Hydrol. Earth Syst. Sci. Discuss., 10, 9689–9720, 2013
www.hydrol-earth-syst-sci-discuss.net/10/9689/2013/
doi:10.5194/hessd-10-9689-2013

Hydrology and Earth System Sciences Discussions
Open Access

The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: testing outcrop hydrogeological parameters with independent borehole data

B. Rogiers[1,2], K. Beerten[1], T. Smeekens[2], D. Mallants[3], M. Gedeon[1], M. Huysmans[2,4], O. Batelaan[2,4,5], and A. Dassargues[2,6]

[1]Institute for Environment, Health and Safety, Belgian Nuclear Research Centre (SCK-CEN), Boeretang 200, 2400 Mol, Belgium
[2]Dept. of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200e – bus 2410, 3001 Heverlee, Belgium
[3]Groundwater Hydrology Program, CSIRO Land and Water, Waite Road – Gate 4, Glen Osmond SA 5064, Australia
[4]Dept. of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[5]School of the Environment, Flinders University, G.P.O. Box 2100, Adelaide SA 5001, Australia

[6]Hydrogeology and Environmental Geology, Dept. of Architecture, Geology, Environment and Civil Engineering (ArGEnCo) and Aquapole, Université de Liège, B.52/3 Sart-Tilman, 4000 Liège, Belgium

Received: 28 June 2013 – Accepted: 9 July 2013 – Published: 23 July 2013

Correspondence to: B. Rogiers (brogiers@sckcen.be)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Outcropping sediments can be used as easily accessible analogues for studying subsurface sediments, especially to determine the small-scale spatial variability of hydrogeological parameters. The use of cost-effective in situ measurement techniques potentially makes the study of outcrop sediments even more attractive. We investigate to what degree air permeameter measurements on outcrops of unconsolidated sediments can be a proxy for aquifer saturated hydraulic conductivity (K) heterogeneity. The Neogene aquifer in northern Belgium, known as a major groundwater resource, is used as case study. K and grain size data obtained from different outcropping sediments are compared with K and grain size data from aquifer sediments obtained either via laboratory analyses on undisturbed borehole cores (K and grain size) or via large-scale pumping tests (K only). This comparison shows a pronounced and systematic difference between outcrop and aquifer sediments. Part of this difference is attributed to grain size variations and earth surface processes specific to outcrop environments, including root growth, bioturbation, and weathering. Moreover, palaeoenvironmental conditions such as freezing-drying cycles and differential compaction histories will further alter the initial hydrogeological properties of the outcrop sediments. A linear correction is developed for rescaling the outcrop data to the subsurface data. The spatial structure pertaining to outcrops complements that obtained from the borehole cores in several cases. The higher spatial resolution of the outcrop measurements identifies small-scale spatial structures that remain undetected in the lower resolution borehole data. Insights in stratigraphic and K heterogeneity obtained from outcrop sediments improve developing conceptual models of groundwater flow and transport.

HESSD
10, 9689–9720, 2013

Testing outcrop hydrogeological parameters

B. Rogiers et al.

1 Introduction

Compared to core drilling for sample collection and analysis, outcropping sediments are easily accessible analogues for studying subsurface sediments. This outcrop analogue concept has been extensively applied in the oil industry for the analysis and modelling of reservoirs (e.g. Flint and Bryant, 1993; McKinley et al., 2004) resulting in various tools to characterize geological facies geometries, their connectivity and continuity (Pringle et al., 2004), and to create 3-D virtual outcrop models (Pringle et al., 2006). The concept has also been used with small-scale outcrops in unconsolidated material (e.g. Teutsch et al., 1998; Bayer et al., 2011), collecting both hydraulic and geophysical data. Most of these studies are more concerned with defining the geological facies geometry rather than determining the corresponding hydrogeological parameters and hence direct quantification of these parameters and certainly a comparison with the corresponding subsurface parameters is often lacking.

In slightly dipping unconsolidated stratigraphic settings, a very limited number of facies are generally encountered in a single outcrop. The information contained within such lithofacies type potentially represents key stratigraphic features and hydrogeological parameters for building conceptual groundwater flow models. Furthermore, different outcrops may represent different parts of a stratigraphic or landscape succession series (Beerten et al., 2012). The combination of several outcrops can then be used to obtain a composite picture of an aquifer system containing the same or at least similar sediments. As demonstrated by Rogiers et al. (2013a), the use of a hand-held air permeameter is a very accurate and cost-effective approach for quantifying hydraulic conductivity (K) and its spatial variability in situ on outcropping sediments. The question that remains however is how representative the obtained outcrop parameters are for the actual subsurface sediments.

In first instance, the outcrop sediments may differ in some aspects from their subsurface equivalents as a result of slightly differing depositional contexts, e.g. with respect to the position in the basin (palaeogeographical conditions). Inherently, this problem is

HESSD
10, 9689–9720, 2013

largely circumvented by comparing outcrop and subcrop sediments from one and the same formation.

Secondly, the outcropping sediments could also be influenced by post-depositional processes such as surficial weathering and compaction due to slightly different overburden sedimentation and erosion histories. During the initial loading of sands, a rapid increase of packing density and soil strength is expected due to grain reorganisation (Pettersen, 2007). As packing becomes tighter, further packing will be increasingly more difficult to achieve, each packing level is more stable than previous levels and deformation is permanent. This process should be visible in the porosity, bulk density and eventually K data of a progressively compacted material. Overconsolidated sands should however not show dilation properties, and unloading would thus have little effect. However, the amounts of silt and clay present throughout the Neogene aquifer sediments might initiate such dilation properties. Moreover, dissolution of certain mineral phases or framework grains by meteoric water might also enhance permeability, as shown by Lambert et al. (1997).

The objectives of this paper are therefore (i) to test whether the hydraulic conductivity and its spatial heterogeneity in outcrops obtained through air permeametry are comparable to those of nearby aquifer and aquitard sediments, (ii) to evaluate major differences between outcrop and aquifer sediment K heterogeneity including the transferability of information from outcrop to aquifer sediments, and (iii) to discuss the scale effect and overall outcrop parameter representativity for use in groundwater modelling. For this purpose the results from the outcrop study by Rogiers et al. (2013a) are compared with more standard borehole core analyses and pumping test results. Moreover, grain size analyses are used to verify the similarity between outcrop and subsurface sediments. In a final step, we provide possible explanations for the observed differences in K behaviour and options on how to integrate air permeametry-based data with existing knowledge available from borehole and pump test analyses in view of developing more reliable groundwater flow models.

