Exercise – Quality & Accuracy Assessment

Introduction

The only way to assess the quality of geodata is by investigation of the accuracy of the method used for data capture. In principle, this can be done by comparing the geodata to reality, but reality can only be realised by measurement...

So eventually you will be comparing the methods against each other.

In such case it can be hard to judge from which of the two geodata the error comes.

Of course, in a case where you have geodata, which you trust and believe are as good as — or better than — your own data, you can use them as basis for comparison.

Differential GPS (D-GPS) For example, if you are possessing some Differential GPS positions (D-GPS) with a known high accuracy.

Another example, if you are going to digitize a forest management plan in 1:25 000 and you have the ownership boundary from a cadastral map in 1:5 000, the latter can be used for assessment of your (spatial) digitizing quality.

Objective

<u> Aim</u>:

In this exercise you will:

- Be introduced to one method to assess
 the quality of your digitized geodata using
 a template of relatively high accuracy.
- Learn the importance of keeping metadata.



Task

In this exercise you will play the part of a project manager.

The task of the project is digitizing wetland areas in Denmark for purposes of environmental protection.

The biggest challenge you are facing is access to geodata.

You have still not decided for the image that will serve as background during digitizing.

High-resolution orthophotos are expensive and nothing your project can afford.

Instead you might have to rely on cheaper alternatives.



Accuracy

SDFI = The Agency for Data Supply and Infrastructure Styrelsen for Dataforsyning og Infrastruktur (Formerly known as: Geodatastyrelsen (GST), in Danish) As your budget is tight, you would like to use the cheapest geodata available that still gives you good enough accuracy. But what is accurate?

To tell if the features you have digitized are accurate, you will need a reference to compare with.

You have decided that your 'role model' will be the feature geodata called the Top10DK, provided by the SDFI.

The Top10DK is a series of thematic geodata best suited for maps at scale 1:10 000 and smaller.

If the features you have digitized deviate too much from the Top10DK 'template' this would be an indication that the image you used as background for digitizing is inadequate (and that you need a more accurate background image).

There are multiple methods to check whether one set of geodata diverts from another.

The method you are going to use involves creating multiple ring buffers (Fig. 1).

INPUT OUTPUT

Fig. 1 Creating a Multiple Ring Buffer around a polygon feature.

Procedure

In brief, this is what you are going to do:

- A. You will create multiple ring buffers around your 'template'.

 In this case your 'template' will be the Top10DK feature of the Kastellet (Fig 2).
 - The choice of the Kastellet is simply because it takes some precision to digitize the contours of this land-to-water feature.
- B. Then you are going to digitize your own polygon of the Kastellet using an image as a background.
- C. Finally, you will compare how much of your digitized polygon that falls within each of the buffers of the Top10DK Kastellet polygon.
- D₁. Conclusion

 If the most part of your digitized polygon falls within the *closest* buffers of the Top10DK Kastellet polygon you may conclude that the background image will serve you good for digitizing the wetland areas.
- D₂. If not, you may need to find a more accurate background image.
 - Generally this means finding a background image of higher resolution (= more expensive).



Fig. 2 A Multiple Ring Buffer around the Top10DK polygon of the Kastellet.

Red nuances indicate buffers closer to the outline of the Kastellet feature.

- A. Creating a Multiple Ring Buffer
- 1. Start a new ArcGIS Pro project.
- 2. a) From I:\SCIENCE-IGN-CGD-UVMAT\GIS_course\Exercises\ArcGIS_Pro\Exercise_Quality_AccuracyAssessment\Quality_AccuracyAssessment.gdb\:

 Copy the Kastellet_Top10DK feature class. (Import or Export features are alternative techniques.)
 - b) Paste it into your own project geodatabase.

Multiple Ring Buffer

- Use your skill to search and locate the Multiple Ring Buffer tool from the Geoprocessing pane.
- 4. Open the Multiple Ring Buffer dialog (Fig. 3).
 - Input Feature: Select Kastellet _Top10DK
 - Output Feature Class:
 Navigate to a location where you would like to save the new feature class that will be generated.
 Give it a name.
 - Distances: Type a number in the Distance field.
 - → Click on a surface next to the field.

Repeat for all numbers in figure 3, red box.

The negative values are representing the buffers which are going to be directed inwards, towards the centre of the Kastellet polygon.

- Buffer Unit: Meters
- Accept the defaults → Run.

A new layer appears in the Contents pane and the Map view. (Be patient. It may take a moment to process.)

