3D modelling of particle-size distributions in the shallow subsurface: Zeeland, The Netherlands

Sam ROBERSON¹, Gert-Jan WELTJE¹

(1) Department of Applied Geology, Institute of Civil Engineering and Geoscience, Delft
University of Technology, Delft 2628 CN, The Netherlands,
s.l.roberson@tudelft.nl

Peer-reviewed IAMG 2011 publication

[doi:10.5242/iamg.2011.0076](https://doi.org/10.5242/iamg.2011.0076)

Abstract

A method for 3D interpolation of complete particle-size distributions is presented and applied to the Dutch province of Zeeland. Particle-size data are integrated into a common format using a cubic-spline interpolation scheme to create distributions with twenty-three particle-size categories ranging from 64 mm to 0.2 μm . Interpolation uses linear regression of log-ratio transformed particle-size distributions to estimate percentage frequency values for each of the twenty-three particle-size categories for each grid cell. The voxel model, has a spatial resolution (x,y,z) of 100 by 100 by 0.5 m and consists of ~ 1.7 billion voxel cells. This contains percentage frequency estimates for twenty-three particle-size categories. A full range standard particle-size statistics, e.g. mean, sorting, skewness and cumulative percentiles, can be calculated from these data.

1 Introduction

Knowledge of the physical properties of the subsurface are of considerable importance in the context of building and infrastructure development, groundwater management and resource extraction. In areas of high population density this knowledge becomes increasingly important as multiple uses of the subsurface develop. This scenario is present in the Netherlands, where the Dutch population density is among the highest in the world (487 km^{-2} in 2011), with more than 85% of the Dutch land surface being developed (van der Meulen and others, 2005). These activities put considerable pressure on use of the shallow subsurface. Water management comes high on the list of priorities in the Netherlands, where the half of the land surface lies more than 1 meter below sea level. Successful development of groundwater models requires detailed input of subsurface properties, including lithology and permeability.

A relatively small number of studies have attempted to directly model the spatial distribution of particle-size data in either two or three dimensions. Approaches have been far from unified, with the range of techniques being almost equal to the number of studies. The parameters interpolated, the interpolation techniques and the range of particle-sizes used have varied. For example, McArthur and Parsons (2005) focused on ordinary kriging of percentage values for a single particle-size category; Asselman (1999) used 2D block Kriging for interpolating grain size distribution parameters as mean, sorting and skewness; whereas Hengl and others (2007) used multinomial regression and regression-kriging of classification categories to derive soil texture estimates. One of the first attempts at interpolating complete particle-size distributions in three dimensions was made by Gruitjers and others (2005), who attempted to characterize the sediment facies underlying a modern river channel confluence by kriging cumulative percentage-frequency values of twelve particle-size categories.

1.1 Aims

This article presents the application of a three-dimensional interpolation scheme to predicting particle-size distributions in the Dutch province of Zeeland. The methods used to integrate a range of different particle-size data are outlined. These include: (i) modelling particle-size distributions for a common set of particle-size categories; (ii) the identification of the optimal number of particle-size categories to characterize the data, and (iii) log-ratio transformation of particle-size distributions to prepare the data for interpolation.

Figure 1: A map of borehole locations in Zeeland and the position of grid zone B5. Position of Zeeland in The Netherlands is shown inset.

1.2 Geological setting

The province of Zeeland is located at the southern extent of the North Sea Basin in the southwest of the Netherlands. The shallow subsurface geology consists of gently northerly dipping Tertiary and Quaternary strata, of which marine, estuarine and fluvial sediments comprise the majority of the geology. Holocene tidal channels, tidal flats and lagoon sediments alternating with peat beds, coastal shore face and dune deposits comprise the upper part of the sequence. Holocene deposits are up to 20 m thick in the northwest of the province. The Holocene deposits consist of clastic coastal and marine clastic sediments, inter-bedded with organic peat members.

