COMPARISON OF BASIS FUNCTIONS FOR THERMAL ERROR COMPENSATION BASED ON REGRESSION ANALYSIS – A SIMULATION BASED CASE STUDY

It is a well-known problem of milling machines, that waste heat from motors, friction effects on guides, the environment and the milling process itself greatly affect positioning accuracy and thus production quality. An economic and energy-efficient method of correcting this thermo-elastic positioning error is to gather sensor data (temperatures, axis positions, etc.) from the machine tool and the process and to use that information to predict and correct the resulting tool center point displacement using regression analysis. This paper compares multilinear characteristic diagrams, B-spline characteristic diagrams, Radial Basis Function fitting and Wavelet fitting in general and also in the context of thermal error compensation. The demonstrations are made using FEM simulation data from a machine tool demonstrator. The results show that all of the above kernel types, if properly used, are able to create good compensation models. However, high-dimensional multivariate analysis usually only works by adding grid structures and regularization.

1. INTRODUCTION AND STATE OF THE ART

THERMAL EFFECTS are one of the main causes of positioning errors in machine tools [1]. They are caused by shifting temperature distributions in the machine tool, which lead to thermo-elastic deformations. These temperature fields are shaped by heat sources and sinks, such as waste heat from the cutting process, friction from guides and bearings, power loss from motors, coolants and the environment. Other relevant factors which influence the time-dependent temperature distribution inside a machine tool are the heat transfer coefficients and the thermal capacity which influence the rate at which heat is transferred and stored. It is this multitude of influences that makes thermal issues so difficult to handle [2].

The compensation of thermo-elastic deformations involves the prediction or measurement of temperature or deformation fields and using them to offset and thus correct the tool center point (TCP) position [3]. Compensation strategies can be subdivided into white-box, grey-box and black-box methods. White-box methods model the entire physical process.
of thermo-elastic deformation with all relevant aspects and parameters, although with
simplifications and limited resolution. Most of these methods use finite element simulations.
A detailed description of such a simulation based method can be found in [4]. Black-box
methods completely or at least mostly ignore the underlying physical effects and instead use
(large) databases to model the input-output behaviour of the system. Prominent are
Regression Analysis [5] and Artificial Neural Network (ANN) based methods [6, 7]. Grey-
box methods lie somewhere in between as they use some simplified representation
of the physical model to compute the input-output behaviour. This can, e.g., be done by
using transfer functions to compute deformations from the power loss of motors and other
components. Transfer functions (PT₁, PT₂, etc.) can for instance be used to approximate the
time-dependent thermo-elastic behaviour of simple structures [8]. Categorizing methods
into white, grey and black box methods is not unequivocal and it can be argued that all
models are grey box methods.

Regression Analysis, or more accurately Multiple Regression Analysis (MRA),
usually maps several temperature inputs directly onto the TCP displacement, based on
sample data derived from measurements or simulations [9]. The often-used least-squares
approach minimizes the sum of the squared error eₖ across all training data samples:

\[ \sum_{k=1}^{n_{\text{data}}} e^2_k = \sum_{n=1}^{\infty} (Y_k - f(X_{1,k}, X_{2,k}, ..., X_{d,k}))^2 \rightarrow \min! \]  

where: the \( X_{i,k} \) are the \( d \) inputs, \( Y_k \) the real output and \( f \) the desired regression model.

There are a large number of different regression models and most of them can
theoretically be used for thermal error compensation. Examples are characteristic diagrams
[10], B-Splines [11], Radial Basis Functions (RBFs) [12], wavelets [13] or even single
analytic functions (e.g. multivariate polynomial functions).

There are a few works comparing different compensation strategies. In [14] an MRA
and an ANN model are used to compute time-variant thermal error components using
varying spindle speeds and temperature history. In [15] characteristic diagram and structure
model based correction are compared using measurements on a parallel-kinematic
demonstrator. There is, however, no comprehensive examination and comparison of the
many different MRA methods within the context of thermal error compensation.

