

Locally Adaptable Non-parametric Methods for Estimating Stand Characteristics for Wood Procurement Planning

Jukka Malinen

Malinen, J. 2003. Locally adaptable non-parametric methods for estimating stand characteristics for wood procurement planning. *Silva Fennica* 37(1): 109–120.

The purpose of this study was to examine the use of the local adaptation of the non-parametric Most Similar Neighbour (MSN) method in estimating stand characteristics for wood procurement planning purposes. Local adaptation was performed in two different ways: 1) by selecting local data from a database with the MSN method and using that data as a database in the basic k-nearest neighbour (k-nn) MSN method, 2) by selecting a combination of neighbours from the neighbourhood where the average of the predictor variables was closest to the target stand predictor variables (Locally Adaptable Neighbourhood (LAN) MSN method). The study data used comprised 209 spruce dominated stands located in central Finland and was collected with harvesters. The accuracy of the methods was analysed by estimating the tree stock characteristics and the log length/diameter distribution produced by a bucking simulation. The local k-nn MSN method was not notably better than the k-nn MSN method, although it produced less biased estimates on the edges of the input space. The LAN MSN method was found to be a more accurate method than the k-nn methods.

Keywords local non-parametric estimation, MSN method, stand characteristics, wood procurement planning

Author's address University of Joensuu, Faculty of Forestry, P.O. Box 111, FIN-80101 Joensuu, Finland **E-mail** jukka.malinen@joensuu.fi

Received 2 October 2001 **Accepted** 27 December 2002

1 Introduction

In order to evaluate cutting possibilities in available stands, custom oriented wood procurement requires more accurate information than what is available on the characteristics of a marked stand (cutting area). Better and more accurate information on reserves of marked stands would result in more optimal allocation of cutting operations. With the help of additional information sawmills may reduce trimming losses, avoid unmarketable raw materials and steer the flow of wood to the best secondary processing destinations (Uusitalo 1995).

Information on marked stands should contain log length-diameter distributions produced by a simulator with the relevant bucking controlling parameters. Although forest management planning information on marked stands may be available, in most cases it is not directly useful for bucking simulation (Räsänen 1999). Pre-measurements made while buying may be accurate enough, but these methods (Lemmetty and Mäkelä 1992, Uusitalo 1995) have turned out to be too expensive or not sufficiently reliable.

Tommola et al. (1999) used a non-parametric k-nn regression method to estimate the characteristics of a marked stand from data measured by means of log-measurement instruments installed at various sawmills. The results were encouraging, although estimates could be achieved only for previously used log length-diameter distribution constraints. For a situation where values or demand change, the method is not appropriate.

Malinen et al. (2001) developed an application of k-nn MSN-inference (Moer and Stage 1995) for estimating characteristics of a marked stand using a stem database. The stem database was collected while harvesting stands with modern harvesters, which collect diameters for every 10 centimetres from cut stems and save this detailed data into a specific file. The application developed was found to be a useful and flexible tool for predicting stand characteristics. The usefulness of the k-nn MSN-method compared to previously used k-nn regression methods lies in automated weighting calculation, where canonical correlation produces a weighting matrix to be used as a neighbourhood selection. The weighting matrix summarises the best (i.e. maximally correlated)

multivariate linear relationship between design attributes and planning attributes. Furthermore, when comparing to the method used by Tommola et al. (1999), this method has the additional benefit that the stem database estimates could be produced for new log constraints and their influence on the expected outcome could be tested.

Although non-parametric methods have been found to be useful tools in many forestry situations (Haara et al. 1997, Maltamo and Kangas 1998, Tommola et al. 1999, Malinen et al. 2001), there are some disadvantages in the use of these methods. The question of the optimal shape of the selected bandwidth or the number of neighbours to use is difficult because the answer usually depends on the shape and location of the data in the input space. In general, the approaches used in such situations vary from the simplest case, which ignores the variation of the function throughout the input space, to more complex algorithms, which attempt to select a number of neighbours appropriate to the interpolation scheme and the local properties of the function. This is not simple. At the boundary of the predictor space, the neighbourhood is asymmetric and models tend to be highly biased. Bias can also be a problem in the interior if the predictors are non-uniform or if the regression function has a substantial curvature. These problems are particularly severe when the predictors are multidimensional (Hastie and Loader 1993). By increasing the number of neighbours we can enhance the local validity of a model but at the same time the bias of results increases. This is the classic bias/variance dilemma, which greatly complicates the use of non-parametric methods.

