International Journal of Applied Earth Observations and Geoinformation 112 (2022) 102894

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

LR B-splines to approximate bathymetry datasets: An improved statistical criterion to judge the goodness of fit

Vibeke Skytt [a,*], Gaël Kermarrec [b], Tor Dokken [a]

[a] SINTEF, P.O. Box 124 Blindern, 0314 Oslo, Norway
[b] Institute for Meteorology and Climatology in Leibniz Universität Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany

ARTICLE INFO

Keywords:
LR B-splines
Information criteria
AIC
Surface fitting
Optimality
Bathymetry dataset
T-distribution
Full span refinement strategy

ABSTRACT

The task of representing remotely sensed scattered point clouds with mathematical surfaces is ubiquitous to reduce a high number of observations to a compact description with as few coefficients as possible. To reach that goal, locally refined B-splines provide a simple framework to perform surface approximation by allowing an iterative local refinement. Different setups exist (bidegree of the splines, tolerance, refinement strategies) and the choice is often made heuristically, depending on the applications and observations at hand. In this article, we introduce a statistical information criterion based on the t-distribution to judge the goodness of fit of the surface approximation for remote sensing data with outliers. We use a real bathymetry dataset and illustrate how concepts from model selection can be used to select the most adequate refinement strategy of the LR B-splines.

1. Introduction

Approximating observations with mathematical surfaces allow reducing millions of points to a comparably compact representation. Prominent applications include the processing of point clouds from terrestrial laser scanners in geodesy (Kermarrec et al. 2020), of bathymetry observations (Skytt et al. 2017), elevation data (Mitasova et al. 2005), or of point clouds from turbine blade (Bracco et al. 2020).

Most software packages use interpolating methods of irregularly spaced data into a regularly spaced grid (raster representation), which is unfavourable in case of noisy observations. Approximation strategies using Radial Basis Function (RBF, Skala et al. 2020) necessitate often to solve huge linear systems of equations and may lead to several local optima for the parameter estimations. Thin plate splines are a variant of RBF, and used for interpolation (see Keller and Borkowski 2019 for an application to earth gravity modelling). The method can be extended for approximation problems (Sprengel et al. 1996) but may be computational demanding so that a reduction of the point clouds is mandatory (Majdisova and Skala 2016). Another method uses Non-uniform rational B-splines (NURBS): this provides an intuitive and tractable scheme for solving a mathematical point cloud approximation and is widely used in computer graphics and geometric modelling (Piegl and Tiller 1997). Unfortunately, the tensor product-based structure yields many superfluous parameters to fit the topological requirement (Li et al. 2016) so that the approximated surface may oscillate in case of scattered and noisy point clouds with missing data and curvature changes (Bracco et al. 2018).

These issues can be addressed by locally refining the spline space. To that end, different approaches have been proposed including Hierarchical B-splines (Forsey and Bartels 1995), T-splines (Sederberg et al. 2003), Truncated Hierarchical B-splines (Giannelli et al. 2012), and locally refined (LR) B-splines (Dokken et al. 2013). LR B-splines were shown to provide well-behaved mathematical surfaces for remote sensing applications: they are the focus of the present contribution, following investigations of Skytt and Dokken (2022) on various point clouds. A comparison between different methods is not straightforward - see Stangeby and Dokken (2021) for an intent -, and will depend on some chosen criteria (numerical stability, goodness of fit, number of degrees of freedom, computational time, abbreviated as CT). Such a comparison is outside the scope of this paper.

Within the context of adaptive fitting, a refinement is performed when the error between the noisy point cloud and the approximated surface exceeds a predefined tolerance at a given iteration step of the algorithm. The final LR B-spline surface depends on this tolerance, on the bidegree of the spline space, but also heavily on the method used for refining it. Each approximation setup -called “model” in the following- has its own strengths and weaknesses, may it be the CT needed to reach a given accuracy, the number of coefficients to estimate, the maximum

* Corresponding author.
E-mail addresses: Vibeke.Skytt@sintef.no (V. Skytt), kermarrec@meteo.uni-hannover.de (G. Kermarrec), Tor.Dokken@sintef.no (T. Dokken).

https://doi.org/10.1016/j.jag.2022.102894
Received 2 March 2022; Received in revised form 13 May 2022; Accepted 24 June 2022
Available online 30 June 2022

error, or the number of points outside tolerance. Usual criteria to judge the goodness of fit are the root mean square error (rmse) or the mean average error (mae) between the mathematical approximation and the point cloud. Unfortunately, these performance indicators are not optimal as they depend strongly on the number of observations. Additionally, small values may be linked to a fitting of the point cloud's noise, which should be avoided if possible. We identify a need for a new criterion that would balance the number of estimated coefficients of the surface with the rmse of the residuals of the approximation. Such a quantity would avoid overfitting. In this contribution, we introduce the penalized model selection criterion called Akaike Information Criterion to solve that challenge (AIC, Akaike 1973; Burnham and Anderson 2002). Unfortunately, the major assumption for the derivation of the basic AIC is the normal distribution of the residuals of the approximation. This is clearly unrealistic when dealing with real observations: outliers will be unavoidably present (Skytt et al. 2017; Skytt and Dokken 2022). In such cases, the t-distribution is more favourable than the normal one. For that reason, we propose a "t-distribution based AIC" and show its potential to select an appropriate mathematical model to approximate remotely sensed point clouds. To illustrate the new methodology, we focus on the choice of two variants of the Full span refinement strategy within the framework of LR B-splines. We will make use of a bathymetry dataset for a real application and combine the output of the different criteria, including the newly developed AIC to analyse the result of the surface approximation. The same methodology can be used to select the optimal tolerance or the bidegree of the splines.

