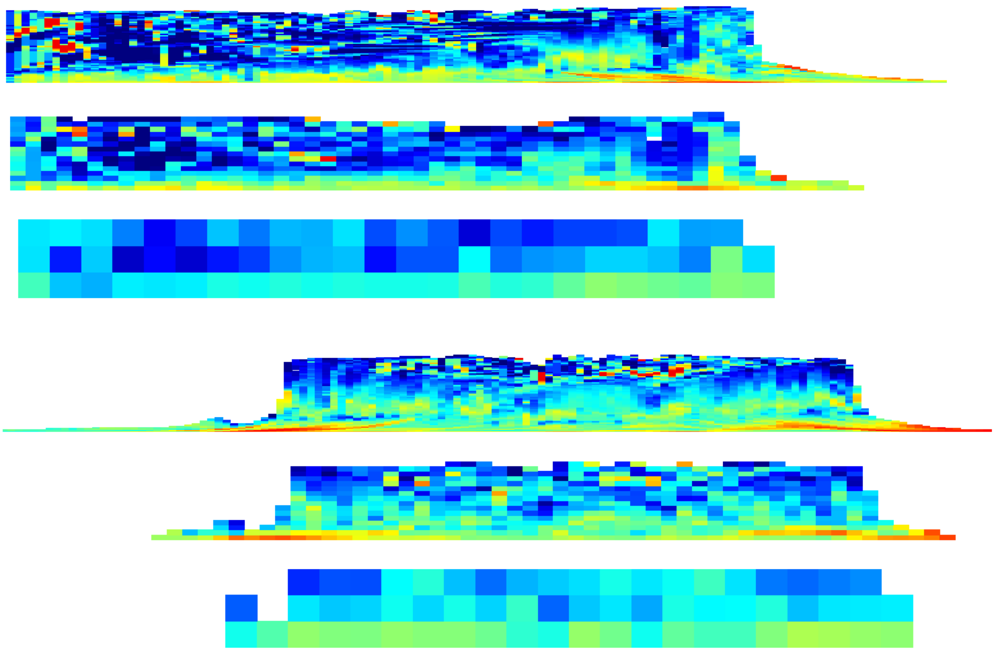
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Consequences of changes in grid resolution for process model based variogram estimation



Note no SAND/13/20

Authors Jacob Skauvold
Ragnar Hauge

Date December 7, 2020

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modelling. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Consequences of changes in grid resolution for process model based variogram estimation**

Authors	**Jacob Skauvold** `<jacob.skauvold@nr.no>` **Ragnar Hauge** `<hauge@nr.no>`

Date	December 7, 2020
Publication number	SAND/13/20

Abstract

This supplementary report examines how variogram estimation is affected by a change in the resolution of the resampled grid.

Keywords	Variogram estimation, process model, Delft3D, gaussian random field, resampling, upscaling
Availability	Internal, Equinor
Project number	190549
Research field	Reservoir modeling
Number of pages	16

1 Assessing grid resolution sensitivity

In our main project report [1] we investigated cropbox placement sensivity. Here, we examine how variogram estimation is affected by a change in the resolution of the resampled grid.

We do this by estimating variograms at the same locations in the same delta, varying how the resampling is done, and comparing the resulting estimates.

1.1 Grid resolution levels

We consider three different levels of resolution.

1. **Normal**. The default output resolution of the resampling routine in the variogram estimator. The lateral grid cell size is the same as the Delft3D output grid, while the vertical grid cell size is $0.25\,\mathrm{m}$.
2. **Geogrid**. Double the normal horizontal grid cell size. Vertical grid cell size: $1\,\mathrm{m}$.
3. **Simgrid**. Four times the normal horizontal grid cell size. Vertical grid cell size: $5\,\mathrm{m}$.

We vary the vertical resolution by passing different values of the input parameter `resample_dz` to the variogram estimator. To vary the lateral resolution, we begin by using standard resampling to create a grid with the desired vertical resolution and the normal (fine) lateral resolution. If the desired lateral resolution is lower, we then aggregate horizontal neighborhoods of either two by two (normal $\rightarrow$ geogrid) or four by four (normal $\rightarrow$ simgrid) cells.

Note that we could have done the vertical coarsening the same way as the horizontal coarsening, by first resampling onto a regular grid with a vertical resolution of $0.25\,\mathrm{m}$, and then aggregating three-dimensional neighborhoods of 2^3 and 4^3 cells.