In addition to the bias/variance dilemma, non-parametric methods that depend on the notation of 'neighbourhood' generally scale unfavourably to high dimensions. The reason for this behaviour comes from the non-intuitive effect that in high dimensional spaces all data points are approximately the same distance away from each other, thus destroying the discriminative power of neighbourhood relations. Given this 'curse-of-dimensionality', the ability of non-parametric methods to succeed seems to be limited.

Many practical applications demand transferability and therefore parameter tuning is out of the question. Although the term non-parametric refers

to a methodology without parameters, there are often plenty of parameters to be considered.

An ideal non-parametric method for this practical problem needs to eliminate redundancy in the input data, detect irrelevant input dimensions, keep the computational complexity low and achieve accurate function approximation and generalisation. A route to accomplish these goals can be sought from local models, which adapt without external parameters in each case according to the local input space.

Local adaptation of non-parametric methods has been the subject of several studies. Yang (1981) defined a new type of nearest neighbour regression estimate using the empirical distribution function with the predictors to define the window over which to average. This has the effect of forcing the number of neighbours to be the same both above and below the value of the predictor of interest.

Friedman (1994) introduced a flexible metric nearest neighbour classification method, which estimated the local relevance of each predictor variable, or their combinations, for each individual point to be classified. That information was then used to separately customise the metric used to define the distance from that point to its nearest neighbours. The procedure introduced is a hybrid between a regular k-nn method and tree-structured recursive partitioning techniques popular in statistics and machine learning.

While the performance of non-parametric methods is influenced by the curse-of-dimensionality, much research has focused on reducing this effect (Hastie and Tibshirani 1996, Schaal et al. 1998). If globally high dimensional data has locally only low dimensional distributions, it is advantageous to perform a local dimensionality reduction before further processing the data.

Hastie and Tibshirani (1996) proposed a locally adaptive form of nearest neighbour classification, discriminant adaptive nearest neighbour classification (DANN), by using local linear discriminant analysis to estimate an effective metric for computing neighbourhoods. They determined the local decision boundaries from centroid information and then shrunk the neighbourhoods in directions orthogonal to these local decision boundaries and elongated them parallel to the boundaries.

Schaal et al. (1998) examined several techniques for local dimensionality reduction. They found that locally weighted partial least squares seems to be the most favourable technique for local dimensionality reduction for high dimensional regressions, even outperforming the statistically appealing probabilistic factor analysis.

Despite the fact that developed local dimensionality reduction techniques are theoretically able to exploit low dimensional distributions, they quickly become computationally unfeasible and tend to be numerically less robust. Vijayakumar and Schaal (1998) proposed a method called locally adaptive subspace regression, which pre-processes data using local principal component analysis.

The aim of this study was to develop and test properties of two local adaptation methods for non-parametric estimation. Local adaptation methods were compared to the k-nn MSN-method (Malinen et al. 2001) by using harvester collected stem data to predict stand characteristics. The first locally adaptable non-parametric method presented uses the k-nn MSN-method and a sub-region of the data for estimating new local weighting for the predictor variables. Another method uses MSN-distance to calculate a combination of neighbours that minimises the distance between target stand and the weighted average of the predictor variables.

2 Material

This study used harvester collected stem data. It was measured and stored by Finnish forest enterprises in order to be used as the stem database prototype (Table 1). Due to different measurements and data collection in enterprises the data also includes information which was not usable as actual study material. The uniform study data consist of 209 stands located in central Finland. 151 stands were dominated by Norway spruce and 58 by Scots pine. Due to the small amount of Scots pine and birch data only the Norway spruce data was used in the study.

Stand variables of this material were calculated as means and sums of standwise measurements. Other variables such as forest site, location and

Table 1. Mean stand characteristics in study data.

	Min	Mean	Max
Stand area (ha)	0.3	2.3	20
Stand age (y)	58	93	160
Basal area (m ²)	0.1	14.2	47.4
Basal area mean diameter (cm)	12.9	25.8	35.0
Height of basal area median tree (m)	11.38	21.1	28.2

stand development class were registered when these stands were originally pre-measured.

3 Method

3.1 Non-parametric Methods

Non-parametric regression methods predict the value of the variable of interest as the weighted average of the values of neighbouring observations, the neighbours being defined with the predicting variables (e.g., Härdle 1989, Altman 1992). Before non-parametric methods can be implemented the following aspects must be considered: the distance function used, the smoothing parameters and the weighting function.