HESSD
10, 9689–9720, 2013

2 Materials and methods

2.1 Hydrogeological setting and outcrop analyses

Rogiers et al. (2013a) proposed a methodology to characterize small-scale K variability from outcrops, and at the same time obtain outcrop-scale equivalent K values. This methodology relies on air permeability measurements that are converted to saturated K values using the empirical equation from Iversen et al. (2003), and a subsequent numerical upscaling step. This methodology was tested on five outcrops from three key formations of the Neogene Aquifer in north-eastern Belgium (from top to bottom): the Mol Formation (the abbreviation Fm will be used in the subsequent discussions), sandy and clayey parts of the Kasterlee Fm, and the clayey and sandy parts of the Diest Fm. For these five formations additional geological and hydrogeological data is available from a recent characterization campaign (Beerten et al., 2010) of the shallow aquifer sediments in Mol/Dessel (up to about 40 m depth), including seven cored boreholes (Fig. 2 in Rogiers et al., 2013a). This lithostratigraphical succession and its main characteristics are presented in Fig. 1. Apart from the minimum and maximum unit thickness obtained from this recent characterization campaign, typical borehole cores are displayed for each unit, as well as a typical cone penetration test (CPT), grain size and glauconite content profile through most of the units. The most striking features are the high clay and fine silt contents within the aquitard represented by the clayey part of the Kasterlee Fm, the sudden increase of the glauconite content in the sediments below this unit, and the contrast in coarse sand content between the upper and lower aquifers separated by the aquitard.

In addition to the individual air permeameter measurements (measurement scale of several cm^2) and their statistics, the measurement grids were numerically upscaled to obtain equivalent horizontal and vertical K values at the scale of the outcrop (i.e., typically several m^2; Rogiers et al., 2013a). Moreover, the air permeameter results were validated using independent constant-head laboratory permeameter measurements on $100\,cm^3$ ring samples taken from the same outcrop measurement grid. Therefore, the

HESSD
10, 9689–9720, 2013

K data obtained from the outcrops is deemed accurate and unbiased. An overview of all outcrops characterized by air permeameter measurements within the study area is provided by Rogiers et al. (2013b).

2.2 Constant head K measurements

To characterize the aquifer sediments' hydraulic conductivity variability, multiple undisturbed 100 cm^3 ring samples (with diameter of 53 mm) were taken from contiguous borehole cores (Beerten et al., 2010). The ring samples were pushed in the cores in horizontal or vertical direction, for characterisation of respectively horizontal or vertical K. The gathered data enclose several hundred hydraulic conductivity measurements on such 100 cm^3 ring samples from 7 cored boreholes, representing 350 m of core material. Two samples were taken each 2 m, for horizontal and vertical K, but the anisotropy at the sample scale was generally negligible (Beerten et al., 2010). The average thickness of the Mol and Kasterlee Formations in these boreholes is respectively 20 and 10 m. The highly stratified clayey part of the Kasterlee Fm – coarse sand layers alternate with heavy clay lenses with thickness varying from less than a cm to several cm – varies in thickness from 2 to 6 m. The Diest Fm is not penetrated fully by the cored borehole, but was characterized on average across 15 m.

All 100 cm^3 ring samples were analysed in the lab using the constant head method (Klute, 1965), using a low-pressure device for coarse material and a high-pressure device (approx. 6 bar) for the clay material expected to display low K values (see Beerten et al., 2010, for more details). Total porosity was also determined for most core samples, as well as bulk density; volumetric moisture content was measured for the outcrop samples. The methodology is similar to that used by Rogiers et al. (2013a) to validate the outcrop air permeameter measurements.

HESSD
10, 9689–9720, 2013

2.3 Grain size measurements

A sedigraph or a combination of standard sieving and a suspension cylinder (European standard EN 933-1) was used to quantify respectively 20 and 8 grain size fractions of the borehole core samples. All samples were prepared by removing carbonates and organic matter. Clay samples were analysed with the first method, after removing particles larger than 250 µm by sieving. For more details on the data, the reader is referred to Beerten et al. (2010) and Rogiers et al. (2012).

Grain size analyses of outcrop samples were performed by laser diffraction with a Malvern Mastersizer (Malvern Instruments Ltd., UK). This method consists of monitoring the amount of reflection and diffraction that is transmitted back from a laser beam directed at the particles, and quantifies 64 grain size fractions. Each sample was divided into 10 sub-samples by a rotary sample splitter to enable repeated measurements on a single sample, and all samples were measured at least twice. The final result was based on the average grain size distribution of all sub-samples. Note that particle sizes are expressed as the size of an equivalent sphere with an identical diffraction pattern.

2.4 Pumping tests

Step drawdown, constant discharge and recovery tests were performed at different locations within the study area, including some of the borehole locations. The transient groundwater head observations were interpreted with analytical as well as numerical models (Meyus and Helsen, 2012). Results from these large-scale tests are used here to illustrate the scale effect for hydraulic conductivity determination on subsurface sediments, and to compare such large-scale measurements with the numerically upscaled K values for the outcrops.

HESSD
10, 9689–9720, 2013

2.5 Variography

The experimental variograms are all fitted with spherical models, using a weighted least squares approach. Two approaches are tested: (1) treating outcrop and borehole data sets separately (variogram models for the outcrops are taken from Rogiers et al., 2013a), and (2) using a pooled data set which combines both outcrop and borehole data. In the latter case equal weight is given to both datasets in the least squares fitting. In the former case individual experimental variogram points are weighted according to the number of point pairs they represent. The initial variogram parameters for the nugget, total sill and range were respectivley set to the overall minimum semivariance, the data variance, and the maximum lag distance. In certain cases singular model fits occured due to non-uniqueness (data does not allow to discriminate between different equivalent models, e.g. pure nugget vs spherical model with zero range). The responsible parameters were then fixed at their initial value, before re-initialising the model fitting procedures. All variography was performed with the gstat package (Pebesma, 2004).

3 Results and discussion

3.1 Grain size distributions

Prior to comparing K values obtained from different measurement methods, a comparison is made between grain size distributions for the outcrop sediments and aquifer materials collected from cored boreholes (Beerten et al., 2010). This evaluation is necessary to verify if the outcrop and aquifer sediments represent the same lithostratigraphical units, and to highlight possible discrepancies between both to inform the comparison of their corresponding K values.

Overall there is good correspondence between outcrop/aquifer grain size distributions for the sandy part of the Kasterlee Fm and clayey part of the Diest Fm (Fig. 2a–

HESSD
10, 9689–9720, 2013

c), with a somewhat larger fraction of fines (i.e. between 2 and 22 µm) for the outcrop samples. Van Ranst and De Coninck (1983) suggested that post-depositional weathering of glauconite material, a green iron-rich clay mineral, might increase the relative amount of fines. Kasterlee Formation samples collected from boreholes contain glauconite up to a few percent, but for the Diest Formation it is at least 10 to 20 % (Beerten et al., 2010). The disintegration of the glauconite fractions in the outcrops could thus have increased the fines content.

