Area uncertainty

Miha Kadunc, Marija Vidmar

November 2010

Contents

1 Introduction

We would like to make an estimation of area uncertainty based on an uncertainty of a point. We will study different properties of polygons which represent parcels in real world. For basic mathematical analysis we will first study triangles because all other polygons can be studied as an analogy to them, and they represent the simplest of all poygons.

The general idea is that the position of the vertices of the polygon is not accurate but uncertain and distributed according to some distribution. To make a good approximation of the real world uncertainty we picked two-dimensional, radial symmetrical Gaussian distribution (normal distribution). An example for three different triangles can be seen on Figure 1. The uncertainty of any vertex is an input parameter and as output we will study area uncertainty (that can depend on area or perimeter) and limits within our model is a good description.

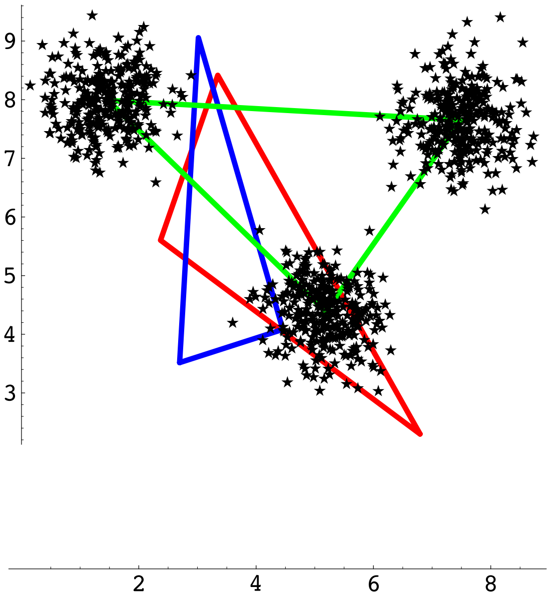


Figure 1: Different triangles and radial symmetrical normal distribution of uncertainty of a point around vertices

2 Mathematical background

We will use some basic mathematical properties of a rectangle: area A:

$$A = \frac{1}{2}[x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)], \tag{1}$$

and perimeter p

$$p = a+b+c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} + \sqrt{(x_3 - x_3)^2 + (y_3 - y_2)^2} + \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2}, \tag{2}$$

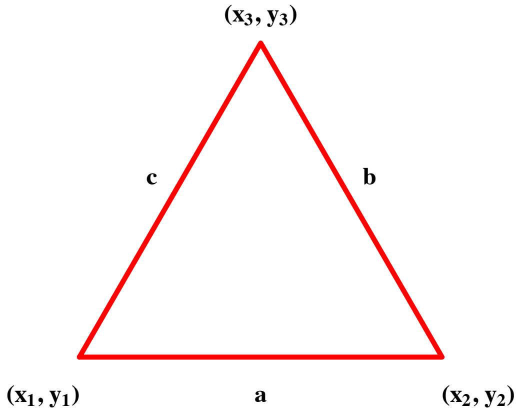


Figure 2: Triangle with sides and vertices

where x_i, y_i are the coordinates of vertices and a, b, c are the sides of the triangle as it can be seen on Figure 2.
Area uncertainty of a polygon can be defined analytically [1] as

$$\sigma_{AN}^2 = \frac{1}{4}\sum_{i=1}^{N}[\Delta y_{i-1,i+1}^2 + \Delta x_{i-1,i+1}^2]\sigma_0^2, \tag{3}$$

where σ_0 is the uncertainty of a vertex (we assume that is is the same for all vertices), and (x_i, y_i) are the exact values of the coordinates. In case of triangle the above equation can be simplyfied into

$$\sigma_{AN}^2 = \frac{1}{4}[(y_3 - y_1)^2 + (x_3 - x_1)^2 + (y_1 - y_2)^2 + (x_1 - x_2)^2 + (y_2 - y_3)^2 + (x_2 - x_3)^2]\sigma_0^2. \tag{4}$$

3 Numerical analysis - study of triangles

We want to compare analytical area uncertainty (4) with *numerical area uncertainty* that we calculate from the distribution of vertices. For each triangle we perform Monte Carlo simulation – we randomly pick a point around each vertex according to normal distribution N times. For each index i we join the three randomly picked points into a triangle and calculate area and after having N values for area we can calculate area distribution and area uncertainty. We repeat this for M different triangles.

The interval where we pick points is limited to [0, a], $a = 1$ and we choose $N = 10^5$ and $M = 1000$ unless differently said. For vertex uncertainty σ_0 we choose $\sigma_0/a = 0.01$.

The area distribution for one particular triangle can be seen on Figure 3 (a). The red line represents the average area value ($\bar{A}$), which is the same as the area, calculated from the exact value of vertices A [equation (1)]. On Figure 3 (b) we can see the distribution in a histrogram form. We can see that the area distribution is also a *normal distribution*. Numerical area uncertainty – σ_N – can be interpreted as the *standard deviation* of this distribution.

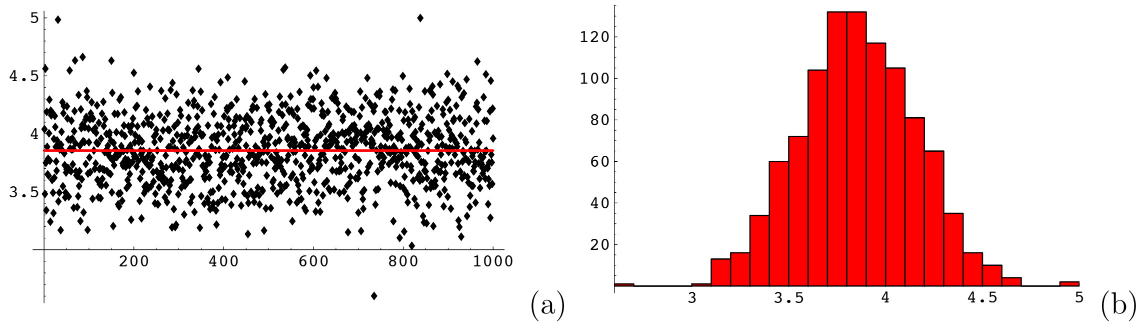
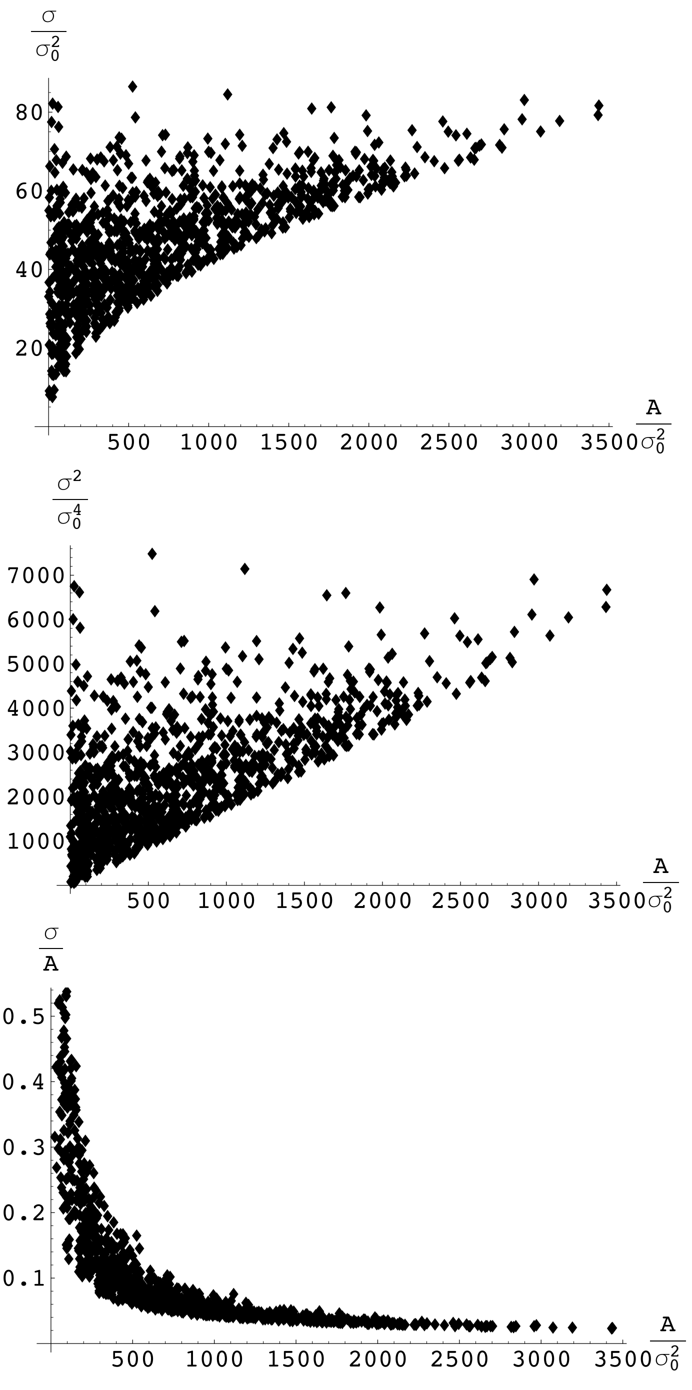


Figure 3: (a) The distribution of numerically calculated areas around analytical area (red line); (b) histrogram of the distribution - it is the normal distribution and the standard deviation correspondes to analytical area uncertainty [equation (4)].

3.1 Area uncertainty dependancy on area

We would like to know what area uncertainty σ_N depends on. First we look at the correlation between area uncertainty and area on Figure 4 (a). We can see that area uncertainty is limited downwards – when the area is large enough we can not expect uncertainties to be small; but at one particular area we can get many different values of uncertainties. On Figure 4 (b) we see area uncertainty squared – σ_N^2 – here too we can see the bottom line but nothing specific can be said about area uncertainty at a given area.

It can be easily seen that we get the same result if one side of a triangle is parallel to x axis. From now on we will study triangles, defined with vertices $\{\{0, 0\}, \{x_2, 0\}, \{x_3, y_3\}\}$ unless otherwise stated. From Figures 4 we can see that the area alone is not a very good parameter when studying general shapes of triangles (or other polygons). At the same area values we can get many different values of area uncertainties. We can assume that the shape of the triangle could also be an important property to study.
However, if we observe relative area uncertainty σ/A, we can see the potential trend – Figure 4 (c). For smaller areas we have big relative area uncertainty (REA) and as the area gets larger the relative area uncertainty gets smaller. This result will be studied in much more detail on the case of 3 rectangles.



(a)

(b)

(c)

Figure 4: (a) Area uncertainty dependancy on area – $\sigma_A(A)$; (b) the square of area uncertainty dependancy on area – $\sigma_A^2(A)$; not much can be said about the correlation of both, at one particular area we can have many values of uncertainty; (c) relative area uncertainty dependancy on area is apower law. All variables are represented in non-dimensional form.

3.2 Area uncertainty dependancy on perimeter

On Figure 5 we can see the area uncertainty dependancy on perimeter of the triangle $\sigma(p)$. The trend is limited between two linear borders that can be interpreted as two limit shapes of triangles - equilateral triangle and a special type of isosceles triangle that has one side much shorter than the other two sides.

Analytically we can see that from equation (4):

$$\sigma(\text{equilateral}) = \sqrt{\frac{3a^2}{4}}\sigma_0 = \sqrt{\frac{3p^2}{36}}\sigma_0 = \frac{p}{2\sqrt{3}}\sigma_0,$$
$$\sigma(\text{isosceles}) = \sqrt{\frac{2a^2}{4}}\sigma_0 = \sqrt{\frac{p^2}{8}}\sigma_0 = \frac{p}{2\sqrt{2}}\sigma_0,$$

where in the equilateral case we insert $p = 3a$ and in the isosceles case $p = 2a$ (the short side being practically 0). The bottom border are the equilateral triangles , the upper border the isosceles triangles.

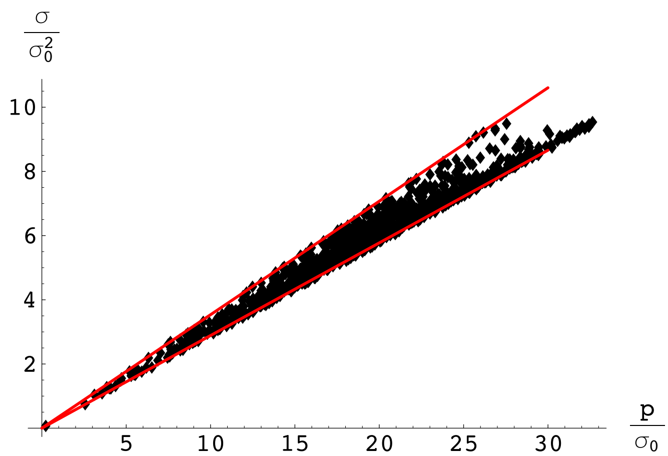


Figure 5: Area uncertainty dependancy on perimeter: it is limited between two linear borders, the upper representing the isosceles triangles and the lower representing the equilateral triangles. All variables are represented in non-dimensional form.

Next thing we can observe is how area uncertainty depends on area at a constant perimeter. We believe this will give us some new information about an impact of the shape of a triangle on area uncertainty. For one particular triangle we can see the behaviour on Figure 6. The area uncertainty is smallest at the biggest area, at small areas we have two borders: the upper and the lower.

From all cases we can find out the ones with the smallest area uncertainty and we can draw a few – Figure 7 (a). We can see these are the triangles who are shaped like equilateral triangles or are similar to equilateral triangles. The same can be done for the case of small area and great area uncertainty – we look at the triangles, represented by the points in the upper left part of graph from the Figure 6. They are shown on Figure 7 (b). As we see these are the ones, similar to isosceles triangles with one side much shorter

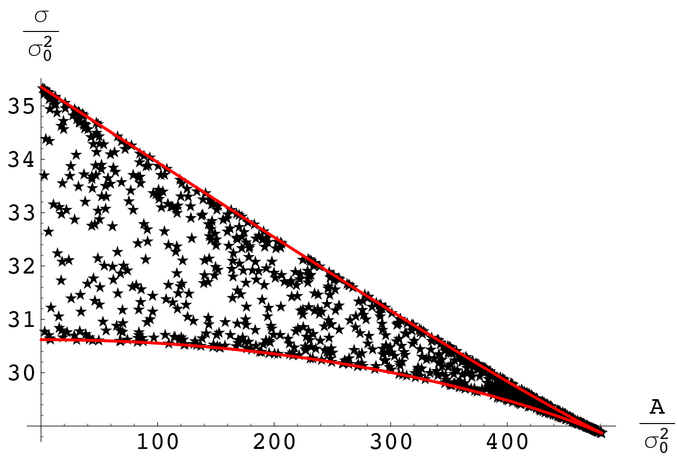


Figure 6: Area uncertainty dependancy on area at constant perimeter; the smallest uncertainty belongs to the triangles with the biggest area – these are the equilateral-like ones; the biggest uncertainty belongs to another special type of trinagles - isosceles-like with very small angles between the longer sides.

than the other two.

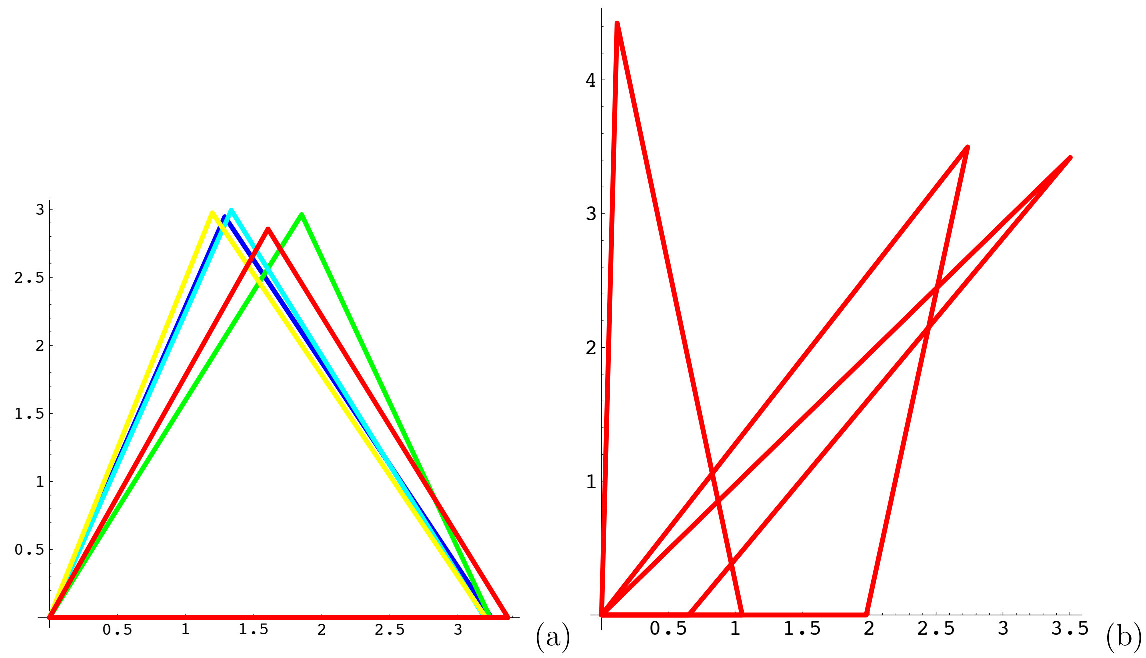


Figure 7: (a) Triangles with the smallest area uncertainty – the ones that lie on the right side of the graph from Figure 6 – they are similar to equilateral triangles; (b) triangles with the biggest area uncertainty – they are similar to isosceles triangles with one side much shorter than the other two.

We can observe the borders in more detail; for that purpose we do not generate the triangles randomly but we observe only equilateral and isosceles triangles – we want to learn more about convergence to analytical boders. On Figure 8 (a) we can see the convergence for an interval of small ratios vertex uncertainty/side ($\sigma_0/a \sim 0.01$). The special case of isosceles triangle was generated with the ratio between the shorter and the longer side $k = 10$. We can see that the convergence is good.

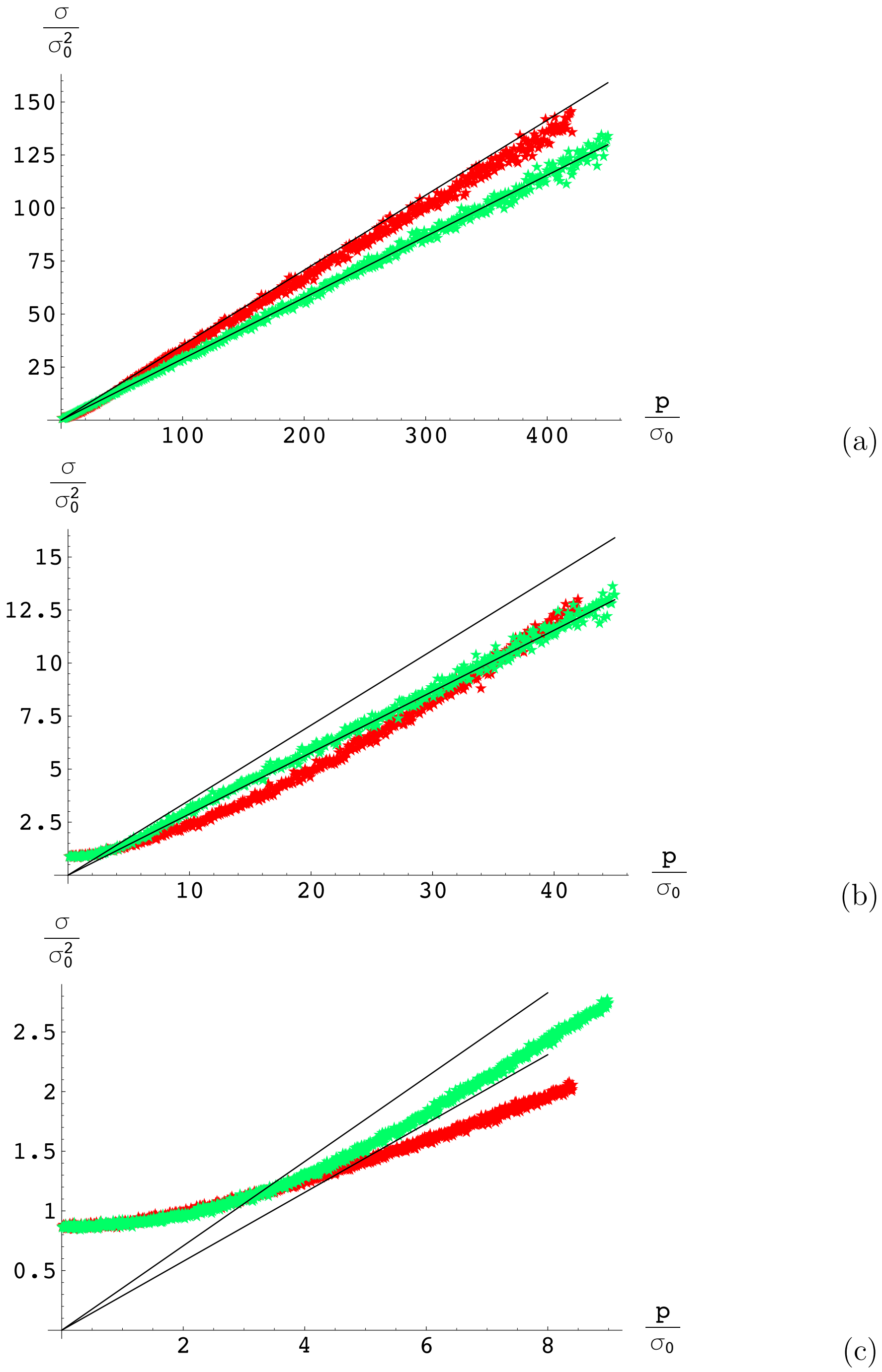


Figure 8: Convergence to the borders of equilateral triangle (green line) and iscosceles triangle (red line): the dependancy of the area uncertainty on perimeter – $\sigma(p)$; (a) small ratio $\sigma_0/a \sim 0.01$; (b) bigger ratio $\sigma_0/a \sim 0.1$; (c) large ratio $\sigma_0/a \sim 0.5$; a is the size of the side

Now we enlarge the uncertainty of the vertices on ratio $\sigma_0/a \sim 0.1$. – Figure 8 b. We can see that the equilateral triangles still converge, but the isosceles are already under the limit of convergence. We can also see that for small values of p/σ_0 the numerical model isn't good anymore. If we numerically calculate uncertainty with statistcs from Monte Carlo we can never get to uncertainty that would equal to zerro (even if the sides are very small and analytically uncertainty limits to zerro). That can be even better seen on Figure 8 (c), where the ratio $\sigma_0/a \sim 0.5$ which is quite large.

