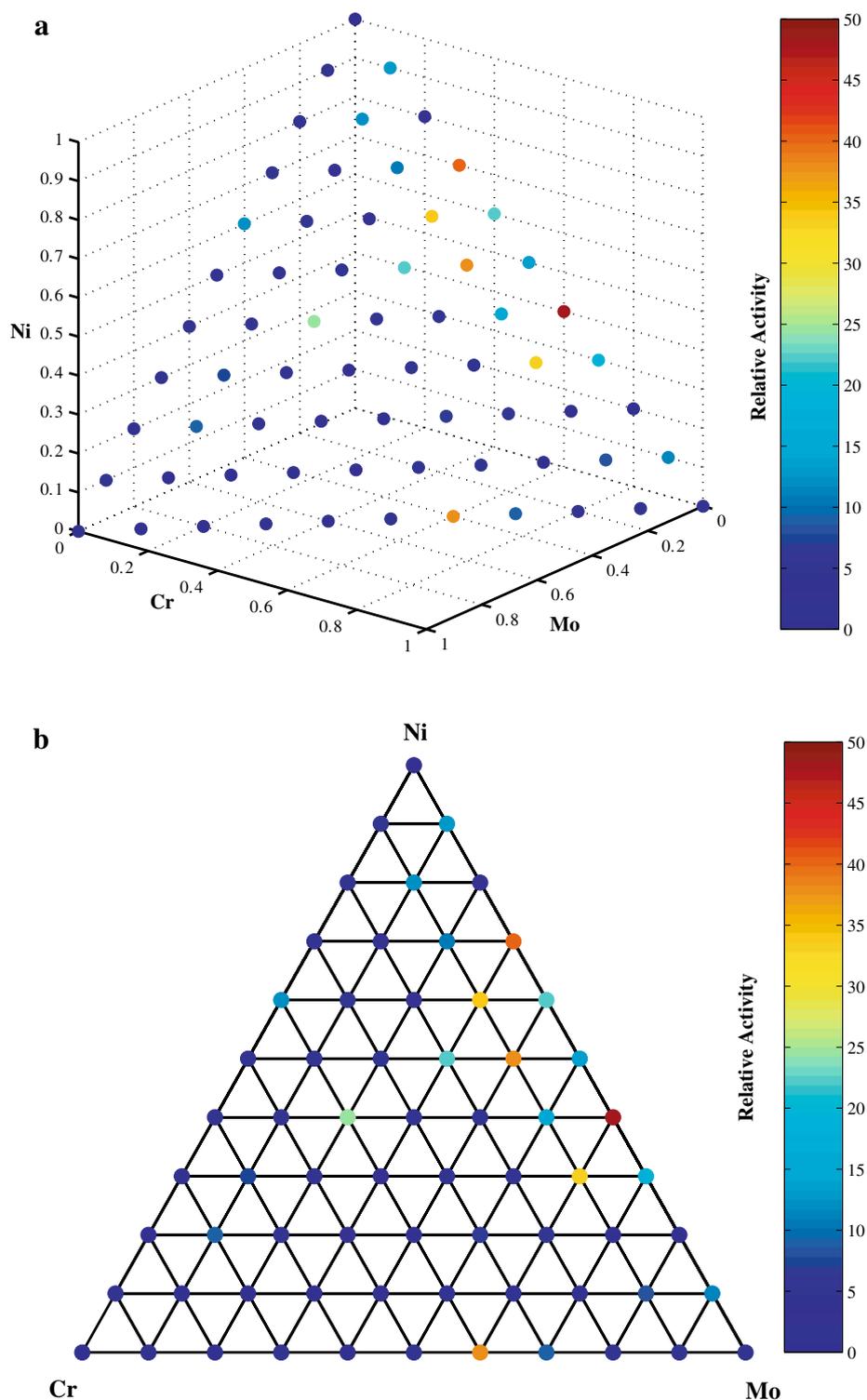
approximating curve should not exhibit jumps or sharp bends but is assumed to be of rather smooth shape.