The reminder of our contribution is as follows: in section 2, we briefly review the concept of surface refinement with LR B-splines as well as the principle of model selection with AIC. In section 3, we approximate a real bathymetry point cloud and illustrate how to interpret the minimum of the proposed AIC in combination with more usual criteria. We conclude with some recommendations on the refinement strategy when using LR B-splines for surface approximation.

2. Methodology

In this section, we present the main concepts needed to understand the adaptive refinement with LR B-splines. The reader should refer to the dedicated contribution for the corresponding detailed derivations (Dokken et al. 2013).

2.1. Adaptive refinement with LR B-Splines

LR B-splines can be viewed as a generalization of univariate non-uniform B-splines. We assume that the reader has some knowledge about B-splines, see, e.g., Piegl and Tiller (1997). We focus on the bivariate case (2D), where knotline segments take over the role of knots from univariate B-splines, and knotline segments are assigned multiplicity.

- The starting point is a tensor product B-spline space. From this, we initiate the LR-mesh and a collection of tensor product B-splines spanning the spline space, from now on denoted the LR B-splines (see Fig. 1 for an example of a LR-mesh).
- The LR-mesh is successively refined by inserting new knotline segments in the initial LR-mesh.
- An LR B-spline is required to have minimal support meaning that the B-spline cannot be decomposed into two or more B-splines with a smaller support in the mesh. The support is the domain where the B-spline has a value different from zero.

2.2. Refinement strategies

The approximation of a parametrized scattered point cloud by an LR B-splines surface is done iteratively where a requested accuracy measure is not met. The surface is refitted at each iteration step, or level. When there exist some points of a mesh cell with a distance to the surface larger than a given tolerance, at least one B-spline having a support that overlaps this element must be split. This action increases the number of degrees of freedoms available for the surface approximation, i.e., the number of coefficients, and is called a "local refinement". We will focus on two related strategies of refinement of LR B-spline surfaces for scattered point cloud following Skytt et al. (2022). They are called the Full span strategies. Here elements are split either in four (i.e., both directions) halving the element width and height, or in two (i.e., one direction). These strategies are called FB and FA, respectively. In this contribution, we propose to judge if FA is more appropriate than FB within the context of surface approximation with LR B-splines. The potential number of new coefficients at each iteration step is much fewer for FA than for FB. Thus, for FA, more iterations are expected to reach an acceptable accuracy. In Fig. 1, we have highlighted these differences by drawing meshes obtained from the two methods and focusing on domains where the number of inserted lines slightly differs for FA and FB.

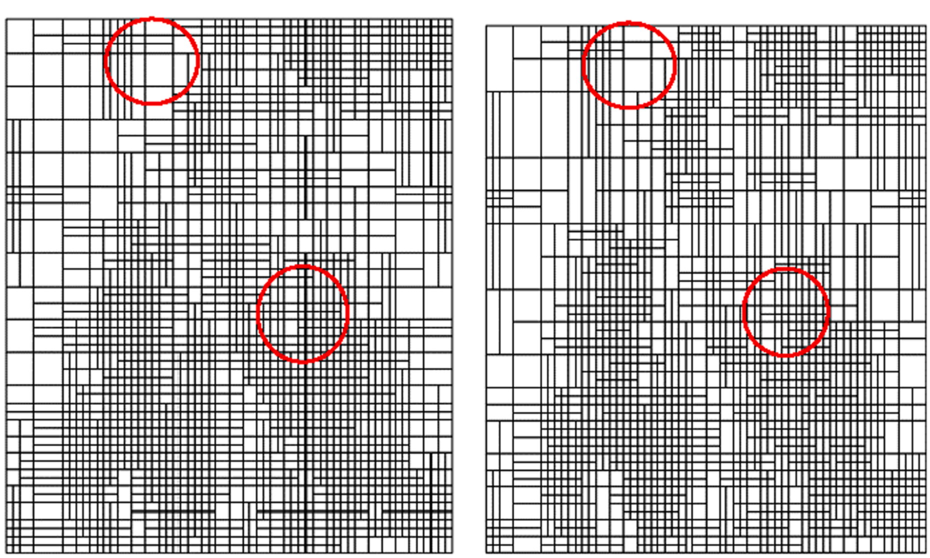

Fig. 1. Meshes for the two refinement strategies: left FB and right FA. The meshes relate to the approximations in Section 3 after 3 and 6 iterations, respectively. The red circles focus on two domains of interest where differences are visible (right).

2.3. Addressing linear dependence for Full span strategies

As for the general formulation of LR B-splines, linear dependence can occur also in the case of biquadratic Full span strategies. However, the refinement configuration resulting in linear dependence in this case is very special, we have not come across linear dependency using the biquadratic Full span strategies. However, to handle linear dependency should it occur, the following procedure can be followed:

1. As meshlines are inserted incrementally, the maintenance of linear independence can be immediately confirmed after each meshline insertion by using the hand-in-hand principle, see Dokken et al. (2013). This procedure checks that the increase in the number of tensor product B-splines matches the increase in the dimension of the spline space spanned. If this is not the case a T-joint related to the last meshline inserted triggered this linear dependence.
2. Using the Peeling algorithm from the paper above we can identify a collection of candidate tensor product B-splines for the linear dependency relation, and candidate regions to which they belong. As linear dependence is detected, at least one region that intersects the last meshline inserted have at least 7 tensor product B-splines nested inside (and 5 T-joints inside). Here we refer to Patrizi and Dokken (2020) for more details.
3. The linear dependency is eliminated by extending a meshline ending in a T-joint so that it crosses such a region. Candidate meshlines here are either (i) the last meshline inserted, (ii) another meshline ending in a T-joint inside the region(s) or (iii) a meshline ending in a T-joint outside the region that will split the region(s) if extended.
4. The algorithm returns to the situation before linear dependency occurred and check which alternative solves the situation best, if any. If the initial extension of the candidates doesn't solve the situation, further extensions of the candidate lines are made incrementally until a solution is found. In case that an extension of the meshline that triggered the linear dependence at each end connects to the boundary of the mesh, linear dependence will be guaranteed to be resolved. However, alternative and better candidate extension will in most case be found well before this occurs, i.e., extensions that meet a preexisting T-joint and eliminate this. As the Full span strategies split the width and height of elements in two, there is a limited number of possible constant parameter values of meshlines at each refinement level. Consecutively, there is a good chance that that the extension can join to an existing T-joint.

