

Deriving adequate sample sizes for ANN-based modelling of real estate valuation tasks by complexity analysis

Sabine Horvath^{a,*}, Matthias Soot^b, Sebastian Zaddach^c, Hans Neuner^a, Alexandra Weitkamp^b

^a TU Wien, Department of Geodesy and Geoinformation, E120-05 Research Division Engineering Geodesy, Wiedner Hauptstraße 8, 1040 Vienna, Austria

^b TU Dresden, Professorship of Land Management, Hülse-Bau, Südflügel, Helmholtzstr. 10, 01069 Dresden, Germany

^c Ministry for Internal Affairs and Sport of Lower Saxony, Department 44, Lavesallee 6, 30169 Hannover, Germany

ARTICLE INFO

Keywords:

Real estate valuation
Artificial neural network
Complexity analysis
Adequate sample size

ABSTRACT

Property valuation in areas with few transactions on basis of a linear regression fails due to a not sufficient number of purchasing cases. One approach which is enhancing the available data set is to evaluate these purchasing cases together with a neighbouring submarket. However, it leads to non-linearities. Consequently, non-linear models for a cross-submarket real estate valuation are required to obtain reasonable results. We focus in this contribution on non-linear modelling on basis of artificial neural networks (ANN). A prerequisite for these procedures is an adequate sample size. We present a new approach based on the aggregation of submarkets additional to the markets with few transactions at the expense of increasing complexity of the model required. The cross-submarket ANN estimation aims to reach accuracies comparable to local property valuation procedures in a first step and in further consequence to enable a reasonable estimation in areas with few transactions. We introduce an extended Kalman filter (EKF) estimation procedure for the ANN parameters and compare it to the standard optimisation procedure Levenberg Marquardt (LM) as well as to the multiple linear regression. Thus, German spatial and functional submarkets are aggregated. For the spatially aggregated data set, the ANN estimation leads to improved results. The ANN estimation of the functionally aggregated data appears deceptively simple due to too small samples not representing the sampling density. The question arises, what are adequate sample sizes regarding the complexity of the unknown relationship. We propose a model complexity analysis procedure based on resampling and the structural risk minimisation theory and derive a minimum sample size for the spatially aggregated data. Only for the EKF computations, this minimum sample size is reached due to less variance of the ANN estimations. Generally, the EKF computation leads to a better ANN performance in contrast to LM. Finally, the spatial cross-submarket ANN estimation reaches accuracies of local property valuation procedures.

1. Introduction

Real estate economy and real estate valuation are key factors of every big economy. The total value of all real estates e.g., in Germany is estimated four times the GDP (real estate value: 13.3 trillion Euro, GDP: 3.3 trillion Euro, source: ZIA Germany¹) and is rising significantly faster than the GDP in the last years (statista²). Because of the importance of the real estate values for the economy an accurate real estate valuation is necessary for a transparent market to avoid bubbles.

Property valuation is frequently acting in areas with few

transactions. The available purchase cases are not sufficient for a mathematic-statistical evaluation. In practice, hedonic procedures with corresponding linear models have become established. An aggregation of functional or spatial submarkets is not expedient for these procedures, as the functional relationships can no longer be adequately modelled linearly. This requires non-linear procedures. The paper, therefore, examines the applicability of artificial neural networks and, in connection with this, the necessary size of the required sample. This will be compared with conventional methods.

In this work, we focus on comparison factors for the comparison

* Corresponding author.

E-mail address: sabine.horvath@geo.tuwien.ac.at (S. Horvath).

¹ <https://www.zia-deutschland.de/marktdaten/bedeutung-der-immobilienbranche/>

² Statista Entwicklung der Hauspreise in Deutschland in den Jahren von 2000 bis 2019 (Index of Real estate prices) vs. Statista Index des Bruttoinlandsprodukts in Deutschland von 1991 bis 2020 (Index of GDP).

valuation approach e.g., legally regulated in Germany in § 15 (2) of the valuation code (ImmoWertV). The comparison factors are normalised with the size of the real estate (total price divided by area of living space). These comparison factors are normally publicised in market reports for different spatial and functional submarkets and used for the valuation of real estates by appraisers. In markets, with few transactions, no comparison factors are available because derivation isn't possible in these markets. Methods like artificial neural networks can derive data for these regions and need to be improved in this sense.

Two objectives are targeted with the proposed learning method in this study: First, an automated cross-market valuation method shall be generated. Second, the appraisal of areas with few transactions shall be enabled (Soot et al., 2018). To analyse aggregated data sets with non-linear relationships in these areas where single market appraisals fail, we consider artificial neural networks (ANN) as a learning method. We focus on ANN because of its universal approximation ability. The standard tool in real estate valuation praxis is the multiple linear regression (MLR) and it will serve as a basis for comparison.

2. State of art and research gap

Numerous studies have demonstrated a non-linear structure of the hedonic price function in real estate context (Lisi, 2013; Fan and Xu, 2009; Mimis et al., 2013; Din et al., 2001; Worzala et al., 1995; Halvorsen and Pollakowski, 1981). However, the MLR, which is established, is not the optimal method for estimating the hedonic price function.

Tay and Ho (1992) are among the first to use ANN for evaluation tasks. In their study, they compare the performance of MLR with ANN in estimating the purchase prices of apartments in Singapore and conclude that ANNs are a useful alternative to the MLR model. Various studies conducted over the next two and a half decades show that ANNs are superior to MLR in real estate valuation. Mimis et al. (2013) use internal physical (structural quality and quantity) and external environmental characteristics (neighbourhood, transport accessibility) of the properties to predict house prices in Athens. They compare the results of the regression analysis with those of an ANN and conclude that the ANN approach is superior to the MLR in most cases because it can handle non-linear relationships. Morano et al. (2015) apply ANN to the valuation of property market values using the example of residential property in a narrow spatial submarket in Bari (Italy). The interpretation of their results is questionable because the sample size of 90 transactions used to train the ANN is far too small. A recent comparative study of various machine learning algorithms applied on housing rental prices of social media data (online housing rental websites) in Shenzhen is treated by Hu et al (2019). They integrate machine learning and the hedonic model. Six different methods including ANN are compared. They evaluate two data sets with approximately 20.000 samples and use 56 determinants. No further information about the specific models is given. The best performing methods are the tree-based bagging algorithms followed by ANNs.

Nguyen and Cripps (2001) refer for the first time to correspondingly required amounts of data to achieve a good prediction with ANN and to the importance of the quality criteria to be chosen for the assessment of the approaches. Zurada et al. (2011) compare several learning methods (incl. ANN) with regression methods (incl. MLR) and conclude that for homogeneous data sets the non-traditional regression methods (including support vector regression, additive regression and M5P regression trees) perform better and, in the case of heterogeneous data sets and overlapping clusters, learning methods provide a better prediction. They also show that up to then, too small data sets have been used for modelling with learning methods. In other scientific fields, the application of machine learning algorithms to small sample sizes is a topic. Zhang and Ling (2018) developed a strategy to apply machine learning to small data sets in materials science. They use crude estimations of properties as inputs in kernel ridge regressions to reduce the noise level of the original data. Thereby, they achieve a good estimation

Table 1

Formulated research question and hypotheses of the study.

Research Question	Are aggregation strategies appropriate to generate a sufficient database for ANN-based modelling of non-linear real estate valuation tasks?
Hypothesis 1	The complexity of an input-output relationship is approachable. A minimum sample size corresponding to a specific problem complexity is derivable for ANN modelling.
Hypothesis 2	The cross-submarket estimation using ANN leads to accuracies comparable to local property valuation procedure.
Hypothesis 3	When applying ANN the optimisation procedure EKF performs better than the standard method LM.

by machine learning without an increased dimensionality and rise in costs for further measurements.

Artificial neural networks are treated several times in the field of real estate valuation over the last three decades. However, often only a few data samples are used and no information about the model selection process is given. In a first investigation (Horvath et al., 2019), we dealt with the influence of the data set size on the estimation quality. We showed that in case of non-linear relations and an adequate sample size ANN performs better than the MLR (in terms of a smaller RMSE). That was true for the property interest rate in Lower Saxony. Approximately 2900 purchasing cases and an ANN model consisting of four nodes in one hidden layer led to a promising result outperforming the MLR. For another data set, consisting of only approximately 300 purchasing cases, already the model selection showed that linear models are advantageous. Therefore, we tested an approach to enlarge this data set by a simulation preserving the data structure and reached for an appropriate number of samples beneficial results in terms of a smaller RMSE. However, there is still a need for large data sets. A more realistic approach to the enlargement of data sets is to aggregate submarkets, which will be treated in this study. ANNs have only been applied to selected functional and spatial submarkets. National studies (especially related to German submarkets) on the applicability of ANNs in valuation have been missing so far. In this context, it is particularly necessary to investigate the performance of ANN compared to established methods (MLR) in local assessments. Still open is what number of data samples is sufficient for a beneficial real estate valuation by ANNs and if a criterion exists, which allows a statement about the appropriateness of sample sizes. Alwosheel et al. (2018) investigate these last points in the field of discrete choice analysis. They derive a specific minimum sample size based on the learning curve obtained for various sample sizes. We additionally use the statistical learning theory (see Section 5) and consider it for different estimation methods.

As a result, the main research question of this study deals with the generation of an adequate database for ANN-based modelling of real estate appraisals in Germany and is formulated in Table 1. The amount of needed data samples depends on the difficulty/complexity of the unknown input-output relationship. The reconstruction of a complex target function requires a complex approximation function. Consequently, the complexity of a relationship is a decisive criterion, if a database is sufficient or not. The more complex a target function is the more data is needed to model it. A classical measure of complexity is the sampling theorem, which states that a function can be recovered from samples if the sampling frequency is at least twice the largest frequency of a signal (Cherkassky and Mulier, 2007). However, it bases on knowledge about the process, noise-free data and a fixed sampling rate. Consequently, there is a need for an adequate complexity measure. This results in hypothesis one (Table 1) – complexities of input-output relationships are approachable and minimum sample sizes for ANN modelling are derivable.

The aim of the ANN modelling of the aggregated data is to enable the appraisal of areas with few transactions in future. Therefore, ANNs shall achieve similar results as local standard valuation methods, as formulated in the second hypothesis (Table 1).

Furthermore, we use a combination of an ANN with an extended Kalman filter (EKF) to estimate the ANN parameters due to faster convergence behaviour and superior estimation quality in comparison to standard optimisation. The standard optimisation method Levenberg Marquardt (LM) and the multiple linear regression (MLR) serve as a basis for comparison. This leads to the hypothesis three (Table 1), stating that these methods are beneficial.

The investigations in this paper base on two different data-samples. One sample contains an aggregation of functional markets within a regional market. The second sample aggregates one functional market from several regional markets. In the following sections the methodical fundamentals (Section 3), containing the used approaches, as well as the design of this study are presented. Afterwards, the data and their basic processing are shown (Section 4). The complexities of the data sets are discussed (Section 5) and the different used methods are compared to local property valuation procedures in Section 6. Finally, the results are summarised (Section 7) and an outlook is given (Section 8).