3.1.1 Distance Function

Non-parametric methods are critically dependent on the distance function. The relative importance of the predictor variables in generating the distance measurement depends on how the predictor variables are scaled. However, a distance measure does not need to satisfy the formal mathematical requirements for a distance metric. By altering the distance function using scaling predictor variables it is possible to tailor the method used to a particular planning objective.

There are many ways to define and use distance functions (Scott, 1992):

- Global distance functions. The same distance function is used in all parts of the input space.
- Query-based local distance functions. In each query the form of the distance function with its param-

eters is set by an optimisation process that typically minimises cross validation errors or related criteria. Stanfill (1987) refers to this approach as a uniform metric.

- Point-based local distance functions. Each stored data point has a distance function associated with it and the values of corresponding parameters. Stanfill (1987) refers to this approach as a variable metric.

In this study the distance in all methods presented is measured using the Most Similar Neighbour Inference (Moeur and Stage 1995). The similarity function used in the MSN method is the generalised Mahalanobis distance (Mahalanobis 1936). Canonical correlation analysis provides a unified multivariate approach to the computation of the weighting matrix in the distance function by summarising the relationship between a set of search attributes and a set of design attributes. The MSN similarity measure derived from canonical correlation analysis is:

$$D_{ij}^2 = (\mathbf{X}_u - \mathbf{X}_j) \Gamma \Lambda^2 \Gamma' (\mathbf{X}_u - \mathbf{X}_j)' \quad (1)$$

$\begin{matrix} 1 \times p & p \times p & p \times 1 \end{matrix}$

where \mathbf{X}_u is the vector of the known search variables from the target observation, \mathbf{X}_j is the vector of the search variables from the reference observation, Γ is the matrix of canonical coefficients of predictor variables and Λ is the diagonal matrix of squared canonical correlations.

In the MSN similarity function, the weighting matrix simultaneously weights the elements of the search variables according to their predictive power, while incorporating the covariance between the elements of the design attributes.

3.1.2 Bandwidth Selection

Smoothing bandwidth parameter h defines the scale of the range over which generalisation is performed. There are two ways to use this parameter (Atkeson et al. 1997):

- Fixed bandwidth selection (kernel method): h has a constant value. The distance of the farthest neighbour possible is constant.
- Nearest neighbour selection (k-nn method): h is set to be the distance to the k :th nearest data point.

The distance of the farthest neighbour depends on the density of nearby data.

These two ways to select bandwidth can be used globally or locally. In global bandwidth selection h is set globally by an optimisation process that typically minimises cross validation error over all the data. In the local bandwidth selection process h can be set on each query or each stored data point has bandwidth h associated with it.

A fixed bandwidth and a weighting function that goes to zero at a finite distance can have large variance in regions of low data density (Atkeson et al. 1997). This problem is present at the edges or between data clusters. In addition, a fixed bandwidth smoother cannot have any data within its span, which leads to undefined estimates. Using a nearest neighbour based bandwidth fixes this problem. More generally, nearest neighbour methods tend to produce less dramatic swings in variance than a fixed bandwidth (Cleveland and Loader 1994).

3.2 Comparison of Estimation Methods

3.2.1 *k*-nn MSN Method

The *k*-nn MSN method is based on distance-weighted nearest neighbour estimation, where *k* most identical stands are used for predicting the characteristics of the target stand. The *k*-nearest neighbour MSN method can be summarised as follows.

- 1) Obtain search variables. Complete coverage of search variables on all parcels.
- 2) Determine design variables. Concentrating on the relationship of search and design variables allows the MSN analysis to be tailored to a particular set of interesting characters.
- 3) Estimate the canonical coefficients and correlations from search and design variables for the MSN distance function.
- 4) Select the number of nearest neighbours to be used.
- 5) Perform the *k*-nearest neighbour MSN analysis, selecting the most identical neighbours from the reference data to assign to each target observation, while minimising the squared distance function.
- 6) Form estimates from selected most identical neigh-

bours using weighted averages.

The optimal number of neighbours used in the *k*-nn MSN method can be determined by using a cross validation method, for example by minimising the RMSE of certain characteristics. The minimisation of RMSE emphasises local accuracy of estimates more than average accuracy of estimates.

3.2.2 *Local k*-nn MSN Method

Although *k*-nn methods are more resistant to the curse-of-dimensionality than might be expected, they are not immune to it (Friedman 1994). The curse-of-dimensionality can be reduced by taking advantage of the fact that at least locally correlations between search attributes and design attributes may not vary with equal strength or in the same manner in all directions. By choosing locally proper weights for features we can give the most influence to those directions in which the correlation is biggest and scale insignificant directions to zero. The resulting shape of the neighbourhood will be thereby elongated along the de-emphasised directions and constricted along the influential ones.