We can get an analog information if we observe when numerically calculated area (the average of all triangles obtained with Monte Carlo) starts to differ from the analytical area. On Figure 9 we see this happens when the ratio between perimeter and area is $p/\sigma_0 \approx 10$.

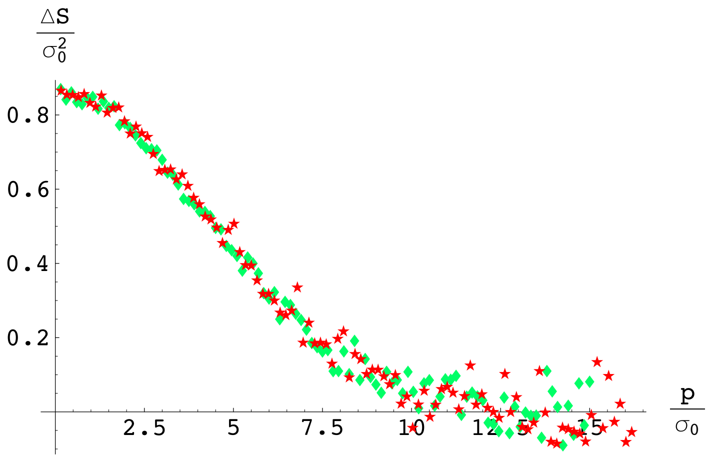


Figure 9: The difference between numerically and analytically calculated area that depends on perimeter – practically this expains at which ratio p/σ_0 we can no longer trust the numerical calculation. The red points represent the equilateral case, the green ones isosceles. We can see this happens at ratio $p/\sigma_0 \sim 10$ – when uncertainty of a vertex is one third of a side of a triangle (in case of equilateral triangle). Based on this result we will only use sides greater than 3 σ_0 in future simulations.

3.3 Uncertainty of uncertainty

We would like to know how uncertain is area uncertainty. We know that area uncertainty represents the standard deviation of the area distribution. Now we want to see how the area uncertainty changes with the number of triangles included in a Monte Carlo simulation.

On Figure 10 (a) we can see how the average area uncertainty behaves as we make M, the number of triangles, greater. When M is small, the distribution is not well defined and area uncertainty changes its value a lot; when M gets bigger, the numerically calculated uncertainty converges towards the analytical value of area uncertainty [equation (4)]. On Figure 10 (b) we can see the uncertainty of the area uncertainty – this too at the small number of triangles changes a lot but than converges. Another thing we can observe is the distribution of the area uncertainty – Figure 11, which is also normal. From all these we can conclude that the uncertainty of the area uncertainty is a standard deviation of area uncertainty.

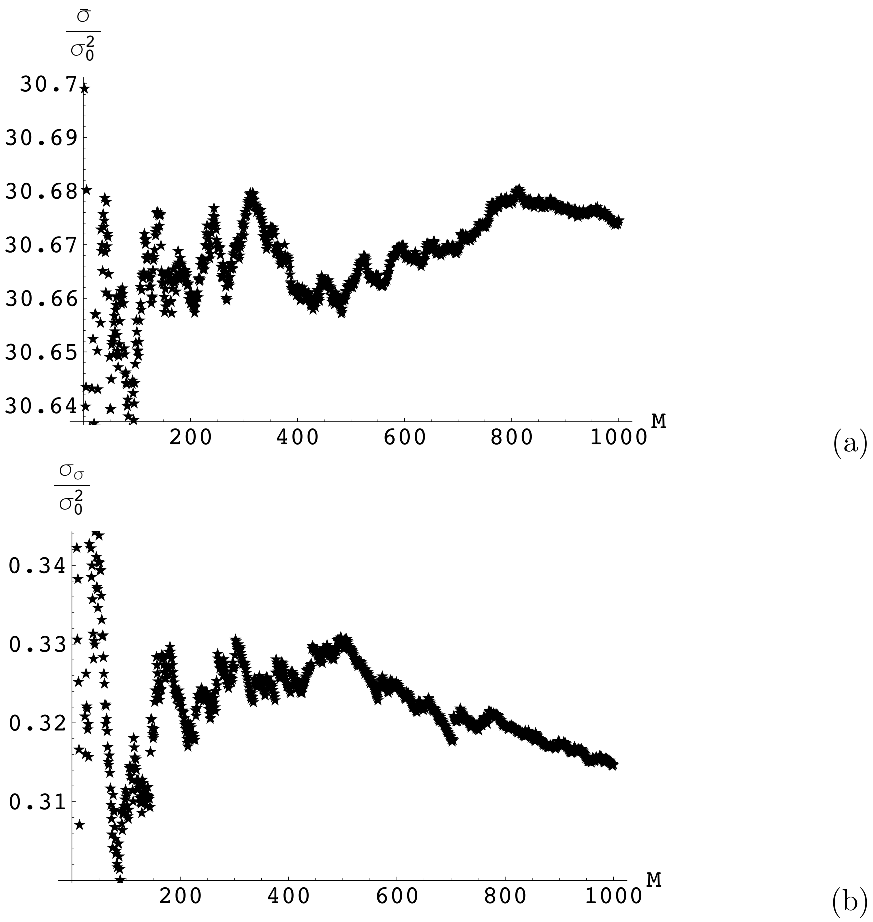


Figure 10: (a) The convergence of area uncertainty towards the analytical value when M – the number of triangles in Monte Carlo simulation – gets bigger; (b) the uncertainty of area uncertainty. Both examples are done for one randomy chosen triangle.

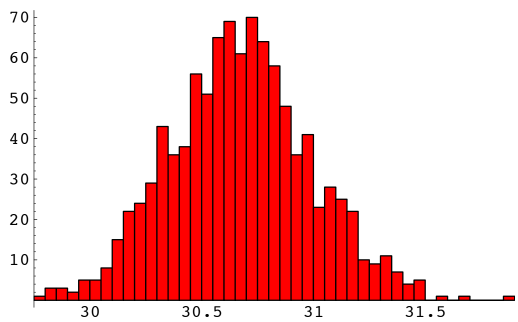


Figure 11: The distribution of area uncertainty is also a normal distribution and therefore the uncertainty of area uncertainty can be interpreted as a standard deviation of uncertainty.

3.4 Buffer analysis

Untill now we have analyzed area uncertainty based on vertex uncertainty. Another perspective is the buffer analysis. For the width of the buffer we take the same value as we usually took for the uncertainty of the vertex, σ_0. Inside and outside of the triangle we draw the buffer as it can be seen on Figure 12.

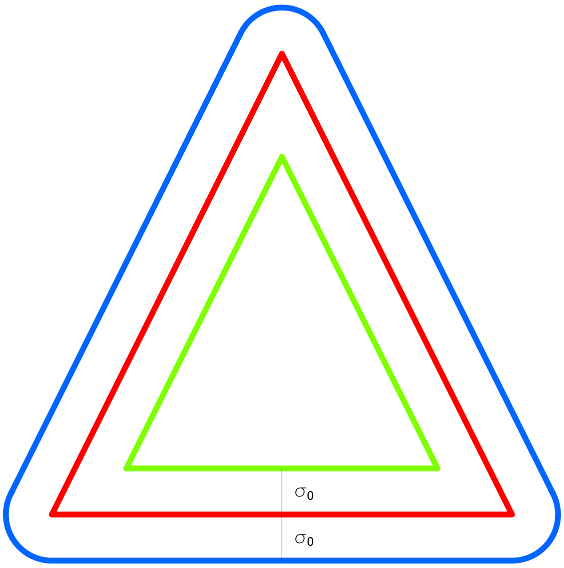


Figure 12: Triangle with an outer buffer (blue line) and inner buffer (green line)

We will observe two variables: *buffer area* A_B, that is the area of the outer and inner buffer, and *full area*, that is the area of the triangle together with the area of outer buffer, A_F. We will compare buffer area A_B with area uncertainty σ_A and full area A_F with the average area $\bar{A}$ of the triangle. To do that we will observe the correlation between variables. For the calculation of σ_A and $\bar{A}$ we still use Monte Carlo simulation – for each of the M triangles we do N perturbations while buffer is calculated only once for each of M triangles.

On Figures 13 we can see the comparisons for three different ratios σ_0/a. On Figure 13 (a) we see the correlation between area uncertainty σ_A and buffer area A_B, corr(σ_A, A_B) at $\sigma_0/a = 0.01$. On Figure 13 (c) we observe the same variable, just with ratio $\sigma_0/a = 0.1$ and on Figure 13 (e) $\sigma_0/a = 0.5$. From these Figures we see something similar to what we have seen on Figures 5 and 8 – there exists a trend, limited between two limit cases of triangles which are shaped as equilateral triangle and isosceles triangle. At large ratios σ_0/a that corresponde to large buffer width we can see that we no longer have the linear trend.

On Figures 13 (b), (d) and (f) we see the correlation between average triangle area $\bar{A}$ and full buffer area A_F. At small ratios σ_0/a the correlation is linear but when the ratio (and buffer width) gets bigger, we no longer have the linear trend. That makes sense - when the buffer area is comparable to triangle area, some small values of area can never be obtained.

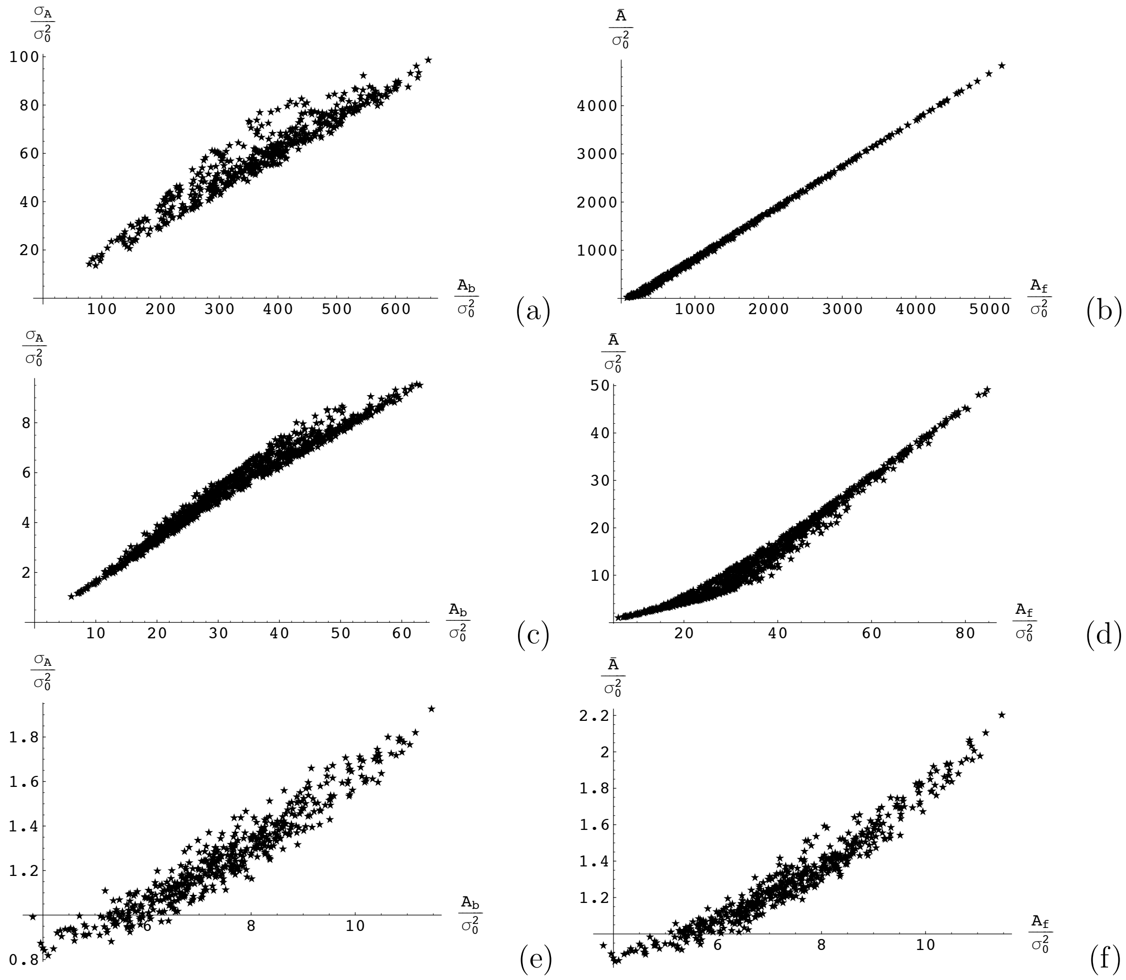


Figure 13: Correlation between area uncertainty σ_A and buffer area A_B - (a), (c), (e); correlation between full area A_F(outer buffer + triangle) and average triangle area $\bar{A}$- (b), (d), (f); from top to bottom we change the ratio between the point uncertainty and the side of a triangle – $\sigma_0/a = 0.01$ for (a) and (b), $\sigma_0/a = 0.1$ for (c) and (d), $\sigma_0/a = 0.5$ for (e) and (f).

From Figures 13 we can see that buffer is a variable, correlated with variables we have used so far – average area $\bar{A}$ and area uncertainty calculated from vertex uncertainty σ_A – but it does not give us new information about the problem of area uncertainty.

3.5 Effective area uncertainty

Untill now we have calculated area uncertainty as *standard deviation* – with an iteration

$$\sigma_A = \sqrt{\frac{1}{N}\sum_N (A - \bar{A})}. \tag{5}$$

Besides this uncertainty we can also look at the *effective uncertainty* – the only difference is that now in iteration we do not observe the difference between temporary and average area value but the differences between the temporary and analytical area value.

$$\sigma_{eff} = \sqrt{\frac{1}{N}\sum_{N}(A - A_{AN})}. \tag{6}$$

On Figure 14 a we can see σ_{eff} and analytically calculated uncertainty σ_{AN} [equation (4)] – they both depend on the length of a side a. We study this on the case of isosceles triangle with a constant perimeter $p = 2$ and ratio vertex uncertainty/side $\sigma_0/a = 0.01$. Analytical uncertainty is the concave curve, σ_{eff} is the one that differs from it (red points). The best similarity is at minimum of both uncertainties, the worst is on the borders of the interval, at very small and very big a/σ_0. This result is the same as things discovered so far – the most problematic (the ones with the biggest area uncertainty) are those (isosceles) triangles which have one side very short or very long – at $a \to 0$ the ones that can be seen on Figure 7 (b), at $a \to p/2$ the ones with a large obtuse angle between equally long sides. The smallest uncertainty is at $a = p/3$, which means equilateral triange.

On Figure 14 (b) we see the difference between numerically calculated area from analitically calculated area, on Figure 14 (c) we can see the absolute difference between both areas $\Delta A = A_{NUM} - A_{AN}$. In this example too the biggest difference is on the borders, in case of very small or very big angles at the top of isosceles triangles.

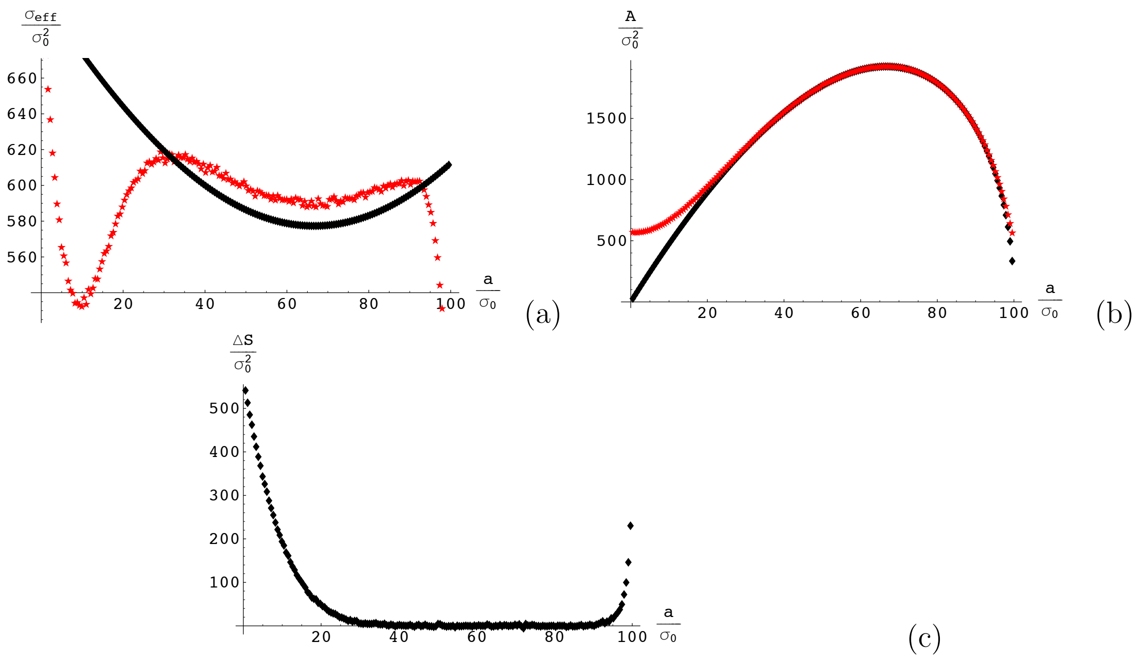


Figure 14: (a) Effective uncertainty σ_{eff} (red curve) in comparison to analytically calculated uncertainty σ_{AN} (black curve) for the case of isosceles triangles with a side a and constant perimeter $p/\sigma_0 = 200$; (b) comparison of numerical (red line) and analytical area (black line); (c) The difference between both areas.

4 Numerical analysis - study of rectangles

So far we have seen some properties about area uncertainty from the study of triangles. Now we would like to compare some results with the study of rectangles, first in general and then for some specific cases in more detail. We use the exact some procedures and formulas for simulation.

On Figure 15 we can see how area uncertainty depends on area. We can see that in general absolute area uncertainty grows with area and that at one particular area we can have different values of area uncertainty. Figure 15 is similar to Figure 4 (a) – area uncertainty dependancy on area for triagles.

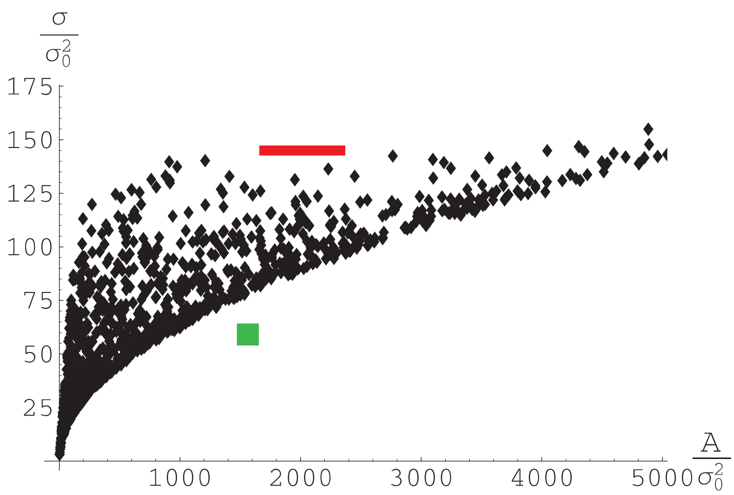


Figure 15: Normalized area error (area uncertainty) dependancy on normalized area for a rectangle; σ - area uncertainty, A -area, σ_0 - vertex uncertainty; In general, area uncertainty grows with area; at one particular area A/σ_0^2 we still can have different values of σ and the small ones correspond to the most regular shape - square like and the bigger ones correspond to elongated rectangles.