Due to synthesis and measurement errors, the catalyst data cannot be considered as absolute values but as values plus a certain amount of noise and also outliers. Therefore, the use of approximation techniques instead of interpolation approaches to model the catalytic surfaces appears to be more suitable.

## Data

For our modeling purpose we chose a data set consisting of 66 ternary heterogenous catalysts containing Cr, Mo and Ni in 10%-wise variation without replication. The activity values for the oxidation of propene can be seen in Fig. 2. Among these catalysts the main activity occurred for mixtures containing Mo and Ni in nearly equal contents and little Cr. The data point containing only Cr and Mo as well as the point having the same Cr content but containing more Ni show higher activity indicative of local maxima. Both points were suspected to be outliers due to measurement errors since there are no more active samples within their direct neighborhoods. To prove this, 50 replicates including 2–3 repetitions were prepared and analyzed together with more local samples (5 and 10% spacing) in the direct neighborhood. The data show no significant differences between the data points and their direct neighborhood, which identifies the two original measurement points in the left middle and the middle of the bottom of Fig. 2b as outliers. In the following sections the results of three different models are discussed and their behavior in coping with possible outliers is stressed.

**Fig. 2** **a** Ternary composition spread illustrated in three dimensions. The color corresponds to the measured activity at 350°C. **b** Ternary composition spread projected onto two dimensions. The color corresponds to the measured activity

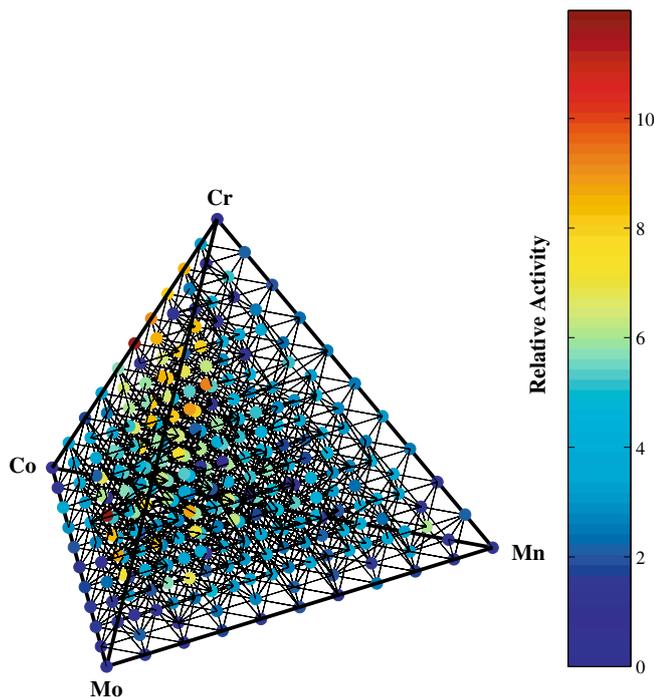


For our approach, we have chosen variations of three parameters (ternary composition spread) for demonstration, because these can be visualized easily by projection onto a triangular plane (see above). Parameter sets of higher order are clearly of interest and a tetrahedral presentation (quaternary composition spread) has also been shown above. A more attractive visualization approach is pre-

sented below based on the modeling of the parameter space of quaternary mixed oxide catalysts.