This is the new map layer containing the multiple buffers around the Kastellet polygon (Fig. 2, above).

Notice! The rings in Figure 2 and 4 have been actively symbolized.

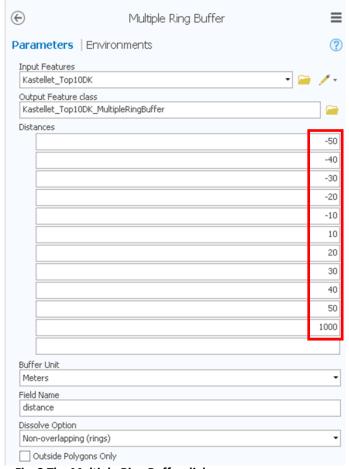


Fig. 3 The Multiple Ring Buffer dialog.

The tool is generating several buffers extending away from the input feature like the 'layers of an onion'.

Buffer Distance

Let's have a closer look at the multiple ring buffer geodata you just generated.

5. Zoom in closer to the true Kastellet area.

6. Position the Kastellet Top10DK map layer on top of the multiple ring buffer map layer (Fig. 4a). You may now see that the buffers are like layers of an onion around the Kastellet Top10DK polygon feature.

7. Open the Attribute Table of the multiple ring buffer map layer (Fig. 4b). There are the 'onion layers' listed as separate polygons, each with its buffer distance.

Or, should we say buffer interval, because each polygon is actually representing just one 'onion layer'.

For example, the '20' buffer distance polygon is representing the buffer at **10-20m** distance from the *Kastellet Top10DK* polygon feature.

- 8. To get even more familiar with your multiple ring buffer layer, you are going to calculate the areas of each of the 'onion layers'.
 - a) Create a new column in the Attribute Table. Name it 'Area'.
 - b) Use your skills to define an appropriate data type for the column. (Hint! Since you are going to measure area, a numeric data type allowing decimals is a good choice).
 - c) Use your skills to calculate the area of the polygons. (Hint! The Geodatabases & Attribute Tables exercise)



a)

)	OBJECTID	Shape	Shape_Length	Shape_Area	distance
	1	Polygon	1309.05042	53931.651424	-50
	2	Polygon	2735.922718	13671.647997	-40
	3	Polygon	2976.204956	14886.458881	-30
	4	Polygon	3211.657713	16051.911714	-20
	5	Polygon	3435.505169	17189.294542	-10
	6	Polygon	3676.144894	36844.938656	10
	7	Polygon	3853.716182	19265.917067	20
	8	Polygon	3945.478777	19715.42103	30
	9	Polygon	4034.717962	20166.593232	40
	10	Polygon	4129.823314	20667.093432	50
	11	Polygon	9998.640384	4701811.453106	1000

Fig. 4 a) The Multiple Ring Buffer map layer (various colours) and the Kastellet _Top10DK map layer (transparent with a hashed black border).

b) The Attribute Table with the area measurements of the buffer polygons.

Interval

Know your geodata!



This step is only necessary if the output was a shapefile.

Data Type

What can you tell from the area calculations?

Well, obviously three of the polygons are larger than the rest of them (Fig. 4).

The largest are the two extremes; the '-50' and the '1000' polygons:

- The '-50' buffer polygon is representing the core area of the *Kastellet_Top10DK* polygon. That is, the buffer stretching from 50m and inwards, towards the centre point.
- The '1000' buffer polygon is representing the 'shell', stretching 50-1000m.

These are the buffer distances that will capture any extreme digitizing inaccuracies.

In the end, if it turns out that a large portion of your digitized Kastellet polygon falls within these *outer* buffer intervals, you will conclude that the background image you used was not accurate enough. Then you will have to find a more accurate background image for your digitizing operation.

Know your geodata!

• The '10' buffer polygon is actually double the size of the other 'normal-sized' buffer polygons.

Why is that?

The reason is that for the innermost buffer, ArcGIS Pro has calculated the buffer on *both sides* of the *Kastellet_Top10DK* polygon outline.

That is, the '10' buffer polygon is stretching in two directions; <u>0-10m</u> and <u>-10-0m</u>.

(This is a quirk of the ArcGIS Pro software I would say. I cannot see the logic behind this.)

As a result, this buffer area ends up almost twice as large as the other.

The denomination of the other buffer polygons gets a little confusing as a consequence of this.

For example, the '20' buffer polygon (10-20m) is corresponding to the '-10' buffer polygon (-10 to -20m).