1.3 Existing geological models

A number of geological models of the Netherlands currently exist (Weerts and others, 2004; Stafleu and others, 2010). The most recent model produced by Stafleu and others (2010) is a cell-based voxel model that focuses on the province of Zeeland. This high-resolution three-dimensional model consists of 50 million grid cells, with a spatial (x,y,z) resolution of 100 by 100 by 0.5 m, each with estimates of stratigraphy, lithology and lithofacies to a depth of 90 m below NAP (Normal Amsterdam Level). The model pays particular attention to heterogeneous members to model genetically dissimilar lithofacies. Estimates of sand particle-size classes were made using stochastic sequential indicator simulation. Although a frequently used approach to modelling binary indicator variables, stochastic simulation does not adequately capture the dependency relationships between different particle-size fractions within a single voxel. The technique is further limited by using relatively few categorical variables to describe a continuous spectrum of particle sizes. Applying stochastic simulation to binary indicators of particle-size distributions therefore removes a lot of information to start with and then attempts to compensate for data reduction later on, with a high probability of not honouring the original data.

To address these concerns the study presented here attempts to reduce the computational complexity of the spatial interpolation scheme, while at the same time retain as much high-resolution data as possible and capturing their spatial variability in a statistically robust manner.

Figure 2: A 3D block model of Zeeland looking from the south west showing the distribution of lithofacies types. Horizontal resolution is 500 m and vertical resolution is 0.5 m.

2 Methods

This section describes the characteristics of the particle-size data used for the interpolation scheme, the modelling approach used to integrate particle-size distributions into a uniform format, and how an optimum set of particle-size categories were identified. The methods used to prepare the data for interpolation are given, followed by details of the interpolation scheme itself. Some additional notes on how large volumes of data were handled are also provided.

2.1 The Data Set

A large volume of particle-size data are available for the Dutch shallow subsurface, sampled from borehole cores taken by TNO – the Geological Survey of the Netherlands. These are data are collated in three databases: DINO, DINO Qwa and Top Integraal.

The Top Integraal data set contains the most recently analysed samples, with particle-size distributions in thirty-two particle-size categories measured by laser-diffraction analysis from 2mm to 0.1 μm . The proportion of particles coarser than 2 mm are measured by sieving.

The DINO Qwa database contains particle-size distributions measured from c.1980 – 2005. These samples have been analysed predominantly by combinations of sieving, settling columns and laser-diffraction particle-size analysis techniques. These data are highly variable in terms of the number, size and range of particle-size categories given, covering coarse gravels (≥ 16 to < 63 mm) to clays ($< 2 \mu\text{m}$) in up to thirty categories. Historical preference in the Netherlands has been to focus on sand and silt sized particles.

The DINO database contains visual estimates of the relative contributions of different particle size categories, i.e. gravel, sand, silt, clay, peat and sand and gravel median particle-sizes. A total of 82189 particle-size measurements from 1680 boreholes were used: 80107 from DINO, 1447 from DINO Qwa and 635 from Top Integraal.

2.2 Data Integration

2.2.1 Interpolation of particle-size frequency values

In order to use all of these particle-size data within a single geological model, it was first necessary to convert these disparate data formats into a single uniform format with common particle-size categories. Data integration was achieved by interpolating cumulative frequency values at equal-width particle-size categories for each distribution. In order to use visual estimates of particle-size, the estimated percentages were converted into cumulative frequency distributions by summing the relative proportions of each size category (Fig. 3). In order to use estimates of median particle sizes it was assumed that the cumulative percentage frequency values of size-category medians (e.g. the sand median) lay mid-way between cumulative percentage frequency values of their adjacent size categories (see 'Zm' in Fig. 3).