### Support vector machines (SVM)

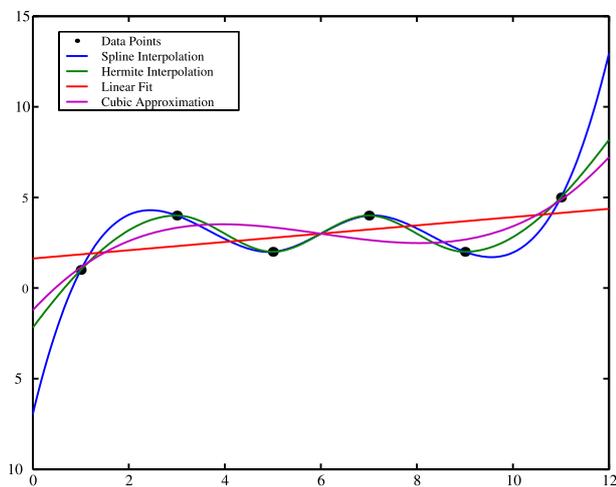
The use of support vector machines was first introduced by V. Vapnik and his coworkers in 1992 [12]. Similar



**Fig. 3** Quaternary composition spread in three dimensions. The color corresponds to the measured activity

approaches have been known in machine learning theory since 1960, but it was not until 1992 that all of the relevant features were put together to define the so-called “maximal margin classifier”, the basic Support Vector Machine. SVM can be used to solve classification problems but also regression problems and are often employed for feature extraction. For a detailed introduction to the theory of SVM and the mathematical background we refer to [13] and [14].

SVM have been applied to our catalytic data to perform regression, i.e. by means of SVM regression. The activities of non-synthesized catalyst samples have been estimated and checked whether calculated activities correlated with



**Fig. 4** Possibilities to interpolate and approximate given data points

measured activities (cf. [15]). The local fitting procedure of SVM seems to be an advantage, since there is no reason to assume that catalytic activity across the complete parameter space can be described by a simple function.

Figure 5 illustrates the application of support vector regression to the dataset introduced in Fig. 2. Among the 66 heterogeneous catalysts selected, Cr, Mo and Ni composition varies from 0–100% with an increment of 10%.

Compared to Fig. 2 it can be seen that the most active regions are well identified by the model. Variation of the resolution of the method can be used to smooth the activity surface. In Fig. 5 the isolated points of activity have been confirmed as outliers (s.a.).

In Reference [15], another ternary data set has been tested and modeled using SVM. After the calculations, several new samples were synthesized and tested by IR-thermography, gas chromatography or mass spectroscopy for certain reactions. Good agreement between predicted and experimental values was found, confirming the validity of the approach.

### Data approximation with multilevel b-splines

As mentioned above, data approximation refers to the problem of fitting a smooth surface through a scattered or uniform distribution of data. This subject is of practical importance in many areas of science and engineering such as physics, geology, chemistry, meteorology, mining etc. All these fields require data approximation or interpolation techniques to determine values at arbitrary positions. This is exactly what will be done with catalyst data and therefore some techniques from this mathematical field are also applied. Publications dedicated to the theory of scattered data interpolation are for example [16, 17], which give good surveys on this topic. For modeling the catalyst data, a multilevel B-spline approach is followed according to [18].

The basic idea of this approximation technique can be seen as follows.

For a three-dimensional application, consider a dataset  $P = \{(x_c, y_c, z_c)\}$  in 3D-space where  $(x_c, y_c)$  is a point in the  $xy$ -plane and  $z_c$  is considered as the corresponding activity value. Let  $\Omega = \{(x, y) | 0 \leq x < m, 0 \leq y < n\}$  be a rectangular domain in the  $xy$ -plane. To approximate the data points, an approximation function  $f$  is formulated as a uniform bicubic B-spline function, which is defined by a control lattice  $\Phi$  overlaid on domain  $\Omega$ . Without loss of generality, it is assumed that  $\Phi$  is an  $(m+3) \times (n+3)$  lattice which spans the integer grid in  $\Omega$ , cf. Fig. 6.