2.4. Iterative approximation

The mathematical surface approximation of some points is performed iteratively. Here we define the surface as with the collection of tensor-products B-splines spanning the spline space of the LR B-spline surface . are the coefficients corresponding to the B-spline . The scaling factors , ensure that the collection of scaled tensor-products B-splines , form a partition of unity.

We define the observation vector of size where , and assume that is parameterized. The vector of the coefficients of the LR B-spline surface gives a rough approximation of the third coordinate. We here outline the workflow for the functional case and thus name the parameter values x and y.

(i) We compute the coefficients with least-squares approximation with a smoothing term (adaptation of Mehlum et al. (1997) to LR B-spline surfaces) for the first iteration levels. The smoothing term in the least-squares approximation ensures that the equation system built is non-singular as it involves also LR B-splines with no data points in their domain.

(ii) If the sizes of the supports of adjacent B-splines differ significantly, the data of the observation vector is very rough, i.e., many coefficients have to be estimated with the least-squares approximation. To face that challenge, the local iterative approximation method (Multilevel B-spline Approximation, known as MBA) can be used after a few iterations (Lee et al. 1997, Zhang et. al. 1998). Here an underdetermined equation system setup allows many solutions. With MBA, a solution is chosen that computes the coefficients of each LR B-spline of the residual surface from residual values and evaluation of LR B-splines at data points. Although less optimal in regular cases, this local method is very stable and has a low memory consumption.

The algorithm is summarized in Algorithm 1.

Each iteration level gives rise to one approximating surface. One of these approximations will be called in the following a "model". Furthermore, we call ε the vector of approximation error or residuals in the z-direction.

Algorithm 1 adaptive surface approximation with LR B-splines

```
Input: point cloud, max number of iterations, tolerance TH
Output: fitted surface, quality parameters
  Generate initial surface (coarse mesh)
  while there exist points I with > TH and max number of iterations not reached
  do
    refine the surface using refinement strategy
    compute the approximated surface (LS or MBA)
    compute quality parameter (residual, mae, points outside tolerance, error term)
  end
```

2.5. Judging the goodness of fit

Judging the goodness of fit of the surface approximation can be made at each iteration step with the following performance indicators:

- The rmse with respect to the approximated surface in the z-direction. This quantity has drawbacks: it does not consider the spatial pattern of the error term . Additionally, it is strongly influenced by the number of observations. A thinned version of a point cloud will give a smaller rmse regardless of the accuracy in each single point. We note that a small rmse is unfavourable if the noise is fitted. In this contribution, we use the mean absolute error (mae) as it is less sensitive to outliers than the rmse. This is favourable and fair when comparing with the proposed AIC with t-distribution, but still does not overcome the aforementioned challenges.
- The maximum error defined as the maximum value of in the z-direction. The number of points outside tolerance or the CT are additional indicators that can be combined to the rmse to judge the accuracy of the approximated surface.

An analysis of the goodness of fit can be made by combining these indicators. This is a rather heuristic way to proceed and these values do not give any indication about the risk of overfitting. To address this drawback, we introduce a new version of the information criterion called AIC, which we adapt to the context of surface approximation. In the following, we review the mathematical concepts behind model selection for the user convenience.

From an observed sample of data of size , we consider possible models called , each of them having a likelihood function specified by the parameter vector of length , which is here the number of estimated coefficients. Unlike a probability, a likelihood has no real meaning *per se*, but is only interpretable by comparing the likelihood of different models. It can be seen as a measure of the goodness of fit to the data (Burnham and Anderson 2002). It is convenient to work with the log-likelihood function for with the estimates , which is defined as . The AIC is a penalized IC defined as . When k models are compared with each other, the model with the smallest IC is chosen, as it minimizes the estimated information loss (Akaike 1973). From now on and for the sake of readability, we will skip the subscript k.

The likelihood function is often taken to be the Gaussian one, assuming the residuals of the surface approximation ε to be normally distributed. Unfortunately, this strong belief can lead to a biased AIC when violated. This compromises the correct and in-dubious determination of the AIC minimum and the choice of the most adequate model among a set of candidates. Points outside the tolerance coming in the surface approximation and/or outliers in the observations themselves are likely to arise in a real case scenario and will be found in the approximation error. The t-distribution (also called student's distribution) gives more probability to observations in the tails of the distribution than the standard normal distribution (McNeil 2006). Consequently, its maximum is lower than for the normal distribution and outliers "fall" in its heavy tails. This makes the t-distribution more appropriate to describe the statistical properties of a real data set, see, e.g., Kargoll et al. (1992). In this contribution, we make use of the t-distribution to compute the likelihood of the model, which is the first term of the AIC. The second term is a penalty term and accounts for the increase of complexity as it becomes larger as the number of coefficients to estimate increases. The proposed AIC using the t-distribution is a global performance indicator: it weights the number of coefficients with the likelihood of the model. It accounts for the specificity of the point clouds and the residuals of the surface approximation which are expected to contain outliers.