The same comparison can be done for relative area uncertainty σ/A - Figure 16. Here we can also see the power law trend. Figure 16 is similar to Figure 4 (c).

Let's compare the dependancy of area uncertainty on perimeter – Figure 17. We again have two borders - the upper represents the elongated rectangle and the lower represents the square - just like in the case of triangles where the upper border was an equilateral triangle and the lower border was an isosceles triangle. Figure 17 is similar to Figure 5.

Another thing we compare is the behaviour of area uncertainty at constant perimeter and different areas. In the case of triangles we have seen the shape on Figure 6. We remember that the smallest area uncertainties corresponded to the regular-like shape – equilateral triangle and the biggest uncertainties correpsonded to the most irregular shape - elongated isosceles triangles and ones similar to them. Area uncertainty dependancy on area at a constant perimeter for rectangles can be seen on Figure 18. This time both graphs are a bit different – on Figure 18 we have just one linear trend while on Figure 18 we had a

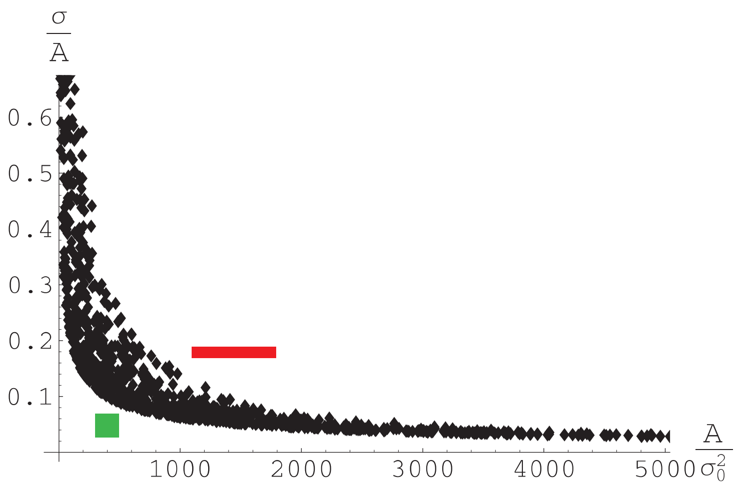


Figure 16: Relative area error dependancy on normalized area for a rectangle; σ - area uncertainty, A -area, σ_0 - RMSE; for parcels, smaller than $A/\sigma_0^2 = 1000$ - that is if $\sigma_0 = 1\text{m}$, $A = 1000\text{m}^2$ - relative area error is more than 10%. At one particular normalized area A/σ_0 we still can have different values of σ/A and the small ones correspond to the most regular shape - square like.

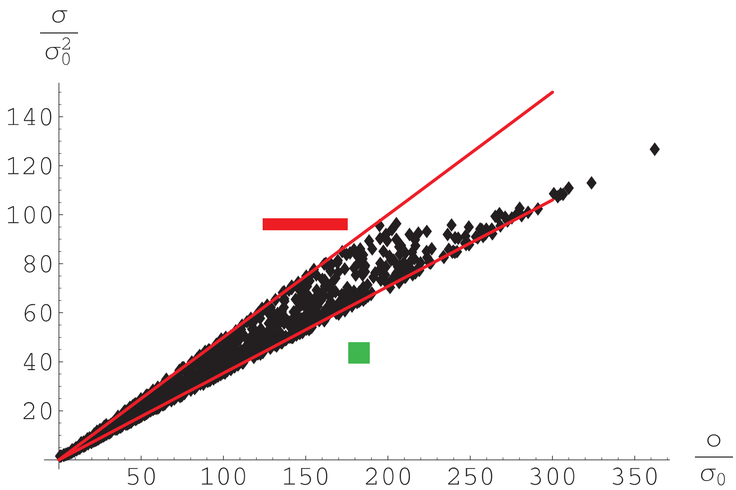


Figure 17: Absolute normalized area error (area uncertainty) dependancy on perimeter for a rectangle; σ - area uncertainty, o -perimeter, σ_0 - RMSE; σ grows with perimeter between two borders - the upper border corresponds to an elongated rectangle and the lower to a square; all rectangles fall in between

greater area - linear trend on top and a curve on bottom and everything in between in the middle. For triangles the linear trend corresponded to isosceles triangles with small angle between equal sides and the curve corresponded to isosceles triangles with big angle between equal sides. Here, in case of rectangles, we do not have two kinds of elongated rectangles - rectangle can either be elongated with large ratio between both sides or more square-like. That is why we have only one trend on Figure 18.

On Figure 19 we can see the relative area uncertainty that depends on area at constant perimeter.

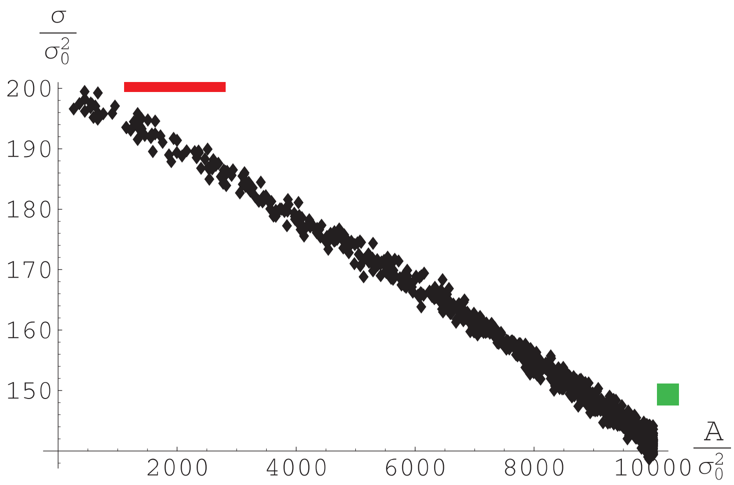


Figure 18: Absolute normalized area error (area uncertainty) dependancy on area at a constant perimeter for a rectangle; σ - area uncertainty, A -area, σ_0 - RMSE; this graph shows dependancy on how elongated the rectangle is; all rectangles have the same perimeter here and the ones with the smallest σ and the biggest A are squares, the ones with the biggest σ and the smallest A are most elongated. Between these two limits is a linear trend.

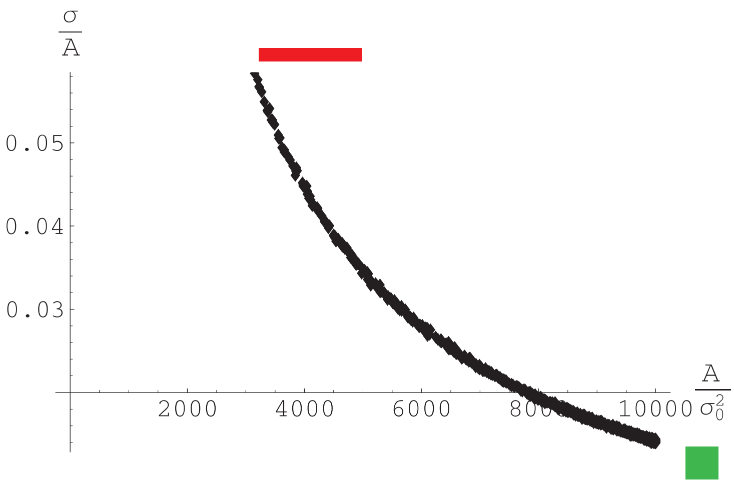


Figure 19: The same as the above graph, just for relative area error - here we can't see the linear trend, so for the study of behaviour at constant perimeter it's better to observe the absolute area uncertainty

To sum up, rectangles have the same or very similar area uncertainty properties as triangles.

5 3 cases of rectangles

Now we want to study three special cases of rectangles which have the same area but different shapes. So far we have observed variables in general, now we want to do some number comparison. Our hypothesis is that area uncertainty largely depends on the shape, not only on area (as we have seen on Figures 4, 15 and 16 we can have many different values of area uncertainty at one particular area when we observe triangles or rectangles in general).

Three cases are: *long rectangle*, Figure 1 (a), that has ratio between longer and shorter side $a/b = 30$, *middle rectangle*, Figure 1 (b), with ratio $a/b = 10$ and *square*, Figure 1 (c), with $a = b$.

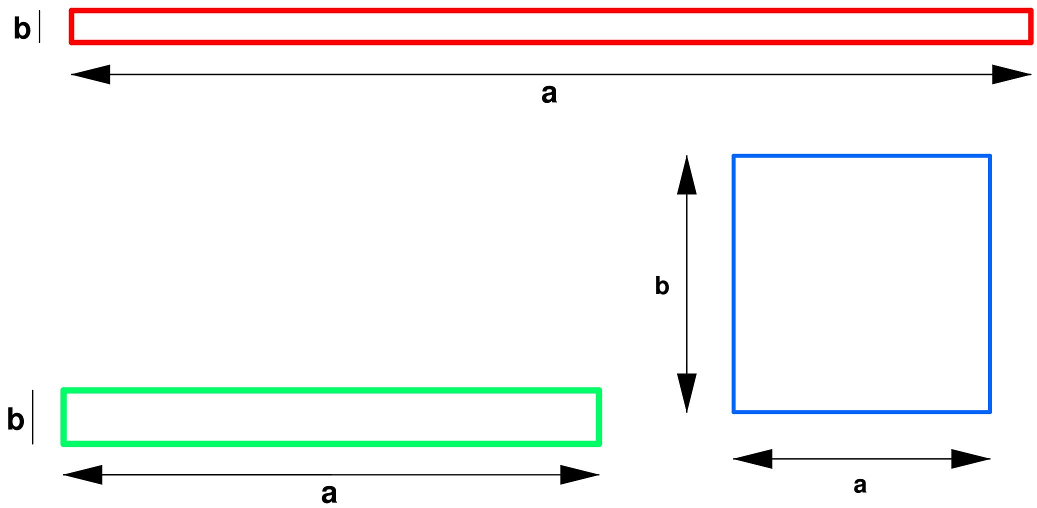


Figure 20: Three types of rectangles: (a) long rectangle, $a/b = 30$; (b) middle rectangle $a/b = 10$ and (c) square with a = b. They all have the same area.

We will scan a whole interval of areas, from a few 10 m^2 to two million m^2. The only limit is the any side of any rectangle shoud not be shorter than RMSE or vertex uncertainty times three ($3\ \sigma_0$). To make information more readable we show area scales in ha= 10^5m^2.

We will study six types of uncertainties:

- DOP (digital ortophoto): the only parameter is uncertainty of a vertex, RMSE. We estimate it on 1m. Set of parameters is (1m, 0, 0).
- DOP + DIG: 2 parameters, uncertainty of a vertex and uncertainty caused by digitalization. Set of parameters is (1m, 0.4036 m, 0).
- DOP+ DIG + INT: 3 parameters, RMSE, uncertainty caused by digitalization and uncertainty caused by interpetation. Set of parameters is (1m, 0.4036 m, 1m).
- ETS1: comparison of two cases, first is DOP + DIG + INT with parameters (1m, 0.4 m, 1m), second is DOP + DIG with parameters (2.5m, 0.4 m, 0m) - in the second case we take bigger RMSE.

- ETS2: comparison of two versions of the same case with parameters (1m, 0.4 m, 1m) for the first version and (1 m, 0.4 m, 0) for the second version. Because DOP unceraínty is the same in both versions we don't use it in calculation.

- OTS: comparison between DOP and on the spot control; first case with parameters (1m, 0.4 m, 1m) and second case with parameters (0.1948 m, 0 m, 0 m).

In first three types (DOP, DOP + DIG and DOP+ DIG + INT) we have only one set of parameters, in last three types we have two sets of parameters - that is because in first three cases we calculate area uncertainty directly while in last three cases we compare two digitalizations.

5.1 Model

Let's look at the six above types in more detail:

In the first case (DOP), to simulate the uncertainty of a vertex we take coordinates of an exact polygon (rectangle) and perform Monte Carlo on them. We calculate the area and the uncertainty of the area for each polygon. The result is the distribution of area that has a Gaussian shape – normal distribution. The average area is limiting towards the analytical value of area [equation (1)] as N, the number of times Monte Carlo was performed, grows.

For digitalization uncertainty (DIG) we used results from authors [2], specifically the relationship between the distribution of error and the turning angle. The distribution can be transformed into normal distribution with $\sigma = 1.58$ for angles $\pi/2$, and for digitalization 1:1000 this gives 0.4036 m. Because RMSE and digitalization uncertainty are not correlated, in the DOP + DIG case we can again use normal distribution with $\sigma = \sqrt{1 + 0.4036^2} = 1.078$.

Uncertainty caused by interpretation (INT): this time we estimate the error around sides of a rectangle. Again, we randomly choose from normal distrubution with $\sigma = 1$, but only once for all vertices. In this way we get an envelope around a rectangle (inside or outside).

For ETS we calculate relative error: now we want to compare two digitalizations. We run Monte Carlo twice and look at the difference between both results. Average area and average uncertainty of area in both cases shoud limit towards the same value (the exact value of the area) but if we look at the difference between digitalizations, we can get an estimate for relative area error.

Mathematically, we define relative area error as

$$\Delta = \frac{A_2 - A_1}{A_1} = \frac{A_2}{A_1} - 1, \tag{7}$$

where A_1 is the area of the rectangle in the first case and A_2 is the area in the second case.

In exact case, where $A_1 = A_2$, $\Delta = 0$. But in Monte Carlo A_{1n} and A_{2n} for the n - th try are probably not the same. We can calculate the uncertainty of relative area error by making a total derivative of the equation (7).

$$\mathrm{d}\Delta = \mathrm{d}\left(\frac{A_2 - A1}{A_1}\right) = \mathrm{d}\left(\frac{A_2}{A_1} - 1\right) = \frac{\mathrm{d}A_2}{A_1} - \frac{A_2}{A_1^2}\mathrm{d}A_1. \tag{8}$$

From here, we get

$$\sigma(\Delta) = \sqrt{\frac{1}{A_1^2}\sigma^2(A_2) + \frac{A_2^2}{A_1^4}\sigma^2(A_1)}. \tag{9}$$

In other words, we have the probability distribution $\mathrm{d}P/\mathrm{d}A_1$ for the first measurment and $\mathrm{d}P/\mathrm{d}A_2$ for the second measurment; now we can calculate the probability distribution of the relative difference between them:

$$\frac{\mathrm{d}P}{\mathrm{d}\left|\frac{A_1 - A_2}{A_1}\right|}. \tag{10}$$

The *standard deviation* of the above distribution equals to the expression (9); when we are performing Monte Carlo, we are calculating σ directly but results are the same as theory predicts.

For OTS we use the same procedure as for ETS.

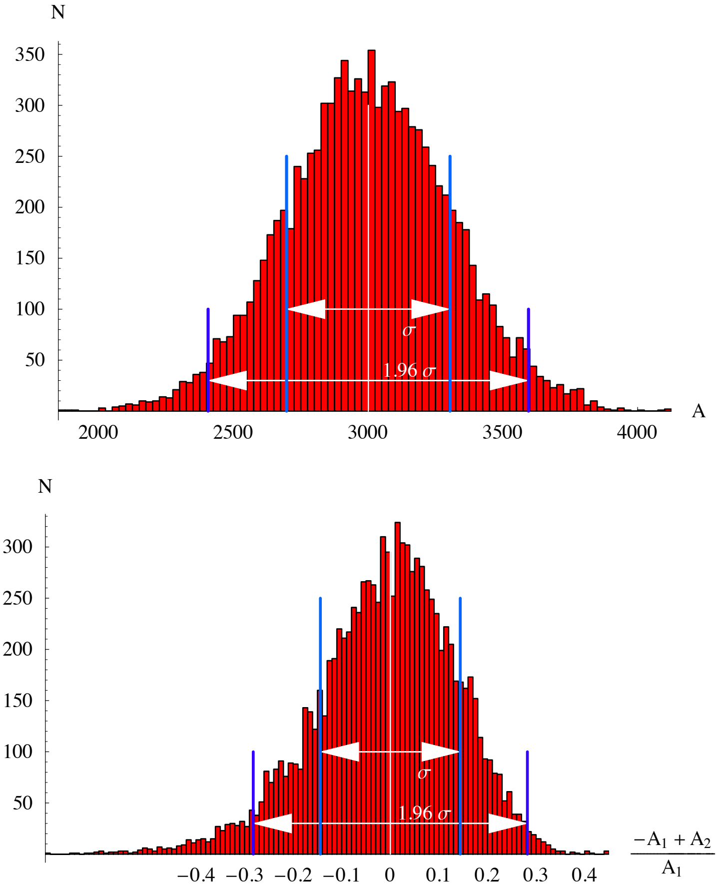
5.2 Results

We will look at the results for our three rectangles from Figure 20. We observe two things: how *relative area error*

$$REA = \sigma/A$$

depends on area and how "2 σ" interval depends on area."2 σ" is actually 1.96 σ (reproducibility limit). We can see area uncertainty – "σ" interval and 1.96 σ – "2 σ" interval on Figures 21. "σ" interval includes 67% of all cases – Figure 21 (a) and "2 σ" interval includes 95% of all cases – Figure 21 (b). The expression relative area error means the same as relative area uncertainty.

On Figure 22 we can see the shape of REA(A). It is a power law and therefore hard to read for greater areas A. Because of that we will use logarithmic scales from now on, the results will be presented as linear functions. In order to make results more readable some data is collected also in tables.



(a)

(b)

Figure 21: (a) Normal distribution of area for DOP type with "σ" and "2 σ" interval; (b) normal distribution of relative difference of areas for ETS type with "σ" and "2 σ" interval;

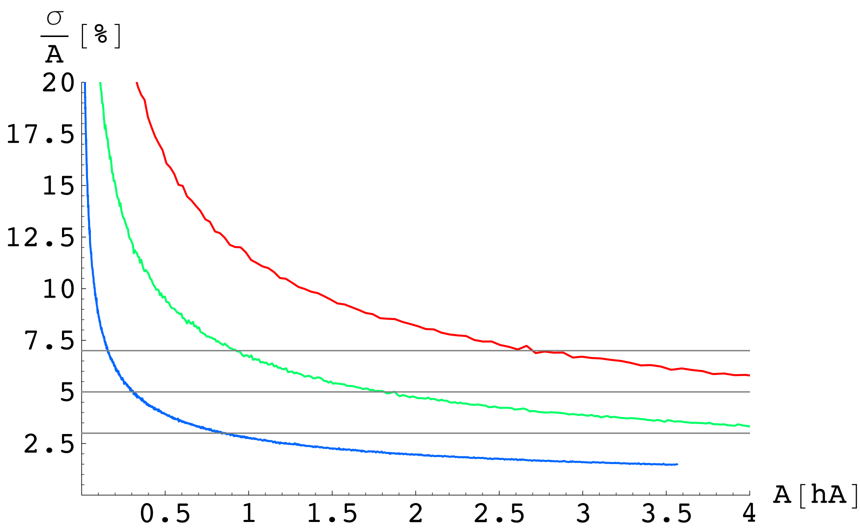


Figure 22: Linear scale example for DOP – for larger areas it is hard to read details

5.3 DOP

On Figure 23 we see relative area error that depends on area. On Figure 24 we see the "2 σ" interval. For small areas uncertainties are around 9% for square, 20% for middle rectangle and more than 30% for long rectangle. For "2 σ" scenario they are exactly 1.96 times value at "σ" at the same area. We can see that more elongated rectangles have much greater relative area error than squares at the same area. In Table 1 (first column) are listed areas (in ha) at REA = 3%, 5%, 7%. In Table 2 (first column) are listed areas where 1.96 σ/A is 3%, 5%, 7%.