In this method the reference data was reduced to consist only of locally important observations by using the *k*-nn MSN method to select the local neighbourhood. This local neighbourhood was then used to calculate a new local weighting matrix, which concentrates on local correlations between search attributes and design attributes. The final *k*-nn MSN analysis was performed by using this local weighting matrix and local reference data.

3.2.3 *LAN* MSN

Large neighbourhoods can affect the bias in boundary situations or in situations where the reference data is distorted. Nearest neighbours may be located on the same side of the input space, even in the situations where observations exist on the other side of the input space. Recognising these situations, adjusting locally the amount of the neighbours and selecting the combination of

neighbours symmetrically can reduce the bias of estimates.

The information used in the selection of the local neighbourhood must be produced from neighbourhood observations and the predictor variables of the target stand. In the proposed LAN MSN method every possible combination of neighbourhoods is examined in turn and the averages of the predictor variables of the neighbourhood combinations are calculated (Eq. 2). Every vector of the average predictor variables of all possible neighbourhood combinations is compared to a vector of the predictor variables of the target stand using MSN metrics and the combination of neighbours which is most identical to the target stand, i.e. minimizing the distance between target stand and average of neighbourhood combination, is chosen to be used in the calculation of estimates.

$$D_{ia}^2 = (\mathbf{X}_u - \mathbf{X}_a) \Gamma \Lambda^2 \Gamma' (\mathbf{X}_u - \mathbf{X}_a)' \quad (2)$$

$\begin{matrix} 1 \times p & & p \times p & & p \times 1 \end{matrix}$

where \mathbf{X}_a is the vector of the averages of the search variables from the neighbourhood combination.

The neighbourhood from which the optimal neighbourhood is selected is limited to 15 neighbours due to computer time consumption. When the size of the maximum neighbourhood (n) is expanded the demand for computer time increases by 2^n .

3.3 Forming Stand Characteristics

The stand characteristics of a target stand were formed according the estimated diameter distribution by selecting actual stems from the stem database using selected neighbour stands. In the selection the probability to pick actual stem from database was weighted according similarity of the chosen neighbour stands and the target stand. The estimated stem population was bucked with the bucking simulator to obtain an estimation of the log length-diameter distribution of the target stand. The log length-diameter distribution was also estimated from the corresponding characteristics of the reference stand by using weighted averages of each log length-diameter class. The

estimated diameter distributions were scaled to the measured basal area of each target stand, which were assumed to be known.

In the bucking simulator the stem was divided into 10-cm sections. For each section the thin-end diameter was obtained from the taper curve. Optimal bucking was performed by using dynamic programming (Bellman 1954) and bucking to demand.

3.4 Comparison of Results

The method used to estimate the accuracy of estimates was cross-validation. A target stand is a stand which was excluded from the reference stands and for which estimates were calculated. Each stand from the reference data, in turn, was used as the target stand.

The search variables (Table 2) were chosen from among commonly measured stand characteristics. According Malinen et al. (2001) the mean tree variables are the most important search variables, while the significance of the other variables is small. The design attributes used were obtained diameters in percentages of 0% (the smallest diameter), 20%, 40%, 60%, 80% and 100% (the largest diameter) of accumulated basal area, a and b of Näslund's height parameters (Näslund 1937) and volume of tree species.

The weights for the reference stands were calculated inversely according to similarity distance (Eq. 3). Thus, the nearest reference stands are weighted according to the similarity distance when target stand characteristics are formed. The weights are scaled between 0 and 1, where value

Table 2. Variables used in selection of the nearest neighbours.

Variables describing site	Variables describing growing stock
Location	Number of tree storeys
Temperature sum	Relative proportion of spruce
Stand area	Basal area per hectare
Stand age	Basal area mean diameter
Forest site type	Height of basal area median tree
Logging method	Volume of basal area median tree
Dominant species	

1 means identical observations. The weight W_{ij} of reference stand i for target stand j , belonging to the nearest stand, was then:

$$W_{ij} = \frac{1}{1 + d_{ij}} \tag{3}$$

where d_{ij} is the distance between target stand and reference stand.