Here we consider the choice of the setup for surface approximation as a statistical model selection problem: our example focuses on the refinement strategies FA and FB described previously.

3. Bathymetry dataset

In this Section, we will demonstrate how the LR B-splines can be used to approximate a noisy bathymetry data set. We will further illustrate how the proposed AIC can be combined with the other performance indicators to judge the accuracy of the approximated surface. We will focus on the choice of the refinement strategy as presented in Section 2. The same methodology can be used to select the most adequate tolerance for refinement, or the bidegree of the spline space.

3.1. Description of the dataset

Here we analyse a mainly subsea point cloud from Søre Sunnmøre in Norway. The acquisition is performed with multibeam sonar. A broad acoustic pulse is sent out from a transmitter and the depth is calculated from the time the sound waves need to reflect off the seabed and return to the transceiver. The width of the simultaneous measurements depends on the sea depth. In shallow water, the width will narrow down resulting in a very dense point cloud; the ship must cross the area several times to provide a good coverage of the seabed.

The data set contains about 9 million points and covers an area of about 7 km^2. The depth varies from −132.31 m to 0.48 m. The point density and the consistency of the point cloud vary throughout the domain. We have selected a subset of this point cloud covering an area of 0.3 km^2 for a detailed study in Section 3.2; it contains about 100 000 points and the depth varies from −88 to −44.3 m. Fig. 2 (left) shows the location of the subset, depicted in Fig. 2 (right).

3.2. Approximation of the small subset of the point cloud

A tolerance of 0.5 m is selected for the approximation of the subset, which corresponds to approximately 2 times the noise level. An initial biquadratic tensor-product B-spline surface of 10 times 10 coefficients approximating the point cloud is computed (iteration 0). Then the iteration is allowed to run until all residuals are smaller than the tolerance. Strategy FB requires 14 and FA 24 iterations and the resulting surfaces have 7742 and 6432 coefficients, respectively. However, most iterations contribute little to an improved approximation accuracy and the risk of fitting noise is present as the iteration step increases. In order to avoid overfitting and do not increase unnecessary the CT, the algorithm should be stopped after an optimal number of iterations and not when all residuals are smaller than the tolerance. We wish to highlight how the cross-analysis of different evaluation criteria, including the proposed AIC with t-distribution, can answer this highly relevant question.

- *AIC:*

We compute the proposed AIC for FA and FB for each iteration step. Following Section 2.4, the minimum of the AIC is assumed to give a statistical indication about the most optimal model for each refinement strategy. This corresponds to the turning point, i.e., the iteration step for which continuing the refinement by increasing the degrees of freedom does no longer lead to a significant improvement of the approximation. The corresponding results are shown in Fig. 3 where the AIC is plotted against the iteration step. For each model, we search for its minimum. This latter is obtained only for the t-distribution after 3 and 6 iterations for the FB and FA strategy, respectively. Here the dotted line in Fig. 3 corresponds to the use of the normal distribution to compute the AIC and the bold line (red for FB and blue for FA) to the proposed t-distribution. This finding highlights the advantage of using the correct distribution to compute the AIC, and find a minimum, see section 2.4.

Results

In Table 1, we interpret the minimum of the AIC together with the usual performance indicators, such as the number of coefficients, the mae, and the maximum distance for a given iteration step. The computations are performed on a stationary desktop with 64 GB of DDR4-2666 RAM. It has an i9-9900 K CPU with 8 cores and 16 threads. A single core implementation is used in the approximation. The approximation functionality is implemented in C++. We focus on the optimal iterations with AIC (3 and 6) found in the previous section, and the iteration step from which the improvement in terms of average distance was shown to be small, i.e., under 0.5 mm (iteration 7 and 15 for FB and FA, respectively). We give additionally the results after the 6th and 14th

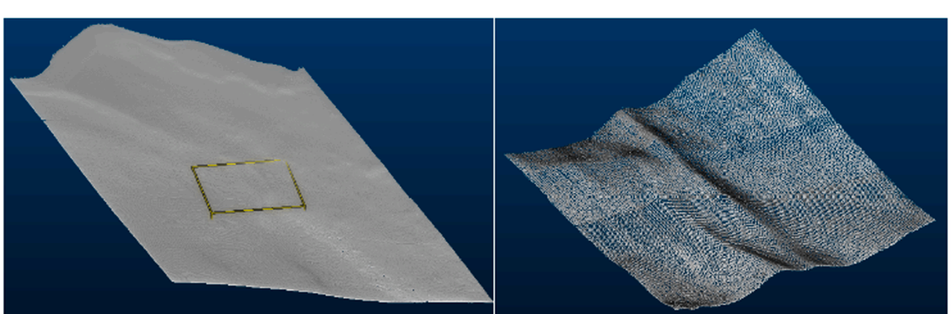

Fig. 2. Full point cloud with indication of the position of the subset of the point cloud (left). The subset is scaled with a factor of 10 in the height direction for the sake of visualization to emphasize the seabed features (right).