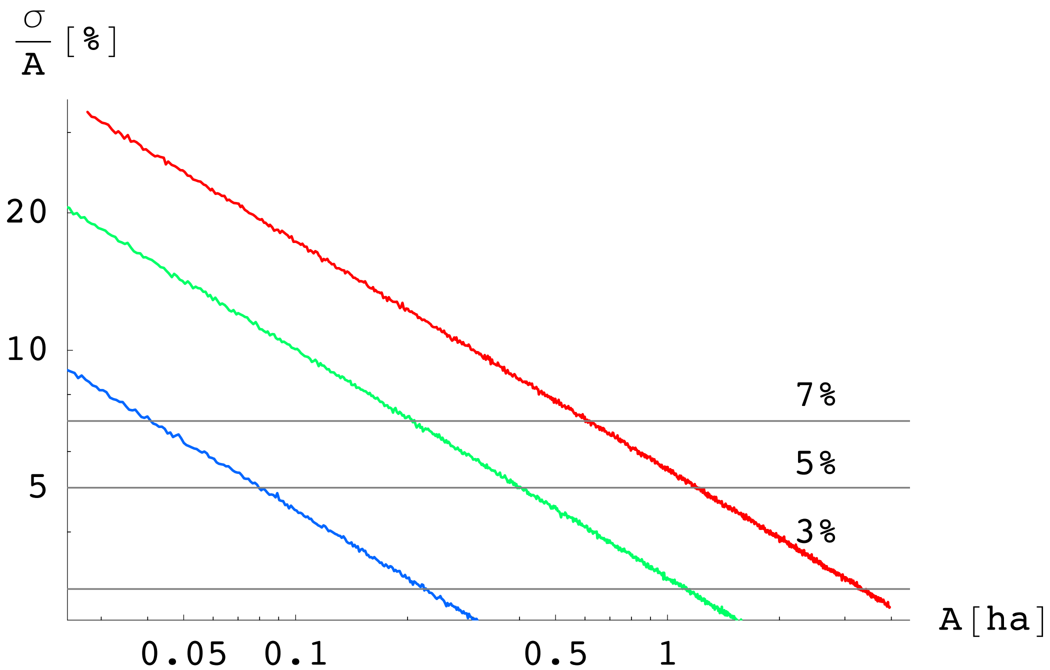


Figure 23: DOP, REA; red line - long rectangle; green line - middle rectangle; blue line - square.

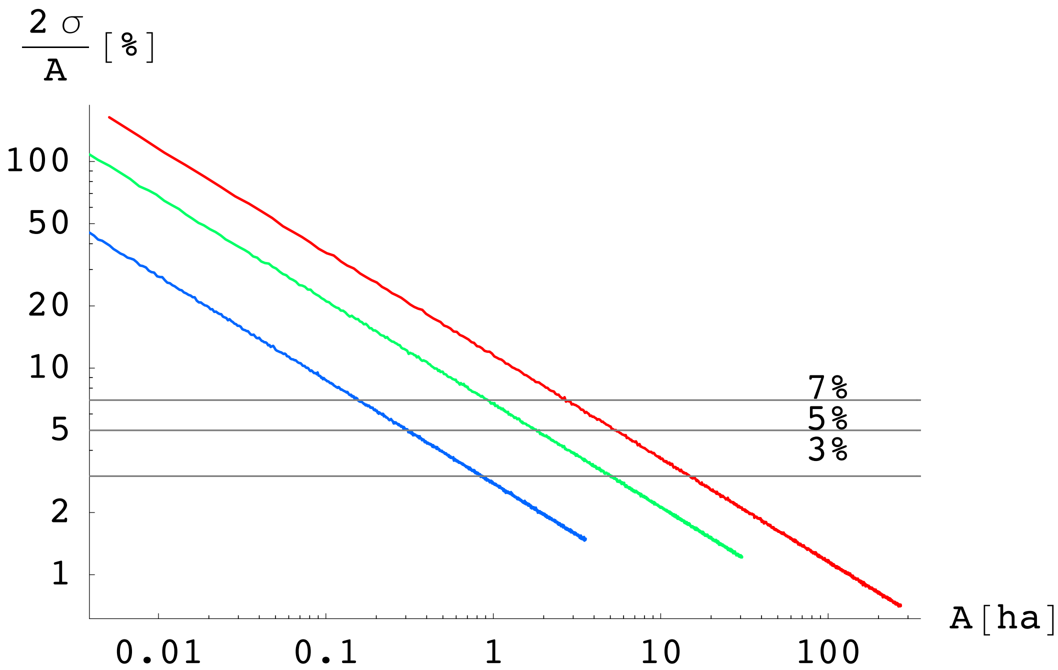


Figure 24: DOP, "2 σ" interval; red line - long rectangle; green line - middle rectangle; blue line - square.

5.4 DOP + DIG

On Figure 25 we see relative area error that depends on area for DOP + DIG scenario. On Figure 26 we see the "2 σ" interval. In Table 1 (second column) are listed areas (in ha) at REA = 3%, 5%, 7%. In Table 2 (second column) are listed areas where 1.96 σ/A is 3%, 5%, 7%.

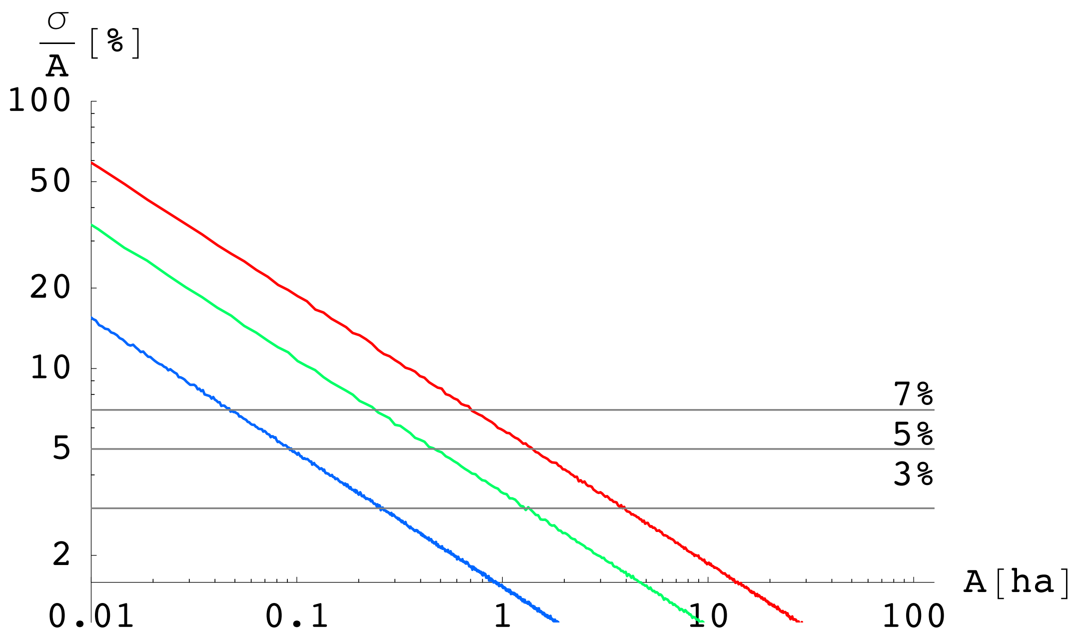


Figure 25: DOP + DIG, REA; red line - long rectangle; green line - middle rectangle; blue line - square.

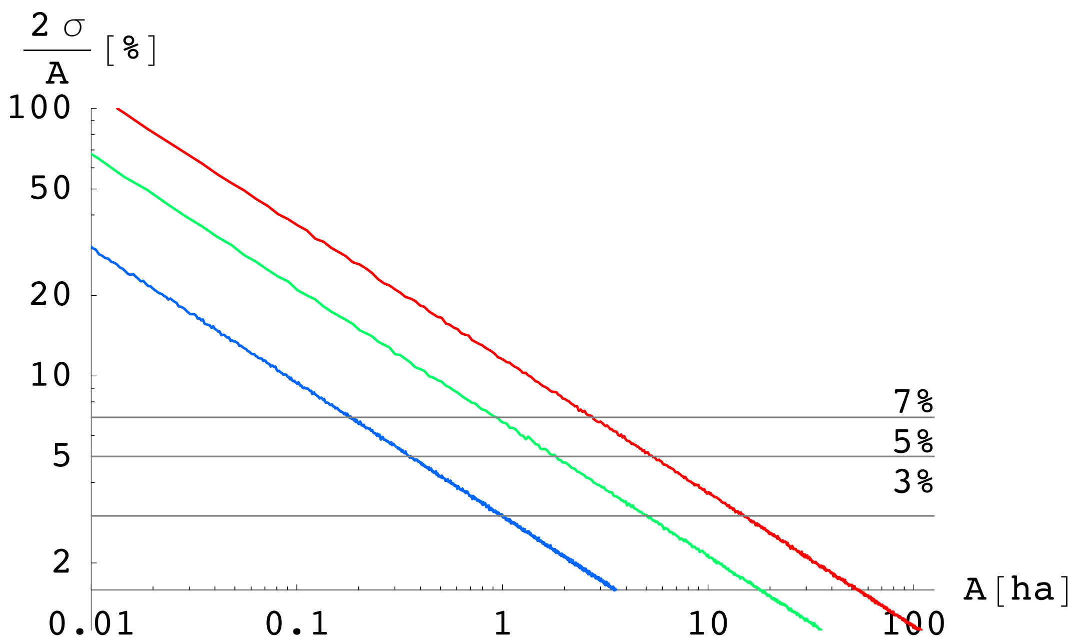


Figure 26: DOP + DIG, "2 σ" interval; red line - long rectangle; green line - middle rectangle; blue line - square.

5.5 DOP + DIG + INT

On Figure 27 we see relative area error that depends on area for DOP + DIG + INT scenario. On Figure 28 we see the "2 σ" interval. In Table 1 (third column) are listed areas (in ha) at REA = 3%, 5%, 7%. In Table 2 (third column) are listed areas where 1.96 σ/A is 3%, 5%, 7%.

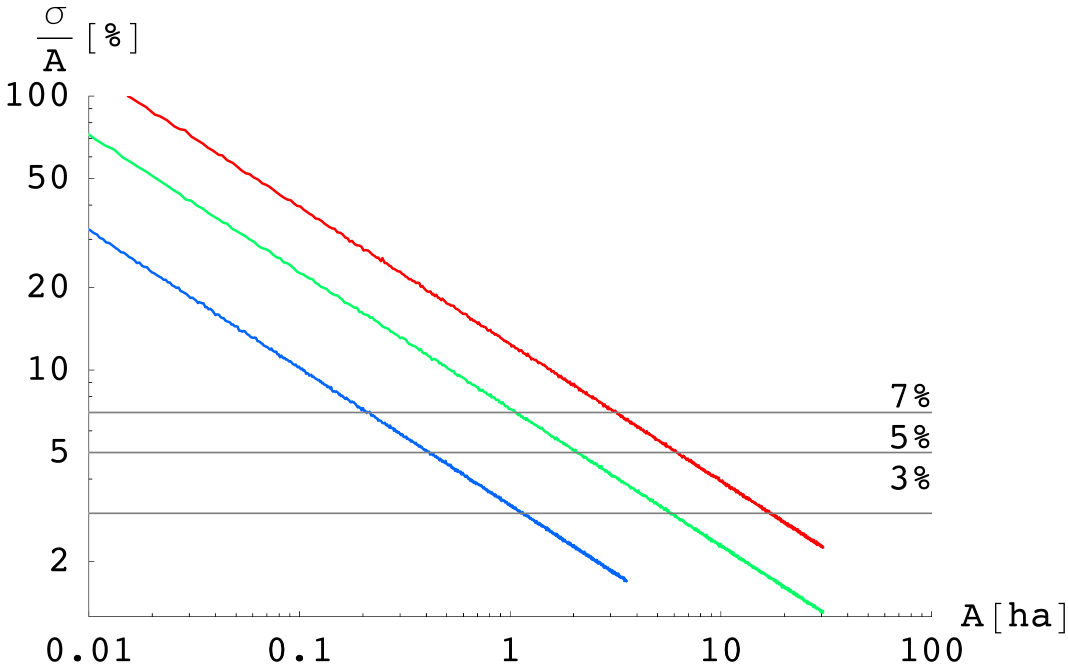


Figure 27: DOP + DIG + INT, REA; red line - long rectangle; green line - middle rectangle; blue line - square.

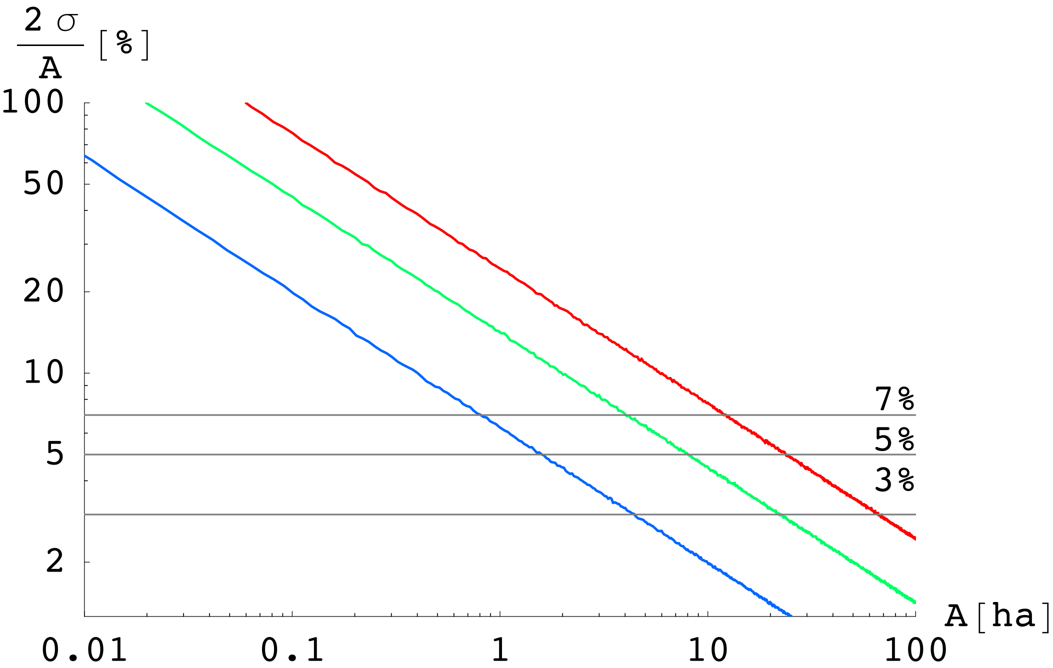


Figure 28: DOP + DIG + INT, "2 σ" interval; red line - long rectangle; green line - middle rectangle; blue line - square.

%	DOP	DOP + DIG	DOP + DIG + INT
3	0.22	0.26	1.16
5	0.08	0.09	0.42
7	0.04	0.04	0.21
3	1.125	1.31	5.86
5	0.40	0.47	2.11
7	0.21	0.24	1.04
3	3.31	3.98	17.2
5	1.20	1.41	6.18
7	0.61	0.71	3.16

Table 1: Area that has bigger relative area uncertainty than the percent on the left in ha = $10^5 m^2$; Upper triplet: square; middle triplet: middle rectangle; lower triplet: long rectangle

%	DOF	DOF + DIG	DOF + DIG + INT
3	0.85	1.00	4.48
5	0.31	0.36	1.62
7	0.15	0.18	0.83
3	5.04	5.09	22.3
5	1.83	1.83	8.11
7	0.92	0.95	4.04
3	14.9	14.9	66.26
5	5.39	5.40	23.87
7	2.71	2.77	12.18

Table 2: Area that has bigger "2 σ" error than the percent on the left in ha = $10^5 m^2$ for "2 σ" interval (95%); Upper triplet: square; middle triplet: middle rectangle; lower triplet: long rectangle

5.6 ETS1

On Figure 29 we see relative area error that depends on area for ETS1 scenario. On Figure 30 we see the "2 σ" interval. In Table 3 (upper half) are listed areas (in ha) at REA = 3%, 5%, 7%. In Table 3 (bottom half) are listed areas where 1.96 σ/A is 3%, 5%, 7%.

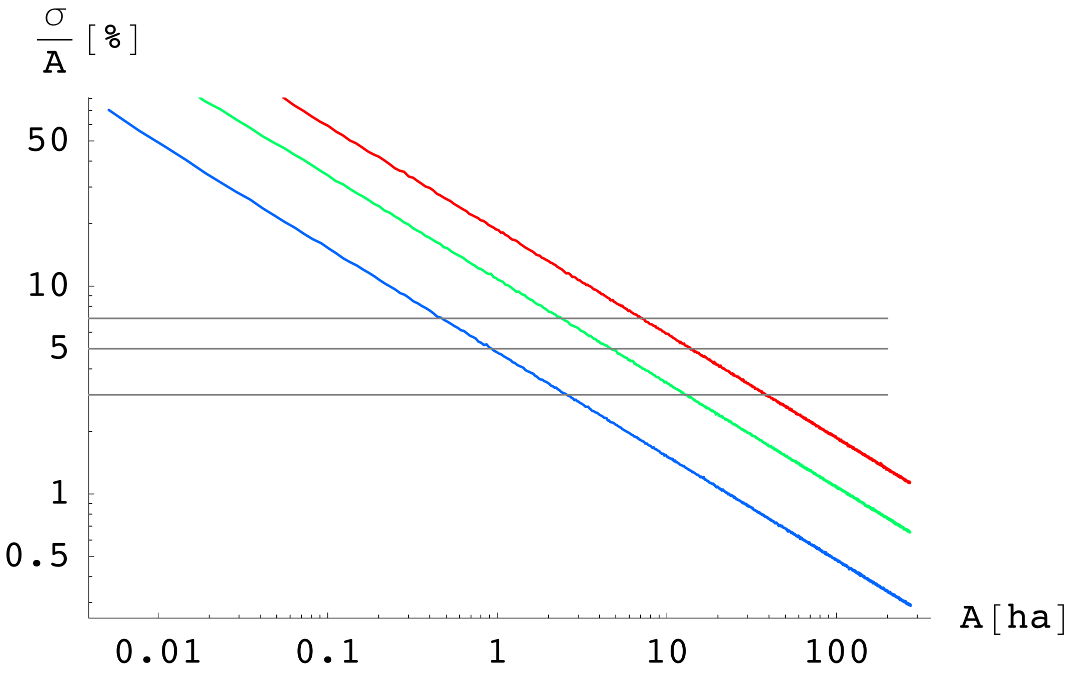


Figure 29: ETS1, REA; First (DOP, DIG, INT) = (1m, 0.4 m, 1m), Second (DOP, DIG, INT) = (2.5 m, 0.4m, 0)

%	Square	Middle	Long
3	2.60	12.99	38.45
5	0.93	4.66	13.83
7	0.48	2.41	7.20
3	9.86	50	148.37
5	3.62	18.01	53.80
7	1.80	9.17	2.35

Table 3: Area that has bigger relative area uncertainty than the percent on the left in ha = $10^5 m^2$ for ETS1 model. Upper half: REA – 1 σ interval (67%), lower half: "2 σ interval (95%)

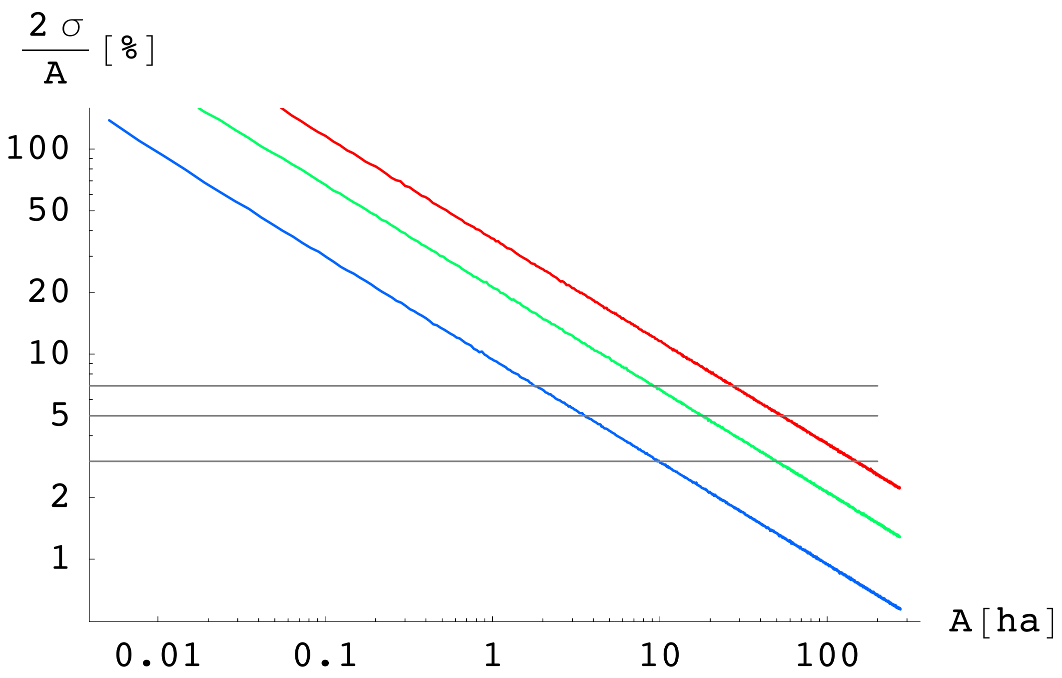


Figure 30: ETS1, "2 σ" interval; First (DOP, DIG, INT) = (1m, 0.4 m, 1m), Second (DOP, DIG, INT) = (2.5 m, 0.4 m, 0)

5.7 ETS2

On Figure 31 we see relative area error that depends on area for ETS2 scenario. On Figure 32 we see the "2 σ" interval. In Table 4 (upper half) are listed areas (in ha) at REA = 3%, 5%, 7%. In Table 4 (bottom half) are listed areas where 1.96 σ/A is 3%, 5%, 7%.