We consider averaging as the preferred upscaling mode, and decimation as an alternative mode. Let $\varphi[i, j, k]$ be an element of a resampled but not yet upscaled porosity field. Denote by ℓ the factor by which the grid should be coarsened horizontally. Neighborhoods of ℓ by ℓ by 1 cells are to be aggregated into single cells. When upscaling by averaging, we define element i, j, k of the upscaled field by

$$\varphi_{\mathrm{avg},\ell}[i, j, k] = \frac{1}{\ell^2} \sum_{\Delta i=0}^{\ell-1} \sum_{\Delta j=0}^{\ell-1} \varphi[\ell i + \Delta i, \ell j + \Delta j, k],$$

and when upscaling by decimation, we define it by

$$\varphi_{\mathrm{dec},\ell}[i, j, k] = \varphi[\ell i, \ell j, k].$$

Table 1. Start and end coordinates for cropbox 1 to 5. See Figure 1 for map view.

Number	x_0	y_0	x_1	y_1
1	5250.0	5600.0	8250.0	8400.0
2	5250.0	2800.0	8250.0	5600.0
3	8250.0	5600.0	11250.0	8400.0
4	5250.0	8400.0	8250.0	11200.0
5	5250.0	2000.0	12250.0	12000.0

1.2 Resolution-dependent input parameters

Certain internal variables in the variogram estimator depend on the grid cell size because they are given as a number of cells or cell widths.

- The "weighting width" `sigma_wt` controls how concentrated or diffuse the weights used in least squares curve fitting should be.
- The initial values of major, minor and vertical correlation range passed to the curve fitting function.
- The upper and lower bounds of major, minor and vertical correlation range passed to the curve fitting function.

To meaningfully compare estimates across grid resolutions, these parameters will have to be scaled inversely to the grid cell size, so that the actual lengths they correspond to remain constant.

2 Estimation on real and synthetic data

We use two sources of input-data for the variogram estimator. The first is the porosity field from a Delft3D realization. The other is a realization drawn from a Gaussian random field with a known variogram and no trend. This allows us to compare the performance of the estimator in a realistic setting where a trend is likely to be present to its performance in an idealized setting where we know the implicit assumptions of the estimator are satisfied.

2.1 Delft3D realization MS2_12

We consider five cropboxes located on the delta in the MS2_12 realization. Table 1 gives upper and lower bounds for the x and y coordinates of these cropboxes, and Figure 1 shows their positions relative to each other and the delta in map view. Boxes 1 to 4 are adjacent squares of equal size, while box 5 is a larger rectangle which covers most of the delta top and delta fringe, and encompasses the four smaller boxes.

To capture the effect of varying grid resolution on the variogram estimation, we

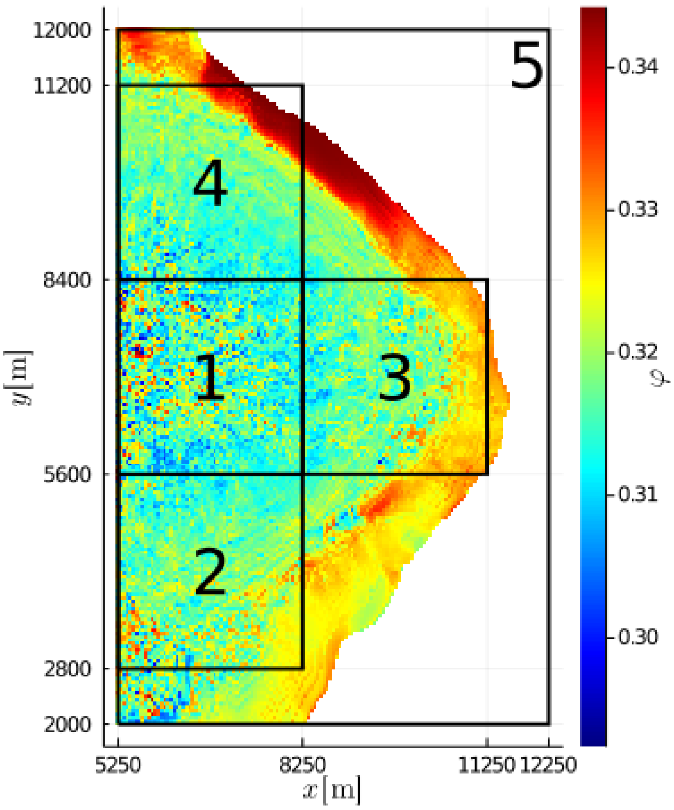


Figure 1. Cropbox outlines. The background porosity field has been resampled inside the region covered by cropbox 5. Displayed porosity values are vertical averages.

