R-05-38

Digital elevation models for site investigation programme in Oskarshamn

Site description version 1.2

Lars Brydsten, Mårten Strömgren
Umeå University

June 2005

Svensk Kärnbränslehantering AB
Swedish Nuclear Fuel
and Waste Management Co
Box 5864
SE-102 40 Stockholm Sweden
Tel 08-459 84 00
+46 8 459 84 00
Fax 08-661 57 19
+46 8 661 57 19

ISSN 1402-3091
SKB Rapport R-05-38

Digital elevation models for site investigation programme in Oskarshamn

Site description version 1.2

Lars Brydsten, Mårten Strömgren
Umeå University

June 2005

Keywords: Digital elevation model, Topography, DEM, GIS, Oskarshamn, Surface ecosystem, Biosphere.

A pdf version of this document can be downloaded from www.skb.se

Abstract

In the Oskarshamn area, a digital elevation model has been produced using elevation data from many elevation sources on both land and sea. Many elevation model users are only interested in elevation models over land, so the model has been designed in three versions:

Version 1 describes land surface, lake water surface, and sea bottom.
Version 2 describes land surface, sediment levels at lake bottoms, and sea bottoms.
Version 3 describes land surface, sediment levels at lake bottoms, and sea surface.

In cases where the different sources of data were not in point form – such as existing elevation models of land or depth lines from nautical charts – they have been converted to point values using GIS software. Because data from some sources often overlaps with data from other sources, several tests were conducted to determine if both sources of data or only one source would be included in the dataset used for the interpolation procedure. The tests resulted in the decision to use only the source judged to be of highest quality for most areas with overlapping data sources. All data were combined into a database of approximately 3.3 million points unevenly spread over an area of about 800 km^2. The large number of data points made it difficult to construct the model with a single interpolation procedure, the area was divided into 28 sub-models that were processed one by one and finally merged together into one single model.

The software ArcGis 8.3 and its extension Geostatistical Analysis were used for the interpolation. The Ordinary Kriging method was used for interpolation. This method allows both a cross validation and a validation before the interpolation is conducted. Cross validation with different Kriging parameters were performed and the model with the most reasonable statistics was chosen. Finally, a validation with the most appropriate Kriging parameters was performed in order to verify that the model fit unmeasured localities. Since both the quality and the density of the data varied greatly over the model grids, the combined overall elevation model has significant variations in quality. The best quality comes from models over land and in shallower parts of the sea near Simpevarp, and the lowest quality comes from the deepest part of the sea.

The elevation model uses the grid projection RT 90 2.5 Gon W and the height system RH 70, and it has a cell size of 10 × 10 metres.

Sammanfattning

En digital höjdmodell över Oskarshamnsområdet har tagits fram med hjälp av punktdata för nivåer över både land och hav från ett stort antal olika datakällor. Flera användare av höjdmodellen är endast intresserade av höjdmodellen över land, varför modellen har konstruerats i tre versioner:

Version 1 beskriver landyta, sedimentytan för sjöar och havsbotten.
Version 2 beskriver landyta, vattenyta för sjöar och havsbotten.
Version 3 beskriver landyta, sedimentyta för sjöar och havsyta.

I de fall där de olika datakällorna inte var i punktform, t. ex. befintliga höjdmodeller över land eller djuplinjer i det digitala sjökortet, har de konverterats till punktform i ett GIS-program. Flera av datakällorna har data som överlappar utbredningen i andra datakällor varför flera tester utfördes för att avgöra om båda datakällornas data eller bara en av källorna skall ingå i det dataset som utgör ingångsdata till interpoleringsproceduren. Resultaten av testerna medförde att för de flesta områden med överlappande data användes endast den datakälla som bedömdes vara av högre kvalité. Datat slogs ihop till en databas med sammanlagt cirka 3,3 miljoner punkter ojämnt spridda över det cirka 800 km^2 stora området. Det stora antalet punkter i databasen medförde att det var svårt att konstruera modellen med en enda interpoleringsprocedur. Området delades därför in i 28 delområden där en modell konstruerades för varje delområde och därefter slogs alla delområden ihop till en stor höjdmodell.

Interpoleringen har utförts i programmet ArcGis 8.3 och dess extension Geostatistical Analyst. Som interpoleringsmetod valdes Ordinary Kriging. Metoden tillåter både en korsvalidering och en validering av höjdmodellen innan interpolering genomförs. Korsvalideringar med olika Krigingparametrar utfördes och modellen med den mest rimliga statistiken valdes. Slutligen utfördes en validering med de mest passande parametrarna för att verifiera att modellen passar även där det inte finns några mätpunkter. Eftersom både kvaliteten på data och datadensiteten varierar stort över modellarean har den sammanslagna stora höjdmodellen också stora variationer i kvalitet. Bästa kvaliteten i modellen återfinns över land och i grunda delar av havet nära Simpevarp, sämre kvalitet i den sydöstra och djupaste delen av havet.

Höjdmodellen har koordinatsystemet RT 90 2.5 Gon W och höjdsystemet RH 70 och har en cellstorlek om 10 × 10 meter.

Contents

1 Introduction

Referenced to a common datum, a regular array of z-values allows a digital elevation model (DEM) to represent a continuous variable over a two-dimensional surface. Typically, digital elevation models describe terrain relief.

The aim of this project is to improve an existing DEM /Brydsten, 2004/ buy using new elevation data and a different interpolation technique.

Many types of surface models – such as hydrological models and geomorphometrical models – use DEM as input data. DEM resolution is the size of DEM cells. DEM interpolates irregular spaced elevation data. In this model, we used the Kriging interpolation method, a geostatistical interpolation method based on statistical models that include autocorrelation (the statistical relationship among the measured points). Kriging weights the surrounding measured values to predict an unmeasured location. Weights are based on the distance between the measured points, the prediction locations, and the overall spatial arrangement among the measured points.

Normally, a DEM has a constant value for sea surface and a constant value for lake surfaces. For the Oskarshamn area, the DEMs have negative values in the sea to represent water depth, constant positive values for lake surfaces represent lake elevations and varying values represent lake bottom elevations.

Input data for the interpolation have many different sources, such as existing DEMs, elevation lines from digital topographical maps, digital nautical charts, and depth soundings in both lakes and the sea. All data are converted to point values using different techniques. The Kriging interpolation was performed in ArcGis 8 Geostatistical Analysis extension.

2 Methods

2.1 Data catch from land areas

Two sources were used to collect elevation point data for land: the existing DEM from the Swedish national land survey (LMV) with a resolution of 50 metres and the SKB DEM with a resolution of 10 metres /Wiklund, 2002/.

The existing DEMs were converted to point layers in shape-format using ArcToolbox in ArcGis 8. The 10-metre model values for lake surfaces have errors. In Lake Frisksjön, for points situated at least 25 metres from the shoreline the z-value has 41 unique values ranging from 2.01 to 2.56 m.a.s.l. Two values dominate these points – 2.31 in the western part of the lake and 2.09 in the eastern part. These areas are separated by a distinct straight north-south line that acts like a threshold in the lake surface at 0.22 metres. This threshold seems to intersect with the border between two adjacent flying transects.