%	Square	Middle	Long
3	0.96	4.87	14.45
5	0.34	1.74	5.09
7	0.17	0.88	2.61
3	3.68	18.86	55.73
5	1.31	6.77	20.04
7	0.69	3.34	10.09

Table 4: Area that has bigger relative area uncertainty than the percent on the left in ha = $10^5 m^2$ for ETS2 scenario; upper half: REA – 1 σ interval (67%), lower half "2 σ" interval (95%).

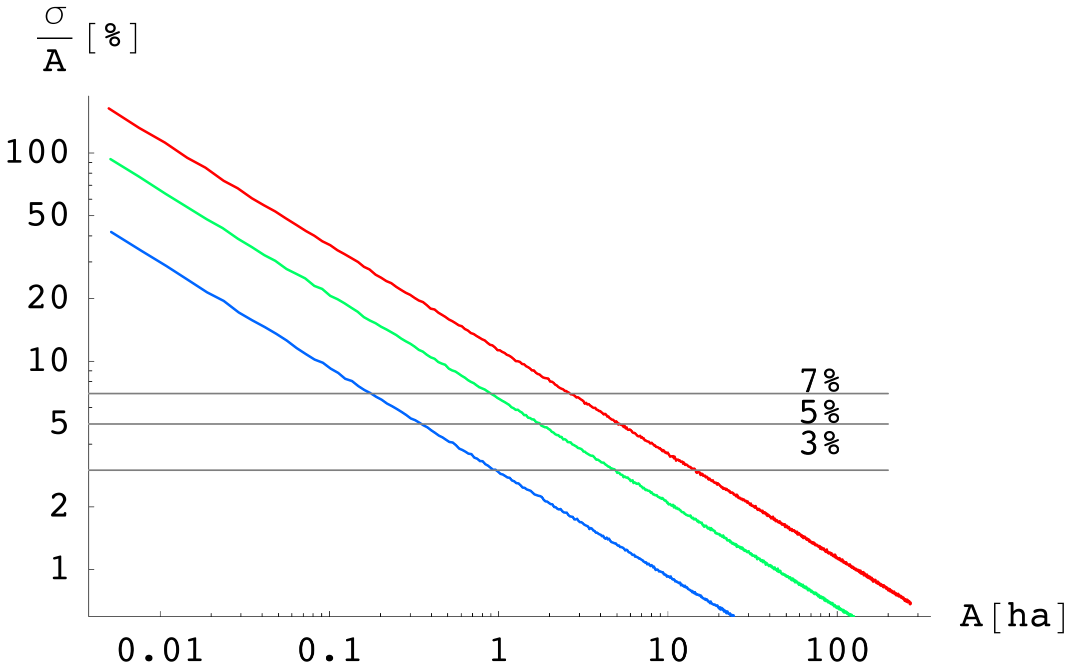


Figure 31: ETS2, REA; First (DOP, DIG, INT) = (0m, 0.4m, 1m), Second (DOP, DIG, INT) = (0m, 0.4m, 0m)

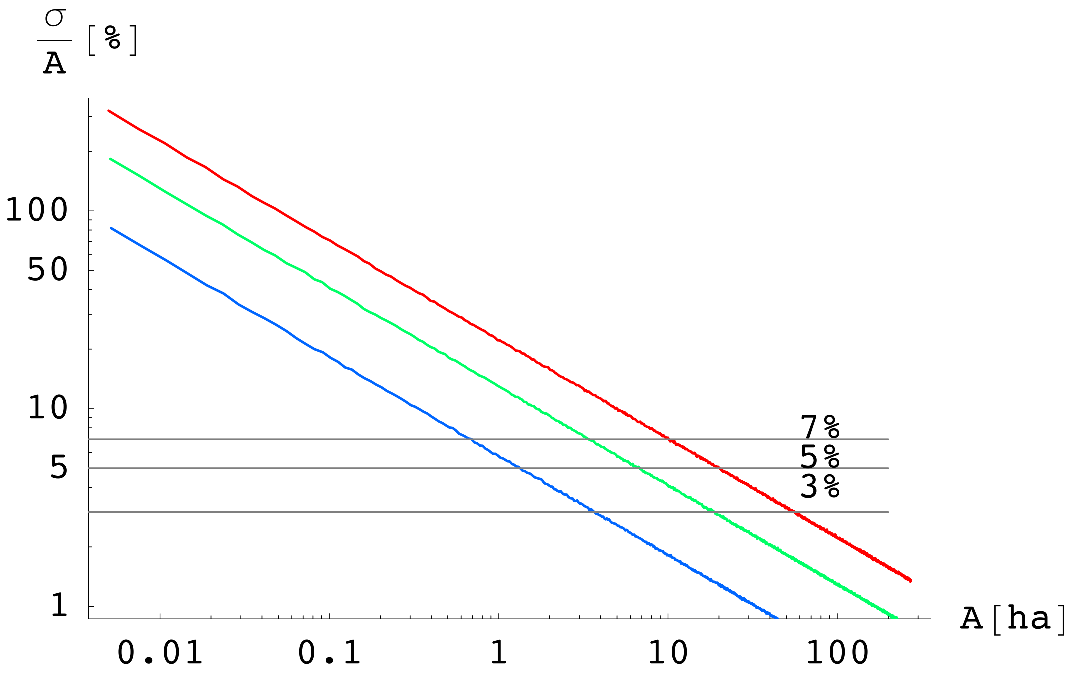


Figure 32: ETS2, "2 σ" interval; First (DOP, DIG, INT) = (0m, 0.4m, 1m), Second (DOP, DIG, INT) = (0m, 0.4m, 0m)

5.8 OTS

On Figure 33 we see relative area error that depends on area for OTS scenario. On Figure 34 we see the "2 σ" interval. In Table 5 (upper half) are listed areas (in ha) at REA =

3%, 5%, 7%. In Table 5 (bottom half) are listed areas where $1.96\sigma/A$ is 3%, 5%, 7%.

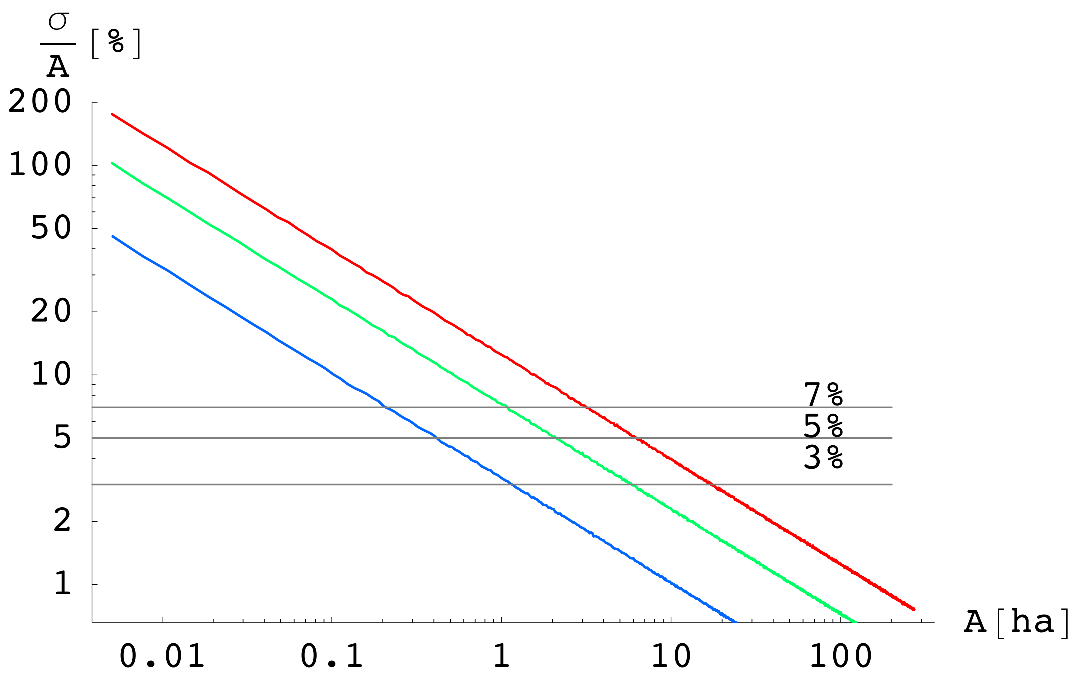


Figure 33: OTS, REA; First (DOP, DIG, INT) = (1m, 0.4m, 1m), Second (DOP, DIG, INT) = (0.195m, 0.0, 0)

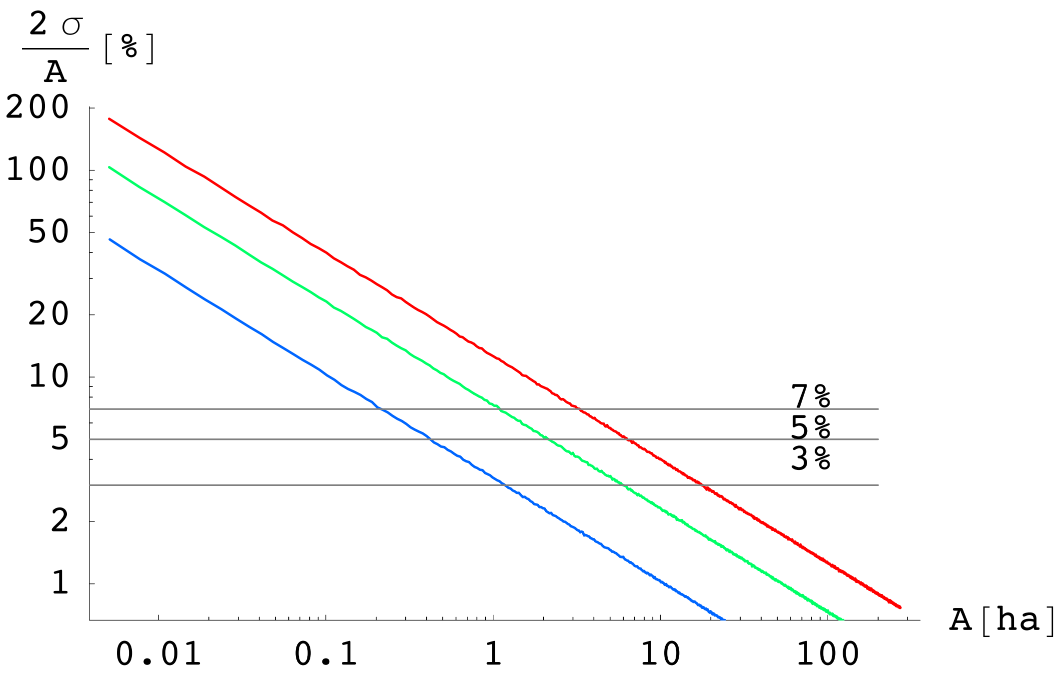


Figure 34: OTS, "2 σ" interval; First (DOP, DIG, INT) = (1m, 0.4m, 1m), Second (DOP, DIG, INT) = (0.195m, 0.0, 0)

%	Square	Middle	Long
3	1.16	5.86	17.47
5	0.42	2.11	6.35
7	0.21	1.12	3.17
3	4.48	22.45	66.514
5	1.63	8.11	24.03
7	0.83	4.04	12.29

Table 5: Area that has bigger relative area uncertainty than the percent on the left in ha = $10^5 m^2$ for OTS scenario. Upper half: REA – 1 σ interval (67%), lower half: "2 σ" interval (95%);

5.9 Comparison of REA and "2 sigma" interval

In previous tables we have compared areas at the same REAs or "2σ" intervals, now let's fix the area and compare REAs and "2σ" intervals.

%	DOP	DOP+DIG	DOP+DIG+INT	ETS1	ETS2	OTS
Square	0.137	0.14	0.32	0.48	0.29	0.32
Middle	0.31	0.32	0.73	1.09	0.65	0.74
Long	0.54	0.59	1.24	1.86	1.13	1.24
Square	1.00	1.07	2.28	3.38	2.06	2.29
Middle	2.26	2.43	5.03	7.53	4.68	5.05
Long	3.80	4.18	8.79	13.19	8.00	8.82
Square	1.41	1.54	3.17	4.74	2.88	3.19
Middle	3.16	3.41	7.17	10.71	6.57	7.19
Long	5.55	5.85	12.35	18.6	11.29	12. 4
Square	2.00	2.15	4.51	6.73	4.10	4.53
Middle	4.48	4.86	10.17	15.24	9.22	10.21
Long	7.65	8.41	17.5	26.11	15.88	17.54
Square	4.46	4.76	10.04	15.15	9.21	10.08
Middle	10.07	10.67	22.90	34.09	20.61	22.94
Long	17.23	18.65	39.24	59.09	36.10	39.52

Table 6: Comparison between different scenarios: REA at 100 ha(first - third line), 2ha (4th - 6th line), 1 ha (7th - 9th line), 0.5 ha(10th to 12th line), 0.1 ha (13th to 15th line)

%	DOP	DOP+DIG	DOP+DIG+INT	ETS1	ETS2	OTS
Square	0.27	0.27	0.63	0.94	0.58	0.63
Middle	0.61	0.62	1.44	2.13	1.28	1.44
Long	1.06	1.15	2.43	3.66	2.22	2.45
Square	1.97	2.10	4.47	6.63	4.04	4.49
Middle	4.69	4.75	9.85	14.76	9.18	9.89
Long	8.19	8.21	17.23	25.85	15.67	17.28
Square	2.76	3.01	6.22	9.28	5.65	6.24
Middle	6.68	6.68	14.04	21.00	12.88	14.10
Long	11.39	11.47	24.20	36.46	22.14	24.30
Square	3.91	4.22	8.84	13.19	8.03	8.88
Middle	9.36	9.51	19.95	29.89	18.07	20.01
Long	16.07	16.49	34.30	51.18	31.13	34.38
Square	8.73	9.32	19.67	29.69	18.06	19.75
Middle	21.01	20.90	44.89	66.81	40.40	44.96
Long	35.97	36.55	76.9	115.81	70.76	77.46

Table 7: Comparison between different scenarios: "2 σ" interval at 100ha (first - third line), 2ha (4th - 6th line), 1 ha(7th - 9th line), 0.5 ha (10th to 12th line) and 0.1 ha (13th to 15th line)

6 Further examples

6.1 3 cases of rectangles - different parameters

In previous chapter we have seen examples for all six types - DOP, DOP + DIG, DOP + DIG + INT, ETS1, ETS2 and OTS for one set of parameters, (DOP, DIG, INT) = (1 m, 0.4 m, 1m) and some variantions of that in ETS1, ETS2 and OTS types. For each type there was a REA (or "1 σ") graph and "2 σ" graph; in addition, some typical area values at 3%, 5%, 7% and some typical REA (relative area uncertainty) at A = 100 ha, 2 ha, 1 ha, 0.5 ha, 0.1 ha were listed.

We would like to compare previous results with a little change in some parameters – only for a few representative graphs and numbers.

- DOP: the only parameter is uncertainty of a vertex, RMSE = 0.2 m. Set of parameters is (0.2 m, 0, 0).
- DOP + DIG: 2 parameters, uncertainty of a vertex and uncertainty caused by digitalization. Set of parameters is (0.2 m, 0.4036 m, 0).
- DOP+ DIG + INT: 3 parameters, RMSE, uncertainty caused by digitalization and uncertainty caused by interpetation. Set of parameters is (0.2 m, 0.4036 m, 1m).
- ETS1: comparison of two cases, first is DOP + DIG + INT with parameters (0.2 m, 0.4 m, 1m), second is DOP + DIG with parameters (0.4m, 0.4036 m, 0) - in the second case we take bigger RMSE.

On Figure 35 we see the linear scale example for DOP. If we compare it to Figure 22, we can see that REA at the same area is smaller in the case of DOP = 0.2m.

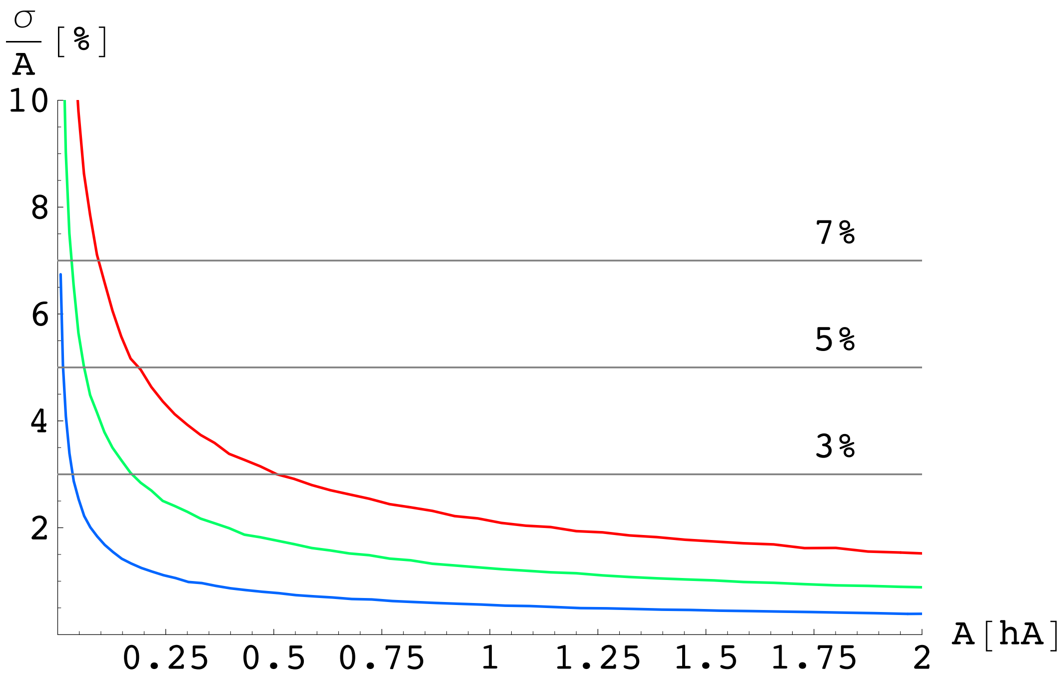


Figure 35: Linear scale example for DOP with RMSE = 0.2 m

This is even more obvious if we look at the logarithmic scale – Figure 36 for DOP and compare it to DOP with RMSE = 1m – Figure 23. For instance, at $A = 0.1$ ha, REA ≈ 1.5% (square) for RMSE = 0.2 m, while at RMSE = 1m REA ≈ 4.5% (square).

On Figure 36 we can see DOP + DIG, REA example for set of parameters (DOP, DIG, INT) = (0.2 m, 0.4 m, 0), on Figure 38 is a DOP + DIG + INT, REA example for set of parameters (DOP, DIG, INT) = (0.2 m, 0.4 m, 1m) and on Figure 39 is a DOP + DIG + INT, "2 sigma" example for the same set of parameters. On Figure 40 we can see ETS1 scenario, "2 σ" example with parameters (DOP, DIG, INT) = (0.2 m, 0.4 m, 1m) for the first case and (DOP, DIG, INT) = (0.2 m, 0.4 m, 0) for the second case.

In Table 8 are listed relative area uncertainties for "2 σ" interval for the above examples.