Let  $\varphi_{ij}$  be the value of the  $ij$ -th control point on lattice  $\Phi$ , located at  $(i, j)$  for  $i = -1, 0, \dots, m+1$  and  $j = -1, 0, \dots, n+1$ . The approximation function is defined in terms of these control points by

$$f(x, y) = \sum_{k=0}^3 \sum_{l=0}^3 B_k(s) B_l(t) \varphi_{(i+k)(j+l)} \quad (1)$$

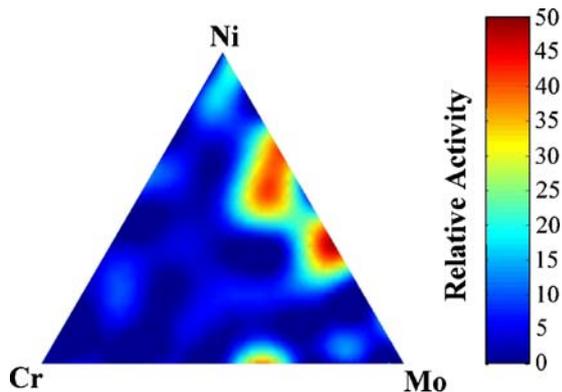


Fig. 5 Estimated activities by support vector machines

where  $i = \lfloor x \rfloor - 1$ ,  $j = \lfloor y \rfloor - 1$ ,  $s = x - \lfloor x \rfloor$  and  $t = y - \lfloor y \rfloor$ .  $B_k$  and  $B_l$  are uniform cubic B-spline basis functions defined as:

$$B_0(t) = \frac{(1-t)^3}{6}$$

$$B_1(t) = \frac{(3t^3 - 6t^2 + 4)}{6}$$

$$B_2(t) = \frac{(-3t^3 + 3t^2 + 3t - 1)}{6}$$

$$B_3(t) = \frac{t^3}{6}$$

where  $0 \leq t < 1$ . They serve to weight the contribution of each control point to  $f(x, y)$  based on its distance to  $(x, y)$ . With this formulation, the problem of deriving function  $f$  is reduced to solving for the control points in  $\Phi$  that best approximate the data points in the least square sense. According to Eq. 1 each point in  $P$  influences  $4 \times 4$  neighboring control points in  $\Phi$  and analogously,  $4 \times 4$  control points of the lattice  $\Phi$  are needed to determine the approximated value at any location in  $\Omega$ .

It is clear that the resulting surface depends on the fineness of lattice  $\Phi$ . As  $\Phi$  becomes finer, the influence of a

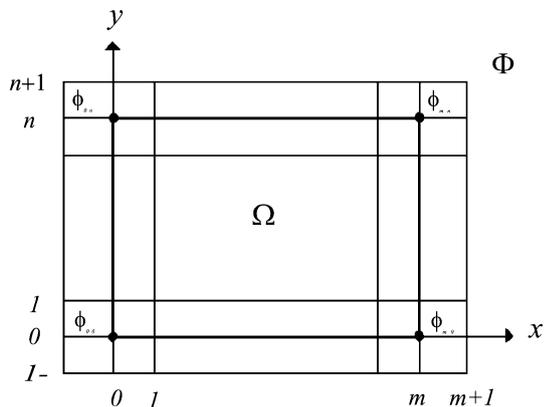


Fig. 6 The configuration of control lattice  $\Phi$  over  $\Omega$  (cf. [18])

data point is limited to smaller neighborhoods. This enables  $P$  to be more closely approximated, although  $f$  will tend to contain local peaks near the data points (danger of local biases). Figure 7 illustrates this behavior.

Using this approach, there is a trade-off between the shape smoothness and accuracy of the approximation function that has to be handled. This is solved by applying a multilevel B-spline approach with a hierarchy of control lattices to generate a sequence of functions  $f_k$  whose sum approaches the desired approximation function. In the sequence, a function from a coarse lattice provides a rough approximation, which is further refined in accuracy by functions derived from fine lattices. When  $\Phi$  is sufficiently fine compared to the data distribution,  $f$  interpolates the data.

The application of this approach to catalyst data is exemplified in Fig. 8, where the same ternary composition spread of catalysts used for SVM above is approximated. Comparing the result to Fig. 2 (discrete data points) reveals that the structure of the data set is well represented. In contrast to Fig. 5 (SVM regression) this approach leads to maxima that tend to be more local than those of SVM.