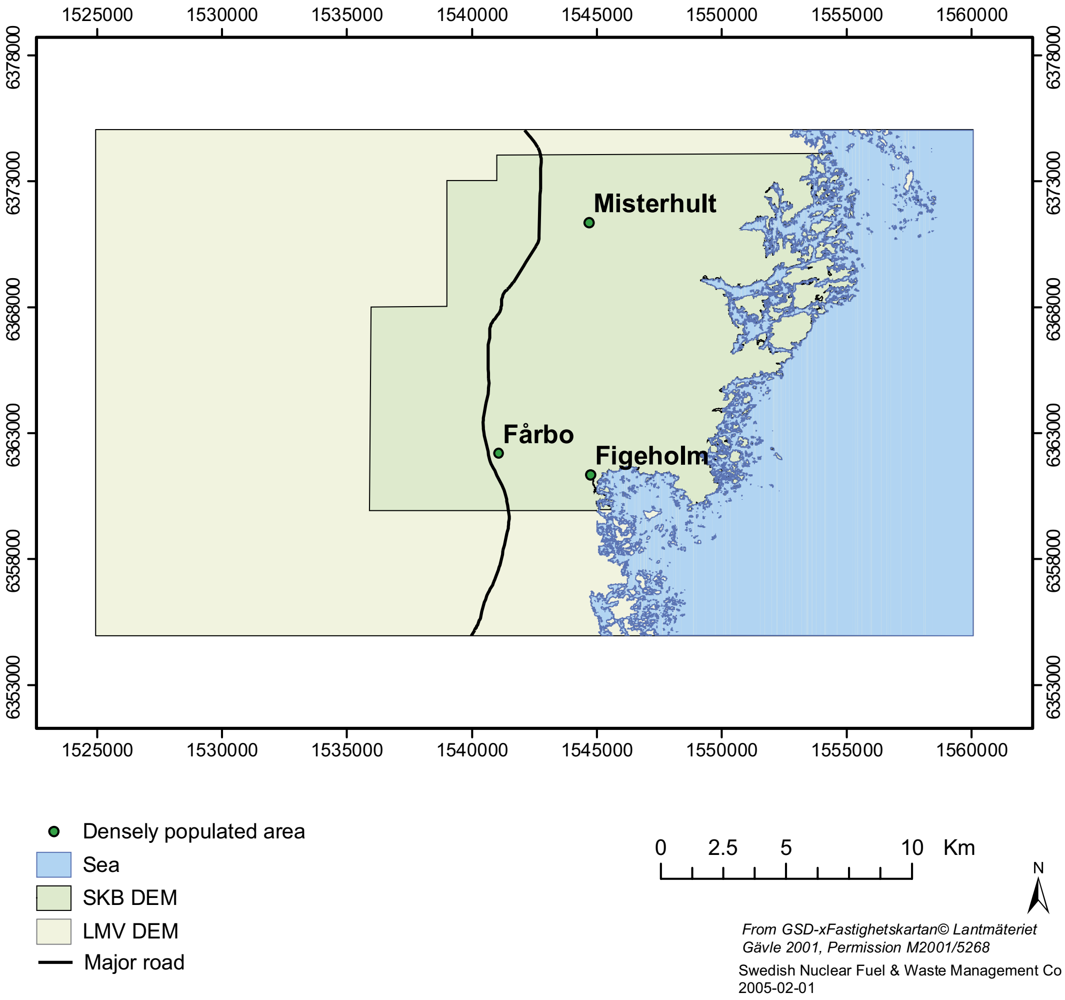

Figure 2-1. Extensions of the LMV DEM and SKB DEM.

The same phenomenon exists in most of the lakes within the 10-metre model extension. All points placed within lakes with levelling measurements (Table 2-1 and Figure 2-2) were replaced by the measured values. It should be noted that these levels are not the mean lake levels, but the levels at each measuring occasion. Because Lake Fjällgöl, in the centre of the map, has not been measured, the mean value for the elevation in the 10-metre model is used.

These two point-layers were merged into one single point layer, and all points placed on the sea surface polygon from the digital localities maps were deleted from the datasets. The final layer is in the Swedish national grid projection (RT 90 2.5 Gon W) and in the Swedish national height system 1970 (RH 70).

Table 2-1. Lake surface elevations for the 5 lakes shown in Figure 2-2 (metres above RH 70).

Lake	Elevation (ma RH 70)
Fjällgöl	21.29
Frisksjön	1.37
Jämsen	25.11
Plittorpsgöl	24.79
Söråmagasinet	2.07

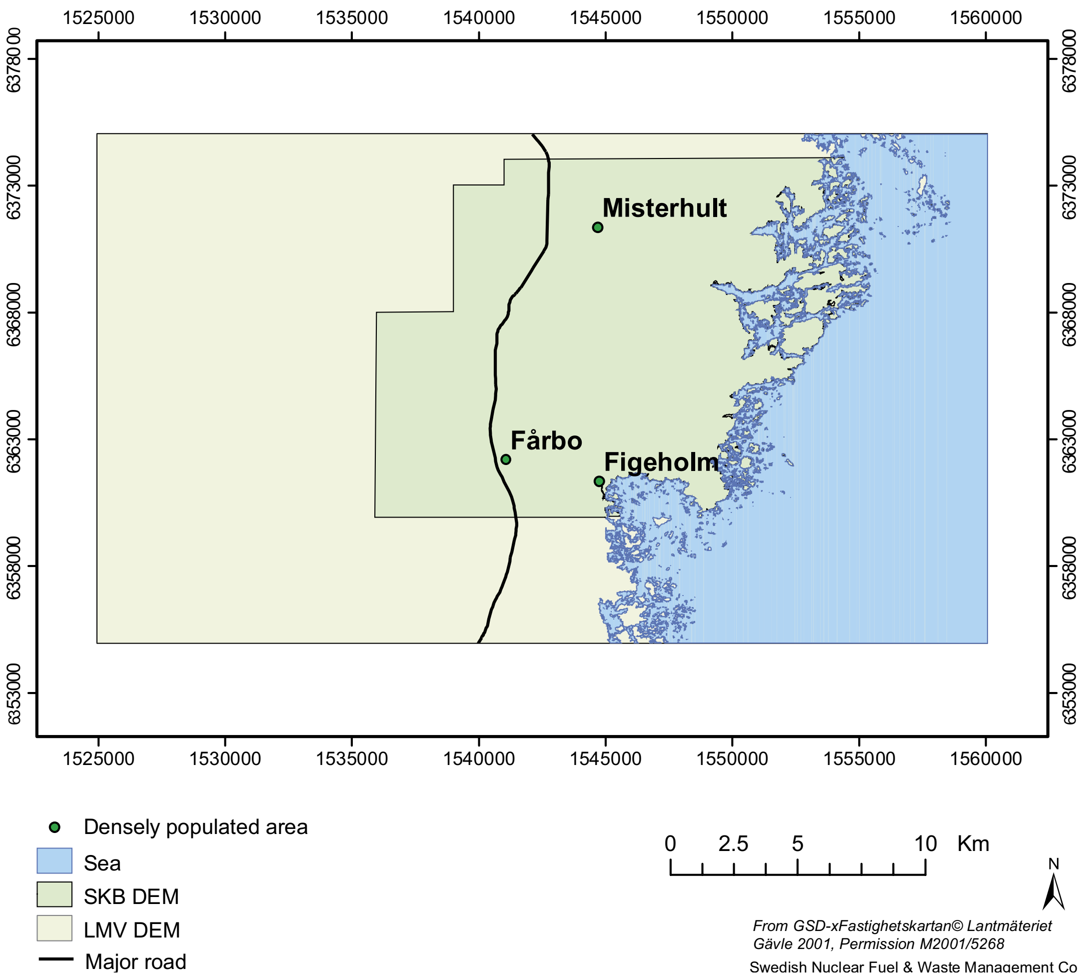

Figure 2-2. Lakes where the SKB DEM points were replaced by measured values.

2.2 Data catch from sea areas in Oskarshamn

Figure 2-3 shows the extensions for elevation data for the sea area. The elevations have been obtained from the following sources:

- the digital nautical chart (the Swedish Maritime Administration), blue area in Figure 2-3,
- detailed depth soundings performed by the Geological Survey of Sweden, SGU /Elhammer and Sandkvist, 2003/, yellow area in Figure 2-3,
- regional depth soundings performed by the Geological Survey of Sweden, SGU /Elhammer and Sandkvist, 2003/, black dots in Figure 2-3,
- depth soundings of shallow bays performed by Marin Mätteknik AB (MMT), red area in Figure 2-3 /Ingvarsson et al. 2004/,
- with DGPS measured shoreline points,
- digitized shoreline points from IR orthophotos, and
- the sea shoreline from the digital nautical chart.

The digital nautical chart has depth lines for 3, 6, 10, 15, 25, and 50 metres. These line objects have been transformed into point objects in ArcView using the Avenue script LineToPoints.avx. The maximum distances between adjacent points were set to 5 metres. The point depths (single water depth values) and symbols for “Stone in water surface” (a plus sign with dots in each corner) and “Stone beneath water surface” (a plus sign) were already stored as points. The water depth for “Stone in water surface” was set to –0.1 metre and for “Stone beneath water surface” to –0.3 metre. The total number of measurements is approximately 380,000.

The SGU depth soundings were delivered to SKB as 141 files in ASCII-format, generally one file for each transect in the survey /Elhammer and Sandkvist, 2003/. The columns in the files consist of x-coordinates and y-coordinates with a resolution of 4 digits (1/10 of a mm) and a z-value with a resolution of two digits. The coordinate system RT 90 and the Z-values are corrected to RH 70. The ASCII-files were merged to one single comma separated ASCII-file using a small program written in Pascal. The total number of measurements is approximately 78,000.