The comparison further illustrates that the clay fraction (< 2 µm) of the clayey part of the Kasterlee Fm is about 20 % lower in the outcrop samples compared to the aquifer material. Since we are dealing with outcrop samples that are close to the surface, post-depositional migration of clay out of the clay lenses (e.g. Mažvila et al., 2008) together with bioturbation in the outcrops is a plausible explanation for the lower clay content in the outcrop. Weathering of clay lenses or drapes close to the surface would be another plausible explanation. For the clayey Kasterlee Fm outcrop, the individual grain size distribution curves (Fig. 2b) indicate a continuous gradation between two extreme cases, i.e. from a clay lens texture (approximately 40 % clay) to coarse sand without fines (> 90 % sand). The corresponding grain size distributions for boreholes show no overlap between the clay and sand samples, an illustration of the existence of two distinctly different materials within the clayey part of the Kasterlee Fm, i.e. heavy clay lenses embedded in coarse sands characterized by a sharp interface (Beerten et al., 2010).

In conclusion, weathering, clay migration, and bioturbation may have influenced the lower end of the outcrop samples' grain size distribution considerably. Furthermore, dissimilarities in palaeogeographic conditions and sediment source regions between the outcrop and borehole locations may equally explain such differences. However, the consistent stratigraphic position of the clayey Kasterlee Fm sediments on top of the Diest Fm and the relatively good correspondence in particle size for the sandy material (i.e. sand layers within the Kasterlee Fm), are sufficient underpinning arguments to support using the studied clayey Kasterlee Fm outcrop at Heist-op-den-berg (for details

HESSD
10, 9689–9720, 2013

of the outcrop see Rogiers et al., 2013a) as surrogate for the clayey Kasterlee Fm aquitard (Gulinck, 1963; Laga, 1973; Fobe, 1995). Additional insight could be obtained from tracing the exact origin and initial composition of the outcrop materials; however, this is beyond the scope of the current paper.

3.2 Hydraulic conductivity distributions

Figure 3 provides a comparison of outcrop and borehole (aquifer) K kernel density estimates of the probability density functions (pdfs) for the five sediments. Statistically significant differences exist for all sediments, with p values for F tests all below $4x10^{-3}$, while the corresponding t tests p values are all below $1x10^{-5}$ indicating statistically significant differences for both the variance and mean. All outcrop pdfs have higher mean K values than their borehole complement. While most outcrop samples display conductivities between 10^{-5} and 10^{-3} m s^{-1}, borehole samples have their most frequent K values between 10^{-6} and 10^{-4} m s^{-1}. Moreover, the standard deviations for the borehole samples are consistently larger than those based on the outcrop samples. The left tail of the pdfs tends to be much larger for the borehole data while the peaks tend to be wider (one to two orders of magnitude for the outcrops versus two to four orders of magnitude for the borehole data), especially for the sandy Kasterlee Fm (Fig. 3b). Relative variability expressed as coefficient of variation (CV) is approximately two times larger for borehole pdfs than for outcrop pdfs (Mol Fm: −13.4 % versus −5.9 %; Kasterlee Fm sands: −24.5 % versus −12.9 %; Diest Fm sands: −23.9 % versus −18.8 %) while it is similar for the clayey parts of the Kasterlee Fm (−23.9 % versus −18.8 %) and Diest Fm (−15.8 % versus −17.4 %). For the borehole data, sampling occurred over a large geographical area (several tens of km^2 versus as little as a few m^2 to at most a few tens of m^2 for the outcrops) and over a much larger depth (up to 50 m) thus having the opportunity to sample a much larger spatial heterogeneity.

Several characteristics typical of heterogeneity in K are however visible in both the outcrop and borehole K distributions. For the sandy part of the Kasterlee Fm (Fig. 3b), a long tail towards low values is present both in the outcrop and in the boreholes, while

HESSD
10, 9689–9720, 2013

the majority of samples is within a much narrower distribution in the outcrop. For the clayey part of the Kasterlee Fm (Fig. 3c), a multi-modal distribution is present for both datasets and representative of samples belonging mainly to clay lenses or sand layers. The clayey part of the Diest Fm (Fig. 3d) displays a similar pdf in both datasets (ratio of borehole to outcrop CV = 0.91), and the sandy Diest Fm data (Fig. 3e) shows the best absolute match in terms of the mean K, although the second peak with lower K values was not observed in the outcrop.

Validation of air permeameter K with core-based outcrop K demonstrated absence of systematic bias in the air permeameter K estimates (Rogiers et al., 2013a). Therefore, differences in K distributions between outcrop and aquifer sediments can be attributed to the scale of investigation (a single outcrop with a typical measurement grid of a few m^2 versus seven ~ 50 m deep vertical transects through the different lithostratigraphical formations, Fig. 1), different evolutionary states of the outcropping and subsurface sediments, and possibly different sedimentation conditions.

3.3 Linear rescaling correction

To investigate the (dis)similarities between the outcrop and borehole data across these five lithological units, the minimum and maximum values are plotted in Fig. 4, with all deciles (10th, 20th, ..., 90th percentile) in between. This shows that linear scaling of the outcrop values to the corresponding borehole distributions is possible for all outcrops. The extreme values are however not always in line with the centre of the distributions (as indicated by the deviation of the overall shape of the first and last line segments). All outcrops exhibit a more or less similar trend for at least part of the data, which is supported by the linear model fit on all minimum, maximum and decile points ($r^2 = 0.7$). The slope, larger than 45°, indicates that the deviation between outcrop and boreholes is larger for low K than for higher K values, which is consistent with the previous observations. The sandy Diest Fm curve lies apart and above the other curves, and is much closer to the 1 : 1 line of perfect agreement. This is as expected

HESSD
10, 9689–9720, 2013

based on the good correspondence in pdfs (see Fig. 3e). In other words, the Diest Fm outcrop is well and truly representative for the entire aquifer unit.

3.4 Porosity and compaction state

Weathering of clay layers at the surface has certainly contributed to produce higher K values for the fine material in the outcrops, but the systematic bias of about one order of magnitude that is also present for the sands remains unexplained.

Trends in porosity or bulk density with depth are very hard to detect in the borehole data due to the extensive layering of different lithologies and grain size distributions at the study area (the same lithology may occur at different depth depending on the geographical location). Moreover, the data from the outcrops are hardly sufficient to prove differences with the subsurface sediments are statistically significant. For example, the mean total porosity for the four Mol and Kasterlee Fm outcrop core samples is 43 % with a mean dry bulk density of 1.52 $g\,cm^{-3}$ (see Rogiers et al., 2013a), while the borehole values of the same two formations (43 samples) are 40 % and 1.60 $g\,cm^{-3}$ (samples between 2 and 28 m below surface). This is consistent with different compaction states, i.e. outcrop samples being less compacted than borehole samples, but the differences remain very small and are only significant for porosity at the 5 % significance level. However, even small differences in porosity can yield large differences in K (see discussion below).