3. Methodical fundamentals

Only with reliable information about the market activity (the price levels and temporal development), the market participants have a chance to assess realistic and sustainable prices for real estates they want to buy or sell. For transparent markets, it is, therefore, necessary that transaction information is investigated with adequate methods to provide information about the market (report about the past, the actual situation and expected further development according to today's knowledge).

In this chapter, we present a short introduction to the current approach of data analysis in practice at examples of the German real estate market. This is followed by the applied learning methods. We present the classic artificial neural networks as well as the ANN estimation in the extended Kalman filter (EKF). As suggested in Section 2, the reachable quality of the ANN estimation depends on the data set size. As a criterion for determining the necessary sample size, the model complexity is introduced with two model selection approaches. The used quality criteria to describe the quality of the results complement the subsection; they enable a comparison of the reached performance by the various estimation methods.

3.1. Current approach in practical real estate valuation

Since the use of computers at the end of the 1970th, hedonic models came to use in real estate valuation in Germany (Ziegenbein, 1978; Pelzer, 1978). The MLR is used as the main approach. The idea behind the MLR is to find model-parameters w_k that describe the target value y with k influencing parameters x_k in a linear model. The error ϵ is reduced with a least-squares approach:

$$y = xw + \epsilon \quad (1)$$

Since then, only a few improvements were implemented, using more advanced criteria for model selection and outlier detection (Soot et al., 2016). In general, the needed accuracy within 10–20% difference can be achieved with this method in transparent markets with a sufficient number of purchase prices.

Because of its linear capacities, it should only be used in investigations where, due to a small span of influencing parameters, a linear approximation is appropriated. However, in regions with few transactions, only a few cases that normally spread wide within spans are available. The linear approximation is questionable in these markets.

Besides the classical linear regression analysis, other approaches have been tested in research backgrounds like collocation (Zaddach and Alkhatib, 2013) or Bayesian approaches (Alkhatib and Weitkamp, 2012; Dorndorf et al., 2017). However, such approaches are not used in practice up to now because of their complexity and necessary effort in data collection. All methods are mainly used in the local analysis (city or

district level) and can only be used if the number of the purchase price is sufficient for a hedonic analysis. Ziegenbein (2010) shows that several approx. 15 purchase prices per influencing parameter are needed to derive accurate data for real estate valuation in the housing sector in a regression analysis. This number applies in a transparent market with a variation coefficient between 10% and 15%. Often this amount of purchase prices is not realised in a specific submarket (Soot et al., 2017). In such regions with few transactions, which are most of the time less transparent due to the lack of data, other approaches have to be used. Due to the wider span of selection criteria (in spatial widening or functional widening) which is necessary for these markets to achieve a sufficient number of purchase prices, the assumption of linearity is not achieved anymore. The non-linearity of influences in MLR can be modelled with different transformation functions of data or polynomial modelling today. Therefore, a knowledge of market behaviour is necessary to choose the correct function for a transformation or a polynomial modelling. The correct function may differ from submarket to submarket according to demand and offer situation in the specific market. For this reason, the matching function needs to be selected in every market, which is difficult without prior knowledge about market behaviour.

3.2. Learning approach and methods

Learning belongs to the class of data-driven methods – meaning to choose a model only on basis of data and no assumptions about the nature of the target function are made. Due to the difficulty of choosing a model only on basis of data, in practical real estate valuations (Section 3.1) mainly linear models are used. As indicated before, it is a quite challenging task to explain and justify a specific target function for nonlinear behaviour. However, a similar model selection process is accomplished in the MLR framework; the essential influencing parameters need to be determined. This will be established either on basis of statistical tests or information criteria. In the following, a more general framework of the learning task and particularly of the model selection is treated. It builds the basis for the next subsections and is taken up again in Section 3.3.

A learning model consists of three components (Vapnik, 2000):

- A generator corresponding to the inputs vector x drawn from an unknown distribution $P(x)$.
- A supervisor returning an output value y to the input vector x according to the unknown conditional distribution $P(y|x)$.
- A learning machine $f(x, w)$ with model parameters w .

Learning is to choose from a given set of functions $f(x, w)$ the one approximating the best the supervisor's response y . The training samples are independent and identically distributed observations drawn from the unknown joint distribution function $P(x, y) = P(y|x)P(x)$.

To select the best approximation to the supervisor's response y , the discrepancy between y and the estimated response by the learning machine \hat{y} , the so-called loss $L(y, f(x, w))$ is computed (Vapnik, 2000). In the regression case, the criteria mean square error or the root mean square error (RMSE) (Section 3.3) are used. The expected value of the loss is the risk functional:

$$R(w) = \int (y - f(w, x))^2 P(x, y) \, dx dy \quad (2)$$

The aim of learning is to minimise the risk of functional $R(w)$. However, the joint distribution function $P(x, y)$ is unknown. The only available information is the training samples that give a finite and incomplete approximation of $P(x, y)$. Applying the inductive principle (a conclusion is made on basis of empirical samples) replaces the risk functional $R(w)$ by the empirical risk $R_{emp}(w)$. The empirical risk minimisation principle is applied (Vapnik, 2000), which is comparable to the minimisation of the residual sum of squares in the MLR case.

$$R_{emp}(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^N (y(n) - f(\mathbf{w}, \mathbf{x}(n)))^2 \quad (3)$$

Applying the estimated empirical model on an independent data set (not used for the model estimation) belonging to the same population is one possibility to give information about the generalisation ability and prediction quality of the estimated empirical model. Further possibilities will be treated in the model complexity part (Section 3.3).

A learning approach consists of an approximating function and an optimisation method (Cherkassky and Mulier, 2007). The performance of ANN as an approximating function strongly depends on the chosen optimisation method. Therefore, the extended Kalman filter (EKF) is used and its advantage over standard optimisation is treated.³ A short introduction to the methods of ANN and EKF is given in the following.

3.2.1. ANN

Artificial neural networks are universal and flexible deployable approximating functions. They can be perceived as a more general form or a non-linear extension of regression models. They are extended by an activation function φ and additionally ordered in layers (Eq. (4)).

A brief description of the functionality of ANNs is given. For an exhaustive introduction, see Haykin (1999).

$$\begin{aligned} e_{m,i}(n) &= y_m(n) - \hat{y}_{m,i}(n) = y_m(n) - f_m(\mathbf{w}_i, \mathbf{x}(n)) \\ &= y_m(n) - \varphi^{(m)} \left(\sum_l \mathbf{w}_{ml,i} \varphi^{(l)} \left(\sum_k \mathbf{w}_{lk,i} \mathbf{x}_k(n) \right) \right) \end{aligned} \quad (4)$$

The function of a simple ANN, consisting of one hidden layer l (NHid) is given in Eq. (4). On basis of K input measures $\mathbf{x}_k(n)$ corresponding to k nodes ($k = 0, \dots, K$) in the input layer and initial model parameters $\mathbf{w}_0 = [\mathbf{w}_{lk,0}; \mathbf{w}_{ml,0}]$ the M output measures $y_m(n)$ of the output layer are computed. The activation functions φ are the basic functions of the ANN (Haykin, 1999); n is the labelled data sample index of the N training data. The computed output measure $\hat{y}_m(n)$ is compared to the observed one $y_m(n)$. The error $e_m(n)$ in the iteration step i is backpropagated through the network to update the parameters \mathbf{w} so that the overall error e_i is minimised (see Eq. (5)).

$$e_i(n) = \frac{1}{2} \sum_m e_{m,i}^2(n) \quad (5)$$

The backpropagation corresponds to linear optimisation methods and therefore offers slow convergence behaviour. Levenberg-Marquardt (LM), a standard optimisation method for ANN, combines the advantages of linear and non-linear optimisation – permanent progress towards a minimum and fast convergence close to a minimum. In Eq. (6) the parameter updating of LM is presented. The Jacobi matrix \mathbf{J} contains the first derivatives of the ANN, the heuristic parameter λ enables the switching between linear and non-linear optimisation behaviour and is steered by the overall error e_i (Eq. (5)) (Hagan and Menhaj, 1994). In this contribution, next to LM the extended Kalman filter (EKF) is used. It is treated in more detail in the next section.

$$\mathbf{w}_{i+1} = \mathbf{w}_i - (\mathbf{J}^T \mathbf{J} + \lambda \mathbf{I})^{-1} \mathbf{J}^T \mathbf{e} \quad (6)$$

3.2.2. EKF

The Kalman filter (Gelb, 1974) is a linear recursive data processing algorithm, which combines a physical problem description with observations. It is mainly used to predict motion or a system behaviour in navigation and control. Due to the extension of the ANN function (Eq. (4)) by a static system (Eq. (7)) another updating formula (Eq. (10)) as in Eq. (6) results. The updating of the EKF leads to a similar form as the LM

³ The Kalman filter is an approximation function for dynamic system modelling. In this context, the EKF is treated as an optimisation method. Further information is contained in Section 3.2.2.

method (Eq. (6)) and under specific assumptions, both methods are even equal, as has been shown in Horvath and Neuner (2019).

The extended Kalman filter (EKF) applies if a non-linear functional is given and is solved by linearising the observation function and estimating the parameters by iterative least-squares. Singhal and Wu (1989) first introduce the combination of ANN and EKF. Detailed treatment is presented in Haykin (2001). The EKF bases on a static system (Eq. (7)) and the ANN is included in the observation equation (Eq. (8)). This problem formulation includes a stochastic description of the system noise \mathbf{p} and the observation noise \mathbf{o} . Both are assumed Gaussian with zero mean and the appropriate cofactor matrix \mathbf{Q} .

$$\bar{\mathbf{w}}_{i+1} = \mathbf{I} \hat{\mathbf{w}}_i + \mathbf{p}, \quad \mathbf{p} \sim N(0, \sigma_0^2 \mathbf{Q}_{pp}) \quad (7)$$

$$y(n) - \mathbf{o}(n) = f(\hat{\mathbf{w}}_{i+1}, \mathbf{X}(n)), \quad \mathbf{o} \sim N(0, \sigma_0^2 \mathbf{Q}_{yy}) \quad (8)$$

On basis of Eq. (7), the ANN parameters $\bar{\mathbf{w}}_{i+1}$ as well as the corresponding parameter cofactor matrix $\mathbf{Q}_{\bar{\mathbf{w}}_{i+1}}$ (Eq. (9)) are predicted and used to update the ANN parameters $\hat{\mathbf{w}}_{i+1}$ (Eq. (10)).

$$\mathbf{Q}_{\bar{\mathbf{w}}_{i+1}} = \mathbf{I} \mathbf{Q}_{\hat{\mathbf{w}}_i} \mathbf{I}^T + \mathbf{Q}_{pp} \quad (9)$$

$$\hat{\mathbf{w}}_{i+1} = \bar{\mathbf{w}}_{i+1} + \mathbf{Q}_{\bar{\mathbf{w}}_{i+1}} \mathbf{A}_{i+1}^T (\mathbf{Q}_{yy,i+1} + \mathbf{A}_{i+1} \mathbf{Q}_{\bar{\mathbf{w}}_{i+1}} \mathbf{A}_{i+1}^T)^{-1} (y(n) - f(\bar{\mathbf{w}}_{i+1}, \mathbf{X}(n))) \quad (10)$$

While the updating of LM is driven by the overall error (Eq. (5)), the implemented EKF estimation bases on the compatibility test under the null hypothesis $e_{i+1} = 0$. The appropriate test value (Eq. (11)) consists of the normalised and decorrelated innovation e_{i+1} and the a priori variance factor σ_0^2 . Eq. (12) shows explicitly the computation of the cofactor matrix of the innovations \mathbf{Q}_{ee} . The test value follows the χ^2 -distribution with $n_{y,i}$ degrees of freedom and is computed in each iteration step i (Horvath and Neuner, 2019).

$$T_{i+1} = \frac{e_{i+1}^T \mathbf{Q}_{ee,i+1}^{-1} e_{i+1}}{\sigma_0^2} \quad (11)$$

$$\mathbf{Q}_{ee,i+1} = \mathbf{q}_{yy,i} \bar{\mathbf{Q}}_{yy,i+1} + \mathbf{A}_{i+1} \mathbf{Q}_{\bar{\mathbf{w}}_{i+1}} \mathbf{A}_{i+1}^T \quad (12)$$

In the case of LM, the heuristic parameter λ is adapted to reach a decreasing overall error. In the case of EKF, the stochastic model will be adapted, if the test value does not lie in the corresponding quantiles of the two-sided test. Thereby, the correct chosen functional model is assumed. If the test value exceeds the upper limit of the confidence interval, the observations will be down-weighted in analogy to dealing with outliers. If the lower limit is underrun, the observations will be up-weighted. This is accomplished by the variance factor q_{yy} ($\Sigma_{yy} = \sigma_0^2 q_{yy} \bar{\mathbf{Q}}_{yy}$). Adapting the observation noise is quite similar to the adaptation of λ .