Simulated Gauss functions have been used to compare the approximation accuracy of SVM to B-splines [15]. The SVM regression yielded a smaller mean squared error (MSE) in this test, although we cannot conclude that in general SVM regression is more suitable to approximate catalyst data since the real function  $f$  is not known as it was in [15]. The B-spline approximation is presently limited by its application to a maximum of three dimensions. Challenges lie in the extension of this method to datasets that range in spaces with more than three dimensions. The formalism for applications to spaces of higher order has not yet been developed. Another possibility could be the use of a polynomial-response surface with interaction terms. For our purpose this is not reasonable due to a lack of data. Furthermore, these techniques cause responses outside the sampled space yielding  $\pm\infty$  instead of zero. Therefore, we have chosen to try Kriging as another approach.

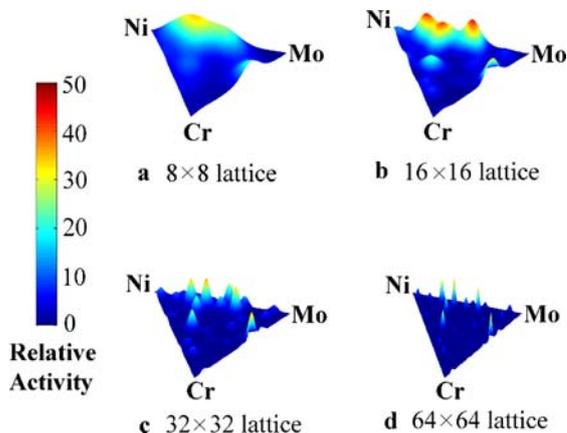


Fig. 7 B-spline approximation using different control lattice resolutions

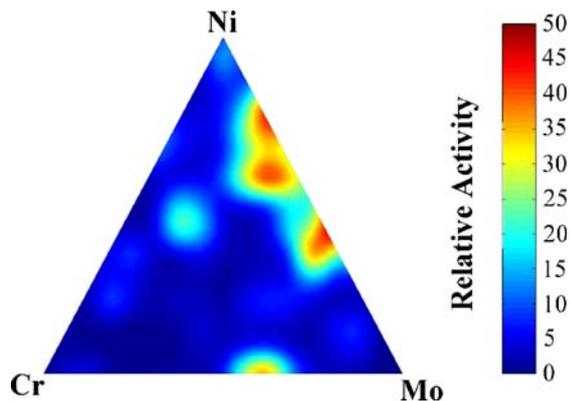


Fig. 8 Estimated activities by a multilevel B-spline approximation

## Kriging

In this section the technique of Kriging is discussed to solve the approximation problem introduced above. The term Kriging is known in the field of geostatistics or spatial statistics and describes modeling by random processes. Kriging was named after a South African mining engineer, D. Krige, who first popularized stochastic methods for spatial predictions, especially for mining. Although originally developed for mining problems, Kriging can also be applied to many other fields and applications with similar problem structure. A brief overview of Kriging is given in this section. For a comprehensive review the reader is referred to the literature [19–21].

A central problem in geostatistics is the reconstruction of a phenomenon over a domain on the basis of values observed at a limited number of points. The main difference to the other techniques described above is that Kriging starts from a statistical model rather than a model of the approximation function. Formally, Kriging appears to be similar to the prediction problem in time series where the signal is to be predicted out of given values of the past, usually at regular time intervals, for some time in the future. With time series, the signal is analyzed first by computing and modeling the spectrum, and then a predictor is designed. Kriging follows a similar approach but in a spatial setting, where there is no general concept of past and future. Within this section the use of Kriging is applied to model catalyst data to be able to predict the activity of untested samples. In contrast to Reference [9], the results of the Kriging model are discussed in comparison to the other approximation techniques described above.