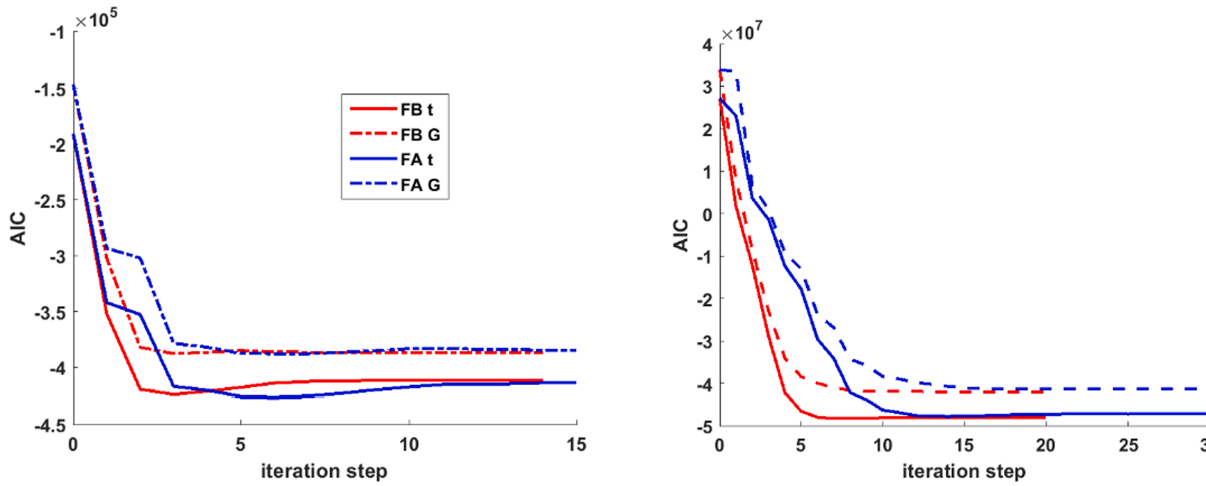

Fig. 3. Left: FA_AIC and FB_AIC with respect to the number of iterations for the two distributions (normal G: dotted line and t for the t-distribution: full line) for the small dataset. The first iteration step is labelled as 0. Right: same as left for the whole dataset.

Table 1
Results of the comparison for FA and FB strategies after a given number of iterations (subset).

Strategy	Iter.	Max. d.	mae	No. out	No. coefs	CT
FB	7	0.665	0.079	24	7 222	0 m0.99 s
FB	6	0.682	0.079	56	6 730	0 m0.89 s
FB_AIC	3	1.262	0.081	359	2 100	0 m0.41 s
FA	15	0.779	0.080	25	6 173	0 m1.45 s
FA	14	0.876	0.080	41	6 007	0 m1.36 s
FA_AIC	6	1.283	0.081	380	1 784	0 m0.62 s

iterations for the sake of completeness and comparison.

Table 1 highlights that the minimum found with AIC is corresponding to an optimal global approximation: it balances the likelihood with respect to the number of coefficients. The average distance (mae) minima for FA_AIC and FB_AIC are rapidly reached after 3 and 6 iterations and do not significantly decrease as the iterations increase. The outputs after 6 and 14 iterations are comparable to the one after 7 and 15 iterations. However, the maximum distance is lower in the latter case: this finding is linked with a local improvement of the surface fitting.

We note that the optimum number of coefficients is less for FA_AIC than for FB_AIC while the accuracy is quite comparable. Furthermore, the minimum of the AIC is reached for the FA strategy (see Fig. 3 left). This latter seems, thus, to be preferable for this point cloud from a statistical point of view. The same holds true after more iterations.

Unfortunately, the FA has a higher CT compared with the FB method. The difference is not too large in our example but may not be negligible for larger point clouds.

Increasing the number of coefficients is a decision that has to be balanced with respect (i) to the details that one wishes to capture as illustrated in Fig. 4 or (ii) the CT: the proposed AIC is, thus, a global indicator.

To illustrate the differences between the levels of refinement, we have plotted the different meshes obtained after 3 and 7 iterations with the FB strategy as well as the corresponding surface approximations, see Fig. 4. Here the optimal iteration level given by the AIC avoids noise fitting over to the price of fewer details in the approximated surface. This can be seen in the domain specified by a blue circle in Fig. 5. Further refinements of the surface have been performed after iteration level 6 and 3 for FA and FB, respectively. Having a look in details at another domain (Fig. 5 red circle), we see that the FA surface fits the point cloud tighter than the FB surface after a low level of iterations. In both cases the distances between the points and the surface are less than the tolerance. FA, which introduces new knot lines in a slower pace than FB and consequently has more information when these knot lines are defined, has been able to position them better. No significant refinement has taken place in this area after the iteration level given by AIC, as the surface is refined only in areas where there are unresolved points. Thus, continued iterations will not improve the accuracy in such areas.

3.3. *Approximation of the complete dataset*

The AIC computed for the whole dataset (9 106 045 points) has a minimum when the t-distribution is used *only*: this is an indication that an adequate distribution is favourable to judge the goodness of fit in a global sense. The minimum of the AIC exists but is weak for the FA strategy after 16 iterations, and for the FB strategy after 9 iterations. This finding is coherent with the results of section 3.2 regarding the differences between FA and FB. Due to the high number of points of the dataset with regard to the number of coefficients to estimate, the second term of the AIC has a low influence, and the AIC becomes dominated by the likelihood of the model, which distribution has to be chosen with care. Here the t-distribution is more appropriate due to the presence of outliers.

Table 2 presents some approximation results at these iteration levels. Despite the difference in magnitudes, most results are similar to those outlined in Table 1. The FB method has lower CT than FA, which is always an important factor when fitting huge point clouds. The surface generated with strategy FA has considerable fewer coefficients than the one from FB. More iterations typically lead to higher CT, but it can be outweighed by a decrease in time due to few coefficients.

We note an unnoticeable decrease in the overall average distance mae for both strategies as the number of iterations increases, which is linked with the weak AIC minimum. In relation to the size of the initial point cloud, the numbers of unresolved points are neglectable for both strategies and iteration levels. Due to the lower number of coefficients, FA would be preferred to FB in most circumstances. However, even if the AIC converges to nearly similar values for both strategies, there is a slight preference for FB for this point cloud. Consequently, we propose to combine the different criteria to judge the goodness of fit for approximating such point clouds; an improved IC based on the local stochasticity of the approximation needs to be specifically developed for such challenging surfaces with a huge number of points.