Chapter 2 introduces characteristic diagram based correction and particularly grid
based models and possible uses of regularisation. Chapter 3 describes B-splines and how
they are constructed. Chapter 4 explains RBFs, some important RBF types and their use for
MRA. Chapter 5 describes some useful wavelet types and shows how they can be used
similarly to RBFs. Finally, Chapter 6 compares the different MRA methods using thermo-
elastic FE simulations of a machine tool column. A summary and outlook on future research
concludes the paper.

2. CHARACTERISTIC DIAGRAM BASED CORRECTION

CHARACTERISTIC DIAGRAMS are a special branch of Multiple Regression
Analysis. They are usually limited to a rectangular grid formed across the domain of the
input variables. In the case of thermal error prediction, usually the minimum and maximum values for each temperature input are chosen and then these intervals are discretized, e.g. equidistantly, to form the grid. Next, a type of kernel function is chosen. Typically, one kernel function is located on each grid vertex and the kernel function has a local support, i.e. is only non-zero in the vicinity of this vertex. An example is the most often used multilinear kernel, which is one at the grid vertex and linearly declines to zero towards its neighbouring vertices, forming a hat shape in 1D and a four-sided pyramid in 2D. More complex types of kernels, such as B-splines, RBFs, wavelets, etc. can be used with little to no changes in the algorithm. With the grid and the kernels on each grid vertex, the rest of the method is quite simple. The approximation is simply the sum of all kernels, where each kernel is scaled by factor. The characteristic diagram is thus computed by calculating all of these scalar factors using a given set of data samples in a least-squares error minimization, see eq. (1). This ultimately results in a linear system which needs to be solved in order to obtain the kernel factors. For this, the linear system must have a rank equal to the number of kernel factors (usually one factor per grid vertex) and therefore a very large number of data samples are needed. Since for most applications, such large datasets are not available, regularization is needed to fill up the rank of the linear system.

The Smoothed Grid Regression (SGR), which is a special type of characteristic diagram based approximation, uses a balance of data fitting and smoothing to create overdetermined linear systems [16]. The assumption of smoothness of the approximated function is used to create additional rows in the linear system.

\[
A^* x = \begin{bmatrix} A \\ S \end{bmatrix} x = \begin{bmatrix} b \\ c \end{bmatrix} = : b^* \tag{2}
\]

where \(x\) are the kernel factors and the initially underdetermined linear system \(Ax = b\) is expanded by adding smoothing equations \(Sx = c\) in order to obtain the overdetermined linear system \(A^* x = b^*\) which is solved by left-multiplying with the transposed matrix:

\[
A^{*T} A^* x = A^{*T} b^* \tag{3}
\]

To illustrate this a bit better, a simple linear kernel in 1D is chosen, where 1D means one input (and one output) variable. Then the data equation for a data point \((t_i, y_i)\) located between the grid vertices \(g_{t,j}\) and \(g_{t,j+1}\) is:

\[
\frac{(t_i - g_{t,j})}{(g_{t,j+1} - g_{t,j})} x_j + \frac{(g_{t,j+1} - t_i)}{(g_{t,j+1} - g_{t,j})} x_{j+1} = y_i. \tag{4}
\]

A simple constant smoothing equation could be:

\[
x_j - x_{j+1} = 0. \tag{5}
\]

A simple linear smoothing equation could be:

\[
x_{j-1} - 2x_j + x_{j+1} = 0. \tag{6}
\]
Higher dimensions, i.e. more input variables, can be obtained by simply multiplying the one-dimensional kernels. The smoothing equations can remain the same and simply need to be repeated for each dimension.

The application of the SGR for machine tool thermal error compensation has been investigated in [10] and demonstrated control-integrated in [15].