This is important to realize, because in the next step you are going to bring such **pairs of** corresponding buffer polygons together.

You will generate a new buffer polygon layer, where for example the 30-40m polygon and the corresponding -30 to -40m polygon are the same polygon.

These two polygons will obviously be separate in space (*i.e.* not be attached to each other). But, they will still be a part of *one* same polygon.

Such a polygon is called a *multipart polygon*.

distance	DistanceID
-50	1000
-40	50
-30	40
-20	30
-10	20
10	10
20	20
30	30
40	40
50	50
1000	1000

Fig. 5 The Attribute Table of the
Multiple Ring Buffer layer.
The 'DistanceID' column is
containing the same IDs for the
corresponding buffer distances.
For example, the DistanceID = 30
is representing the 20-30m buffer
and the -20 to -30m buffer.

Multipart Polygon

Multipart polygons are in fact more common than you may think.

Take Denmark for example. Denmark, like most other countries, has a lot of islands.

In a geodata of the world nations, the nation of Denmark is often defined by just one polygon, even though this (multipart-)polygon constitutes several separate islands.

To create multipart polygons you are going to use the Dissolve tool.

The Dissolve tool requires an ID that recognizes that two separate polygons actually are the same.

- 9. You will create such an ID.
 - a) Open the Attribute Table of the multiple ring buffer layer.
 - b) Use your skills to create a new column.This time you do not need decimal numbers.You will just generate a simple ID number.
 - c) Use your skills to add IDs corresponding to the ones in figure 5, above. (Hint! The Geodatabases & Attribute Tables exercise)
 - d) Make sure that the corresponding buffer polygons has received the same IDs (e.g. the 20-30m and the -20 to -30m buffers has DistanceID = 30) (Fig. 5).

You will now create your multipart polygons.

10. Use your skills to find and locate the Dissolve tool in the Geoprocessing pane.

In the Dissolve dialog:

- Input Features: Select the multiple ring buffer layer
- Output Feature Class: Navigate to a location where you would like to save the new geodata that will be generated → Name it.
- Dissolve Field(s): Select the DistanceID from the dropdown list (You may have another name for your ID column)
- Accept the defaults → Click Run.
 A new polygon map layer appears in the Contents pane and the Map view.
- 11. Use your skills to symbolize the new map layer using the DistanceID column. You will notice that the for example the buffer ranging from 10 to 20m and the one ranging from -10 to -20m will receive the same colour (Fig. 6). The two buffer polygon intervals have become one single polygon feature.

a)

b)

OBJECTID	Shape	DistanceID
1	Polygon	10
2	Polygon	20
3	Polygon	30
4	Polygon	40
5	Polygon	50
6	Polygon	1000

- Fig. 6 a) The dissolved Multiple Ring Buffer layer.

 The outline of the Kastellet_Top10DK
 layer is visible as the hashed black line.
 - b) The Attribute Table of the dissolved Multiple Ring Buffer map layer. Now one polygon feature/row may be representing several buffers.

Multipart Polygon

Dissolve tool

Notice! The rings in Figure 6 have been actively symbolized.

Digitizing

DTK =

Danmarks Topografiske Kortværk

(Topographical Maps of Denmark)

Supply and Infrastructure

SDFI = The Agency for Data

Styrelsen for Dataforsyning og

Infrastruktur (Formerly known as:

Geodatastyrelsen (GST), in Danish)

B. Digitize Kastellet polygon

Remember, as the project manager on a tight budget you opted for the cheapest possible background image that would render an accurate digitizing result.

As the resolution of a background image often go hand-in-hand with the cost you decide to try the topographic map that was developed by the SDFI/GST for use at scale 1:100 000.

This is a pretty inexpensive option.

- 12. Turn off all the layers.
- 13. From the I:\SCIENCE-IGN-CGD-UVMAT\GIS\Geodata\ folder, try to locate the TopographicMap 100k DTK-GST 200x.lyr > Add it to ArcGIS Pro. This is the background image you are going to use for digitizing the Kastellet.

The TopographicMap 100k DTK-GST 200x layer is a topographic map in raster format (Fig. 7).

It was developed for use at a scale around 1:100 000, however, you figure that it might be possible to use it at scale 1:10 000.

14. Zoom to scale 1:10 000 (lower left corner of the Map view). 1:10,000

15. Use your skills to create a new feature class of a polygon geometry type. (Hint! The Data Capture - Digitizing exercise)

Do not forget to give it the ETRS 1989 utm zone 32N projection.