The accuracy of the methods was measured by using relative standard error. The relative root mean squared error (RMSE%) was defined as:

$$RMSE\% = 100 * \frac{\sqrt{\frac{\sum_{j=1}^n (y_{ij} - \hat{y}_{ij})^2}{n-1}}}{\bar{\hat{y}}_i} \tag{4}$$

where n is the number of observations, y_{ij} is the real value of the variable i in stand j , \hat{y}_{ij} is the estimated value of the variable i in stand j and $\bar{\hat{y}}_i$ is the mean of the estimates of variable i .

The bias of the standwise estimates was calculated for the study data as a whole and for the 10% of the stands with the greatest mean height. Although the systematic error has not been a problem with non-parametric methods on a global level, the averaging effect of an increased number of neighbours produces a bias at the borders of the data. The estimations of length-diameter class distribution achieved through bucking simulation were compared to the length-diameter class distribution achieved through bucking simulation of actual stand values. These two distributions were compared using a distribution level (DI):

$$DI = 100 * \left(1 - \frac{\sum_{j=1}^n |D_{rj} - D_{ej}|}{2} \right) \tag{5}$$

where D_{rj} is the length-diameter class distribution of actual output in stand j and D_{ej} is the length-diameter class distribution of estimated output in stand j .

Distribution level is a simple and illustrative variable, which has been used in comparisons between demand and actual output distributions (Lukkarinen and Vuorenpää 1997). Distribution level indicates the similarity of two distribu-

tions. Identical distributions get a value of 100 (%) and the smaller the value is the more different the distributions are. It is also suitable for comparisons between predicted and actual output distributions.

4 Results

The k-nn methods yield more information when the number of reference stands is increased. However, if the number of reference stands is too large, the averaging effect of the method is also increased. In several studies where the k-nn method has been used in forestry the suitable number of reference stands has been set between 3 and 10. In this study the size of the neighbourhood in the k-nn MSN and local k-nn MSN method was set to 5 neighbours to give the best performance within the RMSE criteria (Malinen et al. 2001). A smaller neighbourhood would reduce the bias, but the RMSE would indicate weaker local validity of estimates.

With this data the accuracy of the local k-nn MSN method was not notably better than that of the k-nn MSN method (Fig. 1). Had the study data

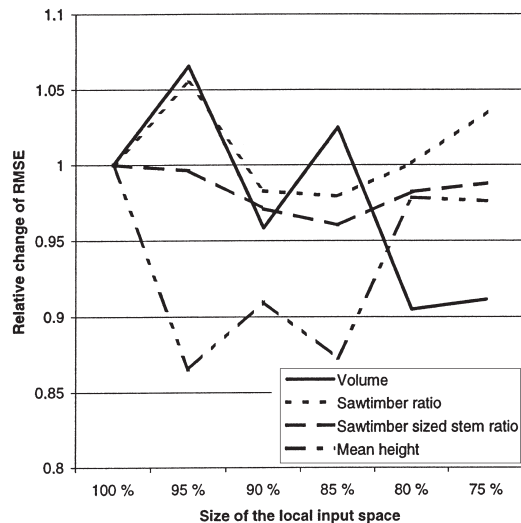


Fig. 1. Relative change of RMSE in the local k-nearest neighbour MSN method with different input space sizes.