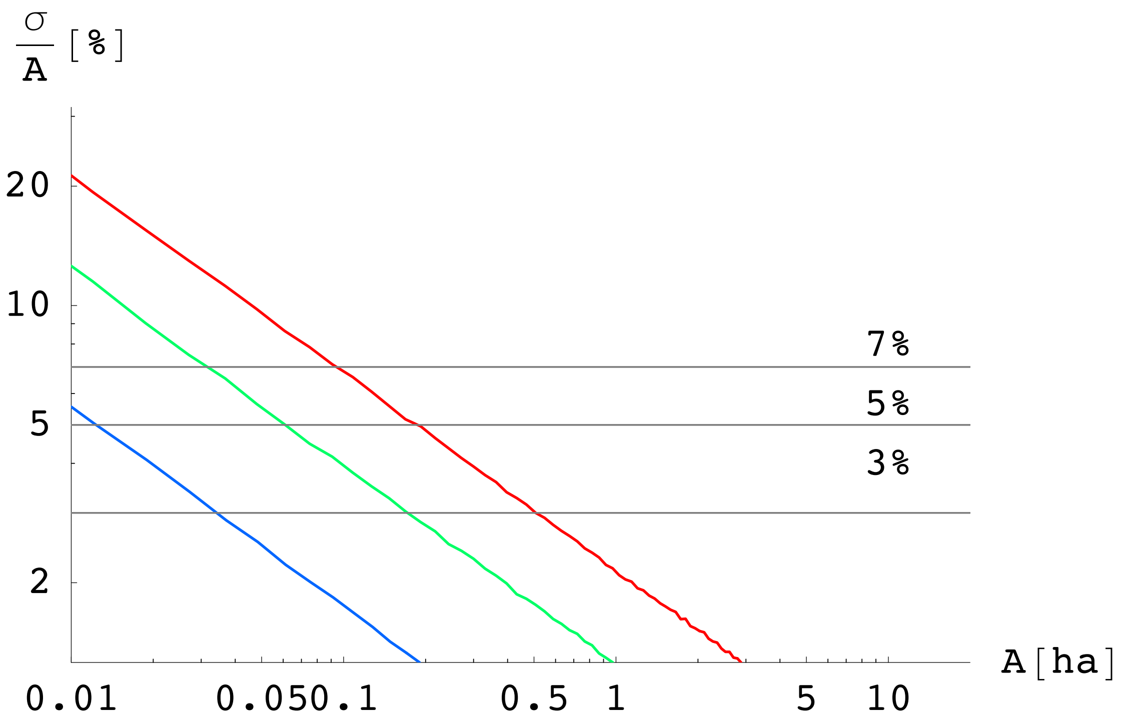


Figure 36: DOP, REA for RMSE = 0.2 m; red line – long rectangle, green line – middle rectangle, blue line – square

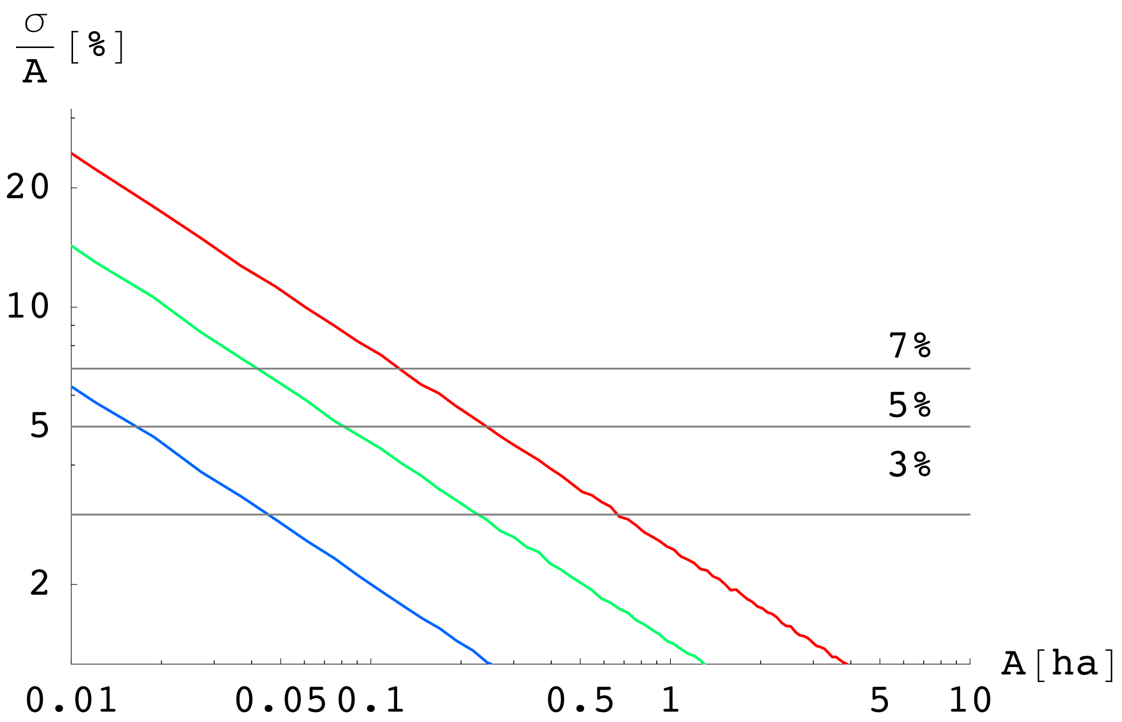


Figure 37: DOP + DIG, REA for RMSE = 0.2 m; red line – long rectangle, green line – middle rectangle, blue line – square

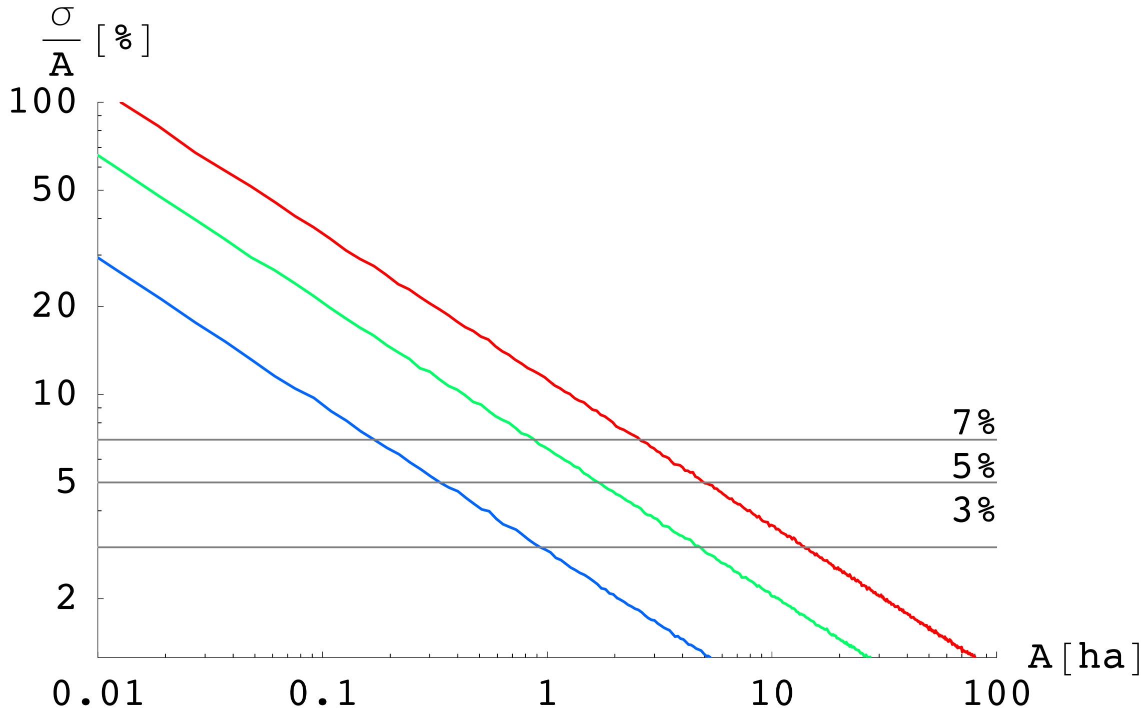


Figure 38: DOP + DIG + INT, REA for RMSE = 0.2 m; red line – long rectangle, green line – middle rectangle, blue line – square

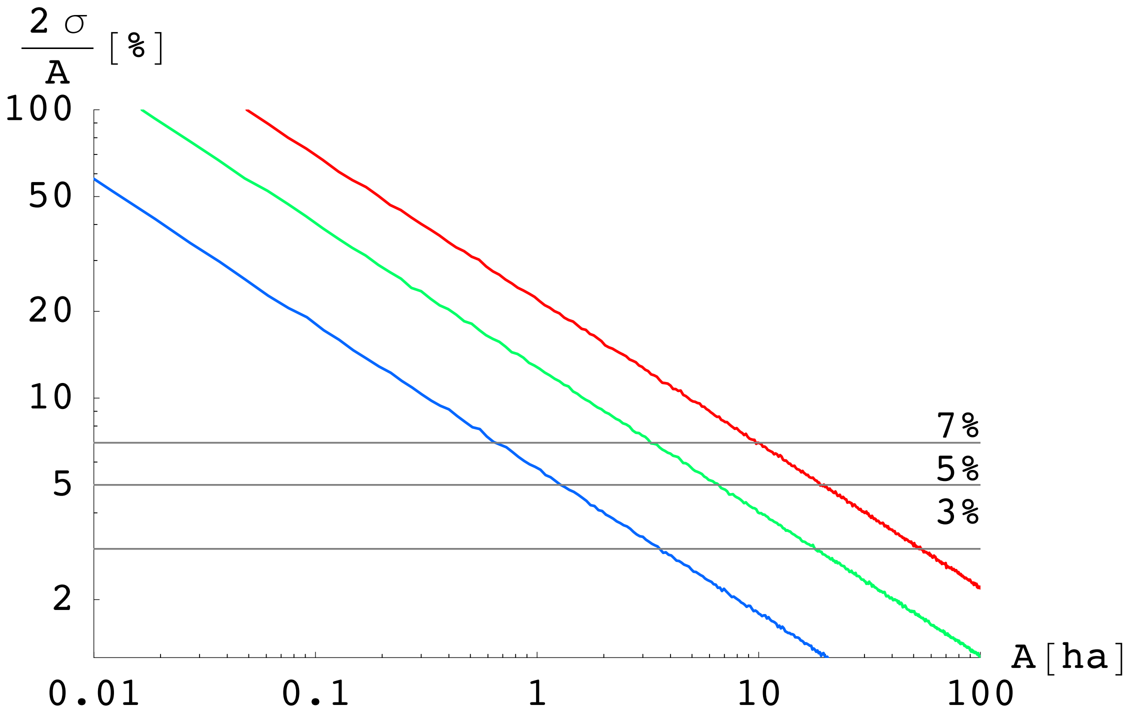


Figure 39: DOP + DIG + INT, "2 σ" interval for RMSE = 0.2 m; red line – long rectangle, green line – middle rectangle, blue line – square

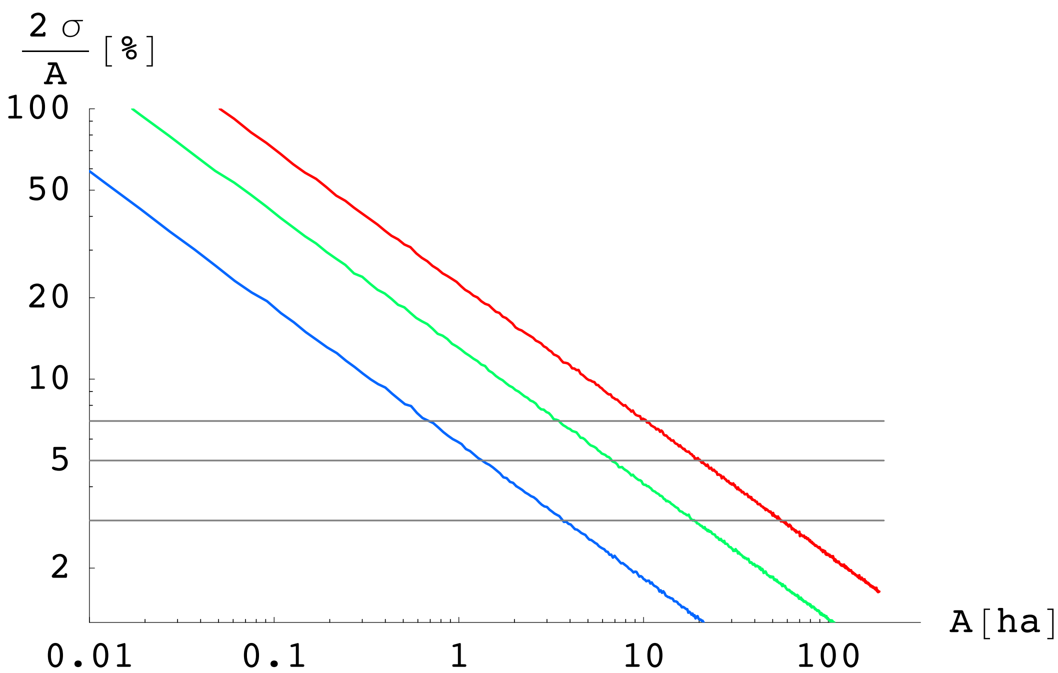


Figure 40: ETS1, "2 σ" interval for RMSE = 0.2 m (first case) and RMSE = 0.4 m (second case); red line – long rectangle, green line – middle rectangle, blue line – square

%	DOP	DOP+DIG	DOP+DIG+INT	ETS1
Square	0.05	0.12	0.56	0.57
Middle	0.12	0.27	1.27	1.30
Long	0.21	0.48	2.18	2.23
Square	0.39	0.87	3.9	4.02
Middle	0.88	1.96	8.9	9.11
Long	1.51	3.41	15.25	15.53
Square	0.54	1.21	5.64	5.75
Middle	1.22	2.77	12.64	12.87
Long	2.08	4.80	21.7	22.08
Square	0.78	1.73	8.0	8.08
Middle	1.76	3.93	18.1	18.43
Long	3.00	6.71	31.0	31.48
Square	1.68	3.76	17	17.54
Middle	3.79	8.62	38.30	39.52
Long	6.60	14.86	66.95	68.20

Table 8: Comparison between different scenarios: "2 σ" interval at 100ha (first - third line), 2ha (4th - 6th line), 1 ha(7th - 9th line), 0.5 ha (10th to 12th line) and 0.1 ha (13th to 15th line).

6.2 Shorter segments

So far we have always studied rectangles (or triangles) with fixed number of vertices – 4 for rectangles or 3 for triangles. Does our simulation give different results if we add vertices to polygons – in analogy to the real world, is it better to make shorter segments?

On Figure 41 (a) we can see an example of a rectangle with four original vertices (green points); on Figure 41 (b) the red points are the added vertices. Now we perform Monte Carlo on all vertices, the old ones and the new ones and compare results.

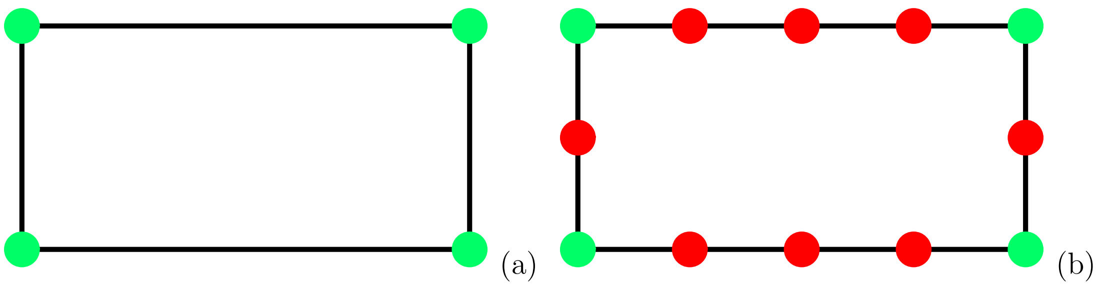
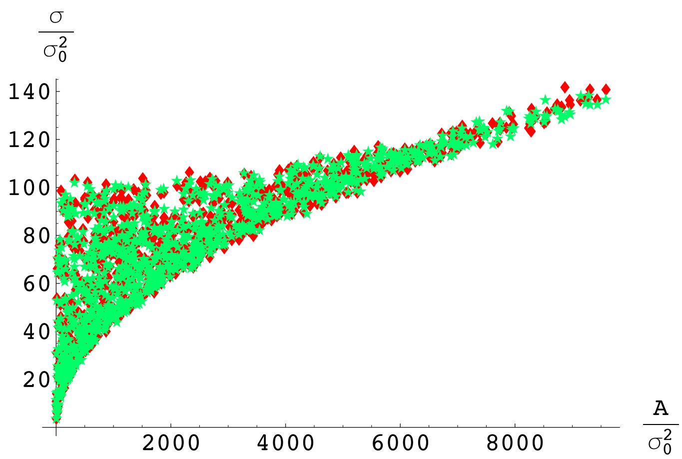


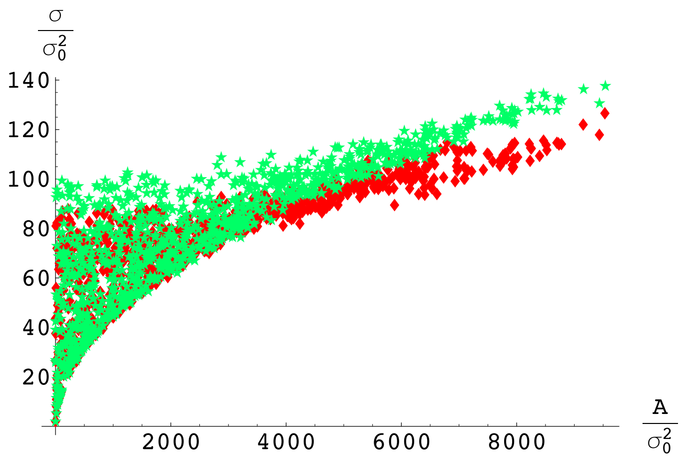
Figure 41: (a) The original rectangle with four vertices; (b) rectangle, transformed into polygon with shorther segments – the red points are the added vertices.

On Figures 42 we can see the comparison between two absolute area uncertainties - the green one represents the original rectangle and the red one the one with shorter segments. On Figure 42 (a) the segment is relatively long in comparison to a (on interval

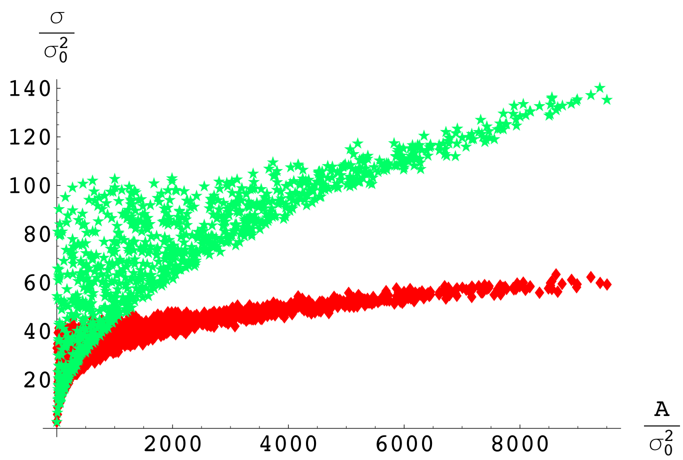
[0, a] we pick side length) – $l/a = 0.8$ and both uncertainties are practically equal. On Figure 42 (b) the segment is middle length $l/a = 0.5$ and we can see that uncertainty is smaller. If the segment is much smaller than the side, like $l/a = 0.1$ on Figure 42 (c) we can see that area uncertainty obviously falls significantly. The segments also can not be too short – the same rule applies to them, two vertices must be at least at 3 σ_0 distance.



(a)



(b)



(c)

Figure 42: Comparison of area uncertainty for original rectangle (green points) and rectangle with segments (red points); length of segments (a) $l/a = 0.8$, (b) $l/a = 0.5$, (c) $l/a = 0.1$, where a is the length of interval from which we pick side length.

Let's compare how shorter segments work for the 3 cases of rectangles. On Figure 43 we can see the comparison for DOP with RMSE = 1 m and segment length $l = 30$ m. The red line as usually represents long rectangle, the green line middle rectangle and the blue line square. The orange line represents long rectangle with segments, the turquoise line middle rectangle with segments and the violet line square with segments.

First thing we can see is that REA for cases with segments is always smaller or at least the same as REA for cases without segments. Not only that – when we enlarge the sides of the ractangles, the length of the segment stays the same - that is why lines that represent segments fall much faster that the lines of the original rectangles. In case of square and middle rectangle we can see that both cases stay the same untill some area – that is untill both sides are shorter than the length of the segment.