An additional variation of the process noise influences the convergence behaviour of the filter positively. Therefore, the process variance factor q_{pp} ($\Sigma_{pp} = \sigma_0^2 q_{pp} \bar{\mathbf{Q}}_{pp}$), which is contained in the prediction of the parameter cofactor matrix $\mathbf{Q}_{\bar{\mathbf{w}}}$ (Eq. (9)), will additionally be adapted in further computation.

3.2.3. Model complexity

In Section 3.2.1, a simple ANN approximation function is described as a model to explain the estimation procedure. However, one decisive step is the selection of an appropriate ANN structure. This model shall represent the unknown input-output relationship $P(y|x)$. The only accessible information is the data itself. The representativeness of the data depends on the complexity of the unknown relationship. In the case of a linear relationship, the minimal number of data samples is two for a one-dimensional problem. The more difficult a relationship is, the more data samples are needed.

The model complexity of an ANN depends on the number of hidden nodes (NHid) l and the type of activation functions. Some of these model

hyperparameters can be met due to the universal approximation theorem (Hornik et al., 1989), which says: An ANN with one hidden layer can approximate arbitrary accurate each continuous function if these nodes are activated by a sigmoidal and the nodes of the output layer by a linear activation function. The determination of the number of hidden nodes in the hidden layer is accomplished in the model selection process. Therefore, within this task, various models with a specified number of hidden nodes are computed and the model leading to the best generalisation, which corresponds to the minimal prediction risk, is selected. One possibility to restrict the computationally intensive model selection is to limit the maximal number of hidden nodes by Widrow’s rule of thumb (Widrow and Stearns, 1985 in Haykin, 1999, p. 230). It relates the number of training samples N with the number of weights W (is derived by the number of hidden nodes) and the permitted classification test error ϵ . Eq. (13) states that N is bounded above by the quotient of W and ϵ asymptotically.

$$N = o\left(\frac{W}{\epsilon}\right) \tag{13}$$

On the one hand, the representativeness of the model is ensured by validating the trained ANN model on basis of data, which have not been used for the training – leading to the resampling approach of cross-validation (CV). On the other hand, the statistical learning theory provides bounds on the prediction risk especially developed for finite samples. The structural risk minimisation principle originates out of it and enables the selection of a model leading to a lower bound of the expected risk.

3.2.4. Resampling

To test the estimated ANN model on its generalisation ability, an independent data set (a test set) is needed (Friedman, 1994). Before, the selection of an optimal ANN model is necessary. Thereby, the same procedure applies – the model shall enable good generalisation ability. Consequently, an independent validation set is demanded. In case of a small number of data samples, cross-validation procedures enable to validate independent data without separating into three parts. Thus, the data is only split into training and test samples.

Cross-validation (CV) means that the training set is divided into n -parts. $(n - 1)$ -parts are used for the training and the remaining part for validation. One extreme version is the leave-one-out CV, where the validation set consists of one sample. The most common version is the n -fold CV using $n = 5$ or 10 . It presents neither an excessively high bias nor a very high variance for the estimated validation error (James et al., 2013). Thereby, the training data is divided randomly in n approximate equal folds. Each n -fold is used for validation once and the result is the mean RMSE (see Section 3.4) overall n separations. The mean training and validation RMSE of each considered model are opposed. Usually, the training error decreases with an increasing number of hidden nodes. In contrast, after a first decrease in the validation error, it starts to raise in case of modelling noise. Consequently, the model leading to the minimal validation error offers the best generalisation. Non-linear cross-validation (Moody, 1994) overcomes variations in error curves and their minima due to data sampling. It bases on the consideration of the same initial parameters w_0 for each of the n -validations. To derive a variation in the initial conditions, a repeated CV on basis of different parameter initialisations w_0 is additionally considered.

3.2.5. Structural risk minimisation

We follow the arguments from Section 3.2, that the empirical risk minimisation (ERM) principle is applied. In this section, the theoretical justification of the ERM induction principle is outlined and a bound for the expected risk according to Vapnik (2000) is derived. The theoretical basis of the ERM induction principle is treated in the statistical learning theory, which provides concepts, proofs and settings of the predictive learning with finite samples. The structural risk minimisation (SRM) is

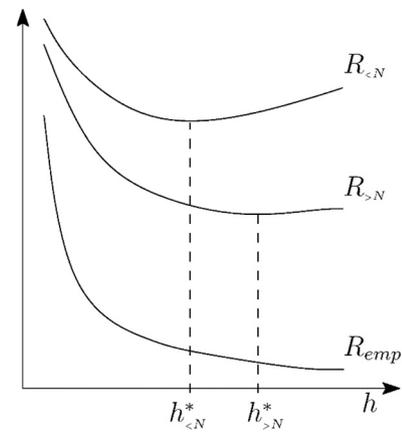


Fig. 1. Based on the empirical risk R_{emp} multiplied by the confidence factor, the bound on the expected risk R is derived. The minimum risk R leads to the appropriate model structure with h^* . Two cases are illustrated – a smaller sample size favouring a less complex model $h_{<N}^*$; a larger number of samples allow more complex models $h_{>N}^*$.

one applicable result of the statistical learning theory and is discussed in the following.

The SRM delivers a basis for choosing an appropriate learning approach. To select the best approximation to the supervisor’s response y , the expected risk $R(w)$ (Eq. (2)) needs to be minimised. Due to the unknown joint distribution function, we minimise the empirical risk $R_{emp}(w)$. The question arises, when does a learning machine, minimising the empirical risk, achieve a minimal expected risk and consequently a good generalisation ability. To reach this aim, the necessary and sufficient conditions for the consistency of a learning process need to be evaluated. The ERM principle is consistent if both the expected $R(w)$ as well as the empirical risk $R_{emp}(w)$ of a set of functions $f(x, w)$ converge to a minimum possible value of the risk (Eq. (14)).

$$R(w_N) \xrightarrow{N \rightarrow \infty} \inf R(w) \tag{14}$$

$$R_{emp}(w_N) \xrightarrow{N \rightarrow \infty} \inf R(w)$$

A necessary and sufficient condition for distribution-independent consistency of the ERM is a finite VC dimension h . Due to its derivation for classification tasks, the VC dimension describes the capability to separate/shatter two classes. The VC dimension h is defined as the maximum number of samples that can be shattered by a set of functions in a binary classification task. No $h + 1$ samples can be shattered by this set of functions. In the case of linear learning machines, the VC dimension results in $h = d + 1$ for a d -dimensional input space and corresponds to the number of parameters. Generally, learning is only possible if the used samples N exceed the VC dimension h . To generalise, the approximating functions should not be too flexible as indicated by the VC dimension as a capacity measure (Cherkassky and Mulier, 2007). Vapnik (1982) derived the bound on the probability of error for algorithms minimising the empirical risk. Three quantities are necessary: the VC-dimension h as a measure of model capacity, the empirical error R_{emp} and the training sample size N (Vapnik et al., 1994). Furthermore, the probability $1 - \eta$ for the computed bound is considered. The practical form of the bound on the expected risk R , already including a standard confidence level derived by $\eta = \min\left\{\frac{4}{\sqrt{N}}, \frac{1}{\sqrt{h}}\right\}$, is:

$$R \leq R_{emp} \left(1 - \sqrt{\tau - \tau \ln(\tau) + \frac{\ln(N)}{2N}} \right)^{-1} \tag{15}$$

with $\tau = h/N$ (Cherkassky and Mulier, 2007). The denominator of Eq. (15) corresponds to a confidence factor. The effect of the confidence factor is illustrated in Fig. 1. The empirical risk R_{emp} decreases with

larger model capacity h ; the confidence factor increases, which results in an increased risk R . For small sample sizes ($N/h < 20$), a small confidence level $1 - \eta$ is derived and consequently, the confidence factor causes a larger bound on the risk $R_{<N}$. If the number of samples grows, the confidence factor reduces proportionally and demands a minimisation of both components – the empirical risk R_{emp} as well as the confidence factor. This procedure is called the structural risk minimisation (SRM) and the VC dimension h is the controlling variable. It represents a trade-off between the quality of the approximation and the complexity of the approximating function (Vapnik, 2000).

Vapnik et al (1994) developed a method to measure the effective VC-dimension for linear classifiers. The VC-dimension of a set of real-valued functions $f(x_n, w)$ is equal to the VC dimension of the set of indicator functions (Cherkassky and Mulier, 2007). Thus, the measurement of the effective VC dimension is applied to the indicator functions:

$$I(x_n, w, \beta) = \theta\{f(x_n, w) - \beta\} \quad (16)$$

If the function output of $f(x_n, w)$ is equal or larger than β , than the indicator function results in 1; if it is smaller than β , it indicates 0. The estimation of the effective VC dimension (Vapnik et al, 1994) bases on the maximal difference between the error rates on two independently labelled data sets, representing the difference between the training and the test error. A theoretical function (Vapnik et al., 1994) is fitted to the empirical error differences for varying data set sizes to estimate the VC-dimension. For a compact description see Cherkassky and Mulier (2007, p. 143).

Vapnik et al (1994) indicate the difficulties of computing the VC dimension for multilayer networks. The VC dimension is derived for methods minimising the empirical risk over the entire set of functions realisable by the method. However, due to the non-linear optimisation, the energy surface has many local minima and the search is confined to a subset of these possible functions. Another point is that the number of local minima changes with the number of samples, which leads to increased variability of the estimated capacity. Thus, a dependence of the VC dimension's estimation on the number of samples is existent. The first drawback can be weakened by a repeated computation of the VC-dimension for different parameter initialisations.