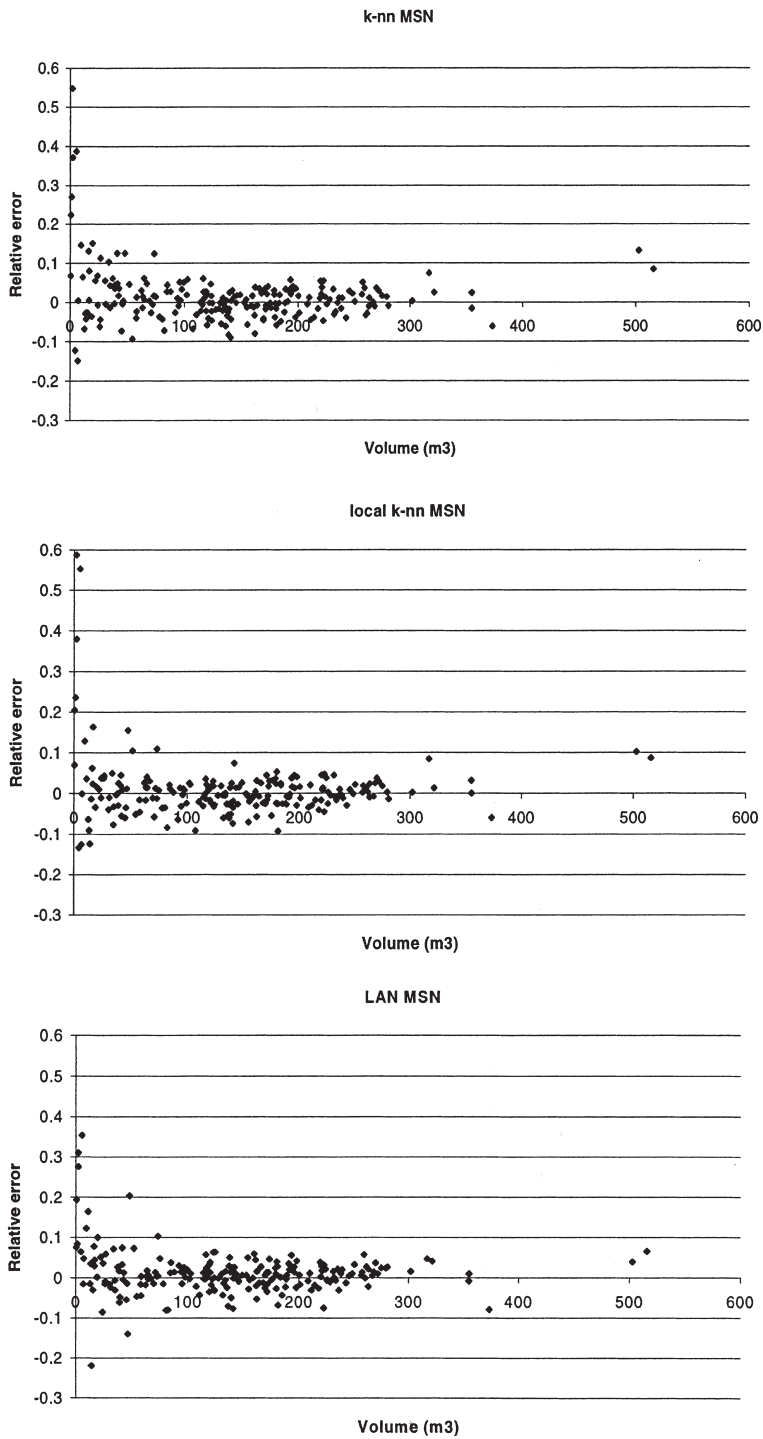


Fig. 2. Relative error of volume for the k-nn MSN method, local (75%) k-nn MSN method and LAN MSN method.

Table 3. Relative RMSE (%) of the k-nearest neighbour MSN method and locally adaptable MSN method.

	Volume	Sawtimber ratio	Sawtimber sized stem ratio	Mean height
k-nn MSN	5.41	7.10	24.09	2.27
LAN MSN	4.03	5.51	24.36	2.28

Table 4. Bias of the k-nn MSN method, local (75%) k-nn MSN method and locally adaptable MSN method.

	Volume (m ³)	Sawtimber ratio (%)	Sawtimber sized stem ratio (%)	Mean height (m)
k-nn MSN	0.821	-0.4307	1.8715	-0.047
local k-nn MSN	0.483	-0.5938	1.4084	-0.093
LAN MSN	0.930	0.1089	2.8511	-0.105

been more heterogeneous, the local data reduction may have yielded more accurate results compared to the global model.

The results of the local k-nn MSN method (Fig. 1) show that improved accuracy of one estimated character may lead to worse accuracy of another character. For example the accuracy of volume and mean height are negatively correlated. It is a question of balancing the weighting of the predictor variables.

The accuracy of the volume and sawtimber ratio was clearly better when using the LAN MSN method than with the k-nn MSN method (Table 3). However, the LAN MSN method did not yield a better accuracy of the sawtimber sized stems ratio and mean height. While relative errors of volume are displayed as a function of volume, it can be seen that the LAN MSN method outperforms the k-nn methods on the edges of the study data (Fig. 2).

Even though the bias of results increases as the number of neighbours increase, the nearest neighbour methods does not produce systematically biased estimates (Table 4). In stands with the greatest mean heights the bias of the mean height was notably lower with the local k-nn MSN method than with the k-nn MSN method (Fig. 3). The local weighting matrix concentrates

on those input space dimensions where the local correlation is greatest and therefore finds the most identical neighbours at the boundaries better than the k-nn MSN method. The LAN MSN method did not yield less biased estimates than the k-nn MSN methods.