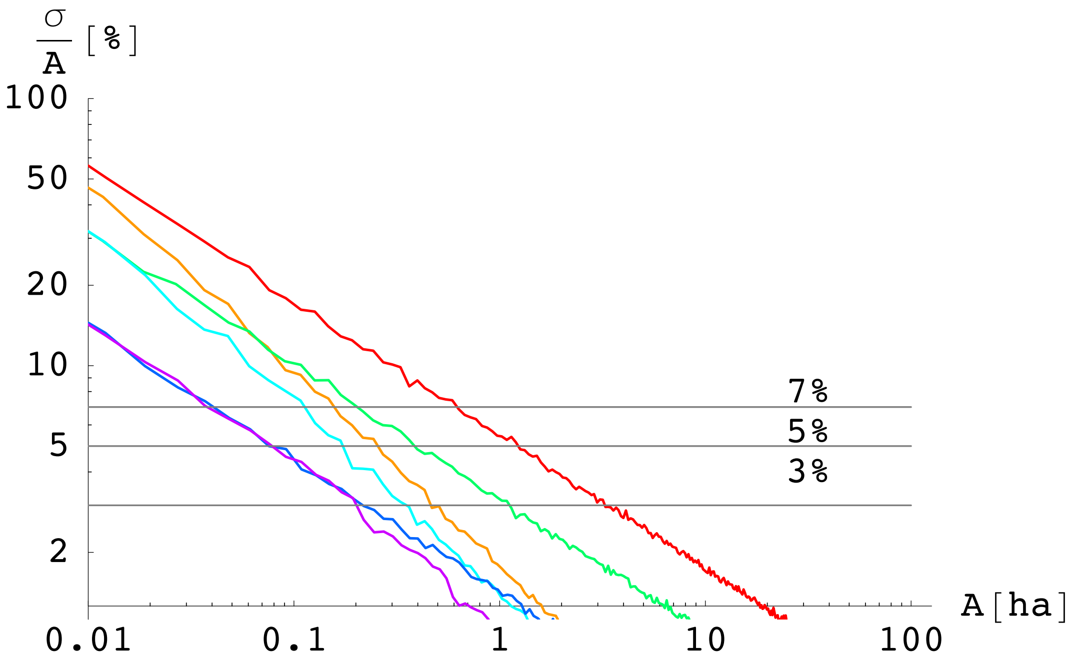


Figure 43: DOP, REA with RMSE = 1m and segment length $l = 30$ m; comparison between the original rectangles and the ones with segments; red line – long rectangle, green line – middle rectangle, blue line – square, orange line – long rectangle with segments, turquoise line – middle rectangle with segments, violet line – square with segments.

7 Conclusion

In this study of area uncertainty we have looked at some properties on which area uncertainty depends on: area, perimeter, area at a constant perimeter etc. For simulation we used Monte Carlo method. First, we have made a general model for triangles where we learned that relative area uncertainty (REA) gets smaller when area gets larger but on the other hand at one particular area we can have many values of REA. This gave us the idea that area uncertainty maybe also depends on shape, not only on area value. We have looked at area uncertainty dependancy on perimeter and from there we have seen that area uncertainty is limited between two special types of triangles: regular-like ones and isosceles-like ones with very small angles between sides of equal length. We have also checked the limit cases where convergence to these two borders is not true anymore and from there we have learned that the length of the side of a polygon should not be smaller than vertex uncertainty times three if we want our numerical model to give proper results. We have compared numerically defined area uncertainty with analytically and have concluded that they are most alike at regular shapes. We have done some buffer analysis but that did not tell us anything new.

We have studied the same properties also for rectangles, the results were comparable. Again we have seen how much area uncertainty (absolute and relative) depends on shape. We have made a comparison for three cases of rectangles and different input uncertainties – comparable to digital ortophoto (DOP), digitalization uncertainty (DOP + DIG), interpretation error(DOP + DIG + INT). Then we made three other types of calculating uncertainty – ETS1, ETS2 and OTS where we compared two digitalizations and observed the difference between them. For each of these types we have compared "1 σ" interval of distribution (equal to analytically calculated area uncertainty) and "2 σ" interval (reproducibility limit). At the end we have expanded basic model with adding vertices and results have shown we can make area uncertainty smaller with this procedure.

References

[1] Chrisman N. R. and Yandell, B. S.: *Effects of point error on area calculations: A statistical model*, Surveying and Mapping, 241 - 246, 1988

[2] Wu, H,, Liu Z. and Lin, L.: *Positional uncertainty of manual digitization vertex based on simlulation test* (Geoninformatics 2008 and Joint conference on GIS and Built Enviroment, 2008).

[3] Hejmanowska, B.: *Validation of methods for measurment of land parcel areas*, 2005 (http://mars.jrc.ec.europa.eu/mars/content/download/800/5268/file/ ValidationMethodFinalVer2_2.pdf)

[4] Hejmanovksa, B.: *Reliability of polygon area measurments for LPIS QA*, 2010 (http://mars.jrc.ec.europa.eu/mars/content/download/1732/9486/file/ S10_Polygons_Hejmanowska.pdf).

[5] Shi, W.: *Principles od modeling uncertainties in spatial data and spatial analyses*, 2010, CRC Press.

Uncertainty of LPIS data or how to interpret ETS results

Supplementary Material II

Grega Milčinski, Marija Vidmar, Miha Kadunc (miha.kadunc@sinergise.com); Sinergise d.o.o.

Uncertainty Composition

In real-life scenarios, more than one process can influence the measurement of a variable. Polygon area measurement, for example, is influenced by the uncertainty in the reference layer (digital ortho-photo), uncertainty in interpretation of the polygon border as well as uncertainty in the digitization.

To arrive at the estimate of the area uncertainty given the contributions of different factors, we need to look at how the final error in area is constructed. The offset of the measured area is the sum of offsets due to the different contributions:

$$dA = dA_1 + dA_2 + dA_3$$

If each of the contributing errors is normally distributed around 0 with RMSE of σ_1, σ_2 and σ_3 respectively and if all the contributing errors are independent, then the estimate for the resulting standard deviation is:

$$\sigma = \sqrt{{\sigma_1}^2 + {\sigma_2}^2 + {\sigma_3}^2}$$

Uncertainty of Relative Difference of Two Measurements

When we estimate the relative area error (RAE) from two measurements of the same polygon, we have to take the uncertainties of both measurements into account. The expression for RAE is straightforward:

$$\mathrm{RAE} = \frac{(\mathrm{A1} - \mathrm{A2})}{A2} = \frac{A1}{A2} - 1$$

Differentiating this expression with respect to both variables A1 and A2 gives the following expression:

$$\mathrm{dRAE} = \frac{1}{A2} dA1 - \frac{A1}{A2^2} dA2 = \frac{A1}{A2}\left(\frac{dA1}{A1} - \frac{dA2}{A2}\right).$$

Assuming small errors, we can set A1=A2 and write the expression for the uncertainty of the relative difference as:

$$\sigma_{\mathrm{RAE}} = \sqrt{\left(\frac{\sigma_{A1}}{A1}\right)^2 + \left(\frac{\sigma_{A2}}{A2}\right)^2} = \frac{\sqrt{{\sigma_{A1}}^2 + {\sigma_{A2}}^2}}{A}.$$

Uncertainty of Polygon Area Derived from Point Position Error

1. Area error produced by independent point position error

The area of a polygon can be written as the sum of areas under individual line segments:

$$A = \frac{1}{2}\sum_{i=1}^{N}(x_{i+1} - x_i)(y_{i+1} + y_i) = \frac{1}{2}\sum_{i=1}^{N} x_{i+1}\, y_i - x_i\, y_{i+1}$$

Note that the indices in the expression should be wrapped cyclically when the last point of the polygon is reached. We can re-write this expression to expose only the terms involving x_i and y_i:

$$A = \frac{1}{2}(\dots x_i y_{i-1} - x_{i-1} y_i + x_{i+1} y_i - x_i y_{i+1} \dots) = \frac{1}{2}(\dots x_i\,(y_{i-1} - y_{i+1}) + y_i\,(x_{i+1} - x_{i-1})\dots)$$

Integration of the uncertain term for x_i times its normal distribution (with σ_{xi}) yields the average contribution of the term to the area, which equals the term itself:

$$\int_{-\infty}^{\infty} (\mathrm{x}_i + \Delta x_i)\,(y_{i-1} - y_{i+1})\,\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{xi}}}\exp\left(-\frac{{\Delta x_i}^2}{2{\sigma_{\mathrm{xi}}}^2}\right) d\Delta x_i = x_i\,(y_{i-1} - y_{i+1})$$

This tells us that the mean area of the polygon with uncertain vertices will be the same as the area calculated from the mean vertex positions. In other words, area calculated from the mean vertex positions is an unbiased estimator of the true polygon area.

To arrive at an estimate of the standard deviation of the area measurement, we have to integrate the square of the area difference between the uncertain and the true polygon. We square the whole sum that is needed to compute the area, but most of the terms involve Δx_i or Δy_i in the first power, and can be removed due to the symmetry of the normal distribution:

$$\begin{aligned}(\dots + (x_i + \Delta \mathrm{x}_i)(y_{i-1} + \Delta \mathrm{y}_{i-1} - y_{i+1} - \Delta \mathrm{y}_{i+1}) - \dots - x_i(y_{i-1} - y_{i+1}))^2 \\ = \dots + {\Delta \mathrm{x}_i}^2((y_{i+1} - y_{i-1})^2 + {\Delta \mathrm{y}_{i-1}}^2 + {\Delta \mathrm{y}_{i+1}}^2) + \cdots\end{aligned}$$

Integral over Δx_i then yields:

$$\begin{aligned}\int_{-\infty}^{\infty} {\Delta \mathrm{x}_i}^2((y_{i+1} - y_{i-1})^2 + {\Delta \mathrm{y}_{i-1}}^2 + {\Delta \mathrm{y}_{i+1}}^2)\frac{1}{\sqrt{2\pi}\sigma_{xi}}\exp\left(-\frac{{\Delta x_i}^2}{2{\sigma_{xi}}^2}\right) d\Delta x_i \\ = {\sigma_{\mathrm{xi}}}^2\left((y_{i+1} - y_{i-1})^2 + {\Delta \mathrm{y}_{i-1}}^2 + {\Delta \mathrm{y}_{i+1}}^2\right)\end{aligned}$$

The term with *y* coordinates of the neighbouring points represents the independent contribution of the ordinate x_i to the uncertainty of the polygon, while the two terms containing the neighbouring points' deltas describe the effect of the interaction of the neighbouring points on the measured area uncertainty.

All these terms will be further integrated to account for uncertainty in all other points, and finally produce the following expression for the total area uncertainty (factor 1/2 comes from the fact that the terms from the expression in the sum will be included at other indices):

$$\sigma_A{}^2 = \frac{1}{4}\sum_{i=1}^{N} \sigma_{\mathrm{xi}}{}^2\left((y_{i+1} - y_{i-1})^2 + \frac{1}{2}(\sigma_{\mathrm{yi-1}}{}^2 + \sigma_{\mathrm{yi+1}}{}^2)\right)$$
$$+ \sigma_{\mathrm{yi}}{}^2\left((x_{i+1} - x_{i-1})^2 + \frac{1}{2}(\sigma_{\mathrm{xi-1}}{}^2 + \sigma_{\mathrm{xi+1}}{}^2)\right).$$

In case of isotropic error ($\sigma_{xi}= \sigma_{yi}= \sigma_i$), this simplifies to:

$$\sigma_A{}^2 = \frac{1}{4}\sum_{i=1}^{N} \sigma_{\mathrm{i}}{}^2((y_{i+1} - y_{i-1})^2 + (x_{i+1} - x_{i-1})^2 + \sigma_{\mathrm{i-1}}{}^2 + \sigma_{\mathrm{i+1}}{}^2)\,.$$

In case all points have the same error ($\sigma_i= \sigma$) the expression becomes:

$$\sigma_A{}^2 = \frac{\sigma^2}{4}\sum_{i=1}^{N} (y_{i+1} - y_{i-1})^2 + (x_{i+1} - x_{i-1})^2 + 2\,\sigma^2.$$

This tells us that the variance (square of RMSE) of area measurement is proportional to the sum of squares of the distances between point's neighbours. The small term outside the sum is significant only in case σ is comparable to the length of the diagonals (but still smaller – see next paragraph).

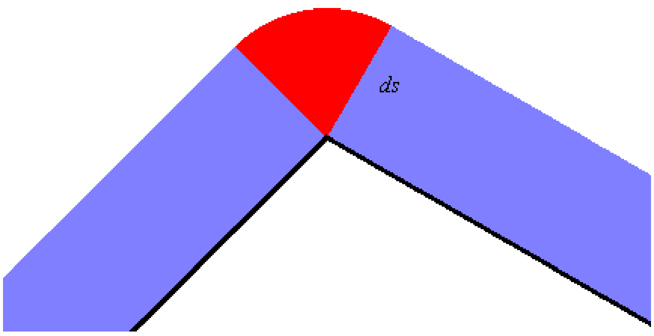
Note that this expression does not take into account extremely thin polygons, where special provisions should be taken in the uncertainty analysis to rule out any combinations of point offsets that would produce invalid geometries, such as self-intersecting polygons, reversed orientations etc.

2. Area error produced by correlated offset from the true boundary

When the measured polygon boundary is offset from the true boundary by distance ds (in this case the individual point measurements are not independent, but are strongly correlated), either to the outside or to the inside of the polygon, the offset of the measured area can be approximated with:

$$\Delta A = l * ds + (N_{\mathrm{out}} - N_{\mathrm{in}}) * \pi * ds^2.$$

Here l denotes the length of the polygon boundary (including holes) while N_{out} and N_{in} denote the number of outer and inner rings, respectively. The first term clearly represents the (signed) area created by offsetting each line segment by ds, while the second term is produced by summation of the areas of circular sectors that fill in the gaps between the offset line segments (see Figure).



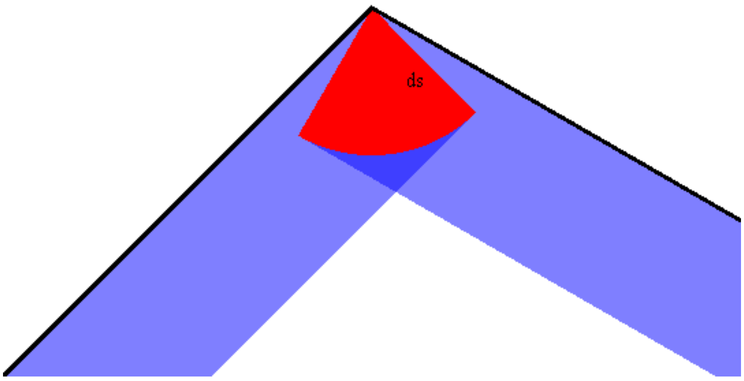
For any closed ring, the sum of angles generating the circular sectors will be 360°, making their total area equal the area of a full circle; for holes, the sum of angles will be negative.

Integrating the expression for the area offset using a normal distribution for ds (with standard deviation σ_s) yields the total uncertainty of the area:

$$\sigma_A = \sigma_s \sqrt{l^2 + 3\,(N_{\text{out}} - N_{\text{in}})^2\, \pi^2\, {\sigma_s}^2}\,.$$

Interestingly, this expression mostly depends on the total length of the polygon boundary (in case the polygon has exactly one hole, this is true exactly), and is only slightly influenced by the number of holes in the polygon (and outer rings for a multi-polygon). In this approximation, the measured area does not depend on the number of points digitized or the shape of the boundary at all.

Note that while the approximation is quite correct for positive offsets and convex vertices, it is somewhat flawed in case of concave angles or negative offsets, as the area subtracted from the polygon is smaller than the area that is covered by the two neighbouring rectangles (see Figure). This is hardly relevant for obtuse angles, but could be quite significant for acute angles.



About RMSE and 95% confidence interval in the normal distribution

Root-mean square error (RMSE or σ or *standard deviation*) is a property of the probability density function (PDF, also called error or probability distribution), that provides a measure of the distribution's width or around its mean or average value ($\bar{x}$). The following equations show the relations for the mean and RMSE, for continuous PDFs (left) and for a list of N imprecise measurements (right):

$$\bar{x} = \int_{-\infty}^{\infty} x \, \mathrm{PDF}(x) \, dx \, ; \qquad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\mathrm{RMSE}^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 \, \mathrm{PDF}(x) \, dx \, ; \quad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i-\bar{x})^2}$$

A normal distribution (also called Gaussian) centred at 0 is only parameterized with its width σ, which also coincides exactly with the distribution's RMSE:

$$NormalPDF(x,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx; \quad RMSE_{NormalPDF} = \sigma$$

Figure 1: Normal distribution centred at 0

It can be seen from the distribution, that the probability is highest around the mean (zero in the case of our figure), but there is also significant probability of a measured value being up to 3 σ away from the mean. The probability of a measured value falling up to one σ away from the mean can easily be calculated:

$$\int_{-\sigma}^{\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = 68.27\%$$

This tells us that for a normal distribution, σ (or RMSE) is about equal to the 68% confidence interval – about 68% of the measured values will be found within 1σ of the mean. The farther from the mean we go, higher percentage of the measured values will fall within the selected interval. If we would like to know how far from the mean we need to go to find 95% of all measured values, the result is easily found by integrating to find the area under the probability density function:

$$\int_{-c}^{c} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = 95\% \Longrightarrow c = 1.95996\,\sigma$$

So, in a normal distribution, 95% of all values will fall in the interval whose boundaries are 1.96 σ away from the mean on either side. This is illustrated on the following diagram:

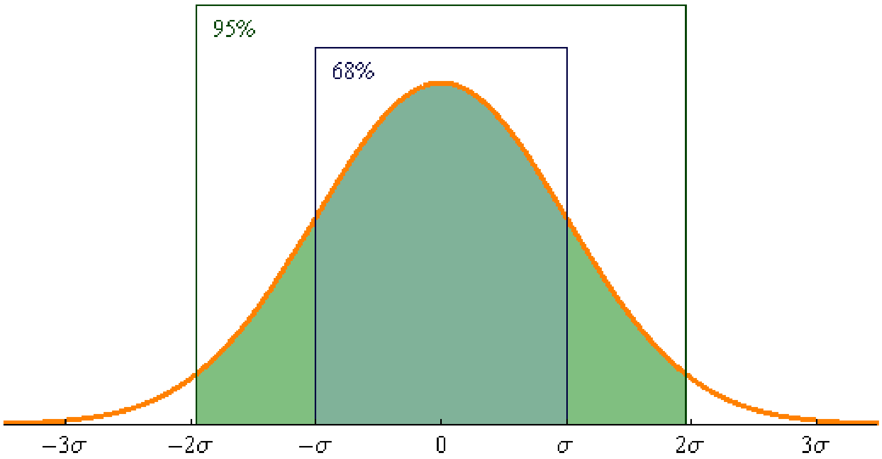
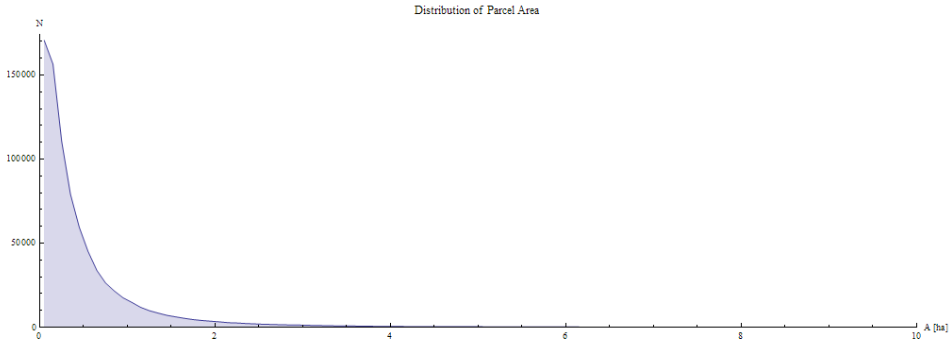
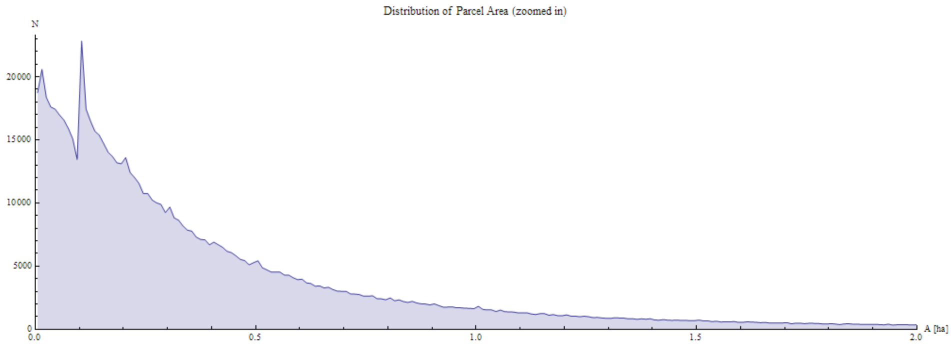


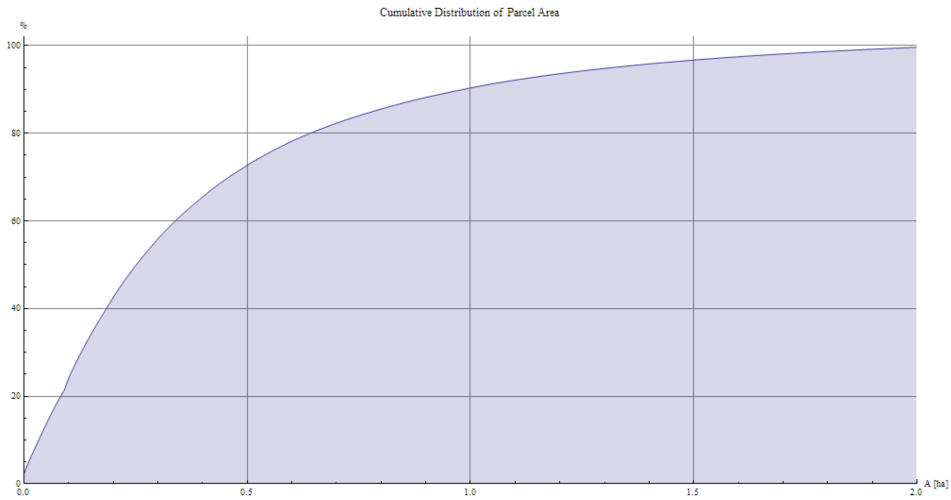
Figure 2: 68% of all the values fall between –σ and σ, and 95% of all the values fall between -1.96 σ and 1.96 σ