3.3. Quality criteria

Various quality criteria can be used to assess the performance and compare the methods presented. The decision was made for the dimensions as following (see Eqs. (17)–(19)) to establish a relationship to former investigations (Nguyen and Cripps, 2001; Zurada et al., 2011).

The Root Mean Square Error (RMSE) is the common quality criterion for the MLR as well as for ANN, as already mentioned in Section 3.2.

The mean absolute percentage error (MAPE) is more resistant to outliers.

A measure of the variation of the MAPE is the Error Below 5% (EB5), which quantifies the number of relative deviations below 5%. The 5% is based on the tolerable estimation error of many investors (Nguyen and Cripps, 2001). The model adapts better to the actual target figures if many relative deviations are less than 5%.

A good estimate is achieved if the RMSE and MAPE are small and the EB5 is large.

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (y_n - \hat{y}_n)^2}{N}} \quad (17)$$

$$MAPE = \frac{\sum_{n=1}^N \frac{|y_n - \hat{y}_n|}{y_n}}{N} \quad (18)$$

$$EB5 = \frac{\text{count}\left(\frac{|y_n - \hat{y}_n|}{y_n} < 0.05\right)}{N} \quad (19)$$

4. ANN modelling of aggregated data

A reliable real estate valuation in areas with few transactions based on data sets is an open research point. If not estimated by experts' knowledge, the usual workaround is to estimate them together with neighbouring local markets. Due to the different behaviour of the markets, non-linearities are present and the MLR is not the appropriate estimation procedure. A decisive criterion for non-linear modelling of real estate valuation data by ANN is the number of data samples. Therefore, the aggregation of various local markets to build an adequate database for non-linear ANN modelling appears as a constructive approach.

On basis of the methodical fundamentals from Section 3, we provide an answer to the formulated research question: Are aggregation strategies appropriate to generate a sufficient database for ANN-based modelling of non-linear real estate valuation tasks? Therefore, two aggregation strategies are considered and the data sets are described in detail. The ANN modelling process of the aggregated data consists of the model selection and the estimation part. These results build the basis for the comparison of the two aggregation strategies and a deeper analysis of the problem complexity in the next section.

4.1. Data aggregation strategies

The number of purchases in a specific market is regularly very small (in Germany). Especially in rural areas and smaller cities, the number of purchases within a short period (1–2 years) is smaller than 100 purchase prices. With this data amount, the statistical analysis to derive comparison factors for the comparison approach does not lead to good results in inhomogeneous markets like the real estate market. For this reason, it is necessary to aggregate data. The first way to do this, is to use data from a longer period (e. g. 10 years). Here the changes that occur in the market have to be modelled. The total amount of data in this region can't be extended to more than 1000 purchase prices in those regions using the temporal expansion.

For this reason, other aggregation procedures have to be adopted. To stay in the same regional submarket with the same local influencing parameters (accessibility, purchase power, etc.) the only option is to aggregate data from different functional submarkets. To combine different submarkets the basic model of influencing parameters should be similar. For this reason, it could be possible to combine different markets like one-family houses, two-family-houses, semi-detached-houses, detached-houses, and villas. These real estates are normally bought for their use. Typical common parameters that can describe the height of the purchasing price in these markets are:

- Standard land value (indicator for locational quality)
- Living space
- Building year, modernisation or remaining usage time
- Quality
- Lot size

Another approach to derive data for regions with few transactions is a regional aggregation. The data from different regions (rural and/or urban areas) can be aggregated to reach a big amount of data. The local phenomena can be modelled with categorical variables. In Lower Saxony, the committee of valuation experts derived a schema of homogeneous regions as so-called housing market regions (OGA NDS 2019). In total, the federal state is divided into 16 regions, where 6 regions are urban areas (cities of Hanover, Brunswick, Wolfsburg, Göttingen, Osnabrück, Oldenburg), 3 regions in the surrounding of the cities, 6 regions in rural areas and 1 region including the islands in the North Sea. Similar regions with a similar density of population and accessibility can be combined to reach a sufficient number of purchase prices for the analysis.

In this paper, we use two different data sets. The first data set is a data set of real estates from different functional submarkets in individual housing from a long-time span. It includes one-family houses, two-family-houses, semi-detached-houses, detached houses and villas for the local submarket of the city of Hanover in northern Germany. The second data set is a data set of one-family houses and semi-detached houses in different regions (cities surrounding areas in Lower Saxony) all over the federal state of Lower Saxony in northern Germany. The data is described in more detail in the following section.

4.1.1. Functional aggregation

The selection of the real estates is only limited by the functional markets. The data set contains different real estates from the city of Hanover with the functional markets: one-family houses, two-family-houses, semi-detached-houses, detached houses and villas. Hanover is the capital city of the federal state of Lower Saxony with half a million inhabitants. Unlike bigger German cities like Berlin Munich or Hamburg, the market in this city is not extremely tense. Nevertheless, the prices in Hanover are also rising.

The used data for this investigation is provided by the public committee of valuation experts in lower Saxony from their purchase price collection. The purchase price collection contains all transactions on the real estate market.

In the pre-treatment the purchase cases that are extremely untypical and differing strongly from the general deviation for the investigated markets are eliminated from the used dataset: Purchases that have living spaces bigger than 450 sqm, a larger lot size than 5000 sqm and standard land value bigger than 600 €/sqm are removed from the sample. These limits are set after investigating the quantiles of the data (boxplot far outliers). In total 1241 purchases remain in the sample.

The descriptive statistic is shown in the appendix (Table A1). All characteristics have a wide range so that nonlinearities can be expected (changing influence within the range). Most of the data have a left-skew distribution, which is common in real estate data. This data set is analysed with the parameters: living space, lot size, standard land value, date of purchase, modernisation.

4.1.2. Spatial aggregation

In the second data set, which also was provided by the committee of valuation experts, we combine different regions from the housing market regions in Lower Saxony. The housing market regions are generated by experts' surveys as homogeneous markets within the federal state of Lower Saxony. We select the regions from the cities Oldenburg, Osnabrück, Wolfsburg, Hanover, Brunswick, Göttingen as well as the surrounding areas of the city of Bremen, Hanover-Brunswick-Wolfsburg (8 submarkets, 7 categories). While the city markets are relatively homogeneous the combination with the surrounding areas is the challenging point here. In these regions, the functional submarket of one-family houses and semi-detached houses are investigated. The transaction took place in a time-span of 8 years.

In the pre-treatment, the standard land value for the data is standardised with local price indexes to one specific data to compensate the economic development of the land-price market (the standard land value has been set i at the moment of purchase). If the standard land value is standardised it only contains the locational quality. The selection of the real estates is only limited by the functional markets. Additional purchases that have living spaces bigger than 450 sqm, larger lot sizes than 5000 sqm and standard land values bigger than 600 €/sqm are removed from the sample. These characteristics are not typical for the real estates in the investigated market. The limits have been set with an investigation in the quantiles (boxplot far outliers) In total 19,606 purchases remain in the sample. The descriptive statistic is shown in the appendix (Table A2).

In comparison to the smaller data set from Hanover, the range is even wider in most parameters. The standard land value and the quality have a more symmetric distribution due to the bigger sample. This is caused

Model selection (4.2.1):

$$\text{NHid} \rightarrow \text{size}(w_0)$$

Model estimation (4.2.2):

$$y_{\text{train}} = f(w_0, x_{\text{train}}) \rightarrow \hat{w}$$

$$\hat{y}_{\text{test}} = f(\hat{w}, x_{\text{test}})$$

Fig. 2. Workflow of the ANN modelling.

by the aggregation of different locational qualities. All local submarkets are typically skew in this parameter but with aggregating of several skew data, the aggregated data set gets symmetrically.

This data set is analysed with the parameters: living space, lot size, standard land value, date of purchase and remaining usage time.

4.2. ANN modelling

This section presents the ANN model selection task as well as the ANN estimation results for the spatial and the functional aggregated submarkets. The workflow of the ANN modelling process is pictured in Fig. 2. In the model selection part, the structure of the ANN will be determined. The adequate number of hidden nodes (NHid) is figured out and thereby the number of weights is set. The model estimation builds on this structure and estimates on basis of training data ($x_{\text{train}}, y_{\text{train}}$) and random parameter initialisations w_0 the parameters \hat{w} . The evaluation of the estimated ANN is accomplished on basis of independent test data ($x_{\text{test}}, y_{\text{test}}$) and the estimated parameters \hat{w} .

Three ANN computation types are treated as follows:

- ANN + LM: Minimisation by LM (Eq. (6))
- ANN + EKF: Minimisation by EKF – the parameter cofactor matrix is updated Q_{ww}^{\sim} (Eq. (10))
- ANN + EKF-dq: Minimisation by EKF – additionally to the updating of the parameter cofactor matrix Q_{ww}^{\sim} (Eq. (10)) the process noise q_{pp} is adapted

The first one is a standard optimisation method for ANN estimation, as mentioned in Section 3.2.1. The standard EKF optimisation only updates the parameter cofactor matrix Q_{ww}^{\sim} . In comparison to LM, the criteria of an updating step (instead of the overall error the test value is used), as well as the adaption term (instead of λ the observation variance factor q_{yy} is used), changed. The EKF version *EKF-dq* additionally adapts the process noise by q_{pp} , which accelerates heuristically the estimation process. Additionally, the multiple linear regression (MLR) is applied to the data to compare and classify the ANN modelling results.

All further processing is accomplished for the three computation types to evaluate if there are some advantages in the performance of one of these methods. The test data is randomly chosen. Based on the recommendations for CV (James et al., 2013), this was set at 15%. The following hyperparameters of the estimation methods have been chosen for all further computations: LM requires an initial learning λ and an updating rate $d\lambda$ (Section 3.2.1). These are chosen with 0.01 and 1.5, which are standard values according to Haykin (1999) and have not especially been estimated in a model selection. For the EKF computations (see Section 3.2.2), also standard values for the hyperparameters q_{yy} and q_{pp} have been considered (Haykin, 2001). For q_{yy} the same values as for the learning rate of LM are used because of its similar behaviour. q_{pp} initially amounts to 0.001 and reduces by a factor 0.96. All computations stop after 200 iterations and a batched version with a

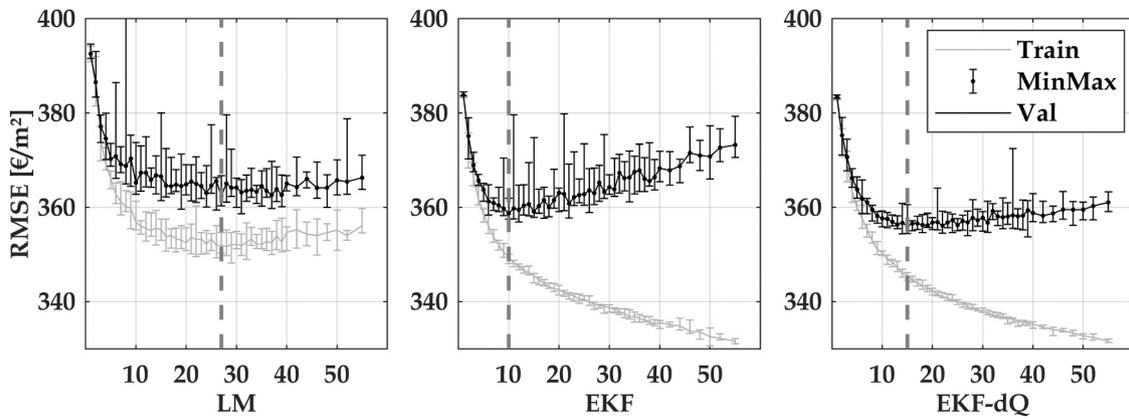


Fig. 4. CV result of the spatially aggregated data set for the computation types LM, EKF and EKF-dq. Minimum and maximum RMSE complement the median training and validation error. The x-axis shows the number of hidden nodes (NHid). The vertical line indicates the location of the minimal median validation RMSE.