The SGU depth soundings were not performed in the shallow bays due to size of the vessel. Therefore, a completing depth sounding using a small boat was performed by the company Marin Mätteknik (MMT) /Ingvarson et al. 2004/. The z-values (water depth) were recorded both with single and multi beam techniques. The total number of depth soundings in shallow bays is approximately 780,000 points.

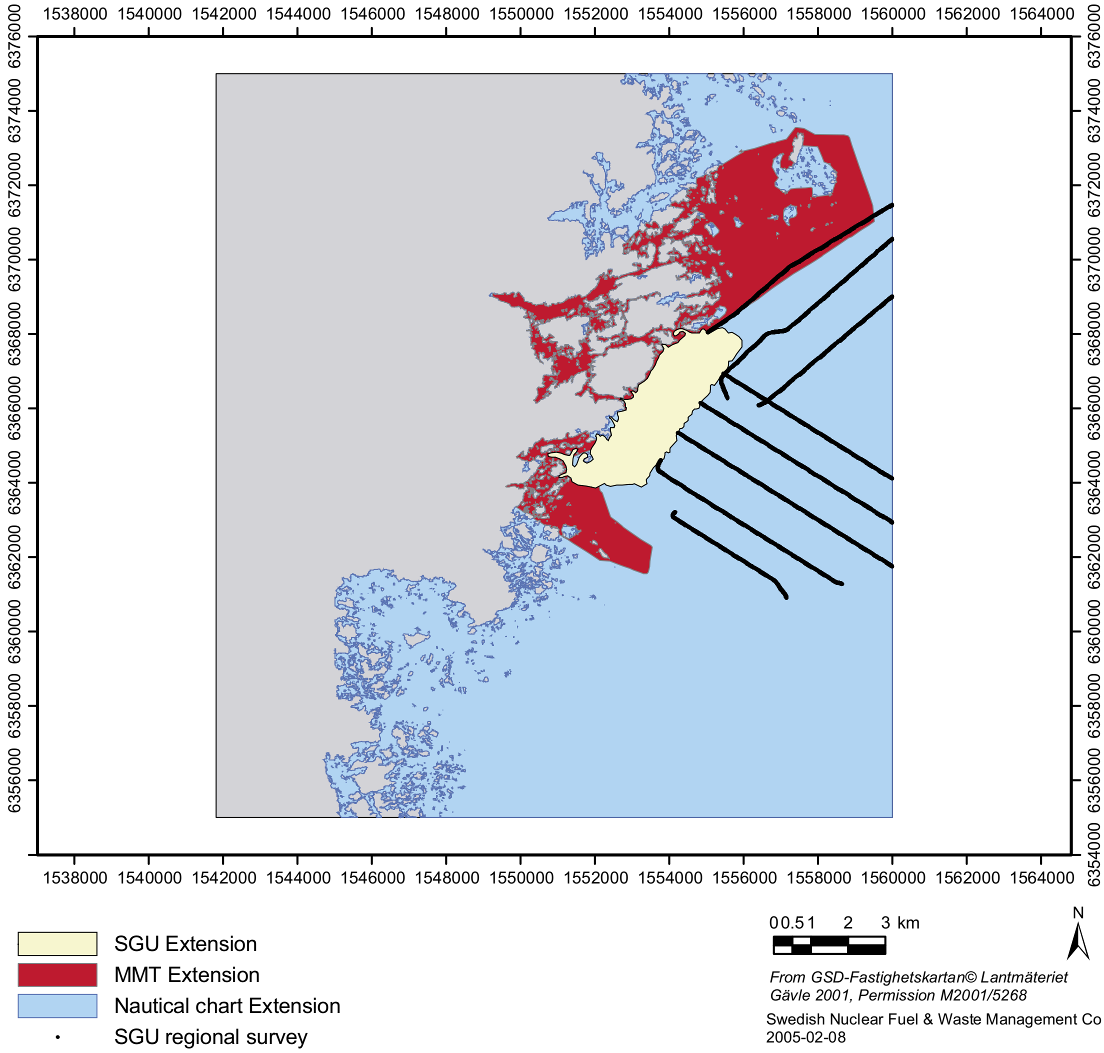

Figure 2-3. *Extensions for different data sources for the sea areas.*

Although a small boat was used in the shallow bay depth soundings, depth values are absent between the shoreline and approximately 0.7 water depth. When using the final DEM in modelling of the modern hydro-geological properties, the DEM of the sea shoreline must be very accurate. Therefore, a measurement of elevation points close to the present shoreline was performed.

There are four opportunities to catch elevation points close to the sea shoreline:

- using the sea shoreline from the digital localities maps (Fastighetskartan),
- using the 0-line from the digital nautical chart,
- manually digitizing the shoreline with the IR orthophotos as background, and
- measuring the sea shoreline by walking the shore with a DGPS.

The accuracy of the sea shoreline from the digital localities maps and the 0-line from the digital chart was tested using GIS and the IR orthophotos. Figure 2-4 shows the result from this test.

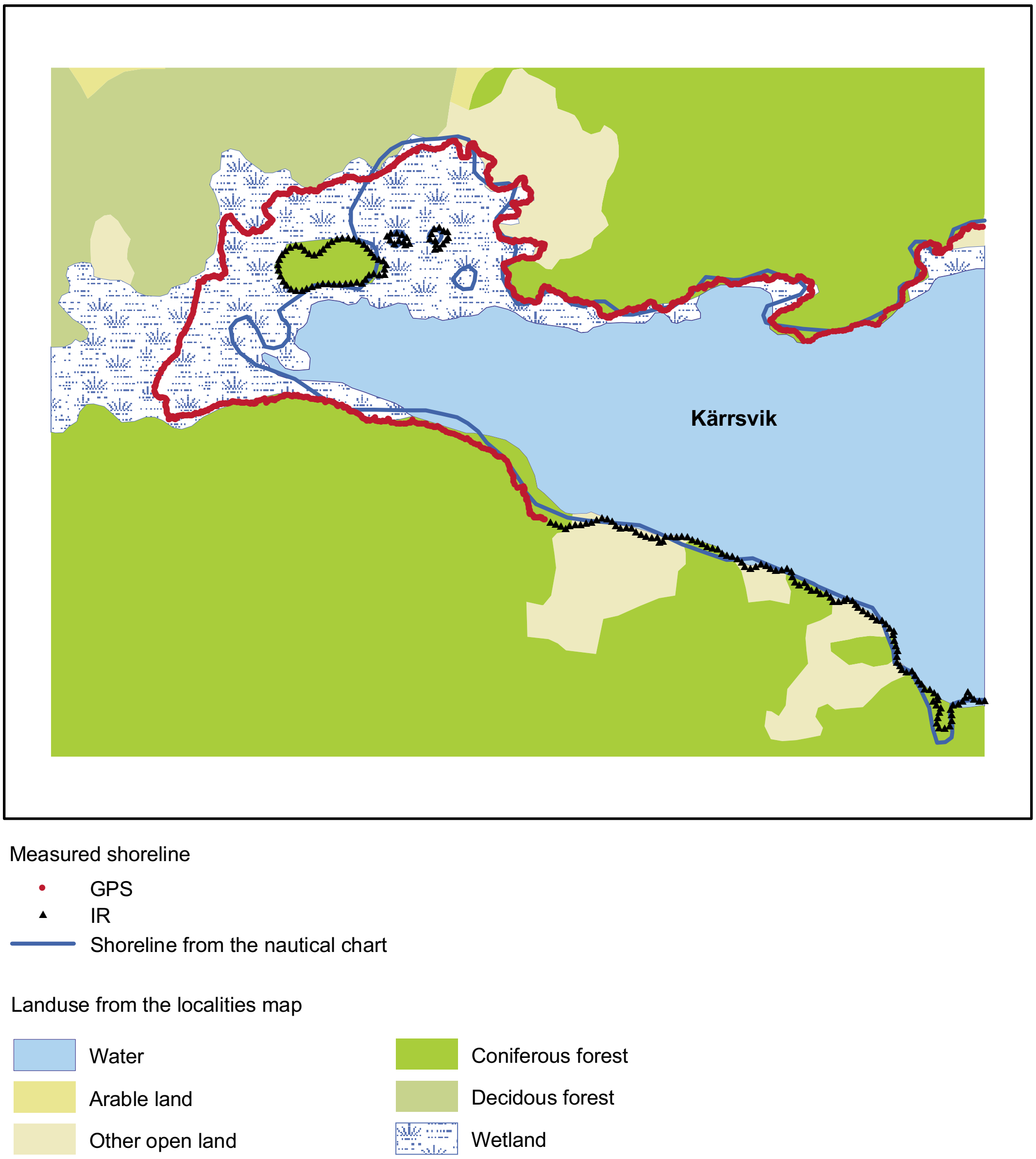

Figure 2-4. Comparison between shorelines from the digital localities map (Fastighetskartan), the digital nautical chart, manually digitized shoreline with the IR orthophotos as background, and measurements done with DGPS by walking the shoreline.