The basic idea of Kriging can be described as follows:

Using the Kriging approach, catalyst data are modeled as a random process

$$Z(x) = Z_0(x) + e(x)$$

with  $x \in D \subset \mathbb{R}^n$  and  $D$  is the experiment's design region.  $Z_0$ , some smoothly varying term, is assumed to depend only on

the position  $x$  and  $e$  is a realization of a random process with zero mean and covariance  $C$ :

$$C(e) = C(x_i, x_j).$$

Let  $S = \{x_1, \dots, x_N\}$  be the set of all experimental samples, i.e.  $S$  contains all measured catalysts. Usually, the covariance is assumed to depend only on distance  $h$  of two samples, i.e.,

$$C(x_i, x_j) = C(h_{ij})$$

For modeling catalyst data, mean and covariance are unknown and have to be estimated out of the data. The prediction of unmeasured samples is done by calculating a *best linear unbiased estimator* (BLUE) for each location. This denotation can be explained as follows:

- Best: estimator of smallest variance among all other estimators.
- Linear: according to its construction, linear combination of all observations.
- Unbiased: no bias (i.e. the expected value of the random error is zero).

Although Kriging is generally considered “unbiased”, this holds only as long as a stochastic process is studied and the correct covariance is used.

Another assumption is made concerning the measured values: data points that are quite close together are assumed to be somehow “linked” to each other and this spatial influence decreases with increasing distance of points until it ebbs away. To analyze this spatial behavior, the covariance between two samples is to be considered. Instead of the covariance, the sample variogram can also be considered to model the spatial behavior of the dataset. The sample variogram is calculated by:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i + h))^2$$

where  $N(h)$  is the number of data points with distance  $h$ . Figure 9 illustrates a typical sample variogram and explains its common descriptors *range*, *sill* and *nugget effect*.

The range can be considered as the distance where covariance disappears and randomness dominates. This means that points that lie more than the range apart from each other are considered to be uncorrelated. The nugget effect describes a discontinuity at the origin. For example, in gold deposits, gold commonly occurs as nuggets of pure metal that might be much smaller than the size of a sample. This results in a strong grade variability in the sample, even when physically very close. In that case, the variogram shows a discontinuity at the origin. In practice, the nugget effect might be due to some kind of microstructure, namely a component of the phenomenon with a range shorter than the sampling distance or due to a structure with a range shorter than the smallest inter-point distance. However,

measurement and synthesis errors also cause a nugget effect.

To calculate the Kriging estimators, the discrete variogram points must be modeled by a variogram function. There exist many approaches and it depends on the problem structure to choose the right variogram function. We refer to [19–21] for the description of different variogram models. By means of the sample variogram, the Kriging system can be set up and solved for the Kriging weights  $w_i$  that assign a weight to each sample value  $x_i$ .

For the sake of simplicity, the derivation of the Kriging system is not discussed in detail here but we refer to [19–21] for extensive studies.

The Kriging estimator at location  $x^*$ ,  $Z(x^*)$  can be derived as a linear combination of all data points together with their calculated weights, i.e.,

$$Z(x^*) = \sum_{i=1}^N w_i x_i$$

where  $N$  denotes the number of given data points. Figure 10 illustrates the result of a Kriging approximation applied to the catalyst data described above. Compared to Figs. 2, 5 and 6 it can be concluded, that all three methods provide comparable models of the activity–composition relationship.

Determining the nugget effect of a Kriging model must be done thoroughly because this parameter controls the extent of approximation. Models with zero nugget effect yield an interpolation of data and the larger the nugget effect the smoother the fitted surface becomes. This enables researchers to adapt the smoothness of the desired surface in a very comfortable way. Another advantage of Kriging lies in its excellent ability also to approximate data in more than three dimensions. The computational effort of Kriging directly corresponds to the number of dimensions. For example, datasets containing four elements lead to Kriging systems with matrix sizes of  $287 \times 287$  entries, which can be easily solved on a PC. This means Kriging can be used