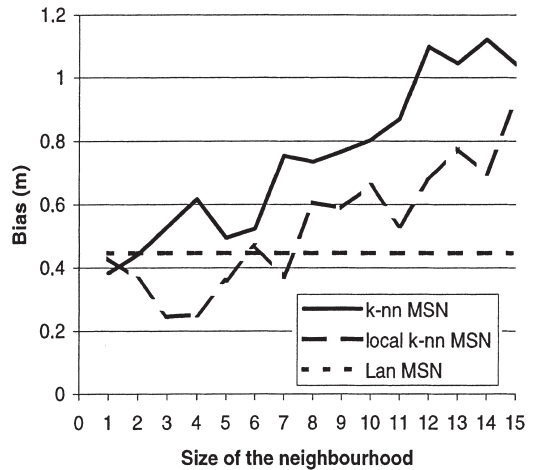


Fig. 3. Bias of mean height for the k-nn MSN method, local (75%) k-nn MSN method and locally adaptable MSN method.

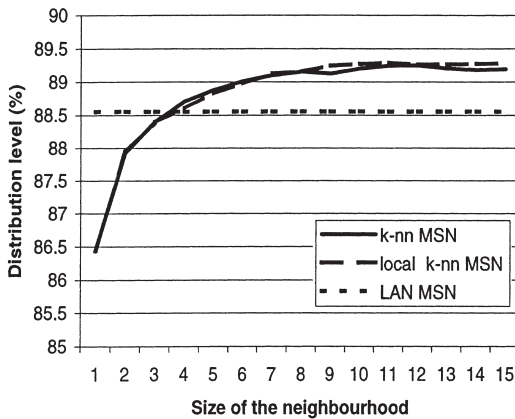


Fig. 4. Distribution levels from the k-nn MSN method, local (75%) k-nn MSN method and LAN MSN method.

Although distribution level is a simple and illustrative parameter when comparing two different log length-diameter distributions, it does not work well if the number of logs is too small compared to the number of different classes (Fig. 4). However, it can be clearly seen that distribution level emphasise more average goodness of estimates than RMSE. At the distribution level errors are weighted equally, independent of the error level, while RMSE is more sensitive to some large errors.

5 Discussion

Local adaptation of a non-parametric MSN method was tested in two different ways. The main principle was to keep the number of tuning parameters low. The method should be transferable to different databases and be automatically updated as harvesters collect new data from harvested stands.

The local reduction of input space in the local k-nn MSN method did not yield expected results. It would be more informative to use larger and more heterogeneous data to find out how local data reduction performs. In the local k-nn method the question of local input space size remains to

be decided with each problem and data before the method can be used.

Another problem with local reduction of input space was that even though the weighting matrix was calculated locally, the local input space was elongated according to global input space properties. It would be more appropriate to find local dimension of input space and use these properties in the calculation of the weighting matrix.

However, the bias of volume and mean height estimates was reduced on the edges of the input space. It would be justified to use the local k-nn method as a method to reduce bias of estimates in the edge situations, where bias is the biggest problem.

The accuracy of the LAN MSN method was clearly better than the accuracy of the k-nn MSN methods. The bias of the LAN MSN method was, quite surprisingly, not at the expected level. Estimates were not less biased than in the k-nn MSN methods.

The use of distribution level in comparison of methods failed due to coincidentally placed log distributions. Even small differences in stem diameter in certain points may produce different log combinations. When the number of bucked stems is too small, these differences affect the usability of distribution level as an indicator of accuracy. According the results of this study, the distribution level improves when size of the neighbourhood grows, i.e. when estimates are more averaging.

Distribution level is extensively used as an indicator of similarity of demand and actual output matrixes in forestry even though it has clear weakness. There are studies (e.g. Kirkkala et al. 2000) where improvements or alternative choices to distribution level have been sought, but the results have not been expected.

In the wood procurement planning context the estimates needed from available stands are complex. The length-diameter class estimate of log distribution contains information of size distribution of trees, volume, height and shape of individual trees, etc. While the estimation method concentrates more on one character of a stand, the other characters may get less weight in estimates. The question of optimal weighting of estimated properties is difficult and is affected by the method itself, but also by the selection of

used input variables and search variables. The effect of different methods on accuracy and bias may be distorted by properties of the data studied and a universally applicable conclusion cannot be made.

In this paper two local adaptation methods for non-parametric estimation were proposed. The results are not unambiguous and neither of these methods can be recommended as being superior in every situation. It would be easy to find situations where these methods are at their best and vice versa it would be equally easy to find situations where they do not work. The usability of these methods depends greatly on the application situation and the data used.