The sea water level at the time for photographing was 0.06 metres, so the distance between the digitized shoreline and the shoreline in RH 70 height system was small. The test shows that both the shorelines in the localities map and the nautical chart have low accuracies, but some localities have higher accuracy for the digital nautical chart. In addition, the test shows that low gradient shorelines are difficult to digitize using IR orthophotos if they are covered with reeds. Therefore, the most appropriate method for catching elevation data close to the zero level is to measure the sea shoreline by walking the shore with a DGPS. This approach is too expensive to use for the whole area, so this was only performed for vegetated shores within the local model area that are difficult to observe using the IR orthophotos. The rest of the shorelines within the local model area were manually digitized with the IR orthophotos as background, and the sea shoreline from the digital nautical chart was used for the rest of the model.

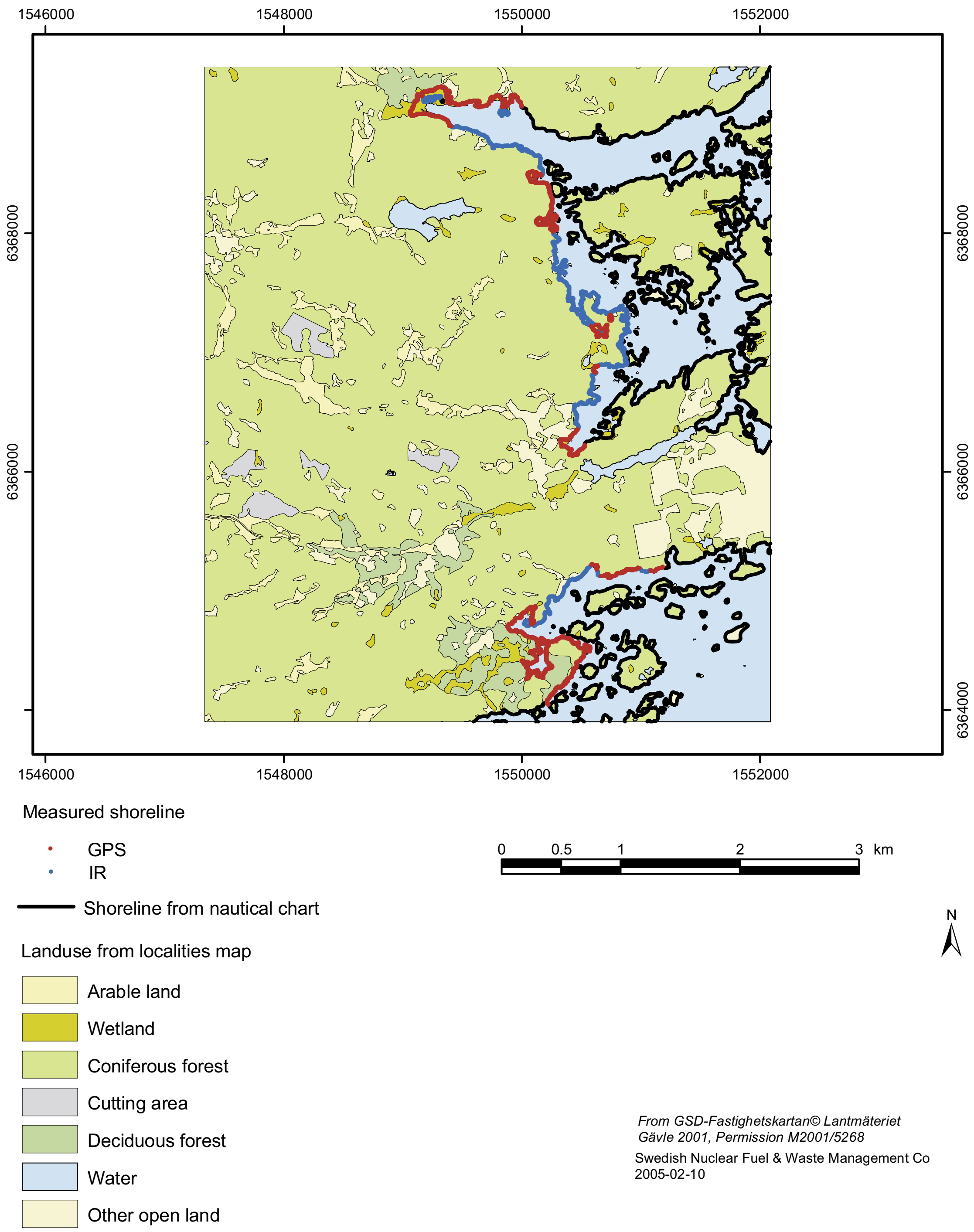

Figure 2-5. *Extensions for different data sources for the sea shoreline.*

During a post-processing procedure, each x/y-record was given a z-value using sea level data from a water level gauge in Oskarshamn. The time resolution of the gauge was one hour. The DGPS measurements were carried out during week 50 of 2004, and during this period the sea water level varied between +0.186 and +0.284 metres in the RH 70 height system.

2.3 Handling data from different data sources that are overlapping

Because some of the extensions of different point elevation data overlap, different tests were performed to determine whether both or only one of the datasets in the overlapping area should be used (Figure 2-6).

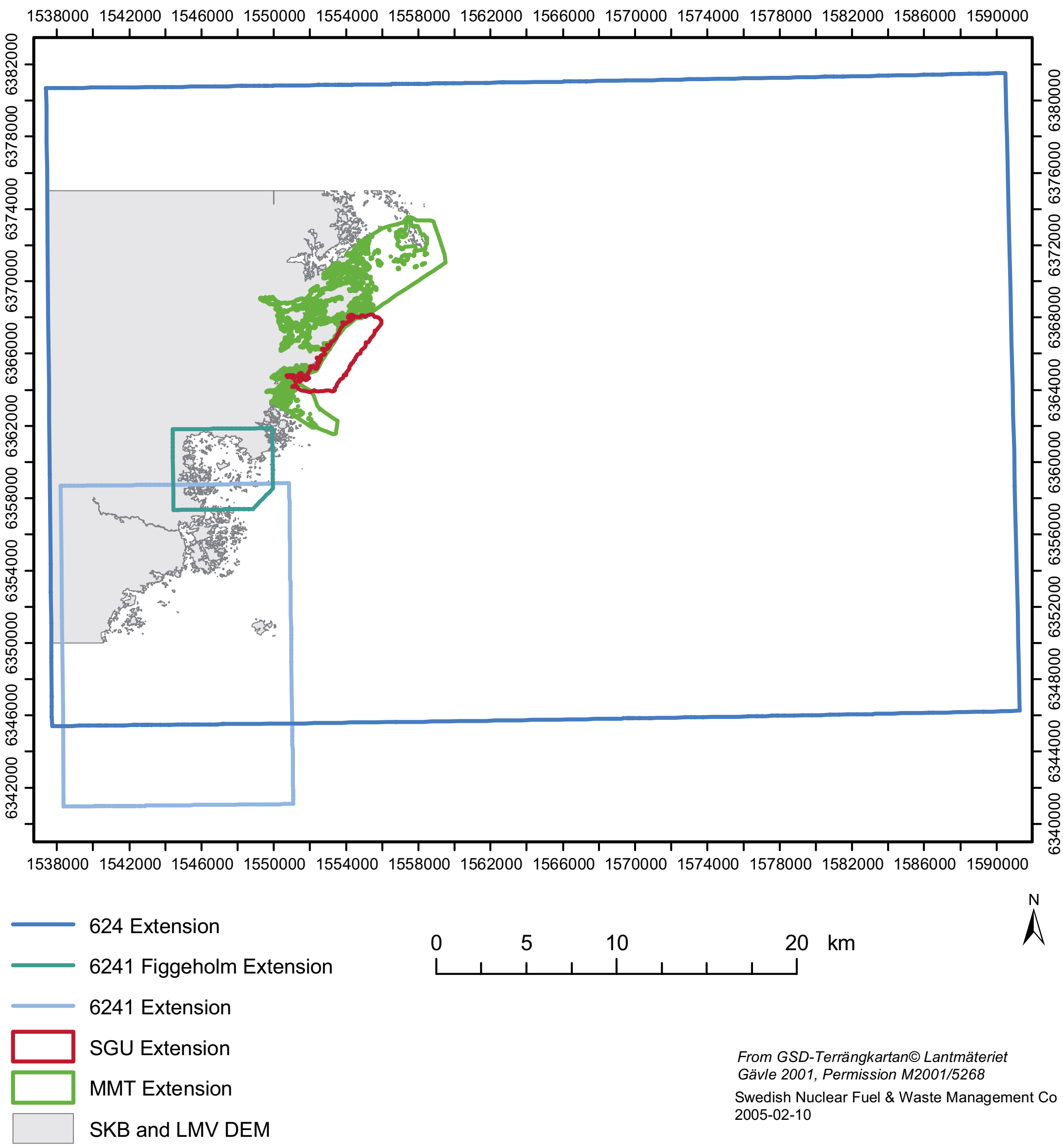

***Figure 2-6.** Extensions of different data sets.*

The tests are based on MMT depth soundings. They are estimated to be the most accurate for sea areas because these tests use modern equipment and use the SKB 10-metre model for land areas. The second most accurate depth measurements are estimated to be the SGU depth soundings. The five tests are as follows:

- the 10-metre model against the 50-metre model,
- the digital nautical chart against MMT depth soundings,
- the digital nautical chart against SGU depth soundings, and
- the SGU depth soundings against MMT depth soundings.

The point elevation data sets were joined against the MMT, SGU, or SKB 10 metre point datasets. This GIS function (point to point join) gives a new attribute with the distance to the closest point in the join to dataset. Points in an actual data set with a distance shorter than 1 metre were selected and the difference in z-value was calculated. If the dataset is classified as accurate as the join to dataset (one metre difference in XY-plane and one metre in Z-value means at least a 45 degree slope), then the differences in Z-values are larger than one metre. This is a rare occurrence. A summary of the test results is shown in Table 2-3.

Table 2-3. Summary results from tests for deciding if one or both datasets will be used for the final interpolation when these datasets are overlapping. Total Nb. = total number of points in the "join from" dataset, Nb. < 1 m = number of points within a distance lower than one metre from a point in the "join to" dataset, Nb. Diff. > 1 m = number of points with a difference in elevation value in the "Nb. < 1 m" dataset that are higher than one metre, Max. diff. (m) = the maximum difference in elevation value between two points in "join from" and "join to" datasets that are situated closer than one metre from each other, and Mean diff. (m) = the average difference in elevation value between all points in "join from" and "join to" datasets that are closer than one metre from each other.

Join from	Join to	Nb. < 1 m	Nb. diff. > 1 m	% error	Max. diff. (m)	Mean diff. (m)
Model 50 m	Model 10 m	79,272	44,464	56	22.1	1.5
Dig. chart	MMT	318	152	48	6.0	1.4
Dig. chart	SGU	80	60	75	12.1	2.5
SGU	MMT	616	47	8	2.3	0.5

The two existing models are overlapping within the 10-metre model extension with approximately 79,000 height values with exactly the same coordinates. If duplicate points are incorporated in the final elevation point data set, the ArcGis program will use the mean value. If there are great differences in z-values among the duplicates, there is a risk for errors in the final model. A statistical analysis was therefore carried out for the duplicates.

The difference in z-values for the duplicate points was calculated. The results of the comparison are presented in Table 2-3. Only about 44% of the duplicate points have a difference in z-value lower than one metre. This means that it is not appropriate to use the mean values of duplicate points. The 50-metre model is evaluated from air photos from 4,600 metre level, while photos from 2,300 metre level were used for evaluation of the 10-metre model /Wiklund, 2002/. Therefore, the 10-metre model will probably be of higher quality. Another reason for using only the 10-metre model points is that abnormal differences in z-values between adjacent points could occur, i.e. a point from the 50-metre model is surrounded by 8 points from the 10-metre model. All duplicate points from the 50-metre model were deleted and not used in the final interpolation.

The tests for the sea depth datasets show that only the depth soundings of shallow bays (MMT) and the SGU depths soundings have low differences in depth values between points situated within a metres distance. All other comparisons produce significant differences.

Based on the test results, the following datasets were used in the final interpolation procedure:

- when the 50-metre model overlapped the 10-metre model, only values from the 10-metre model were used,
- when the digital nautical chart overlapped the SGU depth soundings, only the SGU dataset was used,
- when the depth soundings of shallow bays overlapped the SGU depth soundings, both datasets were used, and
- when the digital nautical chart overlapped the base map, only data from the base map was used.

There are also overlapping areas among different digital nautical charts. Three different charts are used in the data catch:

- Digital chart number 624, an archipelago chart with scale 1:50,000.
- Digital chart number 6241, a special chart with scale 1:25,000.
- Digital chart number 6241_Figeholm, a harbour chart with scale 1:5,000.

A comparison between the three charts shows that the degree of generalization increases from the harbour chart to the special chart, and even more from the special chart to the archipelago chart. Therefore, when the harbour chart overlaps the special chart, only data from the harbour chart is used. When the special chart overlaps the archipelago chart, only data from the special chart is used.

The total number of points in the merged point dataset after deletion of some of the overlapping datasets is approximately 3,330,000.

3 Results

3.1 Construction of the digital elevation models

All elevation point values were collected in two databases, and with these databases new digital elevation models were created. One representing land surface, lake water surface, and sea bottom, and the other one representing land surface, lake bottoms, and sea bottoms.

The DEMs were created with a resolution of 10-metres. The interpolation from irregularly spaced point values to a regularly spaced DEM was done using the software ArcGis 8 Geostatistical Analysis extension. Kriging was chosen as the interpolation method /Davis, 1986; Isaaks and Srivastava, 1989/. The theoretical semivariogram model and the parameters scale, length, and nugget effect were chosen with the extension.

Because of the large size of the merged point file, it was impossible to construct the models by one single interpolation process. Therefore, the model was divided into 28 sub-models (Figure 3-1) that were processed one by one and finally merged together into one single model. Each sub-model was treated with regard to its conditions, i.e. different Kriging parameters were set to different sub-models.

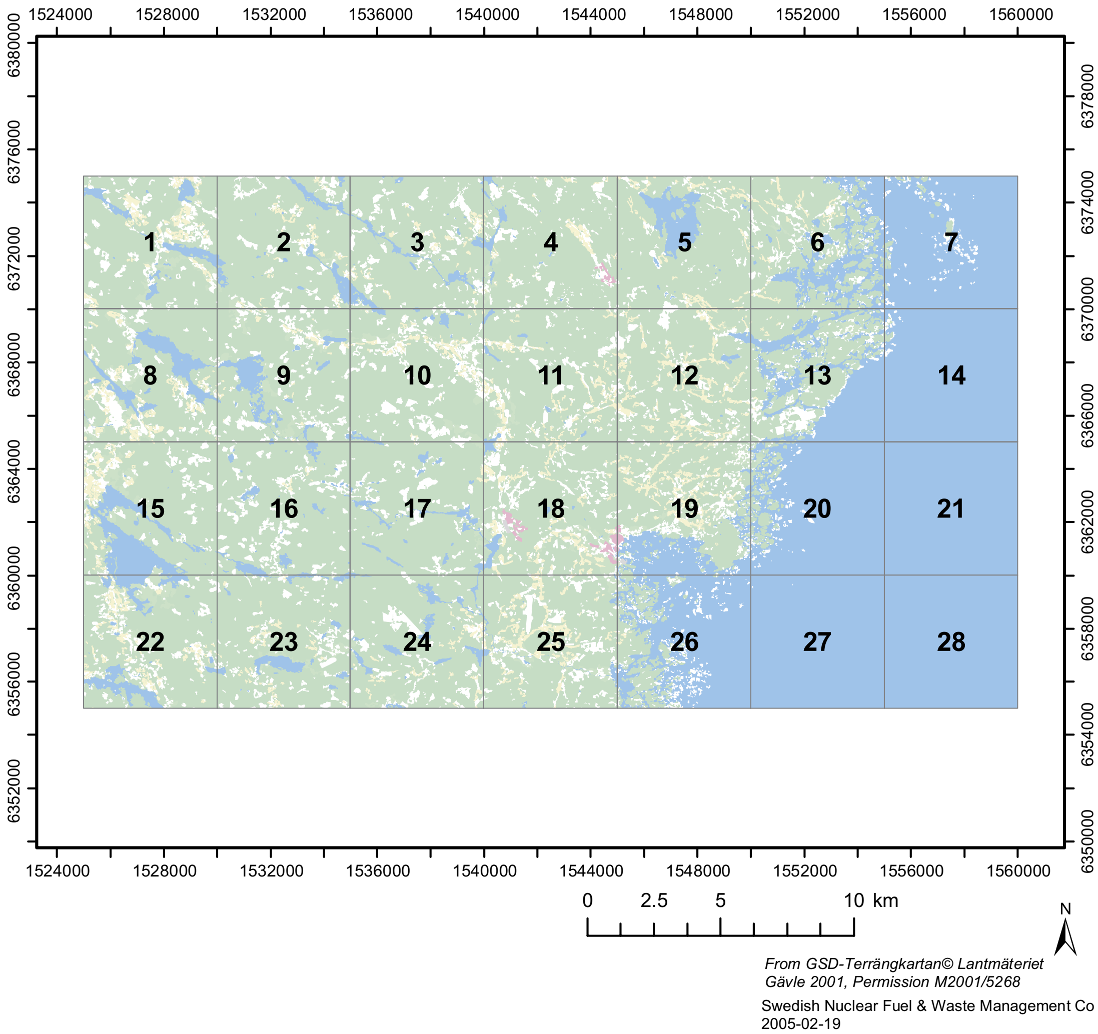

Figure 3-1. The extensions of the 28 sub-models.