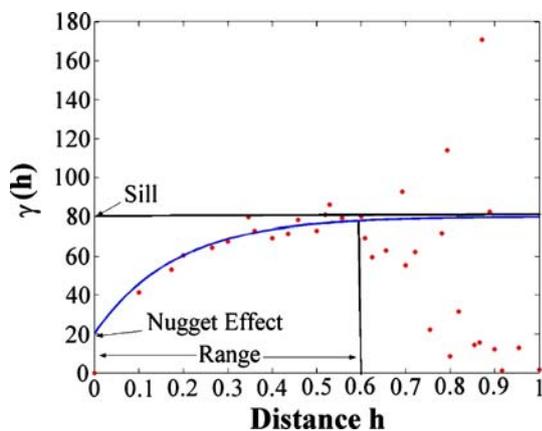


Fig. 9 Sample variogram corresponding to the data set illustrated in Fig. 2 with fitted variogram function

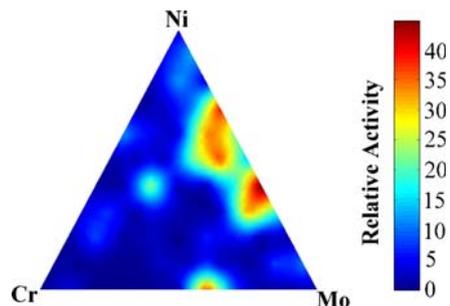


Fig. 10 Approximated surface using the Kriging approach

excellently in more than four dimensions provided that there are datasets available. For more than five elements the synthesizing and screening time of such a catalytic library is still too long to establish valuable datasets.

### Visualization of a quaternary composition spread

According to Fig. 3, a quaternary composition spread can be illustrated as a tetrahedron. The Kriging approach described above is applied to a dataset in four dimensions containing the elements molybdenum, manganese, cobalt and chromium. Fig. 11 illustrates how the structure of the approximated values within the tetrahedron can be visualized by using slice plots. The catalysts were tested for the oxidation of propene at 300°C. We are mainly interested in catalysts that show high selectivity to acrolein among other products such as acetone, propionaldehyde or 1,5-hexadiene.

From Fig. 11 it can be seen, for example, how the nickel content influences the selectivity of the catalysts. The first plot of Fig. 11 illustrates the selectivity of catalysts without any nickel. In the second plot of Fig. 11 all catalysts containing 10% nickel are shown and it can be seen that there is the region of highest selectivity (colored in red). In the following slices of higher Ni-content (Fig. 11) the selectivity to acrolein decreases. Clearly all four elements contribute to the area of highest selectivity. Very important

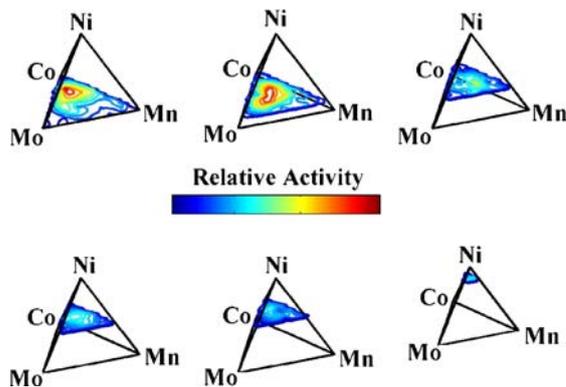


Fig. 11 Visualization of a ternary catalyst system and its approximation using slice plots

here is that the areas of higher activities are supported by the activity of the neighboring points. Points of isolated high activity are most likely outliers, since in all our composition spread studies of the last few years, the surfaces have all been very smooth and continuous. This way of visualization helps to look for trends and redundant elements among quaternary catalysts and enables the researcher to plan the next experiments for a more efficient search of active catalysts. In practice, this visualization can be done more effectively directly on the monitor using the fourth (here Ni) and higher parameters as sliders.