The impact of the degree of compaction on K values was further investigated for the borehole dataset only using total porosity as proxy for compaction, as analyses in literature show that porosity has a high influence on K, given a homogeneous grain size distribution and chemistry (e.g. Bourbie and Zinszner, 1985). On an individual sample basis, it is hard to detect total porosity – K relationships within the borehole dataset, since these are very complex owing to the influence of grain size (Rogiers et al., 2012), sorting, packing and eventually the actual accessible pore throat radii (e.g. Bakke and Øren, 1997; Øren et al., 1998). However, as indicated by the scatterplot in Fig. 5, if total porosity and K are averaged for each formation and for each borehole separately, some statistically significant relationships exist. The slopes of the linear model fits are consistently positive, and in several cases, a change of a few percent in porosity can change K drastically. For instance, a one % decrease in porosity yields a decrease in K of minimum 0.14 and maximum 1.08 $\log_{10}$ units. This is a partial confirmation of the importance of the degree of consolidation and compaction on our K values; corroborating evidence about the effect of grain size, sorting and packing characteristics will be sought in future research.

An additional analysis of the K – depth below surface relationship was performed but did not yield any significant dependencies (results not shown). This is probably due to the alternation of different lithologies and grain sizes with depth, hence obscuring the influence of depth on compaction and thus on porosity and K.

3.5 The scale effect and vertical anisotropy

The representativity of K measurements – whether for outcrop or aquifer sediments – for characterizing a lithostratigraphical unit depends, among others, on the size of the measurement scale (or measurement support) and the spatial extent and lithostratigraphic complexity of the sampled domain. The effect of measurement scale for individual K measurements also impacts the overall variability, as measurements with a larger support volume, like pumping tests, average out the small-scale variabilities (Mallants et al., 1997). It is thus important in the comparison between outcrop and borehole K values to consider such scale-effects.

A comparison between the outcrop data (air permeameter based geometric mean K values and the calculated corresponding equivalent K values) and the subsurface data (borehole core geometric mean K values and the pump test K values) is shown in Fig. 6. It reveals the overall range is smallest for the outcrop data, both at the smallest measurement scale (data for air permeameter measurements spans 5 orders of magnitude versus 8 orders of magnitude for borehole cores) and at the largest scale (calculated equivalent outcrop K values show a range of ~ 2 orders of magnitude versus ~ 5 orders of magnitude for pump tests). It is further evident that the outcrop-based

HESSD
10, 9689–9720, 2013

equivalent K values are systematically higher than the mean borehole core values; a better correspondence is achieved with the pump test values.

Because a pump test represents a large support volume, easily tens to hundreds of m^3, small-scale heterogeneities have much less effect on such large-scale K values, hence the smaller data range. Furthermore, the support volume is commensurate with the computational domains used to calculate equivalent outcrop values. Overall the pump test values are generally only slightly smaller than the equivalent outcrop values, except for the clayey part of the Kasterlee Fm for which the discrepancy is about three to four orders of magnitude. This again emphasizes the need for a correction if outcrop K values are used to inform building conceptual groundwater models. Correction models such as those from Fig. 4 would account for impacts of different compaction and/or weathering processes, especially for the more clay-bearing sediments.

The arrows in Fig. 6 indicate different effects of upscaling for the aquifer and aquitard units. Moving from the sample (cm-scale) to the pumptest-scale (meter-scale) in most cases increases the aquifer geometric mean K values by one order of magnitude, while the outcrop values remain more or less constant when geometric means are compared with effective values. Unlike the other formations, upscaling the clayey part of the Kasterlee Fm data results in a decrease of the average K values, for both K_v and K_h pertaining to the aquifer and for outcrop K_v. This indicates that in both the outcrop and aquifer sediments of this particular lithostratigraphic unit a significant amount of small-scale heterogeneity is present (i.e. clay lenses) which significantly decreases the magnitude of the calculated effective K values.

Faulting could be another process involved enhancing discrepancies between small and large measurement supports. However, this process is considered to be absent as the study area is known as a zone of low seismic and limited tectonic activity (De Craen et al., 2012).

A comparison of the vertical anisotropy values (K_h/K_v) is shown in Fig. 7. The K_h/K_v ratios based on the geometric means of the 100 cm^3 borehole cores lies between 1 and 5. The two lithostratigraphical units with the highest K_h/K_v values are the sandy parts

HESSD
10, 9689–9720, 2013

of the Kasterlee and Diest Fm, which are influenced by some outliers that probably belong to the under- or overlying units. The equivalent outcrop K_h/K_v values are less than the corresponding borehole core anisotropy values, except for the clayey parts of the Diest and Kasterlee Fm. For the latter K_h/K_v increased more than one order of magnitude, when moving from the borehole core to the outcrop scale. The pump test anisotropy values mostly show larger values compared to those from the borehole cores, with a maximum vertical anisotropy of 10. The original Dessel 2 pump test interpretation by Lebbe (2002) yielded K values for the clayey part of the Kasterlee Fm and mentions a vertical anisotropy factor of 190 for part of the aquitard. This value was obtained by inverse modelling of the pump test, but due to a limited drawdown across the aquitard, the optimized parameter values remain highly uncertain. A more reliable estimate was obtained from the regional modelling of the Neogene aquifer and the flow across the aquitard by Gedeon and Mallants (2012). They obtain a vertical anisotropy of 148 by inverse conditioning on regional piezometric observations above and below the aquitard. The high vertical anisotropy determined from the outcrop ($K_h/K_v = 38$) supports these values, and indicates that such large values might be more realistic at larger scales.

3.6 Spatial variability

The vertical spatial variability for the outcrop and borehole data (K_h only) is compared in Fig. 8 and Table 1. For the Mol Fm, the outcrop data overall shows less variability (smaller semi-variance) than the borehole core samples; but correspond well with the experimental borehole variogram at the centimeter to meter scale. The larger total sill for borehole (0.13 + 0.41) compared to outcrop (0.05) is a reflection of the larger variability captured by the borehole data. This larger variability is caused in part by combining two local stratigraphical sub units into the Mol Fm (see Beerten et al., 2010) with thin gravel layers and clay lenses at their interface. The borehole data also displays a larger vertical spatial range (i.e. 20 m) than the outcrop (i.e. pure nugget), owing to samples being collected from a much larger vertical sampling window (up to 20 m) and

HESSD
10, 9689–9720, 2013

multiple boreholes spread over several km^2. As both experimental semivariograms are compatible, fitting the joint dataset improves the variogram model fit considerably.