3. B-SPLINES

B-SPLINES are piecewise polynomial curves [17]. Let \( T := \{ t_i \mid t_0 \leq \ldots \leq t_n \} \) be a set of nodes. Then \( s(t) \) is a spline of degree \( m \) if it is:
- \( (m - r) \) times differentiable on each \( r \)-fold node AND
- a polynomial of degree \( \leq m \) on \([t_i, t_{i+1}]\),
where \( t_{i+1} \) is \( r \)-fold node if \( t_i \leq t_{i+1} = \ldots = t_{i+r} \leq t_{i+r+1} \). Such splines \( s(t) \) with control points \( c_i \) can be expressed as a linear combination of basis functions:

\[
s(t) = \sum c_i \cdot N_{i,m}(t). \tag{7}\]

In 1967 Schoenberg entitled these basis functions \( N_{i,m}(t) \) B-splines [18]. The B-splines of degree \( m \) for all supports \( i = 0,1,\ldots,n \) are defined by the recursion:

\[
N_{i,0}(t) = \begin{cases} 
1 & t \in [t_i, t_{i+1}) \\
0 & \text{else} 
\end{cases} \tag{8a}
\]

\[
N_{i,k}(t) = \frac{t - t_i}{t_{i+k} - t_i} N_{i,k-1}(t) + \frac{t_{i+k+1} - t}{t_{i+k+1} - t_{i+1}} N_{i+1,k-1}(t) \tag{8b}
\]

where: the nodes \( \{ t_j \mid j = 0,\ldots,n+m+1 \} \) are defined by:

\[
t_j = \begin{cases} 
0 & j \leq m \\
 j - m & m + 1 \leq j \leq n \\
n - m + 1 & j > n.
\end{cases}
\]

The number of supports \( (n+1) \) must always be greater than the degree \( m \) of the B-spline. Due to the existence of \( r \)-fold nodes, there may be divisions by zero in (8b), in which case the respective terms are set to zero. The B-Splines of order one through four with six supports are shown in Fig. 1.

For data fitting with B-splines, a set of data points \( \{(x_i, y_i), i = 0,\ldots,n_p\} \) is needed, where the number of points should be large enough to cover all sections of the B-spline and should also be larger, the higher the order of the B-splines. Usually the order two or three is chosen.

The computation of the B-spline coordinates in 2D is as follows:

\[
x(t) = x_0B_{0,m}(t) + x_1B_{1,m}(t) + \cdots + x_nB_{n,m}(t) \quad t \in [0, n - m + 1] \tag{9a}
\]

\[
y(t) = y_0B_{0,m}(t) + y_1B_{1,m}(t) + \cdots + y_nB_{n,m}(t) \quad t \in [0, n - m + 1] \tag{9b}
\]
The 2nd order B-splines in Fig. 1 are the multilinear kernels of the SGR. Embedding B-splines into the SGR is thus relatively easy. Equation (9b) is the data fitting equation for any given data point and can be inserted into the linear system (2). The kernel factors are therefore the supports $y_0, y_1, \ldots, y_n$, which need to be computed by solving the linear system. The only thing left to do is to compute the parameter $t$ for each data point from equation (9a), where the coefficients $x_0, x_1, \ldots, x_n$ correspond to the grid vertices. Higher dimensions can be obtained through the tensor product of B-splines. In 3D, equations (9a), (9b) thus become the following:

$$x(t') = x_0 B_{0,m'}(t') + x_1 B_{1,m'}(t') + \cdots + x_n B_{n,m'}(t') \quad t' \in [0, n' - m' + 1]$$

$$y(t'') = y_0 B_{0,m''}(t'') + y_1 B_{1,m''}(t'') + \cdots + y_n B_{n,m''}(t'') \quad t'' \in [0, n'' - m'' + 1]$$

$$z(t', t'') = \sum_{i=1}^{n'} \sum_{j=1}^{n''} z_{i,j} B_{i,m'}(t') B_{j,m''}(t'') \quad t', t'' as above.$$
4. RADIAL BASIS FUNCTIONS

RADIAL BASIS FUNCTIONS are radially symmetric, real functions, whose value depends on the distance to a fixed central point [19]. They do not generally have a locally compact support and do not even always decrease with increasing distance from their center. Some important RBFs are:

- Gaussian RBF (GS): \( f(r) = e^{-(ar)^2} \)
- Multi-quadratic RBF (MQ): \( f(r) = \sqrt{1 + (ar)^2} \)
- Inverse Multi-quadratic RBF (IMQ): \( f(r) = \frac{1}{\sqrt{1+(ar)^2}} \)
- Inverse Quadratic RBF (IQ): \( f(r) = \frac{1}{1+(ar)^2} \)

The variable name \( r \) is used because the function value depends on the radial distance from the center (here 0). The parameter \( a \) determines the “width” of the RBF. By using a (large) set of RBFs with different centers and sometimes even different widths, most smooth, continuous functions can be approximated quite well. Figure 2 shows the aforementioned RBF types with different parameters.

![Gaussian RBF](image1)

![Multi-Quadratic RBF](image2)

![Inverse Multi-Quadratic RBF](image3)

![Inverse Quadratic RBF](image4)

**Fig. 2.** Four radial basis function types with different width \( a \)
Data fitting with RBFs is usually also done using the least-squares method, see eq. (1). Since RBFs have no compact support, all RBFs need to be evaluated for each data point. Therefore, the resulting linear system will no longer be sparse and thus require more memory and computation time to solve. However, the biggest challenge is in finding a good set of RBFs, i.e. answering the following questions:

- How many RBFs are needed?
- Where should they be located?
- What width should each RBF be given?

Fewer RBFs means a smaller linear system and thus faster computation. More RBFs means finer control over the function shape but also potentially a risk of overfitting. Therefore a compromise is needed. A simple if not very efficient way to solve this dilemma, is to fill the entire domain with RBFs and then remove RBFs until only the truly necessary ones remain.

There are two ways to take the RBFs to higher dimensions. The first is to simply extend the radial symmetry in all dimensions. In 2D, where the input values would normally be $x, y$, one uses the Euclidean distance to the RBF center instead:

$$z(x, y) = z(r) = z(\sqrt{x^2 + y^2}).$$  \hspace{1cm} (11)

The other option is to use a tensor product again, as was done with the B-splines. This approach has the advantage of being able to give the kernels (RBFs) a different width for each dimension. This is particularly useful if the input variables have vastly different domain sizes or even different physical units. As a practical example one temperature sensor near a particularly strong heat source might range from 10° to 70° whereas another might only range from 10° to 40°.

The issue of regularization is theoretically not as relevant for RBFs as it is for B-splines because they do not have a locally compact support. However, since most RBF types decrease very quickly with increasing distance to the RBF center, the linear systems will often be very nearly singular. This can be avoided through careful placement and width adjustment of the RBFs. Alternatively, the RBFs can also be arranged in a rectangular grid formation again, so that smoothing equations can be added, as is done in the SGR, see eq. (5), (6).

5. WAVELETS

WAVELETS are wavelike functions that are primarily used in the wavelet transformation [20]. They are generally not symmetrical and often possess a locally compact support in both the time and frequency domain. Additionally, most have an integral value of zero. They can be used for data fitting in nearly the same way as the RBFs in the previous chapter, except that higher dimensions can now only be achieved using the tensor product. Frequently used wavelets are, e.g.:
- **Haar Wavelet**
  \[ \psi(x) = \begin{cases} 
  1 & 0 < x < 0.5 \\
  -1 & 0.5 \leq x < 1 \\
  0 & \text{else} 
\end{cases} \]

- **Morlet Wavelet**
  \[ \psi(x) = \cos(bx) \cdot e^{-1/ax^2} \]

- **Mexican Hat Wavelet**
  \[ \psi(x) = (1 - x^2) \cdot e^{-1/ax^2} \]

- **Gaussian Wavelet**
  \[ \psi(x) = e^{-x^2} \]

In addition to the parameters mentioned here, wavelets can always be adapted through translation \( t \) and scaling \( s \):

\[ \psi_{s,t}(x) = \psi(s \cdot x - t). \quad (12) \]

Figure 3 shows these sample wavelets with some different parameter settings. While wavelet based data fitting is easier to compute because the local compactness of these kernels creates sparse linear systems, their wave nature can be problematic due to overfitting and reduce the overall global smoothness of the approximation.