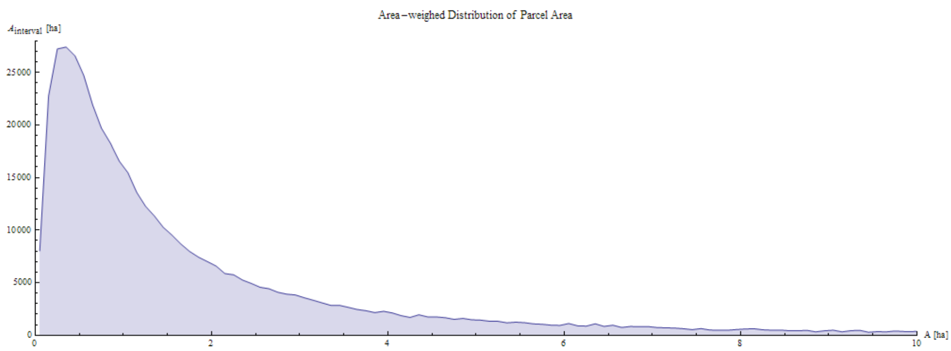
Statistical Analysis of Slovenian LPIS Data

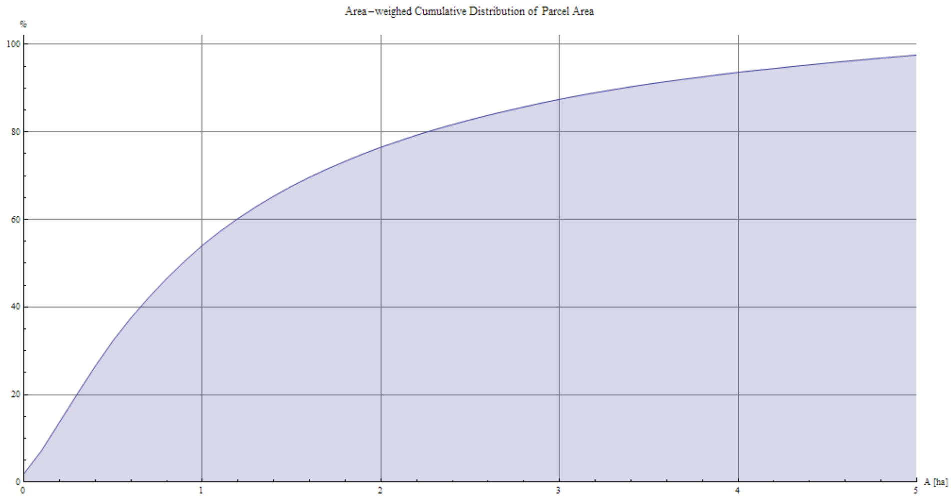
Area of an individual parcel





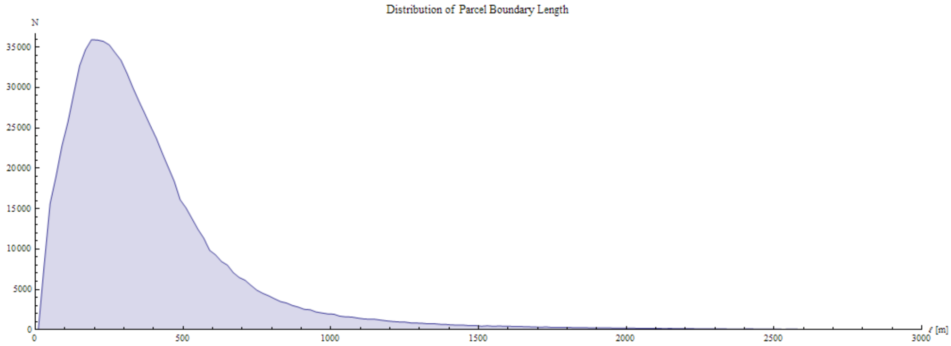


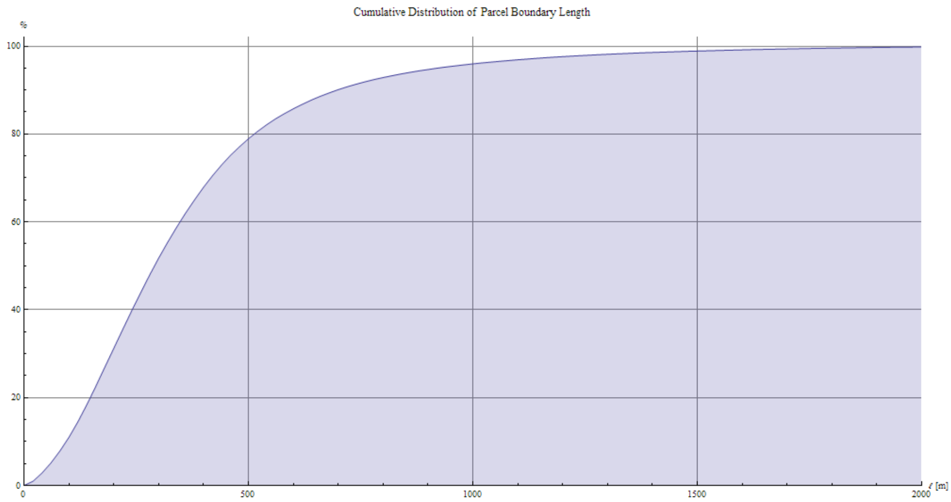


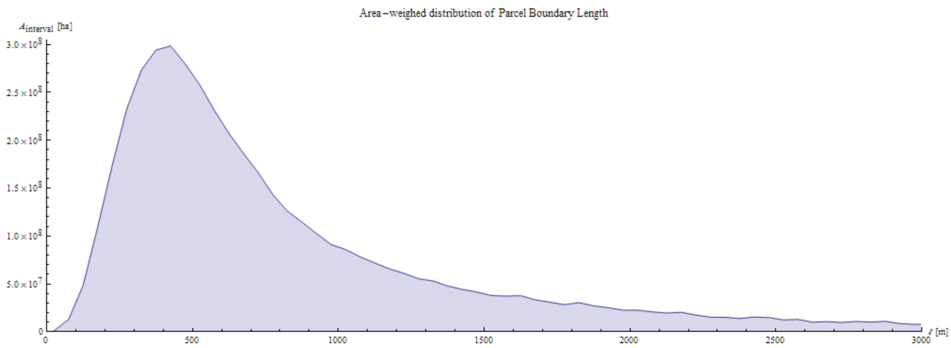


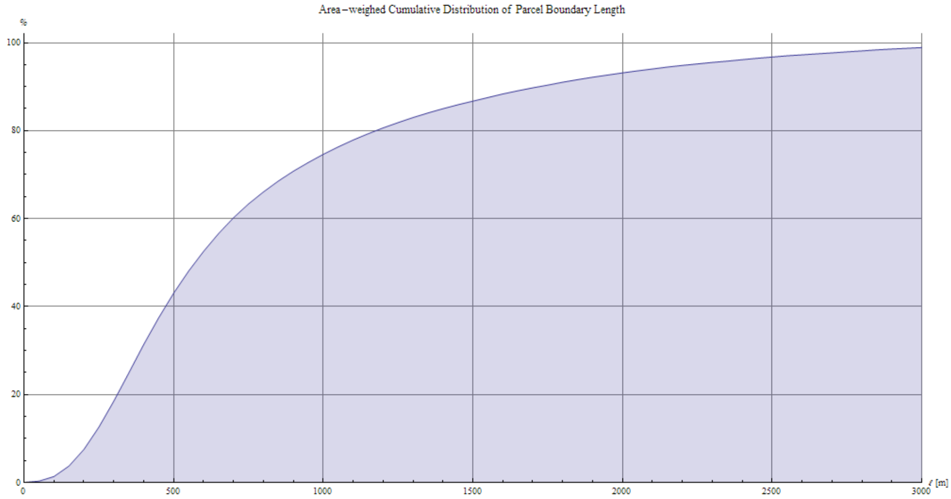
The height of the lines in an area-weighed distribution shows the total area of the parcels which fall into the bin, as opposed to the frequency distribution, which just shows the count. Area-weighed distribution is useful for establishing significance of certain kinds of samples (e.g. parcels with small boundary length) in terms of the sum of area they cover.

Length of the parcel's boundary

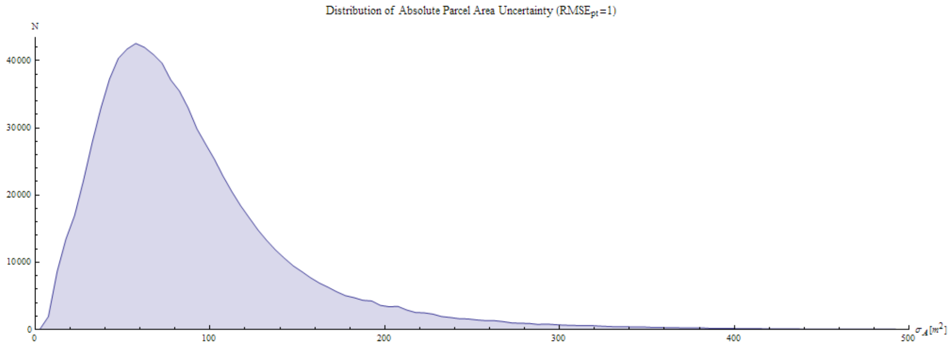


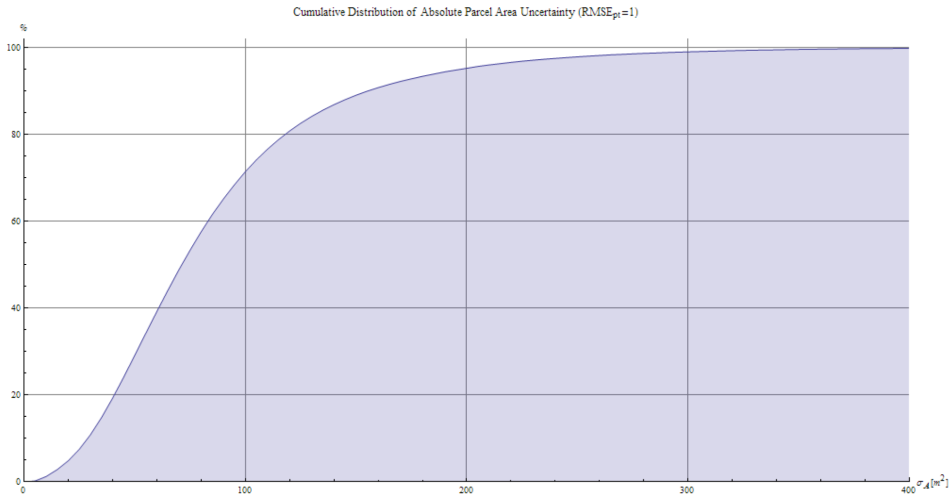


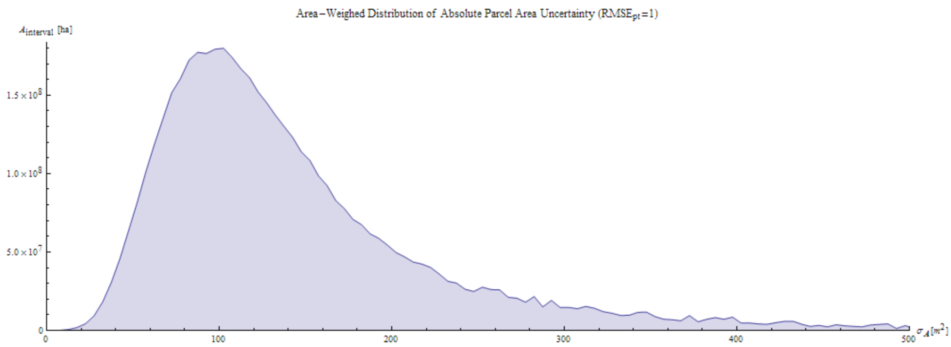


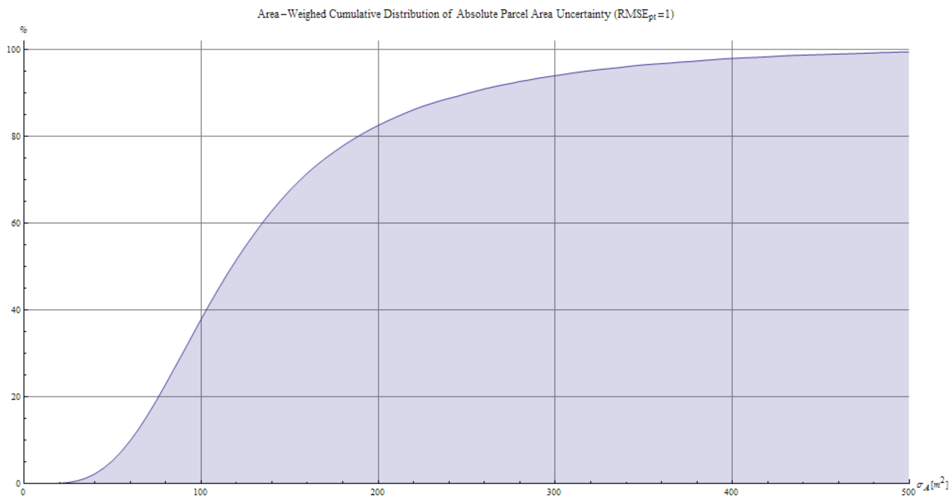


Absolute uncertainty of the parcel's Area

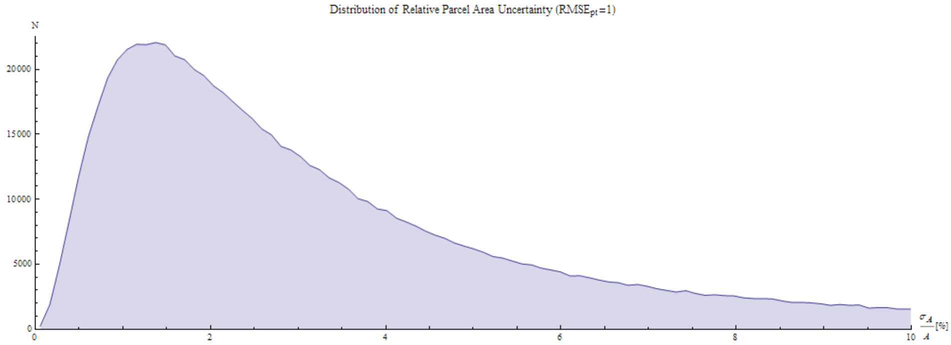


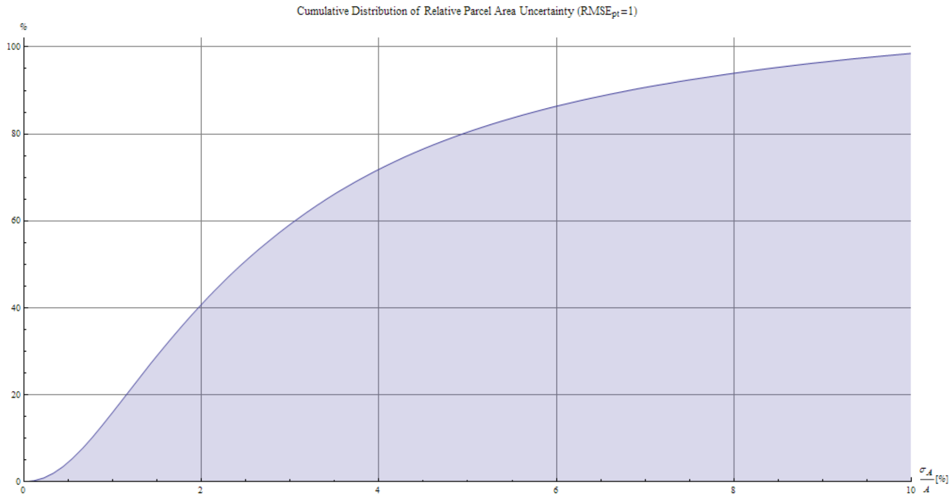


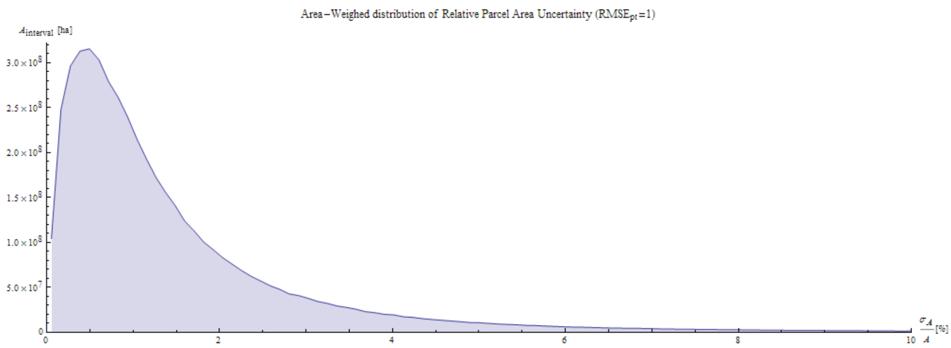

Area−Weighed Distribution of Absolute Parcel Area Uncertainty (RMSE$_{pt}$=1)
$A_{interval}$ [ha]
1.5 × 10^8
1.0 × 10^8
5.0 × 10^7
0
100
200
300
400
500
σ_A[m^2]

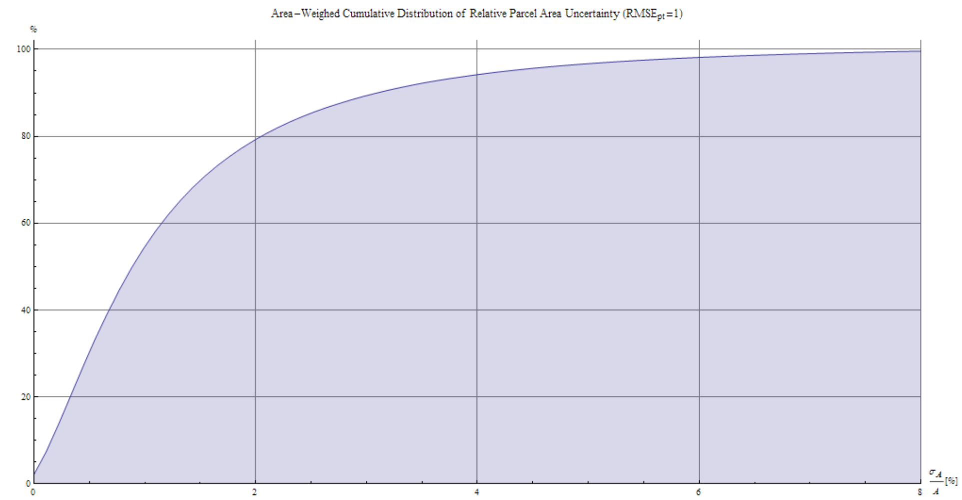

Area−Weighed Cumulative Distribution of Absolute Parcel Area Uncertainty (RMSE$_{pt}$=1)
%
100
80
60
40
20
0
100
200
300
400
500
σ_A[m^2]

Relative uncertainty of the parcel's Area









Effect on the total area of all parcels

Total Area = $\sum \mathrm{A}$	**479 661 ha**
*Average Relative Uncertainty = $\mathrm{AVG}(\sigma_{\mathrm{rel}})$	5.81 %
*Sum of Absolute Uncertainties = $\sum \sigma_{abs}$	7492 ha
*…divided by total area = $\frac{\sum \sigma_{abs}}{\sum A}$	1.56 %
Uncertainty of Total Area = $\boldsymbol{\sigma_{TOT} = \sqrt{\sum \sigma_{abs}{}^2}}$	**9.8 ha**
Relative Uncertainty of Total Area = $\boldsymbol{\frac{\sigma_{TOT}}{\sum A}}$	**0.002%**

* items marked with asterisk are not relevant for the analysis of the total area

Thresholds Applied to 95% Confidence Interval

95% THRESHOLD	0..0.2 ha	0.2..0.5 ha	>0.5ha
>3%	35.83 %	10.22 %	**1.38 %**
>5%	24.02 %	**3.02 %**	0.18 %
>7%	**16.30 %**	1.18 %	0.02 %