Acknowledgements

This research was funded by the Academy of Finland, within the research project 'Databases of timber procurement enterprises and forest mensuration as a basis for operational planning'. The author would like to thank Professor Matti Maltamo and Professor Annika Kangas for their valuable comments on the manuscript and Dr Lisa Lena Opas-Hänninen for revising my English.

References

- Altman, N.S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician* 46: 175–185.
- Atkeson, C.G., Moore, A.W. & Schaal, S. 1997. Locally weighted learning. *AI Review* 11: 11–73.
- Cleveland, W.S. & Loader, C. 1994. Smoothing by local fitting: various issues stimulated by a paper of Fan and Marron. Technical Report, Bell Labs.
- Friedman, J.H. 1994. Flexible metric nearest neighbour classification. Technical Report, Stanford University.
- Haara, A., Maltamo, M. & Tokola, T. 1997. The k-nearest-neighbour method for estimating basal-area diameter distribution. *Scandinavian Journal of Forest Research* 12: 200–208.
- Härdle, W. 1989. *Applied nonparametric regression*. Cambridge University Press, Cambridge.
- Hastie, T. & Loader, C. 1993. Local regression: automatic kernel carpentry (with discussion). *Statistical Science* 8: 120–143
- & Tibshirani, R.J. 1996. Discriminant adaptive nearest neighbour classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(6): 607–615.
- Kirkkala, A., Sikanen, L., Harstela, P., Ruha, T. & Tarnanen, T. 2000. Sakkosegmentoitu tavoitejakauma apteeraustuloksen arvioinnissa ja jakauma-asteen laskennassa. *Metsätieteen aikakauskirja* 1/2000: 59–61.
- Lemmetty, J. & Mäkelä, M. 1992. Suunnittelumittauksen perusteet ja toteutus. Summary: Methods for measurement of a stand for harvest planning. *Metsätehon katsaus* 11.
- Lukkarinen, E. & Vuorenpää, T. 1997. Hakkuukooneiden apteerauksen ohjausta konttorista käsin. *Koneyrittäjä* 1/1997: 12–13.
- Mahalanobis, P.C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences India* 2: 49–55.
- Malinen, J., Maltamo, M. & Harstela, P. 2001. Application of most similar neighbour inference for estimating marked stand characteristics using harvester and inventory generated stem Databases. *International Journal of Forest Engineering* 12(2): 33–41.
- Maltamo, M. & Kangas, A. 1998. Methods based on k-nearest neighbour regression in the prediction of basal area diameter distribution. *Canadian Journal of Forest Research* 28: 1107–1115.
- Moer, M. & Stage, A.R. 1995. Most similar neighbour: an improved sampling inference procedure for natural resource planning. *Forest Science* 41(2): 337–359.
- Näslund, M. 1937. Skogsförsöksanstaltens gallringsförsök i tallskog. *Meddelanden från Statens Skogsförsöksanstalt* 29: 1–169. (In Swedish).
- Räsänen, T. 1999. Runkopankki puunhankinnan ohjauksen apuvälineenä. *Metsätieteen aikakauskirja* 4/1999: 727–729.
- Scott, D. W. 1992. *Multivariate density estimation*. Wiley, New York.
- Shaal, S., Vijayakumar, S. & Atkeson, C.G. 1998. Local dimensionality reduction. In: Jordan, M.I., Kearns, M.J., & Solla, S.A. (eds). *Advantages in neural information processing systems* 10. MIT Press, Cambridge, MA.
- Stanfill, C. 1987. Memory-based reasoning applied to

- English pronunciations. In: Sixth National Conference of Artificial Intelligence. p. 577–581.
- Tommola, M., Tynkkynen, M., Lemmetty, J., Harstela, P. & Sikanen, L. 1999. Estimating the characteristics of a marked stand using k-nearest-neighbour regression. *Journal of Forest Engineering* 10: 75–81.
- Uusitalo, J. 1995. Pre-harvest measurement of pine stands for sawing production planning. University of Helsinki, Department of Forest Resource Management, Publications 9.
- Vijayakumar, S. & Schaal, S. 1998. Locally adaptive subspace regression. *Neural Processing Letters* 7: 139–149.
- Yang, S.S. 1981. Linear functions of concomitants of order statistics with applications to nonparametric estimation of a regression function. *Journal of the American Statistical Association* 76: 658–662.

Total of 23 references