Table 2. Parameters varied in the MS2_12 sensitivity study, and the various valus used for each.

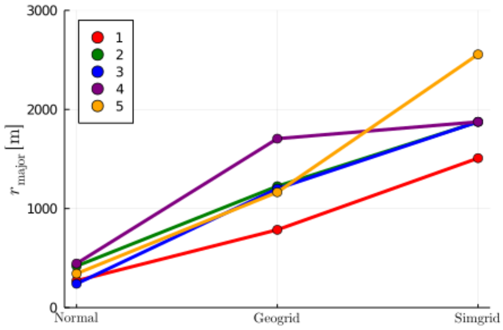
Parameter	Values
sigma_wt	small (10/5/2.5 normal/geogrid/simgrid cells)
	large (40/20/10 normal/geogrid/simgrid cells)
resampling mode	averaging
	decimation
grid resolution level	normal
	geogrid
	simgrid
cropbox	1 to 5 (see Table 1)
variogram type (family)	exponential
	gaussian
	general exponential
	spherical

estimated one variogram for each unique combination of the input parameters listed in Table 2. This yielded a set of 240 estimation results. We present a subset of these estimates here.

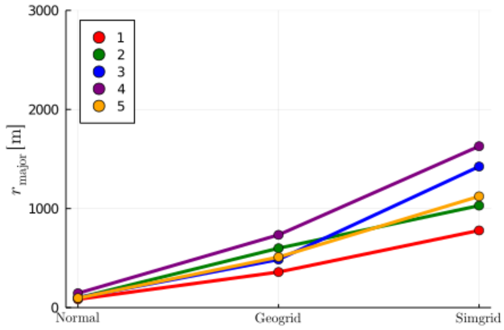
Figure 2 shows how estimates of major correlation range vary with grid resolution when resampling is done by averaging and `sigma_wt` is on the small level. The results are broken down by variogram type (one per panel) and cropbox (one per line in each panel). Ideally, the range estimates would remain approximately constant as the grid becomes coarser. This is clearly not the case in Figure 2. Instead we see a general pattern in which the estimated ranges tend to increase as the resolution decreases. The spherical variogram estimates for cropbox 1 to 4 with simgrid resolution are an exception, as these range estimates are smaller than their higher-resolution equivalents.

Figure 3 tracks changes in estimated azimuths in the same way as for major correlation range in Figure 2. The estimates for cropboxes 1, 2 and 5 are most robust, and least affected by the change in grid resolution. The estimates for cropboxes 3 and 4 are more variable, both within and between grid resolution levels. This kind of variation is expected when the empirical variogram is close to isotropic.

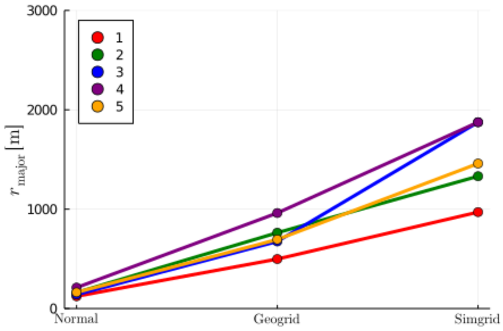
Figure 4 shows empirical variograms for cropbox 1 plotted together with fitted exponential variograms for small and large values of `sigma_wt`. The fitted variograms shown in orange in this figure have the major correlation ranges shown by the red line in panel (a) of Figure 2. With normal grid resolution, the fit is reasonably good in the horizontal directions when `sigma_wt` small, but poor for