For the Kasterlee Fm sands, the borehole and outcrop data show a large difference, which might be due to the rather limited number of borehole core samples identified as the sandy part of the Kasterlee Fm or an increased amount of heterogeneity in the outcrop due to weathering processes. The overall variability (total sill) is more or less similar for both outcrop and borehole data, suggesting that the variability captured by the outcrop samples may be used as surrogate for the variability in boreholes. Despite the presence of spatial correlation in the both datasets, the joint model fit shows a pure nugget because of the high semivariance values for the outcrop data.

The clayey Kasterlee Fm shows the largest spatial variability of all lithological units for both the outcrop and borehole data. While the outcrop shows some spatial correlation, the borehole model shows a pure nugget. The borehole cores show higher variability due to the clay-rich lenses and correspondingly low K values, which are altered in the outcrops, but only the first data point at 0.5 m is contradicting the outcrop data. The joint model fit does reveal their compatibility, and shows spatial correlation up to a few meters. This model might be more useful than the individual variogram models due to the integration of different scales.

Most of the clayey Diest Fm outcrop data seems to be compatible with the borehole core spatial variability. All three model fits show a range of one to two meters, and similar total sills. The sandy Diest Fm also exhibits similar total sill in all three cases, with a larger spatial range for the borehole data. The joint model fit is compatible with that of the borehole data, but shows a higher nugget due to the higher semivariances in the outcrop data.

Overall, the borehole data exhibit larger correlation lengths than the outcrop data. The total sills are mostly similar, except for two cases were the borehole data clearly encompasses more heterogeneity. Three out of five experimental variograms are overlapping at certain locations, indicating that at certain scales both datasets exhibit similar spatial variability. Fitting of the joint datasets results in these cases in more robust variogram models. This indicates that small-scale structural information, such as alternation of relatively thin clay and sand layers, and its effect on spatial variability in K may be preserved in outcrop sediments. Therefore, analysis of outcrop stratigraphy and hydraulic conductivity variability can yield valuable qualitative and quantitative insight about such properties for similar aquifer and aquitard sediments.

4 Perspectives

Despite the limitations of and systematic differences between the outcrop and borehole datasets, we have demonstrated that outcrop studies can provide useful information for developing more reliable groundwater flow and contaminant transport models. Because of the systematic differences observed here between outcrop and subsurface sediments, the obtained outcrop K values are not directly applicable in groundwater flow modelling, unless a correction is applied (Fig. 4). Furthermore, the different K distributions are comparable at least in a relative way, and linear scaling based on deciles was shown to be relatively accurate. In other words, results such as the spatial heterogeneity models, the equivalent vertical anisotropy factors, and relative differences between the different sediments provide us with information useful to guide conceptual groundwater flow model building and constraining model parameterisation.

Potential applications of our findings for building conceptual and numerical models of groundwater flow include (i) where possible highly structured heterogeneity should either be represented explicitly in the models or use should be made of appropriate geostatistical tools (e.g. multiple point statistics) based on detailed structural information visible in and quantifiable from outcrops, (ii) use of the obtained equivalent vertical anisotropy factors can influence conceptual model choices for isotropy/anisotropy for certain units, and the actual value represents a minimum of the parameter range in larger scale groundwater flow simulations (especially in a layered stratigraphical setting), (iii) to avoid over-parameterization, ratios between K values of different units can be fixed during model optimization (e.g., Gedeon and Mallants, 2012) using the ratios

HESSD
10, 9689–9720, 2013

obtained from equivalent outcrop estimates, and (iv) use of the obtained outcrop variogram models can complement information from a larger scale (e.g. boreholes), or be used for small-scale geostatistical simulations for detailed local transport simulations. All these applications will be most beneficial when combined with the traditional borehole coring and measurements and other invasive and non-invasive subsurface characterisation techniques.

5 Conclusions

Analysis of outcrop sediments considered to be analogues for various lithostratigraphical units within a sedimentary aquifer provided a qualitative understanding of aquifer and aquitard stratigraphy and a quantitative estimate about K variability at the centimeter- to meter-scale. Comparison between outcrop and independent borehole core K values revealed significant differences between both datasets. Such differences are believed to be induced mainly by weathering, different palaeoenvironmental conditions and differential compaction, and can be corrected for as was demonstrated on the basis of a linear model. Hence, outcrop information can be used for building better stratigraphic models including determination of spatial structure by variogram fitting for further use in geostatistical simulations. Moreover, the relative variability in K values with similar coefficients of variation for borehole and outcrop K, and the derived anisotropy values are very useful to get a more complete understanding of the heterogeneity within the Neogene Aquifer.

Comparison of outcrop and borehole K values demonstrated the borehole K probability density functions had broader peaks, longer tails towards low values, and the presence of a systematic bias. The reasons behind this discrepancy are manifold, and include weathering of the outcrop sediments and a lesser degree of consolidation and associated stress states in outcrops. Also, measurements performed on outcrops sometimes several tens of kilometres away from the main study site may further invoke differences in K. Grain-size analyses showed that the sediments from the investigated

HESSD
10, 9689–9720, 2013

outcrops and boreholes are similar but not necessarily exactly the same. Clay migration and bioturbation in the outcrop sediments probably contributed to the observed discrepancies, as well as slight differences in palaeoenvironmental settings. The degree of (over)consolidation and stress states might also have an impact, but further research is needed to confirm or quantify this, as trends with the current depth of the sediments are hard to detect due to the alternation of different lithologies.

Based on all data a linear scaling relationship was derived ($r^2 = 0.7$) that permits rescaling of outcrop K values to their subsurface equivalents. For most individual units, the differences between outcrop and subsurface sediments were similar (except for the extremes of the distributions). The sandy part of the Diest Fm however showed a considerably better fit between outcrop and aquifer than the other cases.

In a comparison with K values obtained through other means, outcrop-based equivalent K values were systematically higher than those from pump tests (especially for the clayey part of the Kasterlee Fm), whose support volumes are considerably larger than the simulation domains considered in the outcrops. Mean borehole core samples resulted in the overall smallest K values. Smaller compaction at shallow depth and long-term biophysical weathering processes presumably contributed to outcrop equivalent K values being larger than any other estimate of large-scale K available in this study.