Common to all sub-models are an Ordinary Kriging geostatistical method, a spherical theoretical model, and an elliptical search shape. The parameters that differ between different sub-models are the search size (the length of the major and minor semi-axis of the ellipses), the angle of the major semi-axis, the nugget value, the number of lags, and the lag size. Before the interpolations started, the models were validated both with cross-validation (one data point is removed and the rest of the data is used to predict the removed data point) and ordinary validation (part of the data is removed and the rest of the data is used to predict the removed data). Both the cross-validation and ordinary validation goals produce a standardised mean prediction error near 0, small root-mean-square prediction errors, average standard error near root-mean-square prediction errors, and standardised root-mean-square prediction errors near 1.

Cross validations with different combinations of Kriging parameters were performed until the standardised mean prediction errors were close to zero, but not necessarily the lowest values were always chosen. Because the aim was to determine the most valid model for both measured and unmeasured locations, care was taken to produce low values for the root-mean-square prediction errors and to minimise the difference between the root-mean-square prediction errors and the average standard errors. Different models were compared and the ones with the most reasonable statistics were chosen.

Finally, validations were performed with the most appropriate Kriging parameters in order to verify that the models fit unmeasured locations. Unfortunately, the standardised mean prediction errors and the standardised root-mean-square prediction errors were not calculated for all of the models. The final choice of parameters is presented in Appendix 1 (ordinary validation) and Appendix 2 (cross-validation).

3.2 The quality of the digital elevation models

Figure 3-2 shows the quality of the sub-models as the values of root-mean-square prediction errors that should be low for a high quality model. Sub-models with low quality are those with only data from the digital nautical chart, and those are also the sub-models with lowest point density.

The coordinates of the starting point (upper left corner) were chosen so that the values from the SKB 10-metre DEM were not changed by the Kriging interpolation process; i.e. the central points in the cells in the new DEM coincide with the central points in the SKB 10-metre DEM. The digital elevation model with lake surface values and sea surface values is illustrated in Figure 3-3.

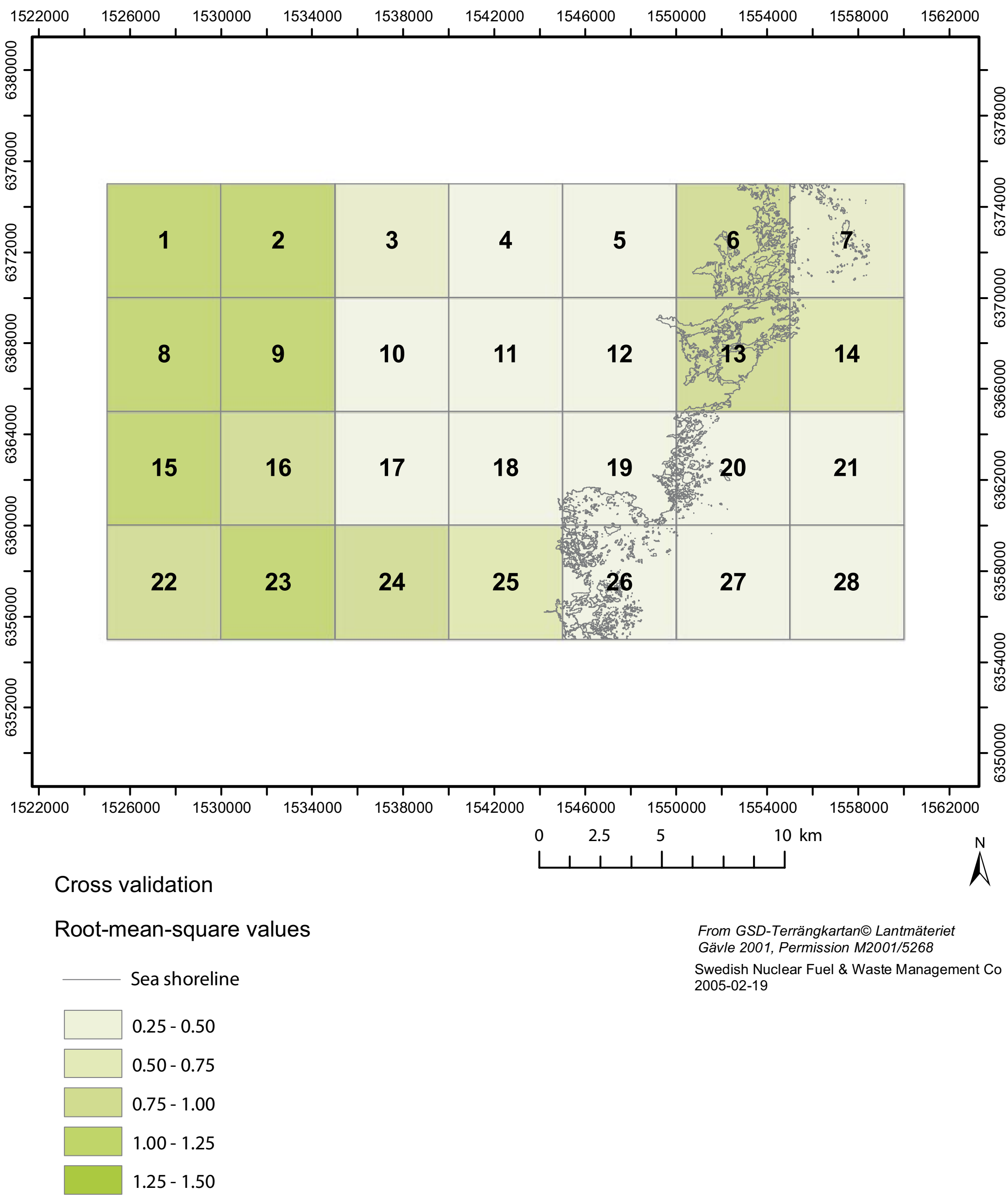

Figure 3-2. *Quality of the sub-models as the values of root-mean-square prediction errors.*