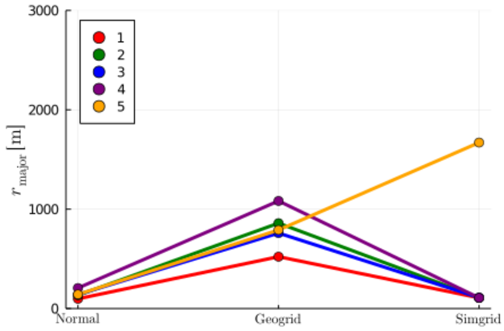
(a) Exponential



(b) Gaussian

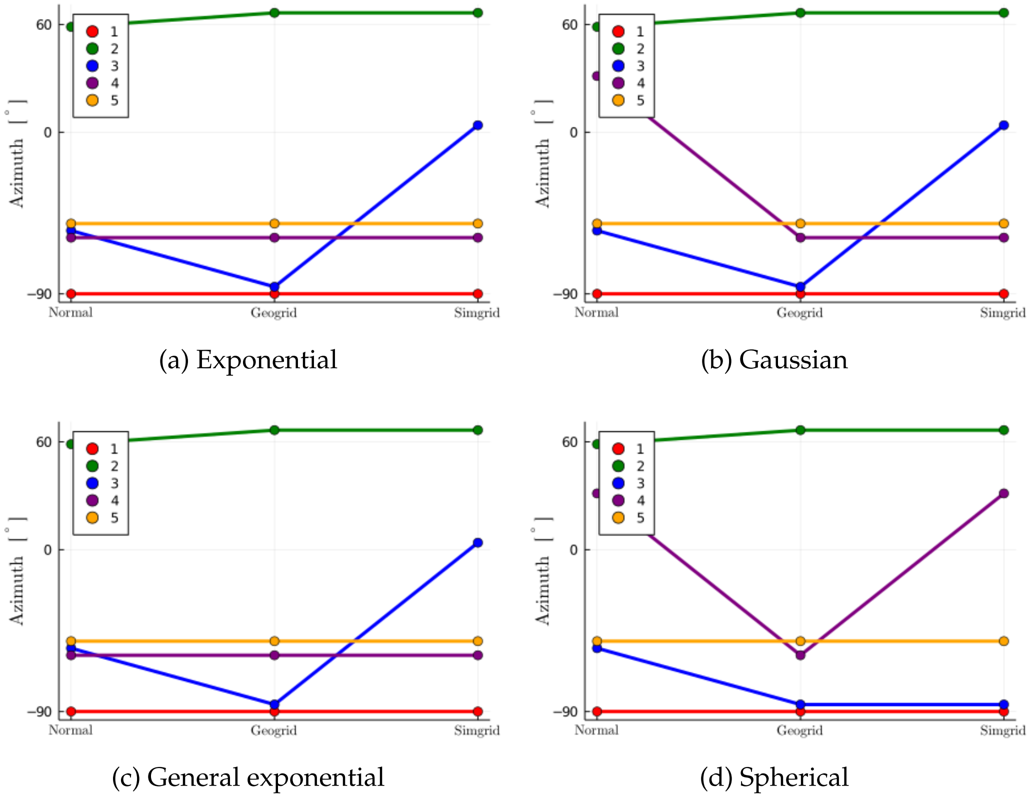


(c) General exponential



(d) Spherical

Figure 2. Estimates of major correlation range plotted against grid resolution level for four different variogram families. Each curve represents one cropbox.



(a) Exponential (b) Gaussian

(c) General exponential (d) Spherical

Figure 3. Estimates of azimuth plotted against grid resolution level for four different variogram families. As in Figure 2, each curve represents one cropbox. Note that all azimuth estimates for exponential and general exponential variograms are identical.

sigma_wt large. For the vertical direction, the fit is acceptable for both large and small values.

For the geogrid resolution, the pattern is similar. The fit is good in the vertical direction, and significatly poorer in the horizontal directions. Again, the best match is achieved for sigma_wt small.

Finally, with simgrid resolution, the thickness of the MS2_12 model only affords one nonzero vertical lag distance. As a consequence the vertical fit is not meaningful in this case. The horizontal fit shows the same pattern seen with finer resolutions. The overall goodness of fit is poor, with a small value of sigma_wt yielding a slightly better match with the empirical curve.

2.2 Synthetic data reference

As a basis for comparison we have run variogram estimation on synthetic data using the same settings and the same three grid resolutions as for the MS2_12 case. Synthetic porosity data were created by simulating a gaussian random field with a known variogram on a regular grid of the same size as the resampled grid on the cropboxes 1 and 5. A linear mapping was applied so that the synthetic field has the same minimum and maximum values as the MS2_12 porosity field in the two cropboxes.