We have used a bathymetry dataset to illustrate the challenge in combining statistical measures to evaluate models approximating a point cloud with highly non-uniform properties. We showed that an improved AIC using the t-distribution to account for outliers provides an

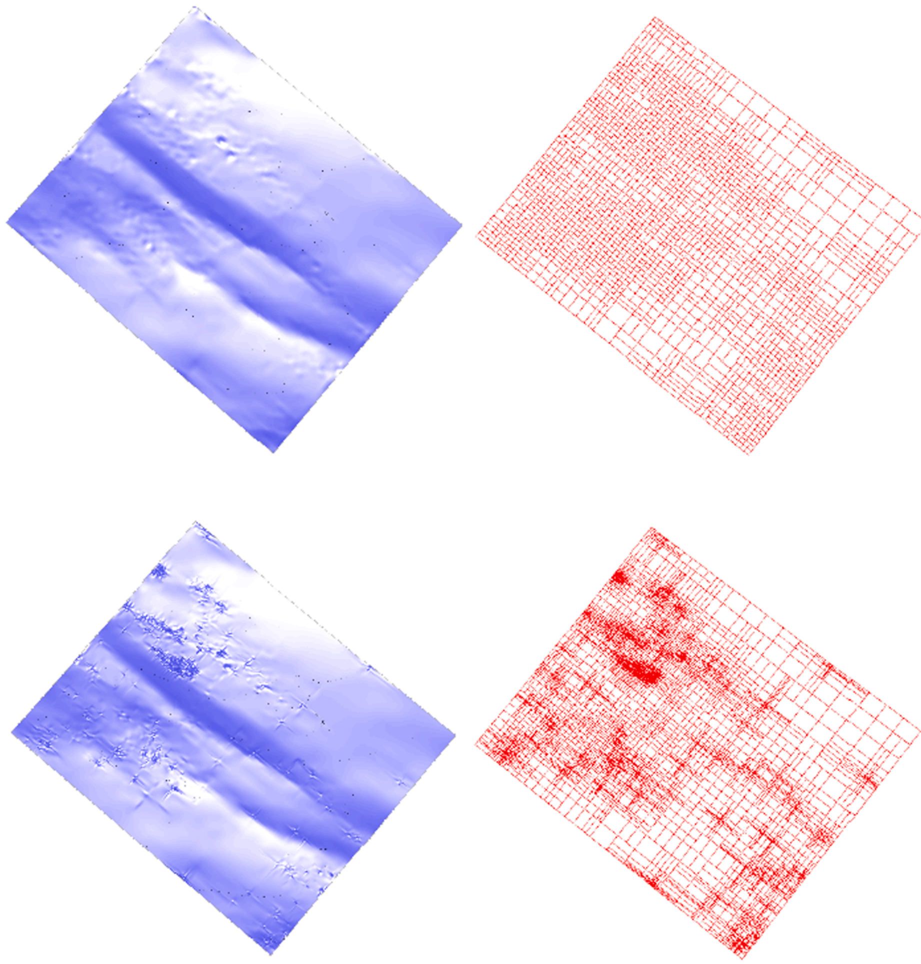

Fig. 4. Surface and the corresponding mesh structure created with strategy FB after 3 (top) and 7 iterations.

indication about the accuracy of the fitting for a given setup. Here we chose to compare the Full span refinement strategies FA and FB. Our fair comparison pointed out the challenges linked with the finding of optimality when an approximation of a large, noisy and scattered point cloud is performed: combining the output of different criteria is indicated.

4. Conclusions

An information criterion is a weighted measure of the quality of a complex statistical model and aims to answer the question how well a model fits some data compared to other models. We have applied it within the context of adaptive surface approximation with LR B-splines. Compared to other local methods for refinement of spline spaces, LR B-splines have the advantage of high flexibility in the definition of new meshlines.

Up to now, judging adaptive surface approximations is performed rather "heuristically" and the choice of the refinement strategy or the decision to stop the algorithm is rather empirical. This may lead to an overfitting, i.e., the fitting of the noise when too many coefficients have to be estimated.

In this contribution, we have introduced some statistical concept in the domain of surface fitting by using concepts from model selection. To that end, we have proposed an adapted version of the AIC to be used within the context of surface approximation of real point clouds. We showed that:

- The proposed AIC can be used it to identify the most adequate model regarding two refinement strategies. To illustrate the methodology, we have chosen Full span refinement strategies (FA and FB) for which the linear dependencies of B-splines can be easily addressed.
- Unavoidable outliers present in the approximation can be accounted for with the proposed AIC. This way, a minimum of the AIC can be found when increasing the iteration step to identify the most adequate model. To that end, we used the t-distribution to compute the likelihood of the model. This new development allows to avoid unnecessary iterations of the surface approximation algorithm and the associated and unfavourable overfitting of the point cloud.
- The model selection strategy has been applied to a real bathymetry data to judge the goodness of fit of the approximation. We computed the AIC as well as more usual criteria such as the mae, the maximum distance and the CT for two Full span refinement strategies. The AIC minimum was reached earlier than for the usual criteria such as the mae, the number of points outside tolerance or the maximum error. The FA strategy was identified as optimal for a small dataset, but this result should be balanced by its higher CT. For the bigger dataset, the FB strategy was weakly preferable.