Model selection:

$$5\text{-fold CV}/t \text{ param. init.} \rightarrow \underset{NHid}{\operatorname{argmin}}(RMSE_{validation}) \quad t = 10$$

Fig. 3. Implementation of the model selection.

Model estimation:

$$t \times \text{ANN parameter estimation} \quad y_{train} = f(w_0, x_{train}) \rightarrow \hat{w} \quad t = 100$$

$$\hookrightarrow \text{ANN prediction} \quad \hat{y}_{test} = f(\hat{w}, x_{test})$$

Fig. 5. Implementation of the model estimation.

size of 80 samples per batch is used.

4.2.1. Model selection

A decisive step in ANN modelling is the choice of an adequate structure. As mentioned in Section 3.3, we only consider an ANN consisting of one hidden layer and therefore, only the number of nodes (NHid) in this hidden layer needs to be determined. The model structure leading to a good prediction ability is selected on basis of cross-validation. The 5-fold CV is applied to the training data. For each candidate model structure, the parameters are initialised 10 times. For each parameter initialisation, a randomised drawing of the five subsets out of the training data is generated and the non-linear CV according to Section 3.2.4 is computed. The results of the ten 5-fold CV are depicted by the median as well as the minimum and maximum value of the RMSE (see Fig. 4). Thereby, the variance is recognisably caused by the different local minima reached in the estimation process due to the different initialisations. The model structure leading to a minimal validation RMSE presents the best generalisation ability. Due to the large variations, the minimum median validation RMSE is used as the selection criterion. In Fig. 3 the procedure of the model selection is summarised.

The small number of samples for the functionally aggregated data set justifies only a small model complexity. According to Widrow’s rule of thumb (Eq. (13)), a maximal number of hidden nodes in the single-digit area is meaningful. Thus, models with a maximum of 10 hidden nodes are considered. A model consisting of only one hidden node (NHid = 1) achieves the minimal validation error for all computation types. The training error decreases typically with increasing model structure. In contrast, the immediate increase in the validation error is peculiar. Comparing the RMSE of the computation types already indicates a smaller minimal median validation error and a more smooth behaviour for the EKF versions due to less variance.

The spatially aggregated data set consists of approximately 16.000 training samples. Therefore, more complex models are theoretically possible. Models consisting of up to 55 hidden nodes are considered (Eq. (13)) and the CV results are pictured in Fig. 4. The validation graphs show that model structures with 27 for LM respectively 10 for EKF and 15 hidden nodes for EKF-dq are sufficient. Again, the EKF computations present a lower median RMSE and exhibit less variance. Additionally, less complex EKF-models achieve these lower RMSE.

To summarise, the functionally aggregated data set only allows a

Table 2

Statistics of repeated (t = 100) model estimation and computing average test errors for various computation types.

	Functional			Spatial		
	Mean (RMSE) [€/m²]	StdMean [€/m²]	<MLR	Mean (RMSE) [€/m²]	StdMean [€/m²]	<MLR
MLR	535	–	–	389	–	–
LM	547	1,3	2/100	376	1,6	70/100
EKF	537	0,9	32/100	429	11,3	20/100
EKF-dq	534	0,1	80/100	360	0,2	100/100

simple model with one hidden node and is equal for all computation types. However, the EKF computations present lower validation errors and less variance. For the spatially aggregated data, the selected models range between 10 and 27 hidden nodes. The same tendencies as for the functionally aggregated data set occur, but the differences between the computation types increase: The EKF computations require less complex models to reach lower validation errors in contrast to LM.

4.2.2. Model estimation

On basis of the selected models in Section 4.2.1 the model parameters are estimated \hat{w} and the models are examined on the not yet considered test data (x_{test}, y_{test}) . Fig. 5 presents a compact overview of the estimation procedure. To avoid solutions on basis of local minima a repeated estimation of the model parameters on basis of different parameter initialisations is accomplished. 100 ANN estimations are computed as a trade-off between deriving reliable results and being computationally achievable.

Table 2 and Fig. 6 present the results of the ANN estimations for the spatially and the functionally aggregated data sets. As a first indication, the parameter <MLR shall be regarded. It specifies if the RMSE and the MAPE (Section 3.4) of the ANN computations (LM, EKF, EKF-dq) are smaller than the MLR and it adds up for all cases (out of t) where it applies. Comparing these values for both data sets, the conclusion can be drawn that the ANN estimation for the spatial aggregation is more

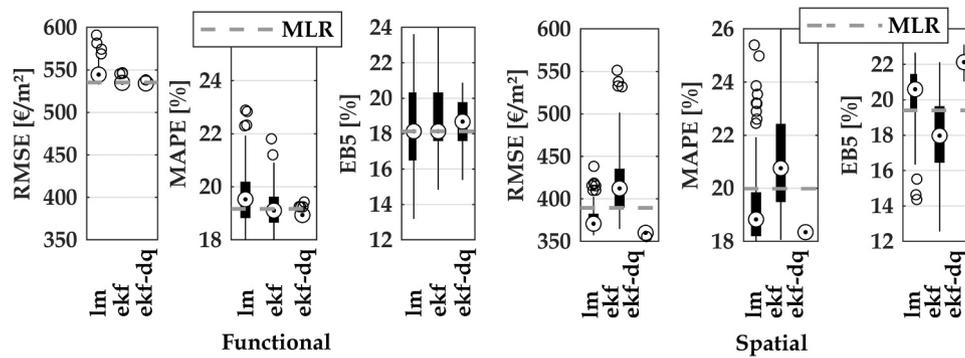


Fig. 6. Model estimation results for the functional (left) and the spatial aggregation (right) on basis of the cross-validated model structures. The dashed line indicates the MLR result. The boxplots summarise the 100 ANN computations for the quality measures RMSE, MAPE and EB5; the circles indicate outliers.

beneficial than for the functional one. However, the more decisive criterion is performance improvement due to the computation methods.

For the spatial aggregation, LM and EKF-dq improve significantly to the MLR. This statement is based on the mean RMSE concerning its standard deviation (StdMean) in Table 2. In the case of the functionally aggregated data set, only EKF-dq yields significant improvements. Fig. 6 clearly shows that the quality measures of the spatially aggregated data deviate much more from the MLR results. While the RMSE strongly depends on the price level of the present market, MAPE and EB5 are better comparable due to its relative character.

The RMSE shows less variance for the functionally aggregated data set. MAPE indicates that the functional aggregation contains less absolute errors than the spatial one. The ratio of errors below 5% is larger for the spatial aggregation signifying a better locally adaption of the approximating function in contrast to the functional one.

The EKF performance of the spatial aggregation is conspicuous. The ANN estimations of the functionally aggregated database on the same parameter initialisations as we have identical model structures in all three computation types. In that latter case, EKF performs better than LM in terms of RMSE and MAPE. Generally, the better performance of EKF is due to the consideration of stochastic information. Simultaneously to the updating of the parameters \hat{w} , the parameter covariance is updated $Q_{\hat{w}\hat{w}}$. This may lead to better handling of the noise. The additional adaption of the process noise (EKF-dq) leads to a reduced influence of the observations. Thus, a constraining of the parameter's update magnitude is achieved. According to (Cherkassky and Mulier, 2007) regularisation controls the complexity respectively smoothness of approximating functions to fit available data. It acts similar as the SRM principle and constrains the empirical risk. This stronger regularisation of the parameter estimation in case of EKF-dq probably circumvents to reach further away from local minima.

The ANN estimation of the spatially aggregated data leads to superior results concerning the MLR. For the functionally aggregated data, the MLR delivers better results in the majority of cases.

5. Complexity analysis

The ANN modelling of the spatial aggregation of submarkets is successful in comparison to the functional one (see Table 2). In this section,

Table 3

Complexity is measured by smoothness (S1) and the average correlation (C2). For the down-sampled spatially aggregated data, 1000 data samples have been 100-fold randomly drawn and therefore mean values inclusive standard deviations are computed.

Aggregation	S1	C2
Functional	699	0.21
Spatial	502	0.22
Spatial (n = 1000, rep = 100)	586 ± 20	0.21 ± 0.01

we analyse causes for the different performances of the ANN in case of the two aggregation approaches. Both data sets have six influencing variables. Additionally, the submarkets are modelled on basis of categories as further inputs to differentiate their different price levels. In the case of the spatially aggregated data, seven categories are added to the inputs, in case of the functional one four categories need to be considered. A different problem dimensionality (10 vs 13 input measures), as well as a different number of training samples (1034 vs 16,650), is encountered.

We assume that the ANN performance for both types of aggregation behaves differently due to a different number of data samples as well as a different complexity of the target functions. From this follows hypothesis one, stating that the complexity of an unknown relationship is approachable and that a minimum sample size corresponding to a specific problem complexity is derivable for ANN modelling.

According to Ho and Basu (2002), an unknown input-output relationship $P(y|x)$ can be complex due to different reasons. First, an unknown relationship is intrinsically complex. Second, the data set at hand does not represent comprehensively the sampling density of the unknown relationship (too small samples). Third, influencing parameters x that are not informative enough to describe the unknown relationship. Thus, additionally to the intrinsic complexity, data sparsity and an ill-defined problem description influence the complexity of a data set. Often it is a mixture of the mentioned causes.

In the following, we approach the intrinsic complexity of the unknown relationships $P(y|x)$ and determine the data sparsity. The chosen inputs are unimproved for this study, as their selection was made with the technical knowledge of experts. Moreover, a critical level of required data samples is derived to get an indication about needed sample sizes for similar complex target functions.

5.1. Complexity measures

Lorena et al. (2018) developed measures of complexity for regression problems. They defined four categories of measures and tested their relevance on the performance of various learning approaches. It turned out that the correlation measures, as well as the smoothness of the outputs in comparison to the inputs, are the most effective in the discrimination of complexity. Therefore, we restrict our investigation to these two measures and explain them shortly.