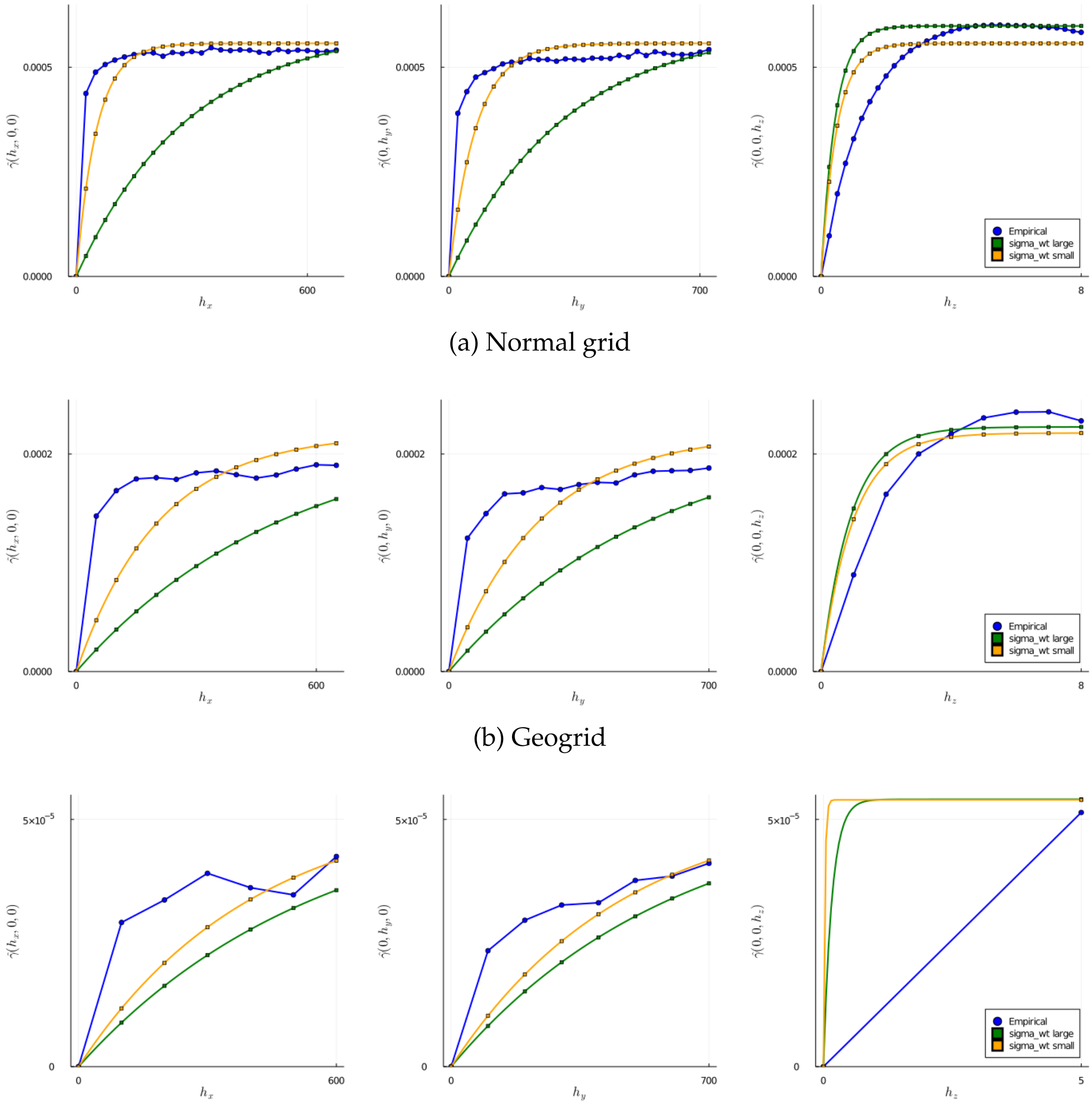
Figure 5 shows major range estimates in the cropbox 1-sized synthetic data case varying in response to coarsening grid resolution, and is comparable to Figure 2. Figure 6 compares empirical variograms and fitted exponential variograms for the same synthetic case, and is comparable to Figure 4. The estimates of major correlation range based on synthetic data stay approximately constant when the grid resolution changes from normal to geogrid. In the transition from geogrid to simgrid, there estimates begin to change in the synthetic case as well.