The proposed AIC accounts for the specificity of the point clouds but remains a global indicator. It gives a first indication about the accuracy of the fitting for large point clouds and helps to avoid fitting of the noise for a given setup. It can be combined with other performance indicators, depending on how "optimality" is defined and with respect to the application at hand. In this contribution, we have applied this methodology and discussed two Full span refinement strategies. It should be

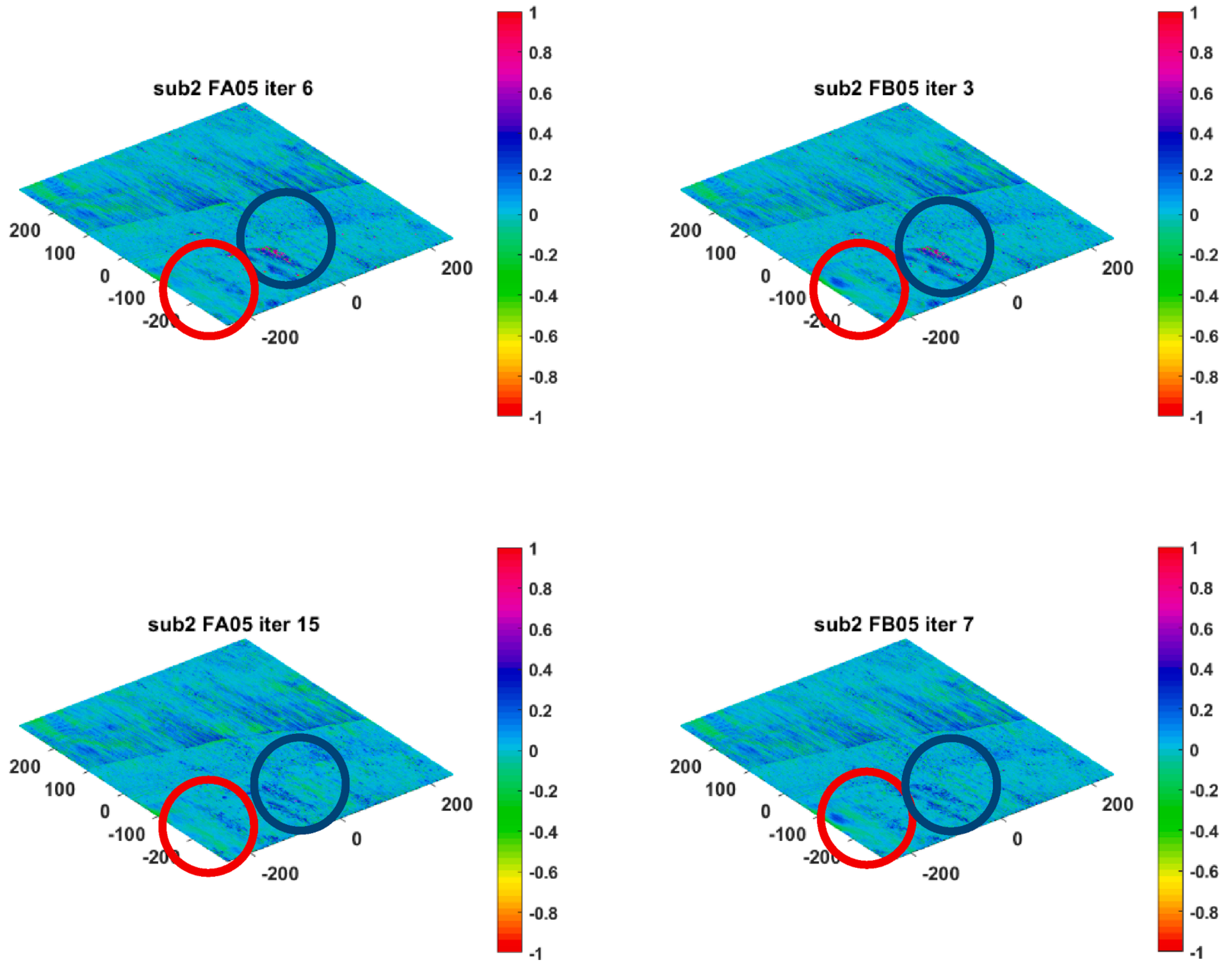

Fig. 5. Approximation error for the models at different iteration step. TH = 0.5. Left: FA, right: FB. Top: AIC optimal refinement step; Bottom: refinement step 7 and 15 for FA and FB, respectively as in Table 1.

Table 2
Results of the comparison for FA and FB strategies after a given number of iterations (complete data set).

Strategy	Iter	Max d.	mae	No. out	No. coefs	CT
FB	15	1.057	0.050	63	203 081	2m46s
FB	14	1.054	0.050	84	202 427	2m40s
FB_AIC	9	1.27	0.050	3 629	157 977	1m39s
FA	26	1.079	0.052	128	159 032	3m37s
FA	25	1.09	0.052	213	158 271	3m29s
FA_AIC	16	2.039	0.052	5 613	108 738	2m8s

noted that there exist refinement strategies such as N_2S_2, see Patrizi et al. (2020) that ensure local linear independence in all elements. These strategies run a process for each hierarchical refinement level where a significant amount of additional meshlines is inserted. This way, local linear independence that are not required for accuracy is maintained. The effect is that each degree of freedom will contribute less to improving accuracy than in the case of the FB or FA strategies. A future study could make use of the AIC within the context of surface approximation to study alternative refinement strategies to the Full span ones. We cite exemplarily structured mesh for LR B-splines, Truncated Hierarchal B-splines and different variants of T-splines.

Similarly, further investigations should be performed using the proposed AIC to determine the most optimal tolerance as well as the bidegree of the splines for larger and more challenging datasets. They would serve an improved understanding of optimality within the context of surface approximation of noisy and scattered point clouds with LR B-splines.