- 16. Add your new, empty feature class/shapefile to ArcGIS Pro.
- 17. Use your skills to digitize the Kastellet (Fig. 7, pink area). (Hint! The Data Capture - Digitizing exercise)

Do not forget to save the polygon.



Fig. 7 The Topographical Map of Denmark, scale 1:100 000 (DTK100). The pink area represents the result of digitizing the Kastellet as a polygon feature.

Accuracy Assessment

Feature To Line

C. Assess accuracy of the digitized polygon

To assess the accuracy of the Kastellet polygon you digitized using the *TopographicMap_100k_DTK-GST_200x* map layer as a background image, you will compare it with the template.

The template is the Kastellet polygon produced for the Top10DK geodataset.

This polygon feature was produced at scale 1:10 000.

Consequently, you may assume that this piece of geodata is more accurate, and you may use it as a 'role model' (As comparison: You digitized a polygon in scale 1:10 000, but the background image was made in scale 1:100 000). Recall, in step A you created a multiple ring buffer around this template polygon.

Before you may conduct the assessment you will need to process the digitized polygon.

You will transform your digitized polygon into a new piece of geodata of (poly)line geometry type instead.

The output of this transformation is that the former polygon border/outline will become a line feature.

The reason is that we are interested in how much of the border/outline falls within each of the multiple ring buffers.

If a large proportion of the border/outline (of the polygon you digitized) falls in the outer buffers of the template polygon you may conclude that your digitized feature was not accurate enough.

Then you will need another background image with better accuracy.

To transform your digitized Kastellet polygon you will use the Feature To Line tool.

18. Use your skills to locate the Feature To Line tool in the Geoprocessing pane.

In the Feature To Line dialog:

- Input Features: Select the digitized Kastellet polygon map layer.
- Output Feature Class: Navigate to a location where you would like to save the new feature class that will be generated →Name it.
- Accept the defaults → Click Run.
 A new line map layer appears in the Contents pane and in the Map view (Fig. 8).
 Make sure it is of line geometry type.



Fig. 8 The Kastellet polygon feature transformed into a (poly)line.

Introduction

Accuracy Assessment

Overlay

Identity

Know your geodata!

To fully understand what the output is, take a minute to study each feature, both in the AttributeTable and the Map view.

Note! → → → → → →

This step is only necessary

if the output was a shapefile.

As described above, you will calculate how much of the digitized line (former polygon) that falls within each buffer interval of the Kastellet template.

To calculate this you will perform an overlay operation. You have used some of the overlay tools in a previous exercise. For this particular purpose, actually a couple of the overlay tools will accomplish the same thing.

In this case, you will opt for an Identity overlay operation (Fig. 9).

19. Use your skills to search and locate the Identity tool in the Geoprocessing pane.

In the Identity dialog:

- Input Features: Select the new (poly)line geodata layer.
- Identity Features: Select the dissolved multiple ring buffer layer.
- Output Feature Class: Navigate to a location where you would like to save the new geodata that will be generated →Name it.
- Accept the defaults → Click Run.
 A new line map layer appears in the Contents pane and in the Map view.

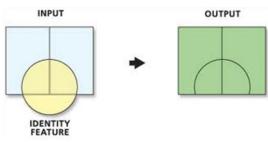


Fig. 9 The Identity overlay operation.

20. Open the Attribute Table of the newly-generated map layer (Fig. 10).

From the table you may conclude that the outcome of the Identity operation is geodata of (poly)line geometry type, and that the lines are divided into groups according to the buffer distances. This looks promising.

What you are interested in is in fact the length of the lines within each of the buffer distances/intervals.

OBJECTID	Shape	FID_Kastellet_DTK100_FeatureToLine	FID_Kastel	Id	FID_Kastellet_Top10DK_MBR_Dissolve	DistanceID	Shape_Length
1	Polyline	1	0	0	1	10	408.982917
2	Polyline	1	0	0	2	20	417.437902
3	Polyline	1	0	0	3	30	229.038082
4	Polyline	1	0	0	4	40	246.334585
5	Polyline	1	0	0	5	50	203.528882
6	Polyline	1	0	0	6	1000	96.7268

Fig. 10 The Attribute Table of the output feature class from the Identity overlay operation.

- 21. To calculate the length of the lines you will need to create a new column.
 - a) Use your skills to create a new numeric column. Make sure to allow for decimals.
 - b) Use your skills to calculate the length of the lines (Fig. 10).