The smoothness of the output distribution (S1) bases on a minimum spanning tree (MST) – an edge-weighted undirected graph with a minimum sum of edge weights. Each data sample corresponds to a vertex of the graph. A computed distance matrix between all pairs of the input data serves as the edge weights. The MST greedily connects examples nearest. This mapping is used to compute the average difference in the corresponding adjacent outputs and to verify if the same smoothness in the output is present. Consequently, a small mean value indicates a simpler unknown relationship and therefore a lower complexity. Out of

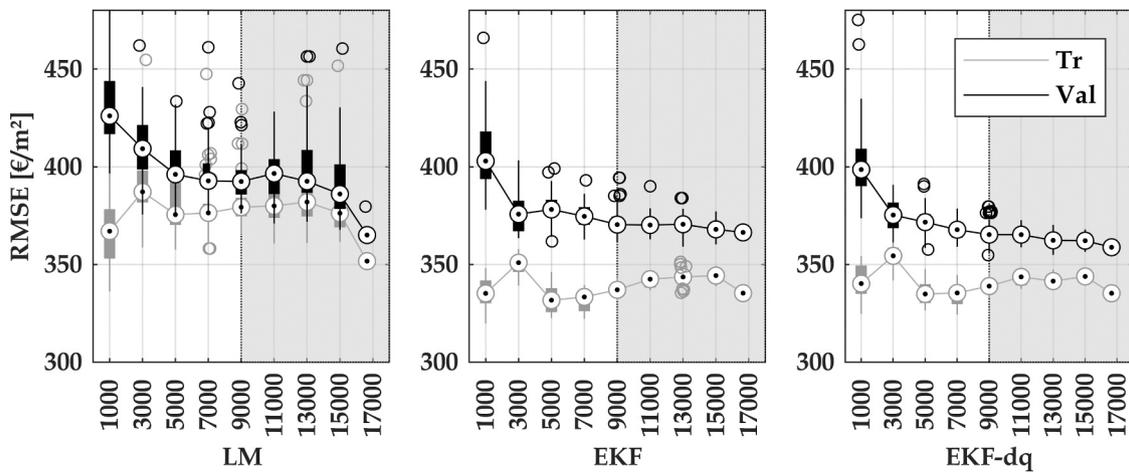


Fig. 7. Model selection results for various sample sizes N_i . It shows the variation of the training and the validation RMSE of the model structure leading to the smallest validation error for each sample size N_i . ANN performance increases and bias reduces up to 7000 samples and a reliable validation is achieved in the grey highlighted area.

the correlations measures, the average correlation (C2) is computed. It describes the average correlation of all inputs concerning the output. The larger the correlation ratio, the more of the output variation is explainable by the inputs and the closer is the unknown relationship to a linear one – and as a result the lower the complexity.

Table 3 presents the complexity measures for the functionally, the spatially as well as the down-sampled spatially aggregated data set for a meaningful comparison. In the latter case, 1000 samples out of the spatially aggregated data set are randomly drawn and the complexity measures refer to mean values of the 100-fold repeated computation. Comparing the down-sampled spatially aggregated data set with the original as well as with the functional one, weak evidence about differences in the problem pr complexity is given. Moreover, the measures confirm the increasing complexity with less data for the spatially aggregated data set, which is an indication for an existent data sparsity.

Data sparsity causes that intrinsically complex problems may appear deceptively simple (Ho and Basu, 2002). This likely applies to the functionally aggregated data. The continuous increase of the validation error lacking a distinct minimum point before the maximum number of hidden nodes is reached, is an additional indicator for this (Section 4.2.1). To verify if the functionally aggregated data set appears deceptively simple, it is again examined on basis of an analogy with the spatially aggregated data set. Therefore, the spatially aggregated data set is down-sampled and a CV-based model selection is accomplished. 1000 samples have been drawn 10 times out of the spatially aggregated data set. Thereby, the samples out of the submarkets are drawn proportional to the overall sample. For each of the ten sub samplings, the CV was accomplished as described in Section 4.2.1. The resulting average model structures comprise one (EKF, EKF-dq) respectively two (LM) hidden nodes. Thus, the down-sampled spatially aggregated data requires nearly the same model complexity as the functional one. The similar required model structures for the small sample size indicates that the functionally aggregated data can only be more complex but not less.

The complexity measures indicate a weak tendency that the functionally aggregated data set is rather more complex than the spatially aggregated one. Comparing the ANN models of the functionally and the down-sampled spatially aggregated data, it can be stated that the functional one is at least equally or more complex. Furthermore, the investigation shows that data sparsity is present: While the complexity measures specify that the down-sampled spatially aggregated data set is more complex than the original data set (Table 3), the ANN model selection indicates that the down-sampled data set requires a less complex model than the original one. The assumption is that if enough data samples are available and consequently no data sparsity is present, both

measures (complexity measures and the number of hidden nodes) should stay constant with varying sample size.

Thus, more samples representing the sampling density are needed for the functional as well as for the spatially aggregated data set to adequately model the unknown relationships.

5.2. Evaluation of minimal sample sizes

We have seen that the ANN model of the functionally aggregated data appears deceptively simple due to too sparse data. The functionally aggregated data set is an intrinsically complex problem demanding an adequate number of samples to reflect the sampling density. Thus, this subsection aims to find out the required number of samples for the complexity at hand. The bound of the true risk (Eq. (15), see Section 3.2.5) relates the expected and the empirical risk by the VC dimension, the number of training samples and a confidence level derived by the training samples. A certain level of the bound can be reached if the number of samples is proportional to the VC dimension. A rule of thumb specifies that the required sample size shall be roughly ten times the VC dimension (Baum and Haussler, 1989). However, this presumes a known model complexity, which we do not have. To get an idea about the required amount of data for both aggregations, the spatially aggregated data set with approximately 16,000 samples is used. For various sample sizes, a model selection is accomplished. The aim is to see if the model structure levels off for a certain sample size because of reaching the complexity of the unknown relationship.

The model selection, i.e. the determination of the number of hidden nodes (NHid), for varying sample sizes N_i is accomplished on basis of the CV method. The samples are randomly drawn 10 times for sample sizes 1000–15,000 in steps of 2000 samples. Thereby, the samples out of the submarkets are drawn proportional to the overall sample. It is complemented by the result of the model selection with all data samples ($N_i = 16650$). Model structures only up to 20 hidden nodes are considered for computational reasons. For EKF and EKF-dq it is sufficient. LM requires more than 20 hidden nodes to reach minimal validation error and a smooth increase to the model selection with all samples. EKF and EKF-dq show a slight levelling off effect for sample sizes starting from 7000/9000 samples up to the complete random sample ($N_i = 16650$). While the number of hidden nodes for EKF ranges between 7 and 15, EKF-dq requires 10–20 hidden nodes. The corresponding figure is enclosed in the appendix (Fig. A2).

For each sample size N_i , the training and validation errors of the optimal model structure (min Val-RMSE) are pictured in Fig. 7. The boxplots illustrate the 10 random samplings for the training and the

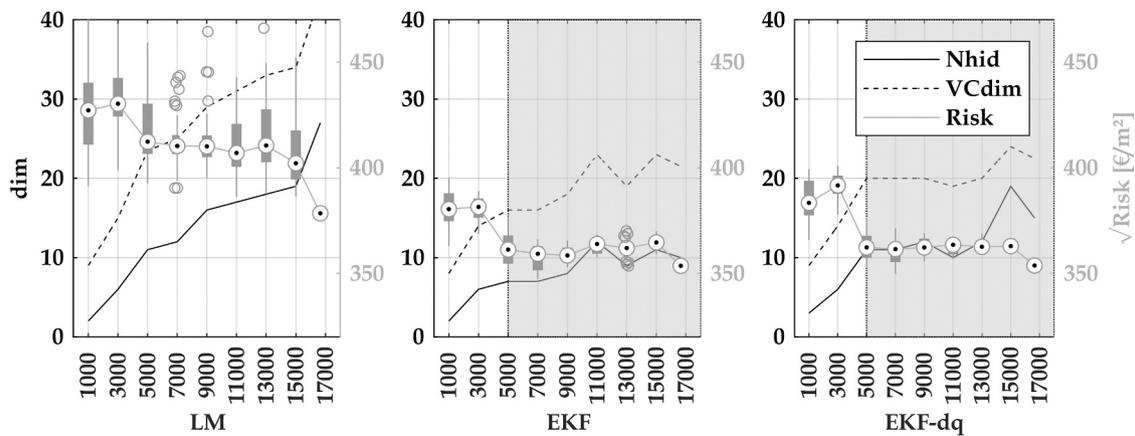


Fig. 8. It presents the model structure (NHid), the correspondent VC dimension and prediction risk for various sample sizes N_i . 5000 samples already lead to a comparable risk as considering more samples in case of EKF computations.

validation data. Noticeable is a change in the slope of the validation error starting by 9000 samples for all computation types. LM probably would need a larger model structure (limited by 20 hidden nodes) and therefore larger validation errors for sample sizes between 9000 and 15,000 samples are present. EKF and EKF-dq show nearly stationary validation errors for 9000–15,000 samples. All computation types show a validation minimum considering 16,650 samples at the cost of a more complex model. However, 9000 samples already lead to comparable validation errors. Smaller sample sizes ($N_i = 1000$ –7000) lead to local adaptations and produce a bias, indicated by the small training errors leading to worse performance on the validation data.

On basis of the identified model structures (Fig. A2) and the corresponding training RMSE (Fig. 7) for the various sample sizes N_i , the model capacities (VC dimension) and the prediction risks are computed and depicted in Fig. 8. The VC dimensions behave similarly as the model structure (NHid) and grow with increasing sample size N_i . The grey graphs in Fig. 8 show the corresponding prediction risk. To quantify the prediction risk (Eq. (15)), the empirical risk (see Fig. 7 - Tr) is multiplied by a confidence factor, which increases with more complex models and decreases with more samples. The EKF computations show a distinctive trend of the risk. Sample sizes N_i with 1000 samples exhibit high prediction risks. However, 5000 samples already lead to comparable risks as for the complete data set. LM exhibits much more variation in the risk and not such a distinctive trend is noticeable.

Due to the large variance in the LM computations, no distinctive tendencies are identifiable. One influencing factor on the ANN performance is the learning rate (Eq. (5)), which has not been varied and investigated. For EKF as well as for EKF-dq, three decisive categories emerge. First, on basis of the expected risk's upper bound (Fig. 8) a critical number of required samples is figured out. ANN estimation delivers reasonable results with minimal 5000 samples for the complexity at hand or comparable ones. Second, the ANN performance improves as well as bias reduces slightly up to 9000 samples (Fig. 7). The third category ranging from 9000 to 16,650 samples is a stable area with minor changes in model structure and validation errors. On basis of the SRM method, the critical sample size is determined with 5000 samples. The resampling technique CV delivers the growing (3000–7000) and the stable region starting by 7000 resp. 9000 samples. The increase between 15,000 and 16,650 samples is an indication for a larger problem complexity, which is not represented in the down-sampled data because of the missing sample density. The results emphasise the strength of the EKF algorithms. It exhibits less variance of the estimation errors and smaller risk bounds as well, requiring a smaller number of samples compared to LM.

Table 4

Average deviations for the realised prices per living space in % for the two test markets.

Spatial submarket	Local valuation methods			Valuation with ANN		
	Market	IPK-Hed	IPK-Comp	LM	EKF	EKF-dq
Hanover	21	19	15	15	20	16
Osnabrück	18	13	10	13	16	12

6. Comparison to state-of-the-art valuation methods

The cross-submarket ANN modelling aims to reach a similar appraisal level as for the local standard valuation methods. Therefore, the ANN computations shall be contrasted with state-of-the-art valuation methods at the local level. The comparison is carried out on basis of additional test data belonging to the spatially aggregated data set. 15 real purchasing cases each are chosen from two local submarkets – Hanover and Osnabrück. Their influencing parameters are used to predict a comparison value for each method. While Hanover is a quite heterogeneous market, Osnabrück presents less variation in the realised prices.