CRediT authorship contribution statement

Vibeke Skytt: Methodology, Software, Visualization, Writing -original draft, Writing -review & editing. **GaëlKermarrec:** Investigation, Methodology, Writing - original draft, Writing - review & editing. **Tor Dokken:** Formal analysis, Wrtiting - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gael Kermarrec reports financial support was provided by Deutsche Forschungsgemeinschaft.

Acknowledgements

This study is supported by the Deutsche Forschungsgemeinschaft under the project KE2453/2-1 and the Norwegian research council under grant number 270922. The data set is provided by the Norwegian map authorities, division Sjøkartverket.

Code availability

The source codes are available for downloading at the link: https://gi thub.com/SINTEF-Geometry/GoTools/wiki/Module-LRSplines2D.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Proceedings of the 2nd International Symposium on Information Theory. Akademinai Kiado, Budapest, pp. 267–281.

Bracco, C., Giannelli, C., Großmann, D., Sestini, A., 2018. Adaptive fitting with THB-splines: Error analysis and industrial applications. CAGD 62, 239–252.

Bracco, C., Giannelli, C., Großmann, D., Imperatore, S., Mokris, D., Sestini, A., 2020. THB-spline approximations for turbine blade design with local B-spline approximations. CoRR abs/2003.08706.

Burnham, K.P., Anderson, D.A., 2002. 2002: Model selection and multimodel inference. Springer, New York.

Dokken, T., Pettersen, K.F., Lyche, T., 2013. Polynomial splines over locally refined box-partitions. CAGD 30, 331–356.

Forsey, D.R., Bartels, R.H., 1995. Surface fitting with hierarchical splines. ACM Trans. Graph. 14, 134–161.

Giannelli, C., Jüttler, B., Speleers, H., 2012. THB-splines: The truncated basis for hierarchical splines. CAGD 29, 485–498.

Kargoll, B., Omidalizarandi, M., Loth, I., Paffenholz, J.A., Alkhatib, H., 1992. An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations. J. Geod. 3, 271–297.

Kermarrec, G., Kargoll, B., Alkhatib, H., 2020. Deformation Analysis Using B-Spline Surface with Correlated Terrestrial Laser Scanner Observations—A Bridge Under Load. Remote Sensing 12, 829.

Keller, W., Borkowski, A., 2019. Thin plate spline interpolation. J. Geod. 93, 1251–1269.

Lee, S., Wolberg, G., Shin, S.Y., 1997. Scattered data interpolation with multilevel B-splines. IEEE Trans. Vis. Comput. 3 (3), 228–244.

Li, X., Chen, F., Kang, H., Deng, J., 2016. A survey on the local refinable splines. Sci. China Math. 59, 617–644.

Majdisova, Z., Skala, V., 2016. A Radial Basis Function Approximation for Large Datasets. SIGRAD.

McNeil, A.J., 2006. Multivariate t-distributions and their applications. JASA 101 (473), 390–391.

Mehlum, E., Skytt, V., 1997. Surface Editing. In: Dæhlen, M., Tveito, A. (Eds.), Numerical Methods and Software Tools in Industrial Mathematics, Birkhäusser, Boston, pp. 381-396.

Mitasova, H., Mitas, L., Harmon, R.S., 2005. Simultaneous spline approximation and topographic analysis for lidar elevation data in open-source GIS. IEEE Geosci. Remote Sens. Lett. 2 (4), 375–379.

Patrizi, F., Dokken, 2020. Linear dependence of bivariate Minimal Support and Locally Refined B-splines over LR meshes, Computer Aided Geometric Design Volume 77, February 2020.

Patrizi, F., Manni, C., Pelosi, F., Speleers, H., 2020. Adaptive refinement with locally linearly independent LR B-splines: Theory and applications. Comput. Methods Appl. Mech. Eng. 369 https://doi.org/10.1016/j.cma.2020.113230.

Piegl, L., Tiller, W., 1997. The NURBS Book, Springer-Verlag, second ed.

Sederberg, T.W., Zheng, J., Bakenov, A., Nasri, A., 2003. T-splines and T-NURCCs. ACM Trans. Graph 22, 477–484.

Skala, V., Karim, S.A.A., Zabran, M., 2020. Radial Basis Function Approximation Optimal Shape Parameters Estimation. In: Krzhizhanovskaya, V., et al. (Eds.), Computational Science – ICCS 2020. ICCS 2020. Lecture Notes in Computer Science, vol. 12142. Springer, Cham.

Skytt, V., Harpham, Q., Dokken, T., Dahl, H.E.I., 2017. Deconfliction and Surface Generation from Bathymetry Data Using LR B-splines. In: Floater, M., Lyche, T., Mazure, M.L., Mørken, K., Schumaker, L. (Eds.) Mathematical Methods for Curves and Surfaces. MMCS 2016. Lecture Notes in Computer Science, vol. 10521. Springer, Cham.

Skytt, V., Dokken, T., 2022. Scattered data approximation by LR B-spline surfaces. A study on refinement strategies for efficient approximation. Geometric Challenges in Isogeometric Analysis, ed. C Manni, H. Speleers, Springer INdAM Series, Volume 49.

Sprengel, R., Rohr, K., Stiehl, H.S., 1996. Thin-plate spline approximation for image registration. Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 3, pp. 1190-1191.

Stangeby, I., Dokken, T., 2021, Properties of Spline Spaces Over Structured Hierarchical Box Partitions. In: van Brummelen, H., Vuik, C., Moller, M., Verhoosel, C., Simeon, B., Juttler, B. (Eds.), Isogeometric Analysis and Applications 2018, Springer, pp. 177-208.

Zhang, W., Tang, Z., Li, J., 1998. Adaptive Hierarchical B-spline Surface Approximation of Large-scale Scattered Data.