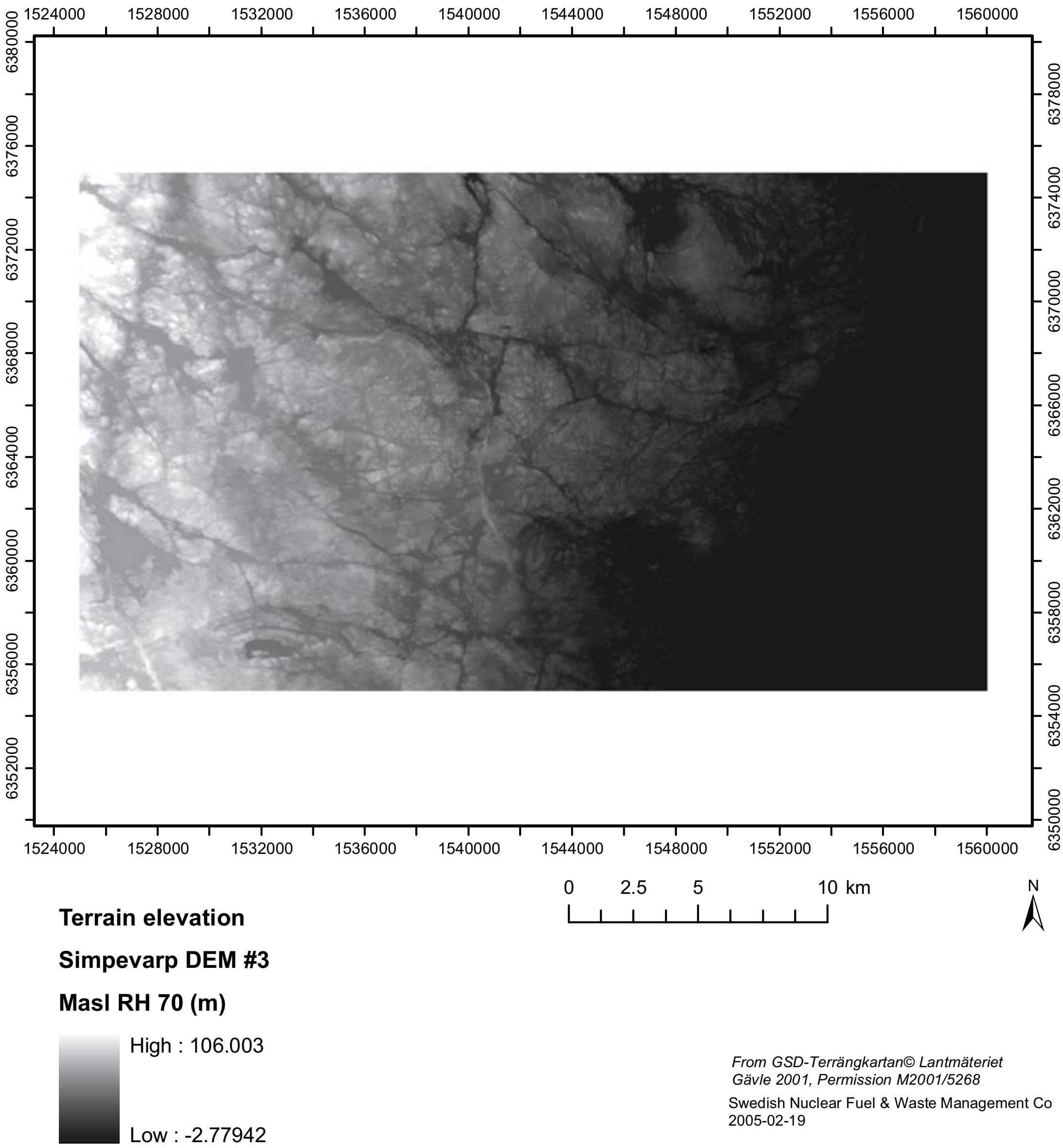

Figure 3-3. *The digital elevation model (Simp_dem #3) with land surface, sea surface, and lake sediment surface.*

The final model had a size of approximately 35 × 20 kilometres, a cell size of 10-metres, 3,501 rows, and 2,001 columns: a total number of grid cells of 7,005,501 and a file size of approximately 28 MB (ESRI Grid format). The extension is 154995 west, 1560005 east, 6375005 north, and 6349995 south in the RT 90 coordinate system. As mentioned earlier, the height system is RH 70.

The area is undulating with narrow valleys situated at bedrock-weakened zones. The range in elevation is approximately 151 metres with the highest point at 106 metres above sea level at the southwest part of the model and the deepest sea point at –45 metres in the southeast part of the DEM. The mean elevation in the model is 24 metres, where 73% is land and 27% is sea. The flat landscape is also shown in the statistics of the slope grid where the mean slope is 2.79 degrees, and 83.0% of the cells have slopes lower than 5 degrees and 14.5% have slopes between 5 and 10 degrees. As expected, almost all of the cells with slope steeper than 10 degrees (2.5%) are slopes along the earlier mentioned narrow valleys or lakeshores.

As a part of the quality control of the DEM, the DEM is simply looked at when displayed with different techniques, such as different classifying schemes or as images with different stretching values. Unfortunately, when displaying the DEM as image with maximum stretching value, the traces from the SGU regional survey can be observed as small darker circular areas (small local depressions) along strait traces (not shown in Figure 3-3). This phenomenon is an indication of errors in the DEM. As mentioned earlier, the depth soundings from SGU are treated as more accurate than the nautical chart, but deleting all the nautical chart values within the SGU regional survey area are not possible, although the distance between each trace in the SGU survey is as long as 1,000 metres. We estimate that the local depressions are existing phenomenon, but they have probably extensions that are too small. With shorter distance between traces in the SGU survey, many of the small depressions will be larger and fewer, i.e. they will be combined.

3.3 Data files delivered to SKB

Following data files are delivered to SKB:

Simp_DEM_1	ESRI Grid format, land surface, lake bottoms, and sea bottoms
Simp_DEM_2	ESRI Grid format, land surface, lake surface, and sea bottom
Simp_DEM_3	ESRI Grid format, land surface, lake bottoms and sea surface
Simp_dem_1.asc	ESRI ASCII Grid format, land surface, lake bottoms, and sea bottoms
Simp_dem_2.asc	ESRI ASCII Grid format, land surface, lake surface, and sea bottoms
Simp_dem_3.asc	ESRI ASCII Grid format, land surface, lake bottoms, and sea surface
Simp_points_1.shp	ESRI Shape format, points for Simp_DEM_1
Simp_points_2.shp	ESRI Shape format, points for Simp_DEM_2

4 References

Brydsten L, 2004. A method for construction of digital elevation models for site investigation program at Forsmark and Simpevarp. SKB P-04-03. Svensk Kärnbränslehantering AB.

Davis J C, 1986. Statistics and data analysis in geology. John Wiley & Sons, New York, p 383–404.

Elhammer A, Sandkvist Å, 2003. In process. Detailed marine geological survey of the sea bottom outside Simpevarp. Svensk Kärnbränslehantering AB.

Isaaks E H, Srivastava R M, 1989. An introduction to applied geostatistics, Oxford University Press, NY, Oxford. ISBN 0-19-505013-4.

Wiklund S, 2002. Digitala ortofoton och höjdmodeller. Redovisning av metodik för platsundersökningsområdena Oskarshamn och Forsmark samt förstudieområdet Tierp Norra. SKB P-02-02. Svensk Kärnbränslehantering AB.

Ingvarson N H, Palmeby A L F, Svensson L O, Nilsson K O, Ekfeldt T C I, 2004. Oskarshamn site investigation: Marine survey in shallow coastal waters. Bathymetric and geophysical investigation. SKB P-04-254. Svensk Kärnbränslehantering AB.

Appendix 1

Validation of sub-models

Sub-model	Lag size	Nb of lags	Regression function	Mean	RMS	Average SE	Mean stand.	RMS stand.	Samples
1	50	12	0.988 * x + 0.910	0.04136	1.526	2.269	0.0174	0.672	2330
2	60	12	0.979 * x + 1.133	–0.00263	1.471	1.987	0.0004	0.736	2330
3	40	12	0.995 * x + 0.153	–0.00339	0.708	1.012	–0.0048	0.543	9604
4	40	12	0.998 * x + 0.072	–0.00076	0.383	0.785			43560
5	40	12	0.997 * x + 0.047	–0.00471	0.368	0.889	–0.0040	0.363	45802
6	10	12	0.981 * x + 0.097	–0.00254	1.146	1.506			49923
7	30	12	0.999 * x + -0.009	–0.00288	0.643	0.746			91479
8	30	12	0.964 * x + 2.383	–0.00382	1.593	2.288	–0.0006	0.695	2330
9	30	12	0.969 * x + 1.713	0.01848	1.448	1.739	0.0101	0.830	2330
10	30	12	0.998 * x + 0.065	0.00137	0.394	0.777			33121
10_2	30	12	0.999 * x + 0.059	–0.00217	0.397	0.764			33134
11	20	12	0.997 * x + 0.090	0.00005	0.324	0.871			58294
11_2	20	12	0.997 * x + 0.094	0.00025	0.324	0.890			58398
12	30	12	0.997 * x + 0.048	0.00031	0.374	0.718	–0.0003	0.688	59631
12_2	50	12	0.990 * x + 0.148	–0.00020	0.461	1.327	–0.0001	0.354	59751
13	30	12	0.990 * x + 0.050	–0.00542	1.194	1.281			72742
13_2	30	12	0.993 * x + 0.046	0.00163	1.257	1.002			72882
14	40	12	1.000 * x + -0.007	0.00593	0.867	0.914			33396
15	40	12	0.974 * x + 1.767	0.02600	1.486	1.905	0.0134	0.778	2330
16	60	12	0.964 * x + 1.944	0.01765	1.242	1.466	0.0121	0.844	2330
17	30	12	0.998 * x + 0.103	0.00257	0.312	0.742	0.0036	0.388	44260
17_2	30	12	0.997 * x + 0.145	–0.00074	0.334	0.850			44331
18	30	12	0.999 * x + 0.037	0.00012	0.270	0.731			56242