In most cases the semivariograms for the outcrop and borehole data are compatible. Only for the sandy Kasterlee Fm the outcrop data clearly shows higher variability than the borehole data. Spatial correlation (i.e. increasing semivariance with distance) is present in most cases, either in the outcrop or borehole data, or both. The clayey Diest Fm shows however a pure nugget effect for both datasets. For the Mol Fm and the clayey Kasterlee Fm both datasets complement each other resulting in more robust semivariogram model fits. For the sandy Diest Fm there seems to be a discrepancy in the range between both datasets.

Given the small number and limited size of the studied outcrops, transfer of information from outcrops to the corresponding aquifer sediments can be improved by

HESSD
10, 9689–9720, 2013

expanding the number of outcrops for the same lithostratigraphical units. In addition, more complementary aquifer information could be collected for developing a depth-dependency in aquifer K that incorporates effects of compaction which could then be used to rescale outcrop K values to sediment values at a given depth. Such information, together with geostatistical parameters, may be used as input or prior information to stochastic flow models.

Next to the quantitative information tested in this paper, information about facies geometry, like the alternating clay and sand layers within the clayey Kasterlee Fm, cannot be revealed easily using available in situ methods, and represents very important qualitative knowledge obtained from outcrops.

Acknowledgements. The authors wish to acknowledge the Fund for Scientific Research – Flanders for providing a Postdoctoral Fellowship to Marijke Huysmans. ONDRAF/NIRAS, the Belgian Agency for Radioactive Waste and Enriched Fissile Materials, is acknowledged for providing the borehole data. Findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of ONDRAF/NIRAS.

References

Bakke, S. and Øren, P.: 3-D pore-scale modelling of sandstones and flow simulations in the pore networks, SPE J., 2, 136–149, 1997.

Bayer, P., Huggenberger, P., Renard, P., and Comunian, A.: Three-dimensional high resolution fluvio-glacial aquifer analog: Part1: Field study, J. Hydrol., 405, 1–9, 2011.

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., and Leterme, B.: Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel, Version 1, Brussels, Belgium: NIRAS/ONDRAF, 2010, 273 pp., NIROND-TR 2009-05 E V1, 2010.

Beerten, K., Deforce, K., and Mallants, D.: Landscape evolution and changes in soil hydraulic properties at the decadal, centennial and millennial scale: A case study from the Campine area, northern Belgium, Catena, 95, 73–84, 2012.

Bourbie, T. and Zinszner, B.: Hydraulic and acoustic properties as a function of porosity in Fontainebleau sandstone, J. Geophys. Res., 90, 524–532, 1985.

De Craen, M., Beerten, K., Honty, M., and Gedeon, M.: Geo-scientific evidence to support the I2 isolation function (geology and long-term evolution) as part of the Safety and Feasibility Case 1 (SFC1). Mol, Belgium: SCK-CEN, 2012, 120 pp., (External Report of the Belgian Nuclear Research Centre; ER-184; CCHO 2009-0940000 Research Plan Geosynthesis), ISSN 1782-2335, 2012.
Flint, S. S. and Bryant I. D. (Eds.): The Geological Modelling of Hydrocarbon Reservoirs and Outcrop Analogues, Blackwell Publishing Ltd., Oxford, UK, 1993.
Fobe, B.: Litologie en litostratigrafie van de Formatie van Kasterlee (Plioceen van de Kempen), Natuurwetenschappelijk tijdschrift, 75, 35–45, 1995.
Gedeon, M. and Mallants, D.: Sensitivity analysis of a combined groundwater flow and transport model using local-grid refinement: A case study, Math. Geosci., 44, 881–899, 2012.
Gulinck, M.: Essai d'une carte géologique de la Campine, Etat de nos connaissances sur la nature des terrains néogènes recoupés par sondages. Symposium Stratigraphie Néogène Nordique (Gand 1961), Mém. Soc. Belge Géol., 6, 30–39, 1963.
Iversen, B. V., Moldrup, P., Schjonning, P., and Jacobsen, O. H.: Field Application of a Portable Air Permeameter to Characterize Spatial Variability in Air and Water Permeability, Vadose Zone J., 2, 618–626, 2003.
Klute, A.: Laboratory measurements of hydraulic conductivity of saturated soil. In: Methods of soil analysis. Part 1, edited by: Black, C. A., Agronomy, 9, 210–220, 1965.
Konikow, L. F.: The secret to successful solute-transport modeling, Ground water, 49, 144–159, 2011.
Laga, P.: The Neogene deposits of Belgium. Guide book for the field meeting of the Geologists' Association London, Belgian Geological Survey: Brussel, 29 pp., 1973.
Lambert, M. R., Cole, R. D., and Mozley, P. S.: Controls on permeability heterogeneity in the Tocito Sandstone (Upper Cretaceous), northwest New Mexico, in: Mesozoic geology and paleontology of the Four Corners Region, edited by: Anderson, O. J., Kues, B. S., and Lucas, S. G., Guidebook, 48th Field Conference, New Mexico Geological Society, 217–228, 1997.
Lebbe, L.: Interpretatie van een pompproef uitgevoerd te Mol, Rapport GROMO2002/07 – RUG – Vakgroep Geologie en Bodemkunde, September, 2002.
Mallants, D., Mohanty, B. P., Vervoort, A., and Feyen, J.: Spatial analysis of saturated hydraulic conductivity of a macroporous soil, Soil Technol., 10, 115–131, 1997.
Mažvila, J., Vaicys, M., and Beniušis, R.: Causes and consequences of the vertical migration of fine soil fractions, Žemes Ukio Mokslai, 15, 36–41, 2008.

McKinley, J. M., Lloyd, C. D., and Ruffell, A. H.: Use of Variography in Permeability Characterization of Visually Homogeneous Sandstone Reservoirs With Examples From Outcrop Studies, Math. Geol., 36, 761–779, 2004.
Meyus, Y. and Helsen, S.: Multi en enkelvoudige pomptests in en rond de nucleaire zone van Mol-Dessel (terreinverkenningen 2008), NIROND-TR 2012-19, 2012.
Øren, P., Bakke, S., and Arntzen, O.: Extending predictive capabilities to network models. SPE Annual Technical Conference and Exhibition, San Antonio, Texas; 324–336, 1998.
Pebesma, E. J.: Multivariable geostatistics in S: the gstat package, Comput. Geosci., 30, 683–691, 2004.
Pettersen, O.: Sandstone compaction, grain packing and Critical State Theory, Petroleum Geosci., 13, 63–67, 2007.
Pringle, J. K., Westerman, A. R., Clark, J. D., Drinkwater, N. J., and Gardiner, A. R.: 3D high-resolution digital models of outcrop analogue study sites to constrain reservoir model uncertainty: an example from Alport Castles, Derbyshire, UK, Petroleum Geosci., 10, 343–352, 2004.
Pringle, J. K., Howell, J. A., Hodgetts, D., Westerman, A. R., and Hodgson, D. M.: Virtual outcrop models of petroleum reservoir analogues: a review of the current state-of-the-art, First break, 24, 33–42, 2006.
Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., and Dassargues, A.: Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks, Math. Geosci., 44, 739–763, 2012.
Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., and Dassargues, A.: The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Quantifying spatial variability in outcrop hydraulic conductivity, Hydrol. Process., in review, 2013a.
Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., and Dassargues, A.: Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium, Geol. Belg., 16, 129–148, 2013b.
van Ranst, E. and De Coninck, F.: Evolution of glauconite in imperfectly drained sandy soils of the Belgian Campine, Z. Pflanz. Bodenkunde, 146, 415–426, 1983.