The realised prices per living space as output by the ANN are compared

- to the values appraised with the comparative value published in the market reports of the local committee of valuation experts (Market),
- to the hedonic model of single markets (IPK-Hed) as well as
- to the sales comparison approach using the mean of the 10 best matching transactions of real estates (IPK-Comp).

The ANN models of Section 4.2.1 are applied and the ANNs base on the same spatially aggregated training samples as for the computations in the previous sections. In Table 4 the performances of the different methods are summarised. Therefore, relative deviations concerning the realised prices per living space are computed and average values for each submarket are tabulated.

All ANN computations (LM, EKF, EKF-dq) lead to smaller average deviations than the appraisals by the market report. LM and EKF-dq lead to similar results. The best prediction of the prices per living space is reached by the IPK-Comp method. Due to the specific choice of purchase cases to derive an appraisal, this good performance is expected. The method uses local, best matching purchase prices and adapt the small remaining differences in all parameters with local regression function. In

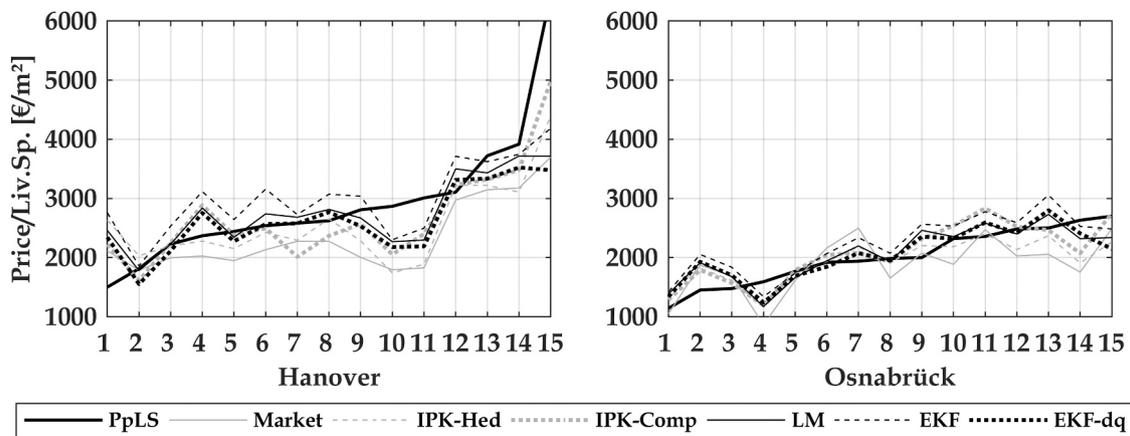


Fig. 9. The realised price per living space (PpLS) is compared to the predicted values of the different state of the art valuation methods (grey graphs) as well as on the different ANN computations (black graphs) for 15 test samples out of Hanover (left) and 15 test sample out of Osnabrück (right). The x-axes show the purchasing cases.

Table 5

Descriptive statistic for the parameter quality in the spatially aggregated data set. Showing the purchase case #10 (dashed line) and #11 (line) from Fig. 9, dataset Hanover.

	Minimum	Maximum	Mean	Median	Distribution
Quality (1 = low, 5 = high)	1	5	2.62	2.5	

case of the Hanover submarket, LM and EKF-dq reach a similar level as the IPK-Comp and a better result than the classical MLR. They are comparable to the hedonic model (IPK-Hed) for the Osnabrück submarket.

It is noteworthy that for these two test areas, regional evaluations with ANNs thus perform at least as well or even better than a local MLR. The ANNs are able to model the local situation in regional analyses at least as well as purely local analyses.

Fig. 9 enables a more detailed analysis of the method performances. It presents the different appraisals for each purchasing case (sorted ascending by purchase price per living space).

In the case of the largest price per living space of Hanover (purchasing case #15), IPK-Comp performs not well, but better than the other methods. The ANN predictions are influenced by all transactions in Hanover in different locations and over a long period. High purchase prices are rather rare, so the algorithm is too pessimistic at this point. However, compared to the IPK-Hed the regional prediction is not significantly worse than EKF. IPK-Hed only uses the closest 300 purchases around the specific real estate and estimates the best local matching hedonic model estimated from 10 parameters to describe the local data. The purchase case numbers 10 and 11 are also noticeable due to the poor valuation using local methods. This can be explained by the low quality of these real estates highlighted in Table 5. The predicted ANN prices for further purchasing cases deviate with a maximum 30%.

A general better appraisal quality can be reached for the Osnabrück market due to its more homogeneous behaviour. Thereby, mainly the second purchasing case leads to a large deviation.

The target of reaching a similar appraisal level as local real estate valuations by a cross-market ANN modelling is met. The ANN computations deliver similar quality levels as it is shown in Table 4. However, the performance differences between LM and EKF-dq are not that distinct. In this respect, ANN seems to be well suited to achieve local

evaluation results with cross-submarket evaluations provided by an adequate spatial aggregation. Thus, they are suitable for further investigations, to obtain statements in locations with few transactions, in which a larger spatially aggregated data set is chosen to generate a sufficiently large sample.

7. Conclusion

Two aggregations of submarkets have been evaluated – a functionally and a spatially aggregated data set. The functional one includes single-, two-family, double houses, terraced housing as well as villas. The spatially aggregated data set comprises single houses of various cities or surroundings in Lower Saxony. Due to the aggregations, the functionally aggregated data set exhibits 1036, the spatial consists of 16,650 training samples.

The number of samples of the functionally aggregated data set is not sufficient for non-linear ANN modelling. The situation is different for the spatially aggregated data set. Here, the ANN modelling leads to superior results in comparison to the MLR. The differences are caused by data sparsity, which means, that the data does not reflect the sampling density of the unknown relationship (not enough data). For this reason, the functionally aggregated data set appears deceptively simple. Consequently, the research question can be answered as following: the aggregation strategies are appropriate and promising to generate a sufficient database for ANN modelling if enough data samples concerning the intrinsic relationship complexity are available.

We try to approach the intrinsic complexity of the unknown relationship by complexity measures and the subsampling of the spatially aggregated data set. We figure out that the relationship of the functionally aggregated data set is equal or more complex than the spatial one. Furthermore, the complexity of the unknown relationship decreases while the ANN model increases for a growing sample size. The

contrary change of the complexity measures and the ANN model structure in dependence of the data set size indicates that the number of samples is not yet adequate to the complexity of the unknown relationship and data sparsity is present.

On basis of the spatially aggregated data, we figure out a minimal sample size concerning the problem complexity at hand. Consequently, the first hypothesis can be confirmed. For the evaluation of the quality of the results, cross-validation and a structural risk minimisation was performed. As a result, it can be stated that for the spatially aggregated data, minimal 5000 samples are needed and a reliable estimation can be achieved using 9000 samples. These results apply to the EKF computations. It has been proven that LM exhibits larger variance and therefore no sample size limits can be derived. Generally, EKF needs less complex models for a similar or better ANN estimation (smaller RMSE). Thus, it can be concluded, that EKF performs better than the standard method LM and hypothesis three can be confirmed.

On basis of the cross-submarket ANN model, accuracies of standard evaluations are achieved, which confirms hypothesis number two. There the differences in the predictions between LM and the EKF computations are not that distinct.

It can be summarised that the aggregation of submarkets is a constructive approach to reach larger sample sizes for non-linear ANN-modelling. Thereby, accuracies of local standard valuation methods are achieved by ANN estimations. Complexity analysis of down-sampled data enables the specification of a minimum sample size for a problem at hand. The higher modelling effort of ANN is only justified if more samples than the derived limit are present for a similarly complex relationship. This applies to the estimation of ANN parameters in the EKF, which generally presents less complex models, lower test errors and less estimation variance.

8. Outlook

Property valuation plays a key role in economies like Germany to keep the market transparent. Therefore, it is necessary to achieve reliable valuation results in all markets. Future work will include the transition of the developed method to further real estate valuation data sets. Thereby, common complexities of real estate valuation tasks shall be figured out and sample size limits will be derived. A controlled aggregation in functional and spatial dimension present a further extension to reach an adequate database for ANN modelling.

It can be observed that the more symmetric data (spatial aggregation) leads to better results than the skew data. As seen in chapter 6,

especially the purchase prices that are on the edge of distribution are at risk for under or overestimation. Further investigation on this issue will follow.

One aim of the non-linear ANN-modelling is to appraise areas with few transactions. These areas need to be included in the data set and the accessible performance of ANN for such purchasing cases needs to be evaluated. Moreover, a deeper analysis of the various computation types will be carried out. The investigations should focus on the selection of hyperparameters and the treatment of influences like outliers.

With the further development of the ANN approach comparison factors for real estate valuation can be provided even in areas with few transactions where the committees of valuation experts can't provide data yet, because of the poor data availability. With the possibility of presenting reliable information in these markets, transparency rises with the advantage of supply and demand are better known. Market participants as well as politicians and administrations are in a better and more secure position to make decisions.

CRedit authorship contribution statement

Sabine Horvath: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Validation, Software, Writing - original draft, Visualization, Writing - reviewing & editing. **Matthias Soot:** Conceptualization, Data curation, Formal analysis, Writing - original draft, Validation, Software. **Hans-Berndt Neuner:** Conceptualization, Methodology, Investigation, Supervision, Writing - reviewing & editing, Software. **Alexandra Weitkamp:** Conceptualization, Writing - original draft, Supervision, Visualization, Writing - reviewing & editing. **Sebastian Zaddach:** Resources, Validation.

Acknowledgement

The investigation in this article is made with data from the committee of valuation experts in the federal state of Lower Saxony. The authors want to thank the authorities for granting the use of the data. Moreover, the authors acknowledge the TU Wien Bibliothek for financial support through its Open Access Funding Program.

Appendix A

Fig. A1, Fig. A2, Table A1, Table A2.

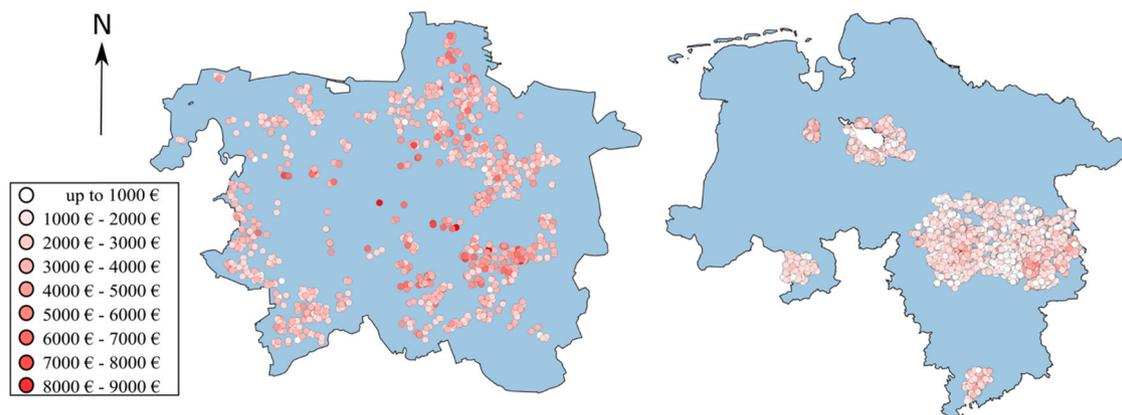


Fig. A1. Spatial distribution of the data set Hanover (left) and the data set of the housing market regions in Lower Saxony (right) showing the price per sqm living space per purchase.