18_2	30	12	0.999 * x + 0.034	0.00118	0.270	0.735			56376
19	20	12	0.997 * x + 0.030	0.00135	0.356	0.653			59601
20	20	12	0.994 * x + -0.030	–0.00401	0.599	0.813			44685
21	10	12	0.999 * x + -0.018	–0.00636	0.453	0.516			4568
22	30	12	0.968 * x + 2.162	–0.00578	1.270	1.559	–0.0056	0.814	2350
23	40	12	0.966 * x + 1.864	0.00213	1.496	1.739	–0.0001	0.857	2350
24	40	12	0.974 * x + 1.005	–0.00610	1.142	1.318	–0.0046	0.730	3996
25	40	12	0.996 * x + 0.067	0.00349	0.912	1.079	0.0036	0.674	5020
26	10	12	0.991 * x + -0.028	–0.01670	0.800	0.535			15563
27	20	12	0.999 * x + -0.017	–0.01685	0.927	0.453			5231
28	50	12	1.000 * x + -0.002	–0.01134	0.200	0.320			3153

Appendix 2

Cross-validation of sub-models

Sub-model	Lag size	Nb of lags	Regression function	Mean	RMS	Average SE	Mean stand.	RMS stand.	Samples
1	50	12	0.988 * x + 0.878	–0.00005	1.458	2.195	–0.0001	0.665	11647
2	60	12	0.981 * x + 1.019	–0.00011	1.372	1.918	0.0000	0.715	11647
3	40	12	0.995 * x + 0.171	–0.00019	0.685	0.980	–0.0002	0.542	48016
4	40	12	0.998 * x + 0.059	0.00000	0.356	0.758	0.0000	0.401	217480
5	40	12	0.999 * x + 0.022	0.00001	0.309	0.643	0.0000	0.403	229008
6	10	12	0.982 * x + 0.090	–0.00334	1.119	1.485	–0.0006	0.818	248998
7	30	12	0.999 * x + –0.007	0.00131	0.641	0.720	0.0015	1.151	456233
8	30	12	0.970 * x + 2.023	0.00015	1.531	2.205	0.0000	0.695	11647
9	30	12	0.969 * x + 1.678	0.00007	1.339	1.676	0.0000	0.799	11647
10	30	12	0.999 * x + 0.056	0.00001	0.360	0.749	0.0000	0.403	165486
10_2	30	12	0.999 * x + 0.056	–0.00004	0.359	0.738	0.0000	0.412	165550
11	20	12	0.997 * x + 0.072	0.00000	0.299	0.842	0.0000	0.355	291132
11_2	20	12	0.997 * x + 0.072	–0.00010	0.300	0.860	–0.0001	0.353	291641
12	30	12	0.997 * x + 0.043	0.00059	0.363	0.693	0.0009	0.722	298150
12_2	50	12	0.990 * x + 0.148	–0.00020	0.461	1.327	–0.0001	0.354	59751
13	30	12	0.991 * x + 0.056	–0.00096	1.180	1.263	0.0007	0.995	363221
13_2	30	12	0.993 * x + 0.059	0.00290	1.272	0.975	0.0035	1.509	363923
14	40	12	1.000 * x + –0.004	0.00209	0.870	0.877	0.0017	1.297	166680
15	40	12	0.977 * x + 1.554	–0.00003	1.359	1.835	0.0000	0.741	11647
16	60	12	0.969 * x + 1.659	–0.00015	1.210	1.411	0.0000	0.857	11647
17	30	12	0.998 * x + 0.093	0.00005	0.293	0.714	0.0000	0.375	221300
17_2	30	12	0.997 * x + 0.128	0.00000	0.307	0.826	0.0000	0.343	221653
18	30	12	0.999 * x + 0.031	–0.00004	0.246	0.706	0.0000	0.354	281181

18_2	30	12	0.999 * x + 0.032	–0.00015	0.248	0.710	–0.0001	0.362	281846
19	20	12	0.998 * x + 0.024	–0.00163	0.291	0.630	–0.0009	0.577	252556
20	20	12	0.995 * x + –0.025	–0.00232	0.499	0.799	–0.0010	0.655	222112
21	10	12	0.999 * x + –0.014	–0.00566	0.360	0.481	–0.0050	0.883	22801
22	30	12	0.971 * x + 1.954	0.000118	1.228	1.501	0.00004	0.8177	11748
23	40	12	0.977 * x + 1.249	0.00019	1.340	1.672	0.0001	0.802	11748
24	40	12	0.981 * x + 0.742	–0.00003	1.056	1.271	–0.0001	0.704	19978
25	40	12	0.997 * x + 0.052	–0.00432	0.863	1.040	–0.0022	0.672	25096
26	10	12	0.999 * x + 0.002	–0.01787	0.376	0.515	–0.0136	0.580	75413
27	20	12	0.999 * x + –0.006	–0.00061	0.309	0.430	–0.0036	0.548	25379
28	50	12	1.000 * x + –0.001	–0.00130	0.263	0.308	–0.0033	0.213	15754

Appendix 3

Sub-model parameters

Common to all sub-models are Ordinary Kriging with a spherical model. The model equation should be read as follows:

Partial sill * Theoretical Semivariogram (Major Range, Minor Range, Anisotropy Direction) + Nugget value * Nugget)

Sub-model	Number of points	Model
1	11647	48.954*Spherical(592.66,592.66,20.9)+0*Nugget
2	11647	33.643*Spherical(592.66,490.24,309.3)+0.019795*Nugget
3	48016	18.947*Spherical(474.13,379.56,302.6)+0*Nugget
4	217480	20.237*Spherical(474.13,391.3,332.4)+0*Nugget
5	229008	16.211*Spherical(474.13,474.13,6.3)+0*Nugget
6	248988	8.2176*Spherical(118.53,114.96,279.0)+1.2774*Nugget
7	456233	30.905*Spherical(355.6,351.71,53.9)+0*Nugget
8	11647	27.505*Spherical(355.6,310.52,308.0)+0*Nugget
9	11647	16.192*Spherical(355.6,318.88,94.3)+0*Nugget
10	165486	15.288*Spherical(355.6,355.6,9.0)+0*Nugget
10_2	165550	14.413*Spherical(355.6,335.67,307.7)+0*Nugget
11	291132	21.774*Spherical(355.6,355.6,9.3)+0*Nugget
11_2	291641	15.209*Spherical(237.07,237.07,9.1)+0*Nugget
12	298150	13.551*Spherical(355.6,283.45,288.1)+0*Nugget
12_2	298748	16.986*Spherical(592.66,409,295.3)+1.155*Nugget
13	363221	29.539*Spherical(592.66,400.59,254.7)+0.95167*Nugget
13_2	363923	22.63*Spherical(355.6,262.96,258.5)+0.30998*Nugget
14	166680	55.733*Spherical(474.13,393.64,48.9)+0*Nugget
15	11647	26.054*Spherical(474.13,431.75,309.1)+0*Nugget
16	11647	10.978*Spherical(355.6,291.63,290.9)+0*Nugget
17	221300	13.435*Spherical(355.6,278.71,275.3)+0*Nugget
17_2	221653	13.599*Spherical(355.6,292.81,269.9)+0.13995*Nugget
18	281181	15.197*Spherical(355.6,350.97,281.4)+0*Nugget
18_2	281846	15.114*Spherical(346.32,344.84,79.0)+0*Nugget
19	252556	8.1024*Spherical(237.07,204.77,253.6)+0*Nugget
20	222112	7.8934*Spherical(226.57,182.11,41.5)+0.31392*Nugget
21	22801	3.6764*Spherical(118.53,103.52,35.2)+0*Nugget
22	11748	13.684*Spherical(355.6,355.6,71.6)+0*Nugget
23	11748	19.221*Spherical(474.13,357.22,272.2)+0.099157*Nugget
24	19978	20.156*Spherical(474.13,474.13,7.8)+0*Nugget
25	25096	13.895*Spherical(474.13,328.58,19.1)+0*Nugget
26	75413	3.4566*Spherical(118.53,118.53,9.0)+0*Nugget
27	25379	4.2684*Spherical(212.24,163.56,337.2)+0*Nugget
28	15754	5.6737*Spherical(592.66,432.31,338.6)+0*Nugget