**Fig. 3.** Four wavelet types with different parameters
6. TEST AND COMPARISON OF KERNEL TYPES

THE TEST AND COMPARISON was performed on validated FEM simulations of the demonstration machine tool Auerbach ACW630, see Fig. 4. The model is presently limited to the machine tool column. It was modelled with two heat sources: the motor of the y-axis at the top and the motor of the x-axis on the machine bed. Since the motors themselves are not part of the FE model, the heat was induced via conduction into the motor flange at the top and via the drive spindle at the bottom of the column. There is also free convection with a constant ambient temperature of 20°C. The column itself is made of grey cast iron and is hollow on the inside with several orifices for faster ambient cooling.

![Fig. 4. CAD model of the Auerbach ACW630](image)

The temperature sensor locations were chosen using an FE based sensitivity analysis [21]. For this test, the four best sensor locations were used, three of which are located at the top near the motor and one at the bottom near the X-spindle. The input data for these four (virtual) sensors as calculated by the FE simulations is shown in Fig. 5.

![Fig. 5. Temperatures of thermo-elastic simulation (optimized locations)](image)

The TCP displacement is largest in the X-direction and slightly smaller in the Y-direction, whereas the Z-direction has almost no error. This is because the waste heat from the Y-motor (top) is significantly larger than that from the X-motor (bottom) and
the Y-motor is attached asymmetrically causing the column to tilt to the side. The displacement of the dominant X-direction is shown in Fig. 6 along with the approximations from selected kernel types.

For RBFs and wavelets, representatives were chosen, although except for the Haar wavelet, all of the above-mentioned kernel types created similarly good approximations. A brief summary of this can be seen in Table 1. For these approximations, 20% of the available simulation data were randomly chosen for model training and then the entire dataset was used for model testing. There is no particular reason to use exactly 20% for training, except that using too much data for training will make it too easy to approximate the test data and using too little will make it too difficult. Either would be bad since the purpose of the investigation is to see whether some methods clearly outperform others or if some are poorly suited for the task. Tests with 33% training data showed very similar results, whereas tests with only 10% roughly doubled the approximation errors, with the RBF methods showing a slightly lower accuracy and the SGR a slightly higher accuracy than the rest.

<table>
<thead>
<tr>
<th>Kernel Type</th>
<th>Min. Residual [µm]</th>
<th>Max. Residual [µm]</th>
<th>RMSE [µm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multilinear SGR</td>
<td>–4.1339</td>
<td>4.2438</td>
<td>1.1507</td>
</tr>
<tr>
<td>3rd order B-Spline</td>
<td>–5.3468</td>
<td>3.0032</td>
<td>1.6122</td>
</tr>
<tr>
<td>4th order B-Spline</td>
<td>–7.8924</td>
<td>3.5978</td>
<td>2.6073</td>
</tr>
<tr>
<td>MQ RBF</td>
<td>–2.3169</td>
<td>6.0562</td>
<td>0.35929</td>
</tr>
<tr>
<td>IMQ RBF</td>
<td>–4.3531</td>
<td>4.829</td>
<td>1.3067</td>
</tr>
<tr>
<td>IQ RBF</td>
<td>–3.9266</td>
<td>3.3931</td>
<td>0.93774</td>
</tr>
<tr>
<td>Mexican Hat Wavelet</td>
<td>–4.3025</td>
<td>4.7964</td>
<td>1.2187</td>
</tr>
<tr>
<td>Gaussian Wavelet</td>
<td>–3.7155</td>
<td>3.6567</td>
<td>0.90733</td>
</tr>
<tr>
<td>Morlet Wavelet</td>
<td>–4.0042</td>
<td>3.6399</td>
<td>0.93619</td>
</tr>
<tr>
<td>Haar Wavelet</td>
<td>–190.9276</td>
<td>10.4849</td>
<td>531.1558</td>
</tr>
<tr>
<td>MLR</td>
<td>–12.0308</td>
<td>10.8224</td>
<td>16.7502</td>
</tr>
</tbody>
</table>
Such minor differences may, e.g., be due to the specific data, the testing procedure or the grid and can therefore be ignored. Both Fig. 6 and table 1 only show the results of the test data approximation. For all kernel functions, including the RBFs and wavelets, the same $5 \times 5 \times 5 \times 5$ grid was used. This refers to an equidistant grid with five support points for each input variable. Linear smoothing (see eq. (6)) was applied for regularisation on all tests. Tests without adding regularisation led to very poor approximations and in some cases to errors due to singularity. As mentioned before, both RBFs and Wavelets can be used without grid and regularisation but this requires an optimized placement and scaling of the kernels, which was not tested here. For reference, the multiple linear regression (MLR) solution was added and while it more or less reproduces the general shape of the deformation curve, it only achieves a much lower accuracy.