The interpolation scheme used to predict cumulative percentage frequencies at equal-width particle-size category boundaries was a constrained cubic-spline algorithm (Kruger, Private correspondence). Cumulative percentiles were transformed in logit-space prior to interpolation in order to linearize the data and remove the restrictions of positive-sum percentage frequency data. The logit-scale is a binary log-ratio transformation, given as:

$$q = \ln \left(\frac{p}{1-p} \right) \quad \text{Equation 1}$$

where p is cumulative probability. The phi-scale was used for particle-size units in order to linearize the particle-size distributions further, as distributions frequently exhibit log-log characteristics. A linear extrapolation regression function was used to predict the out-of-range frequency values that represent the tails of the distributions. Below a cut-off value equivalent to 0.01%, the extrapolation function takes the form:

$$q_i = q_{i-1} + (x_i - x_{i-1}) \frac{\delta q}{\delta x} \quad \text{Equation 2}$$

where q_i is the predicted logit-transformed cumulative-frequency value, x_i is the largest particle in the i^{th} size category and $\delta q/\delta x$ is the slope of the least-squares linear regression model for the logit-transformed particle-size distribution. Beyond the cut-off value, a non-linear extrapolation algorithm is employed, where $\delta q/\delta x$ becomes an iteratively changing slope function given as

$$\frac{\delta q_i}{\delta x_i} = \sqrt[3]{\left(\frac{\delta q_{i-1}}{\delta x_{i-1}} \cdot A \cdot \frac{1}{i} \right)} \quad \text{Equation 3}$$

where A is the slope of the least-squares linear regression model for the logit-transformed particle-size distribution and the initial value of $\delta q/\delta x$ is the slope of last and penultimate data values from the known particle-size distribution.

Figure 3: Construction of a continuous cumulative frequency distribution (red lines) from visual estimates of particle size (black crosses) showing the different transforms used in this paper: (A) cumulative percentage frequency; (B) logit transformation and (C) centred-log ratio transform (clr). G_m = median particle-size of the gravel fraction., Z_m = median particle-size of the sand fraction. N.B. - In C the effect of a zero-replacement scheme on the visually estimated data can be seen by the offset between the visual estimates and the interpolated values. These values do not affect the interpolated scheme.

2.2.2 Data optimization

The optimal number of particle-size categories for the data set was calculated by comparing the difference between raw particle-size distributions and modelled particle-size distributions using the compositional distance metric, which is identical in form as Aitchison's simplicial metric (Aitchison, 2003). A simulation was run for a range of equal-width particle-size categories ($n = 3 - 10000$). The optimal number of classes for particle-size distributions were identified by comparing the results of the simulation with levels of data accuracy. The accuracy of the data was quantified by measuring the compositional distance between repeat measurements of one hundred and forty samples from the Top Integral data set.

Results indicated that particle-size distributions could be best characterized by interpolation to particle-size categories of 1 phi. Distributions were therefore interpolated to twenty-three equal-width size categories ranging from -6 phi to 16 phi.

2.3 Spatial Interpolation Scheme

2.3.1 Data preparation

The position of sediment facies within borehole cores were given as the top and bottom of the facies. Prior to interpolation borehole data were regularised to a vertical resolution of 0.5 m. The regularization process calculated the weighted average of all particle-size distributions contributing to each new 0.5 m cell. Weights were given as the proportional contribution of each sediment facies to the cell.

Following borehole regularization a centred-log ratio transformation (Aitchison, 1986) was applied to each particle-size distribution to avoid the limitations imposed by compositional data with positive sum constraints, thus avoiding issues caused by 'spurious correlations' (Pearson, 1897) within the data during the spatial interpolation. Perhaps more importantly, at least pragmatically, the use of log-ratio data is essential to avoid violations of the positive R^D simplex during interpolation..

2.3.2 Interpolation

Spatial interpolation of log-ratio transformed particle-size distributions onto a regular 3D voxel model was performed using three-dimensional linear interpolation. The model voxel grid has a spatial (x,y,z) resolution of 100 m by 100 m by 0.5 m, and is divided into nine grid zones (B1 to B9), each consisting of 21×10^6 voxels: a total of 1.7×10^9 voxels. The interpolation scheme was further subdivided into geological lithofacies, using the model created by Stafleu and others (2010). This allows particle-size distributions to be interpolated separately for each set of genetically-distinct facies type, avoiding the generation of unrealistic interpolation artefacts across boundaries between geological units.