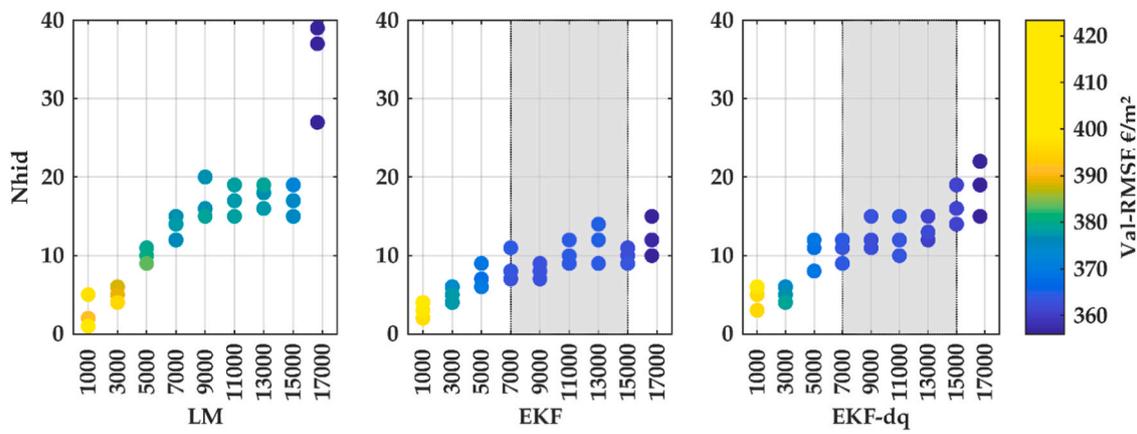


Fig. A2. CV-results for various sample sizes of the spatially aggregated data set. It shows the three model structures leading to the smallest validation error for each sample size N_i . The EKF computations show a slight levelling off effect in the grey highlighted areas.

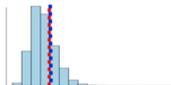
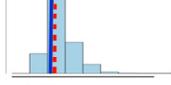
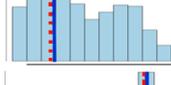
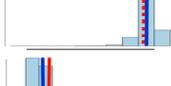
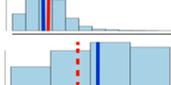
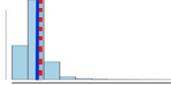
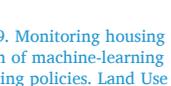
Table A1

Datasheet of the functionally aggregated data set from Hanover.

	Minimum	Maximum	Mean	Median	Distribution
Purchase price per sqm [€/m ²]	437	7292	2341	2208	
Living space [m ²]	43	450	140	132	
Remaining usage time [a]	16	86	35	29	
Building year	1874	2017	1971	1967	
Standard land value [€/m ²]	105	620	258	240	
Date of purchase	1/2011	3/2019	7/2015	7/2015	
Points for Modernization (max. 20)	0	20	4.5	4	
Quality (1 = low, 5 = high)	1	4.8	2.6	2.5	
Lot size [m ²]	60	1856	417	339	

Table A2

Datasheet of the spatially aggregated data set from 8 housing markets in Lower Saxony showing case number #10 (dashed line) and #11 (line) from Fig. 9 in Hannover.

	Minimum	Maximum	Mean	Median	Distribution
Purchase price per sqm [€/m ²]	49	8333	1663	1570	
Living space [m ²]	30	426	142	135	
Remaining usage time [a]	16	69	39	36	
Building year	1550	2017	1975	1976	
Standard land value [€/m ²]	6	585	122	110	
Date of purchase	1/2011	10/2018	3/2015	5/2015	
Points for Modernization (max. 20)	0	20	4.3	4	
Quality (1 = low, 5 = high)	1	5	2.62	2.5	
Lot size [m ²]	84	4736	738	677	

References

- Alkhatib, H., Weitkamp, A., 2012. Robust bayesian regression approach for areas with small numbers of purchases. In: RICS Cobra, 2012b.
- Alwosheel, A., van Cranenburgh, S., Chorus, C.G., 2018. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *J. Choice Model.* 28, 167–182.
- Baum, E.B., Haussler, D., 1989. What size net gives valid generalization? *Neural Comput.* 1 (No. 1), 151–160.
- Cherkassky, V., Mulier, F., 2007. *Learning from Data: Concepts, Theory, and Methods.* Wiley-Interscience.
- Din, A., Hoesli, M., Bender, A., 2001. Environmental variables and real estate prices. *Urban Stud.* 38 (11), 1989–2000. ISSN 0042-0980.
- Dorndorf, A., Soot, M., Weitkamp, A., Alkhatib, H., 2017. Development of a robust Bayesian approach for real estate valuation in areas with few. In: FIG Working Week 2016, Christchurch (New Zealand), 2–6 May 2016.
- Fan, C., Xu, H., 2009. The review on non-linear analysis and forecasting methods of the real estate market in China. *Asian Soc. Sci.* 4 (11), 156–160.
- Friedman, J.H., 1994. An overview of predictive learning and function approximation. In: Cherkassky, V., Friedman, J.H., Wechsler, H. (Eds.), *From Statistics to Neural Networks.* NATO ASI Series (Series F: Computer and Systems Sciences), vol. 136. Springer, Berlin.
- Gelb, A., 1974. *Applied Optimal Estimation.* Massachusetts Institute of Technology, Cambridge.
- Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the marquardt algorithm. *IEEE Trans. Neural Netw.* 5 (No. 6), 989–993.
- Halvorsen, R., Pollakowski, H.O., 1981. Choice of functional form for hedonic price equations. *J. Urban Econ.* 10 (1), 37–49.
- Haykin, S., 1999. *Neural Networks – A Comprehensive Foundation,* second ed. Prentice Hall.
- Haykin, S., 2001. *Kalman Filtering and Neural Networks.* Wiley.
- Ho, T.K., Basu, M., 2002. Complexity measures of supervised classification problems. *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (Issue 3), 289–300.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Netw.* 2 (Issue 5), 359–366.
- Horvath, S., Neuner, H., 2019. System identification of a robot arm with extended Kalman filter and artificial neural networks. *J. Appl. Geodesy* 13 (2), 135–150.
- Horvath, S., Soot, M., Weitkamp, A., Neuner, H., 2019. Artificial neural networks in real estate valuation. *Allg. Vermess. Nachr.* 126 (2019), 199–213, 8–9.
- Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., Cai, Z., 2019. Monitoring housing rental prices based on social media: an integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy* 82, 657–673.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning.* Springer.
- Lisi, G., 2013. On the functional form of the hedonic price function: a matching-theoretic model and empirical evidence. *Int. Real Estate Rev.* 2, 198–207.
- Lorena, A.C., Maciel, A.I., de Miranda, P.B.C., Costa, I.G., Prudencio, R.B.C., 2018. Data complexity meta-features for regression problems. *Mach. Learn.* 107, 209–246.
- Mimis, A., Rovolis, A., Stamou, M., 2013. Property valuation with artificial neural network: the case of Athens. *J. Prop. Res.* 30 (2), 128–143. ISSN 0959-9916.
- Moody, J., 1994. Prediction risk and architecture selection for neural networks. In: Cherkassky, V., Friedman, J.H., Wechsler, H. (Eds.), *From Statistics to Neural Networks.* NATO ASI Series (Series F: Computer and Systems Sciences), vol. 136. Springer, Berlin.
- Morano, P., Tajani, F., Torre, C.M., 2015. Artificial intelligence in property valuations an application of artificial neural networks to housing appraisal. *Adv. Environ. Sci. Energy Plan.* 23–29.
- Nguyen, N., Cripps, A., 2001. Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. *J. Real Estate Res.* 22 (Issue 3), 313–336.
- Pelzer, H., 1978. Ein indirektes Vergleichswertverfahren unter Anwendung statistischer Methoden. *ZfV Z. Geodäsie Geoinf. Landmanag.* 6 (103), 245–254.
- Singhal, S., Wu, L., 1989. Training multilayer perceptrons with the extended Kalman algorithm. In: Touretzky, D.S. (Ed.), *Advances in Neural Information Processing Systems,* p. 133–140.
- Soot, M., Weitkamp, A., Alkhatib, H., Dorndorf, A., Jeschke, A., 2016. Analysis on different market data for real estate valuation – investigations on german real estate market. In: FIG Working Week 2016, Christchurch (New Zealand), 2–6 May 2016.
- Soot, Matthias, Weitkamp, Alexandra, Dorndorf, Alexander, Alkhatib, Hamza, Jahn, Anha, 2017. Different regions with few transactions – an approach of systematization. FIG Working Week, Helsinki (Finland), 29 May – 2 June 2017.
- Soot, M., Zaddach, S., Weitkamp, A., Käker, R., Ziem, M., 2018. Weiterentwicklung der AKS – Implementierung neuer Auswertemethoden und Steigerung der. *ZfV Z. Geodäsie Geoinf. Landmanag.* 1 (143), 24–35.
- Tay, D., K.H. Ho, D., 1992. Artificial intelligence and the mass appraisal of residential apartments. *J. Prop. Invest.* 2 (10), 525–540.
- Vapnik, V., 1982. *Estimation of Dependencies Based on Empirical Data.* Springer, New York.

- Vapnik, V., 2000. *The Nature of Statistical Learning Theory*, second ed. Springer, New York.
- Vapnik, V., Levin, E., Lecun, Y., 1994. Measuring the VC-dimension of a Learning Machine. *Neural Comput.* 6, 851–876.
- Widrow, B., Stearns, S.D., 1985. *Adaptive Signal Processing*. Prentice-Hall.
- Worzala, E., Lenk, M., Silva, A., 1995. An exploration of neural networks and its application to real estate valuation. *J. Real Estate Res.* 10 (2), 185–201. ISSN 0896-5803.
- Zaddach, S., Alkhatib, H., 2013. Anwendung der Kollokation als erweitertes Vergleichsverfahren in der Immobilienwertermittlung. *ZfV Z. Geodasie Geoinf. Landmanag.* 138 (2), 144–153, 2013.
- Zhang, Y., Ling, Ch., 2018. A strategy to apply machine learning to small datasets in materials science. *NPJ Comput. Mater.* 4 (Issue 1), 25.
- Ziegenbein, W., 1978. *Zur Anwendung multivariater Verfahren der mathematischen Statistik in der Grundstückswertermittlung* (Dissertation). Technische Universität Hanover, 1978.
- Ziegenbein, W., 2010. Immobilienwertermittlung. In: Kummer, K., Kötter, T., Eichhorn, A. (Eds.), *Das deutsche Vermessungs- und Geoinformationswesen*. Wichmann Verlag, Berlin, pp. 421–468, 2010. ISBN 978-3-87907-547-8.
- Zurada, J., Levitan, A., Guan, J., 2011. A comparison of regression and artificial intelligence methods in a mass appraisal context. *J. Real Estate Res.* 3 (33), 349–387.