Accuracy Assessment

In figure 10 you may now view the total length of digitized line within a particular buffer interval. In this example, approx. 409m of the digitized line falls within the closest buffer interval ± 10 m of the template border.

You may also compare the line from the digitized polygon with the template buffers visually (Fig. 11).

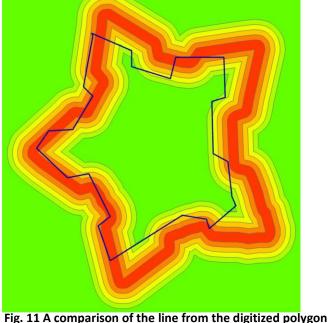
Obviously, the agreement is not too good.

However, you would like a summary that puts numbers to the level of agreement.

Within this project, the level of accuracy was set to: 90% of the border length of a digitized Kastellet polygon should be fall within the ± 10 m buffer interval.

Since ArcGIS Pro has limited capacity to make summaries of mathematical calculation, you will turn over to Excel.

22. Copy the AccuracyAssessment.xls/.xlsx worksheet from the I:\SCIENCE-IGN-CGD-UVMAT\GIS_course\Exercises\ArcGIS_Pro\



I:\SCIENCE-IGN-CGD-UVMAT\GIS_course\Exercises\ArcGIS_Pro\ Fig. 11 A comparison of the line from the digitized polygon ...\Exercise_Quality_AccuracyAssessment\ folder, to your own folder. with the multiple buffers of the template polygon.

- a) Open your copy in Excel > Type the name of the background image and scale you used for digitizing (Fig. 12).
- b) Type the line lengths exactly as they were calculated in step 22 (Fig. 12, red box) Save the file to your folder.

Background	image: DTK100 (Scale	1:10 000)		
DistanceID	Buffer interval	Length (m)	Accumulated length (m)	Fraction (%)
10	± 10m	408,985	408,99	25,5
20	10-20m and -1020m	417,436	826,42	51,6
30	20-30m and -2030m	229,138	1055,56	65,9
40	30-40m and -3040m	246,331	1301,89	81,3
50	40-50m and -4050m	203,555	1505,45	94,0
1000	50-1000m and -50m-	96,605	1602,05	100,0

Fig. 12 The Excel worksheet summary of the fraction of digitized lines within each buffer distance interval of the template.

Notice! → → → → → → →
Perhaps you ended up with an
Attribute Table with fewer than

6 rows of line measurements. How would you explain that?

Conclusion

D. Conclusion

In the Fraction column you will probably find that the closest buffer interval (± 10 m) did not contain 90% of the lines. Therefore, your conclusion must be that you will need a background image of better accuracy. It will cost you some more, but will guarantee you a better quality product.

- 23. a) Use another background image (e.g. the topographic map intended for use at 1:25 000; TopographicMap_25k_) and proceed with the B-D procedures (step 12-22) once more.
 - b) In the Excel worksheet: Type the length measurements in the empty section below your first attempt (Fig. 13)

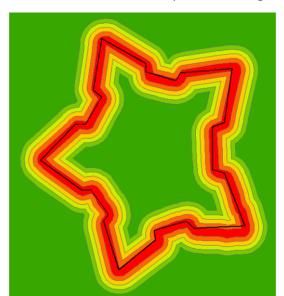
 Save.

If the better resolution topographic background image did not work either, perhaps an orthophoto is an option.

24. Repeat the procedure again, using an orthophoto background image of your choice.

Recall: With a limited project budget, the objective is to minimize the cost-benefit ratio.

- The newer the orthophoto, the higher the likelihood of higher resolution images, but also the higher the price.



Background	image: DTK100 (Scale			
DistanceID	Buffer interval	Length (m)	Accumulated length (m)	Fraction (%)
10	± 10m	408,985	408,99	25,5
20	10-20m and -1020m	417,436	826,42	51,6
30	20-30m and -2030m	229,138	1055,56	65,9
40	30-40m and -3040m	246,331	1301,89	81,3
50	40-50m and -4050m	203,555	1505,45	94,0
1000	50-1000m and -50m-	96,605	1602,05	100,0
Background	image: DTK025 (Scale	1:10 000)		
DistanceID	Buffer interval	Length (m)	Accumulated length (m)	Fraction (%)
10	± 10m	1645,411	1645,41	91,0
20	10-20m and -1020m	161,994	1807,41	100,0
30	20-30m and -2030m		1807,41	100,0
40	30-40m and -3040m		1807,41	100,0
50	40-50m and -4050m		1807,41	100,0
1000	50-1000m and -50m-		1807,41	100,0

Fig. 13 A visual comparison of a digitized line (using the DTK025 as a background image) with the template buffers.