## Conclusions and outlook

Through a systematic preparation of mixed oxides covering the whole composition space by a sol-gel preparation procedure tolerant of composition, the catalytic activity has been determined. It has been demonstrated that catalytic activity for propene oxidation to acrolein with air is a function of chemical composition if the materials studied are prepared by the same method. Three different mathematical methods, support vector machines, B-splines and Kriging have been applied to model the datasets. All methods provide comparable results and can therefore be applied for QCAR. While B-splines can, up to now, only be applied to datasets with functional relationships dependent on up to three parameters, SVM require a larger computational effort. Kriging is favored because it can be applied readily to multiple parameter spaces while its computation is still fast enough. Parts of our future work will deal with the evaluation of the predictive power of each method. Furthermore, replications are now included in all experiments.

This QCAR-modeling provides a new entry into prediction of materials properties. It can be used for materials optimization as well as for the design of experiments in extended high-throughput searches for new materials.

## References

- Potyrailo RA, Takeuchi I (2005) *Meas Sci & Technol* 16:1–4
- Potyrailo RA, Amis EJ (2003) *High-throughput analysis: a tool for combinatorial materials science*, 1st edn. Kluwer, New York
- Takeuchi I, Xiang XD (2003) *Combinatorial materials synthesis*, 1st edn. Marcel Dekker Inc, New York Basel
- Maier WF (2003) *Appl Catal A: General* 254:1–2
- Koinuma H, Xiang XD, Kawasaki M, Chikyow T (2002) *Combinatorial materials science and technology*. Elsevier Science B.V.
- Cawse JN (2003) *Experimental design for combinatorial and high throughput materials development*. Wiley, Hoboken
- Klanner C, Farrusseng D, Baumes L, Lengliz M, Mirodatos C, Schüth F (2004) *Angew Chem* 116:5461–5463
- Farrusseng D, Klanner C, Baumes L, Lengliz M, Mirodatos C, Schüth F (2005) *QSAR & Comb Sci* 24:78–93
- Scheidtmann J, Klär D, Saalfrank JW, Schmidt T, Maier WF (2005) *QSAR & Comb Sci* 2:203–210
- Scheidtmann J, Saalfrank JW, Maier WF (2003) *Plattenbau - automated synthesis of catalysts and materials libraries*. In: Anpo M, Onaka M, Yamashita H (eds) *Studies in surface science and catalysis*, 145. Elsevier, Tokyo, pp 13–21
- Weiss PA, Saalfrank JW, Scheidtmann J, Schmidt HW, Maier WF (2003) In: Potyrailo RA, Amis EJ (eds) *High-throughput analysis: a tool for combinatorial materials science*, 1st edn. Kluwer, New York, pp 125–153
- Vapnik VN (1995) *The nature of statistical learning theory*, 2nd edn. Springer, Berlin Heidelberg New York
- Shawe-Taylor J, Cristianini N (2004) *Kernel methods for pattern analysis*, 1st edn. Cambridge University Press
- Hastie T, Tibshirani R, Friedman J (2001) *The elements of statistical learning*, 1st edn. Springer, Berlin Heidelberg New York
- Stutz B (2004) *Suchstrategien für Optimale Katalysatoren in der Kombinatorischen Chemie*. Diplomarbeit, Universität des Saarlandes
- Franke R, Nielson GM (1991) *Scattered data interpolation and applications: a tutorial and survey* In: Hagen H, Roller D (eds) *Geometric modelling: methods and their application*, Springer, Berlin Heidelberg New York
- Franke R (1982) *Math Comp* 38:181–200
- Lee S, Wolberg G, Shin SY (1997) *IEEE Transactions on Visualization and Computer Graphics* 3:228–244
- Cressie N (1993) *Statistics for spatial data*, revised edn. Wiley, New York
- Rivoirard J (1994) *Introduction to disjunctive kriging and nonlinear geostatistics*. Oxford University Press, Oxford
- Chilès JP, Delfiner P (1990) *Geostatistics—modelling spatial uncertainty*. Wiley, New York, p 1999