Walvoort and de Gruitjer (2001) suggested that spatial trends in compositional data may be obscured by log-ratio transformations. However, these findings contradict several other studies that have demonstrated the advantages of log-ratio regression analyses for compositional data (Aitchison, 2003; Thomas and Aitchison, 2005; Tolosana-Delgado and von Eynatten, 2009). This study takes advantage of the mutual independence of the components of log-ratio transformed compositions. Spatial interpolation is performed on each category separately. This approach has considerable advantages in terms of time needed to process the data as the number of interpolation coefficients that need to be considered at any one time is reduced exponentially. In the example used in this paper, 23 particle-size categories are interpolated independently, which reduces processing time by at least an order of magnitude. For a 3D grid of nearly 2 billion voxels this is a significant saving.

After the interpolation has been performed the estimated log-ratio values are back-transformed into positive constant-sum percentage frequency values. This approach therefore obviates the need for manual adjustments of out-of-range percentage values, a procedure necessary in other studies using percentage-based interpolation schemes (Gruitjers et al., 2005).

2.3.3 Data handling

The subdivision of the interpolation scheme into smaller subroutines and individual particle-size categories allows it to run on relatively low performance 32-bit machines with limited memory capabilities. It should be noted though that any analysis of 21×10^6 particle-size distributions requires at least 3 GB of RAM to be performed, so the use of a 64 bit operating system is recommended. To improve storage efficiency, rather than store empty voxels as 'Nan' values, a hash index for voxels with real numbers was created, providing a reference of their position

within the existing 3D geological model. The hash array was stored alongside the interpolated values; this reduced file sizes by one to two orders of magnitude.

3 Results

The results of particle-size modelling are shown at a horizontal resolution of 200 m and a vertical resolution of 0.5 m for the central part of Zeeland (zone B5) in Figures 4 and 5. Each grid cell in the model contains predictions for twenty-three particle-size categories, ranging from -6 phi to 16 phi (64 000 to 0.02 μm) at 1 phi intervals. These provide high resolution particle-size distributions for the entire voxel model at the threshold of data accuracy (see [Data optimization](#)). From these particle-size distributions it is possible to derive mean, standard deviation and skewness statistics as well as particle-sizes for any percentile values. Modelling is performed using phi units, but results can be output in any units at whatever size intervals are required. End-users can therefore be provided with any particle-size statistics that they wish.

3.1 Lithological model

Figure 4 shows estimates of mean particle-size calculated from the interpolated particle-size distributions consisting of twenty-three particle-size categories. Clays and silts comprise the majority of the surficial deposits (reds through oranges to yellows), predominantly deposited in a tidal environment. The tidal flats are cross-cut by sandy deposits (green through cyans to light blues) that represent tidal channels.

Figure 5 shows the probability of the occurrence of coarse silts; particles smaller than or equal to 63 μm and greater than 31.5 μm . These data show that the majority of surficial tidal deposits have a low probability of coarse silt occurrence (0 to 0.05). At depth, horizontally contiguous bodies of several hundred square metres appear enriched in coarse silt relative to the surficial deposits. Probability estimates like those shown in Fig. 5 can be given for any range of particle-sizes across the model, depending what the end-user is interested in. Care should be taken though in interpreting these values, as they are relative proportions and do not represent absolute mass of sediments present.

Several geological units in the voxel model have insufficient data to inform the interpolation scheme. The absence of these geological units accounts for the 'holes' apparent in both Figures 4 and 5. The majority of boreholes reach a depth of 20 to 30 m and provide a great deal of detail for sediments in the shallow subsurface. However, the number of sediment samples available for the interpolation scheme reduces exponentially with depth. Additionally, the presence of extensive peat members in the subsurface results in a low sediment yield from the borehole cores, making particle-size analysis impractical.

Figure 4: A 3D block model of central Zeeland (Zone B5) looking from the south west showing the estimated mean particle-sizes for all interpolated lithofacies. Horizontal resolution is 200 m and vertical resolution is 0.5 m. See Fig. 1 for location.