In Figure 6 we see that it is not easy to achieve a good fit, even when the data is drawn from a Gaussian random field. In this case, the true vertical range is too long relative to the range of available lag distances to be estimated reliably. The match between the empirical and fitted varigrams in the horizontal directions is reasonably good at normal and geogrid resolution.

3 Analysis and recommendations

Having collected and studied the set of estimates produced by running the variogram estimator on porosity data from a Delft3D model and on synthetic data from a stationary Gaussian random field at three levels of grid resolution, we can make the following observations.

(A1) *There is a grid resolution effect*

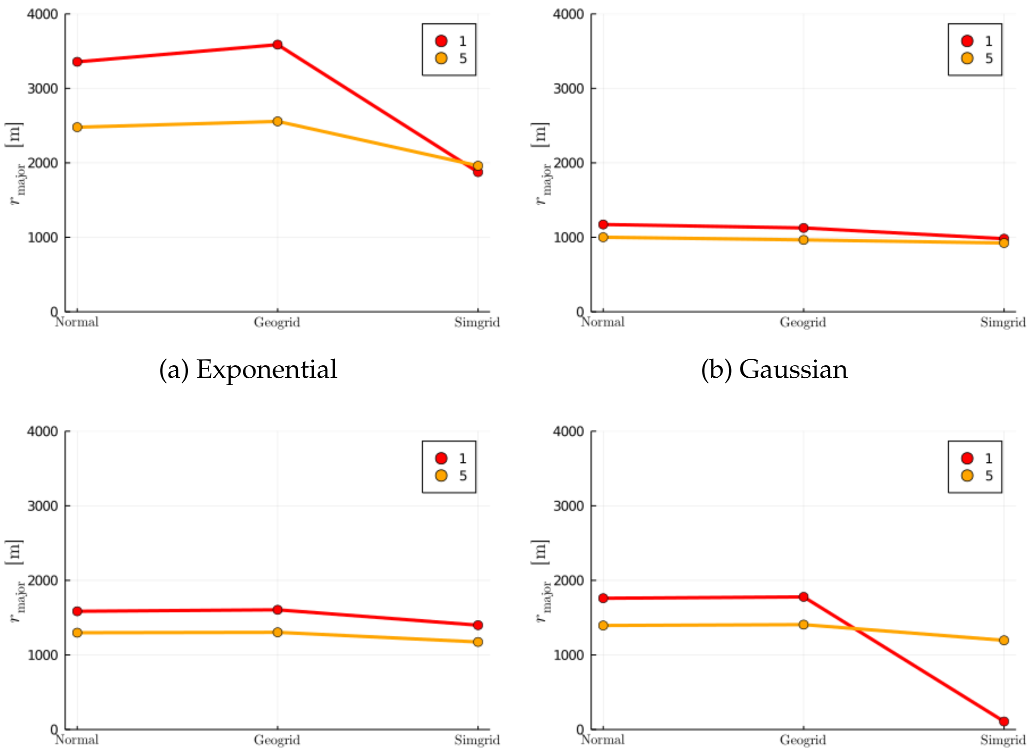


(a) Normal grid

(b) Geogrid

(c) Simgrid

Figure 4. Empirical variograms for cropbox 1. Fitted exponential variograms for `sigma_wt` small and large.