The main conclusion is that for this dataset all of the above-mentioned kernel types have worked equally well for MRA based thermal error compensation, so long as they are properly handled. The use of grid structures and regularization have proven very successful, even in the cases (RBF, wavelets) where they are not technically necessary. These results are expected to translate to most other MRA based thermal error approximations. Therefore, when approaching a new such approximation problem with MRA, it should not matter which type of kernel is chosen. Likewise, when the MRA seems to fail the task, the blame should first be focussed on bad or insufficient data, wrong usage of the kernel types, unsuitable grid structures, etc. rather than the kernel type itself.

If so many different kernel types perform similarly well, the question remains which kernel type should be recommended. W.r.t. the implementation effort, there is also little difference between them although higher order B-splines may be slightly more cumbersome. Computation time is also not relevant. Characteristic diagram based thermal error compensation is extremely fast and pretty much always online-capable, no matter the kernel type. The data fitting is usually time consuming but that is because a large linear system needs to be solved, which is also the case for each kernel type. The authors use multilinear kernels because they are very easy to implement, they are relatively robust w.r.t. overfitting and they make it relatively easy to control the extrapolation behaviour of the characteristic diagram [22].

7. SUMMARY AND OUTLOOK

One way of dealing with thermal errors in machine tools, is by using data-driven black-box approximation algorithms which map inputs, e.g. from temperature sensors, directly onto the TCP displacement. This can be accomplished using Multiple Regression Analysis. Most types of MRA use repetitions of some kernel functions to create approximations over a given domain. Popular kernel function types include B-splines, Radial Basis Functions and wavelets. While B-splines require a fixed grid, RBFs and wavelets can be freely placed within the domain. In many cases, data fitting requires not only a large set of training data but also additional regularization to solve the resulting linear system and obtain good results.
Using simulation data (thermo-elastic FE simulations) from a machine tool column of a 3-axis horizontal milling machine, a large number of kernel functions were tested and compared with respect to their ability to approximate the thermal error. As long as proper regularization was used, almost all of the tested kernel types (except the Haar wavelet) created similarly good results with a root mean square error (RMSE) of around 1 µm.

Further research will be directed towards the use of Regression Analysis on complex 5-axis machining centers under realistic production conditions. These generally require much larger training datasets and more complex approximation models.

This paper has given a basic overview over the available types of kernel functions for MRA and some pointers on how to best employ them for data fitting. While the final test and comparison was only exemplary, it suggests that the attainable approximation quality is mostly independent of the type of kernel function. A validation of these results using a much larger dataset, preferable using measured data from a milling machine is planned for future investigations.

**ACKNOWLEDGEMENTS**

This research was funded by a German Research Foundation (DFG) grant within the Collaborative Research Centers (CRC) /Transregio (TR) 96, Project ID 174223256 – TRR96, which is gratefully acknowledged.

**REFERENCES**