Figure 5: A 3D block model of central Zeeland (Zone B5) looking from the south west showing the estimated probability of particles smaller than or equal to 63 μm and greater than 31.5 μm for all interpolated lithofacies. Horizontal resolution is 200 m and vertical resolution is 0.5 m. See Fig. 1 for location.

4 Conclusions

The methods presented above demonstrate that the estimation of complete particle-size distributions can be made for a large voxel model. The robust integration of particle-size data is a key process in the procedure, as is the identification of the optimal number of particle-size categories using levels of data accuracy to interpret the output of numerical simulations.

The application of a centred log-ratio transformation to particle-size data has considerable advantages in the interpolation step. Firstly, it avoids encountering any spurious spatial correlations caused by the nature of relative positive-sum data. Secondly, the independence of each component of the log-ratio transformed particle-size distribution circumvents the need to perform calculate a large number of interpolation coefficients. This significantly reduces the time required to perform the interpolation step. Thirdly, the estimation of complete particle-size distributions directly allows for the probability of any range of particle-sizes to given.

The results of this study are part of a larger modelling effort involving the subsurface of the entire Netherlands. Investigations into the performance of different interpolation techniques (bi-cubic regression, kriging, thin-plate-spline) are currently underway. These analyses will be though achieved though cross-validation and the application of the techniques presented to 'known' simulated data sets.

References

- AITCHISON, J., 2003. A concise guide to compositional data analysis, in: *Composition Data Analysis* workshop, p. 134.
- ASSELMAN, N.A.M., 1999. Grain-size trends used to assess the effective discharge for floodplain sedimentation, river Waal, the Netherlands. *Netherlands Journal of Sedimentary Research*, **69**(1): 51–61.
- GRUIJTERS, S., MALJERS, D., VELDKAMP, J., 2005. 3D interpolation of grain size distributions in the upper 5 m of the channel bed of three lower Rhine distributaries. *Physics and Chemistry of the Earth*, **30**: 303 - 316.
- HENGL, T., TOOMANIAN, N., REUTER, H.I. and MALAKOUTI, M.J. 2007. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran, *Geoderma*, **140**(4): 417-427.
- MCARTHUR, C.J. and PARSONS, M. 2005. *Spatial Analysis of Sediment Grain Size in the Vicinity of the Canaveral Harbor Ocean Dredged Material Disposal Site*, U.S. Environmental Protection Agency Report: EPA-904-R-05-002, p. 20.
- VAN DER MEULEN, M.J., VAN GESSEL, S.F. and VELDKAMP, J.G., 2005. Aggregate resources in the Netherlands. *Netherlands Journal of Geosciences*, **84**(3): 379-387.
- PEARSON, K. (1897). On a form of spurious correlation which may arise when indices are used in the measurement of organs, *Proceedings of the Royal Society London Series A*, **60**: 489-498.
- STAFLEU, J. MALJERS, D., GUNNINK, J.L., MENKOVIC, A. and BUSSCHERS, F.S., 2010. 3D subsurface modelling of Zeeland, in *Proceedings of the First International Conference on Frontiers in Shallow Subsurface Technology*, Delft University of Technology, The Netherlands, p. 224.
- THOMAS, C. W., and AITCHISON, J., 2005, Compositional Data Analysis of Geological Variability and Process: A Case Study. *Mathematical Geology*, **37**(7): 753-772.
- TOLOSANA-DELGADO, R. and VON EYNATTEN, H. 2009, Grain-Size Control on Petrographic Composition of Sediments: Compositional Regression and Rounded Zeros. *Mathematical Geoscience*, **41**: 869–886.
- WALVOORT D.J.J and de GRUITJER, J.J. 2001, Compositional Kriging: A Spatial Interpolation Method for Compositional Data, *Mathematical Geology*, **33**(8): 951-966.
- WEERTS, H.J.T., SCHOKKER, J., RIJSDIJK, K.F. AND LABAN, C., 2004. *Geologische overzichtskaart van Nederland / Geological map of the Netherlands*. TNO (Utrecht, The Netherlands).