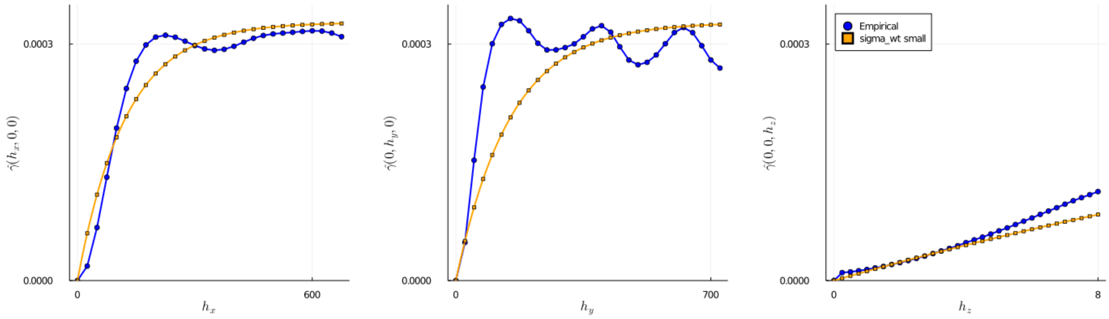
(a) Exponential

(b) Gaussian

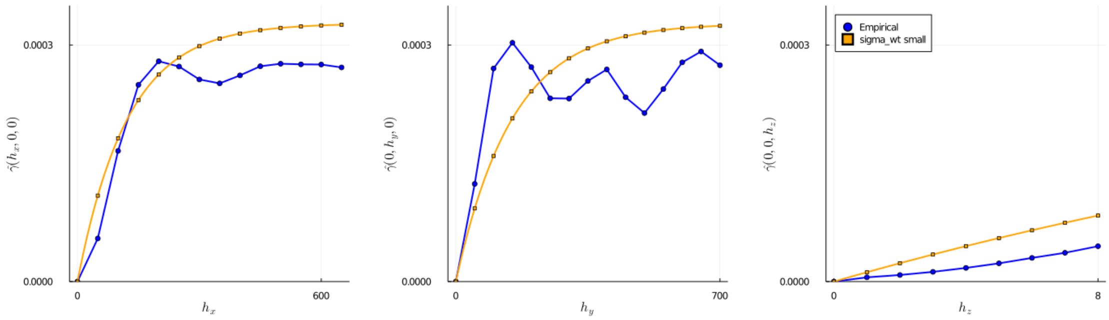
(c) General exponential

(d) Spherical

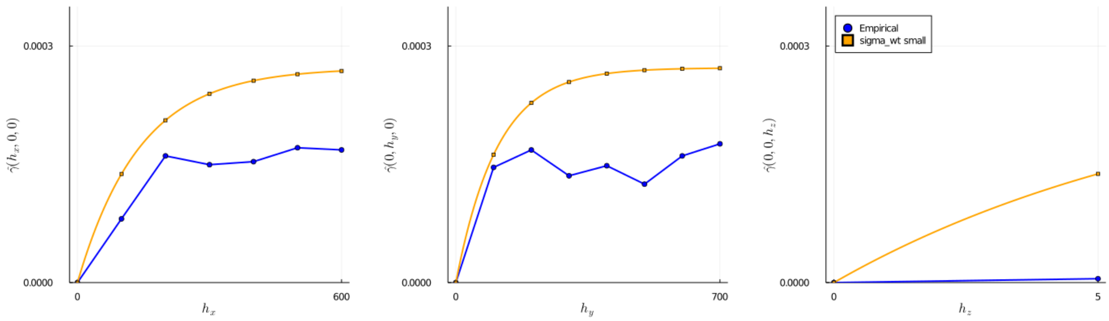
Figure 5. Estimates of major correlation range baed on synthetic data, plotted against grid resolution level for four different variogram families. The dimensions of the synthetic volumes used were chosen to match the dimensions of cropboxes 1 and 5.



(a) Normal grid



(b) Geogrid



(c) Simgrid

Figure 6. Empirical variograms for cropbox 1. Fitted exponential variograms for `sigma_wt` small and large.

Parameter estimates do not always stay the same when the grid scale changes. The direction and magnitude of the change are difficult to predict. When `sigma_wt` is small, estimated correlation ranges tend to increase as the grid gets coarser. When `sigma_wt` is large, the response to coarsening is less systematic, with estimated ranges increasing for some cropboxes and variogram types, while decreasing for others.

(A2) *The grid resolution effect is more pronounced at coarse resolutions*
A change from geogrid resolution to simgrid resolution tends to cause estimates to change much more than a change from normal resolution to geogrid resolution does.

(A3) *Weighting matters*
Estimates are sensitive to the degree of concentration of the weights in the loss function used in least squares fitting, controlled by the input parameter `sigma_wt`. As a source of variation among estimates, weighting sensitivity appears to be at least as important as grid resolution sensitivity.

(A4) *If vertical cell counts are too low, estimates are unreliable*
Estimating the vertical range becomes a problem for coarse grids, because the vertical cell count becomes too low for reliable estimation. In the case of MS2_12, the data volume is about $20\,\mathrm{m}$ thick, and resampling with a vertical resolution of $1\,\mathrm{m}$ (geogrid resolution) gives about ten empirical variogram values, which is enough to deduce the variogram shape somewhat accurately. By contrast, resampling at $5\,\mathrm{m}$ vertical intervals (simgrid resolution) only gives two points of the empirical variogram, which is not enough to reason effectively about the shape of the vertical component of the underlying variogram.