HESSD
10, 9689–9720, 2013

Table 1. Overview of fitted variogram model parameters for the vertical experimental variograms (range = correlation length). The outcrop data is taken from Rogiers et al. (2013a).

Sediment	Parameter	Outcrop	Borehole	Pooled
Mol Formation	Nugget	0.05	0.13	0.04
	Sill	–	0.41	0.41
	Range (m)	–	19.66	12.46
	Type	– – – – Spherical – – – –		
Kasterlee Formation: sandy part	Nugget	0.16	0.00	0.25
	Sill	0.35	0.13	–
	Range (m)	1.36	2.90	–
	Type	– – – – Spherical – – – –		
Kasterlee Formation: clayey part	Nugget	0.40	2.07	0.60
	Sill	0.20	–	1.32
	Range (m)	0.36	–	2.20
	Type	– – – – Spherical – – – –		
Diest Formation: clayey part	Nugget	0.35	0.23*	0.33
	Sill	0.20	0.24	0.14
	Range (m)	2.07	1.17	1.12
	Type	– – – – Spherical – – – –		
Diest Formation: sandy part	Nugget	0.02	0.07	0.10
	Sill	0.18	0.11	0.06
	Range (m)	0.60	13.34*	13.34*
	Type	– – – – Spherical – – – –		

* Fixed during variogram model fit.

HESSD
10, 9689–9720, 2013

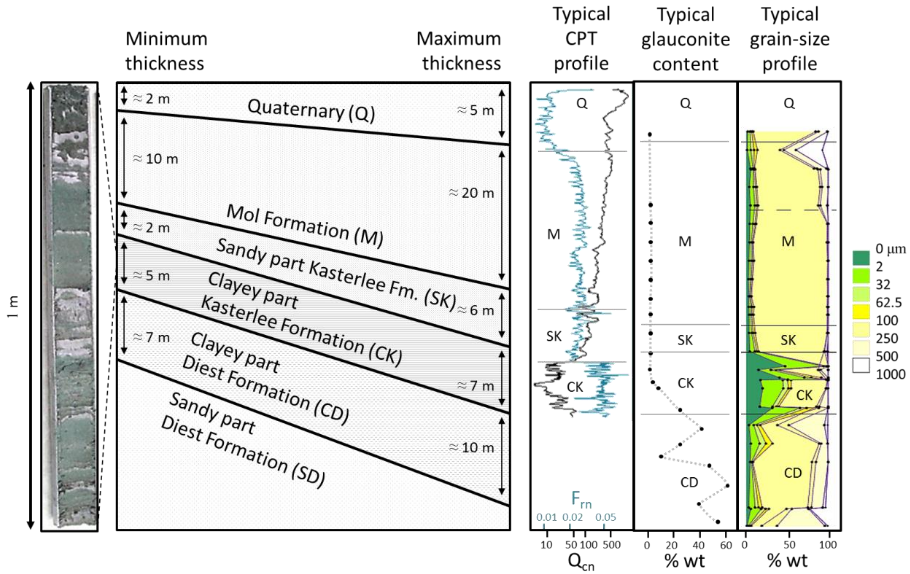

Fig. 1. Overview of the studied lithostratigraphical succession with formation thicknesses, a typical CPT profile (Q_{cn} and F_{rn} are the normalized cone resistance and friction ratios), typical glauconite content (weigth percentage; % wt), and a typical grain-size profile. A picture of a borehole core from the clayey part of the Kasterlee Formation is provided to illustrate its heterogeneity. For more information, see Beerten et al. (2010).

HESSD
10, 9689–9720, 2013

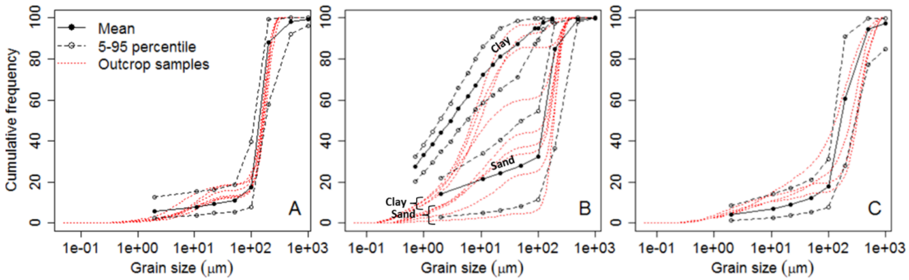

Fig. 2. Cumulative grain size distributions for the outcrop (laser diffraction) and borehole data (mean value and 5–95 percentiles from sedigraph or standard method; Beerten et al., 2010) for **(A)** the sandy Kasterlee Fm, **(B)** the clayey Kasterlee Fm and **(C)** clayey Diest Fm.

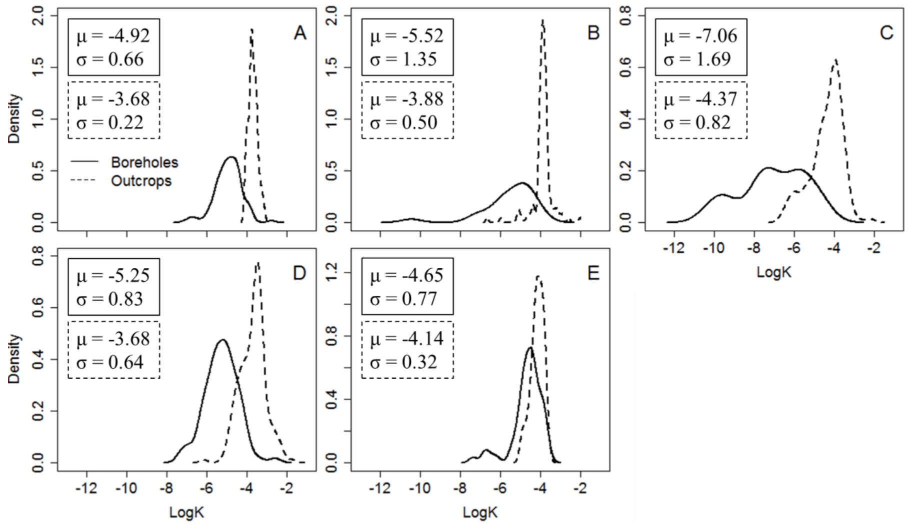

Fig. 3. Comparison between distributions (kernel density estimates of the probability density functions) for air permeameter based outcrop K and constant-head K measurements on undisturbed samples from cored boreholes, for **(A)** the Mol Fm, **(B)** the sandy Kasterlee Fm, **(C)** the clayey Kasterlee Fm, **(D)** the clayey Diest Fm and **(E)** the sandy Diest Fm. Mean (μ) and standard deviation (σ) are given for both data sources.