The lower section of the Excel worksheet shows a summary of the fraction of digitized lines within each template buffer distance interval (blue box).

Eventually, by using a background image of better accuracy, you should have been able to digitize a Kastellet polygon that falls within the quality standards (90% of the digitized line length falls within the ± 10 m buffer interval).

Then you realize that you actually may save a lot of money.

There may be high-resolution background images available for free that you may use for digitizing – Google Earth.

25. Add the Kastellet geodata you converted from a digitized Google Earth kml/kmz file in a previous exercise (Fig. 14).

Transformation may result in lower accuracy.

The fact that the digitized polygon has undergone file format and coordinate system transformation may result in a product of inferior quality, but it is worth an assessment.

26. (Optional step)

Only necessary if your digitized polygon from Google Earth ended up strange. For example, the polygon may also contain some kind of line feature.

One solution can be to try to remove the unwanted line feature. Suggestion:

- a) Use your skills to transform the polygon to a (poly-)line feature.
- b) Enter Edit mode by using the Edit ribbon
- c) Start editing, and delete the unwanted line segments.
- d) An alternative solution is to digitize a new copy of the added Google Earth polygon in ArcGISPro.



Fig. 14 The Kastellet digitized as a feature (purple) in Google Earth.

(Hint! The Data Capture - Digitizing exercise. Use the Snapping environment and the Trace tool.)

Metadata

Metadata?!

Did you by any chance take notes of your digitizing method in Google Earth?

Did you take a note of the background image you used (pixel size, year)? Scale (eye altitude)? Owner/Provider?

What about the original file format and if the geodata went through any transformations?

Anything...?

In GIS everything hangs together.

Every part of the process – from the Data Capture (e.g. digitizing) to the Analysis and the Presentation – relies on that you are well acquainted with your geodata.

The best way to know your data is to keep metadata.

Let's say you received the Kastellet feature class from someone else. How should you know the quality of this piece of geodata? Is the quality and resolution good enough to use for area measurements?

If you actually <u>did</u> scribble down some notes about how your Kastellet feature was generated, please let me know.

I will gladly shake your hand.

On the other hand, if you did not collect any metadata for your Kastellet feature, I strongly suggest that you turn back to the Search Geodata exercise notes to recollect what metadata are, and what they are good for.

-Let this be a lesson ©



25,5

51,6

65,9

81,3

94.0

100.0

91,0

100.0

100.0

100.0

100.0

100,0

100,0 100.0

Exercise

27. Use the Kastellet feature from Google Earth and proceed with the C-D procedures (step 18-22) once more. In the Excel worksheet: Type the length measurements in the empty section below your last attempt (Fig. 15). → Save.

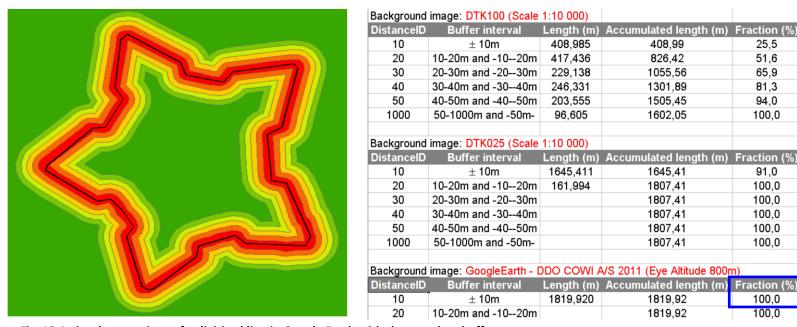


Fig. 15 A visual comparison of a digitized line in Google Earth with the template buffers. In this example, digitizing was performed using an orthophoto from COWI A/S at Google Earth as a background image. The lowest section of the Excel worksheet shows a summary of the fraction of digitized lines within each template buffer

Depending on the scale (Eye Altitude) during digitizing and other factors your results may differ from those in figure 15. What could you have done differently?

28. Hand in (as ONE document):

- The Excel worksheet containing the line measurements like in figure 15.
- Four images showing the overlay comparison between your 4 digitized Kastellet lines and the template buffers (Like the examples in the figures 11, 13 and 15)

Thanks for your attention!