(A5) *The variogram models currently used are inflexible*
In the majority of variogram estimates examined, the fitted parametric variogram matches the empirical variogram quite loosely. The lack of fit seems to be explained in part by a lack of flexibility in the parametric variogram model used, which in turn has several causes. First, we are only using a single variogram function, which means we can only model correlation structure at one scale at a time. Second, the same variogram shape is assumed to apply in the all directions. Third, the sill variance is assumed to be the same in every direction. This is a simple and parsimonious model with a small number of parameters, but it has significant coupling between the vertical range and the major and minor horizontal ranges, and this seems to complicate the curve fitting.

Given these observations, there appear to be several avenues to potentially improving the variogram modeling approach taken in this project. The following points seem especially relevant.

(B1) *Adding flexibility to the variogram model*
There are many ways to relax the parametric variogram model so a tighter fit to the empirical variogram can be achieved. Although flexibility can be increased while sticking to a single variogram function, adding flexibility by nesting variograms seems like the natural next step in this direction. In practice, this would entail defining the variogram to be estimated as a linear combination of a small number, say 2–5, of elementary variogram functions, with some having short correlation ranges and others having longer ranges. Each elementary variogram contributes a certain proportion of the variance of the nested variogram, and these variance contributions must be estimated along with the other parameters. The parameter estimation can be simultaneous, fitting all the elementary variograms at once, or it can be sequential, in which case the long-range components would typically be fitted first, and the fine-scale components last.

(B2) *Decoupling vertical and horizontal correlations*
Using a separable covariance model that factorizes into horizontal and vertical factors would completely decouple the horizontal and vertical components of the variogram, simplifying the curve fitting. This is a strong assumption, with important implications for data conditioning. A partial decoupling, or loosening, is likely preferable. Again, nested variograms can achieve this if one or two of an elementary variogram's correlation ranges are allowed to tend to infinity. In that case, the elementary variogram in question does not contribute to the shape of the nested variogram in the directions where the ranges are infinite.

(B3) *Making use of interactivity*
The variogram estimator created in the project was intended to make the estimation as automatic as possible, with the possibility of manually checking the quality of the estimates by inspecting diagnostic plots comparing the empirical variograms and the paramtric funcions fitted to them. Some approaches to variogram estimation rely heavily on interactive model building, whereby the user adds, removes, adjusts and tweaks variogram components until a satisfactory fit is obtained. While this approach is very different from ours, we should not dismiss out of hand the notion that our estimation procedure could benefit from incorporating more interactivity. Computing the empirical variogram is much more computationally intensive than the subsequent weighted least squares curve fitting. If the heavy computations were done in advance, modifying the estimated parameters and visualizing the

corresponding change in the variogram curve could be done in real time. As could evaluating the goodness of fit of the updated model. This suggests a semi-automatic variogram modeling approach, where a suggested model is fitted in advance, and the user chooses whether to accept the suggestion as it is, or improve it through adjustment. This would be faster than a purely interactive approach, and more robust than a fully automatic approach.

(B4) *Taking trends into account*
The isolation and removal of trends in the input data has been considered to lie outside the scope of the project. That is not to say that there are no trends in the Delft3D models, or that the presence of a trend can be safely ignored. As has been demonstrated in the previous section, the variogram estimator performs differently when run on the MS2_12 data and when run on a realization drawn from a Gaussian random field with no trend. It is reasonable to suspect that the presence of a trend in the MS2_12 realization is at least part of the explanation. Trend removal would entail fitting a slowly varying trend function to the input data, say the porosity field, and subtracting it from the input data, producing a residual field. Variogram estimation would then be carried out normally using the residual field as input data. The estimation of the trend function should preferably be done before cropping, to make best use of the available information in the process model realization. If information about the location of the delta fringe and the main channel's entry point into the model is available, then it should be possible to use this information to support the trend estimation.

References

[1] Vegard Berg Kvernelv, Jacob Skauvold, and Ragnar Hauge. Geomodel parameter estimation from process model analogues. Project report/Note SAND/08/20, Norwegian Computing Center, August 2020.