HESSD
10, 9689–9720, 2013

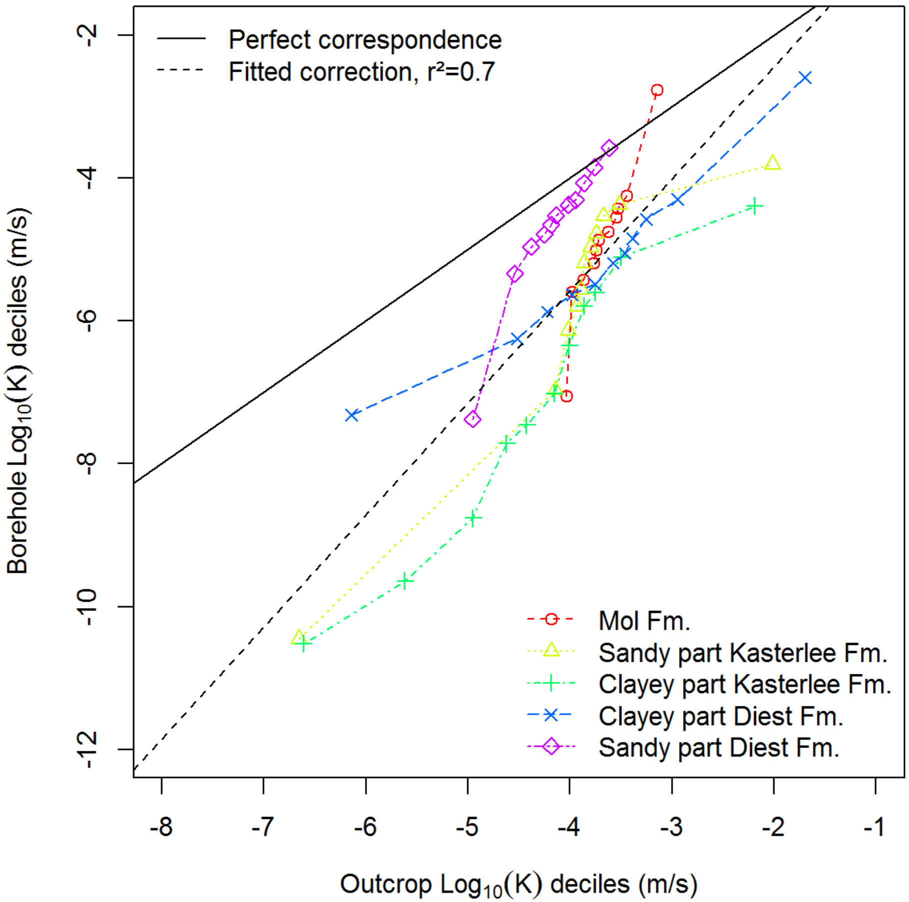

Fig. 4. Outcrops versus borehole $\log_{10}$ (K) deciles, and a fitted linear correction model ($y = 0.6938 + 1.5685x$).

HESSD
10, 9689–9720, 2013

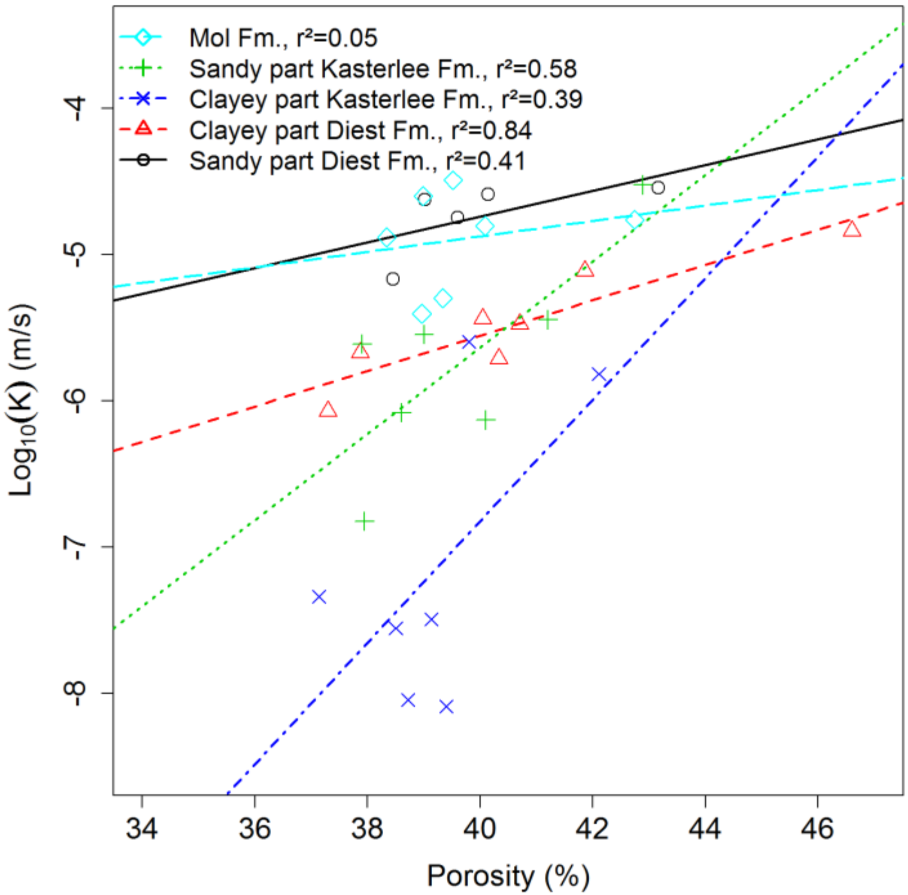

Fig. 5. Scatterplot of $\log_{10}$-transformed hydraulic conductivity *K* versus porosity (borehole dataset only) for the five lithostratigraphical units with corresponding linear model fits. Each data point represents the mean porosity and mean *K* of all measurements pertaining to one formation for one particular borehole.

HESSD
10, 9689–9720, 2013

Testing outcrop hydrogeological parameters

B. Rogiers et al.

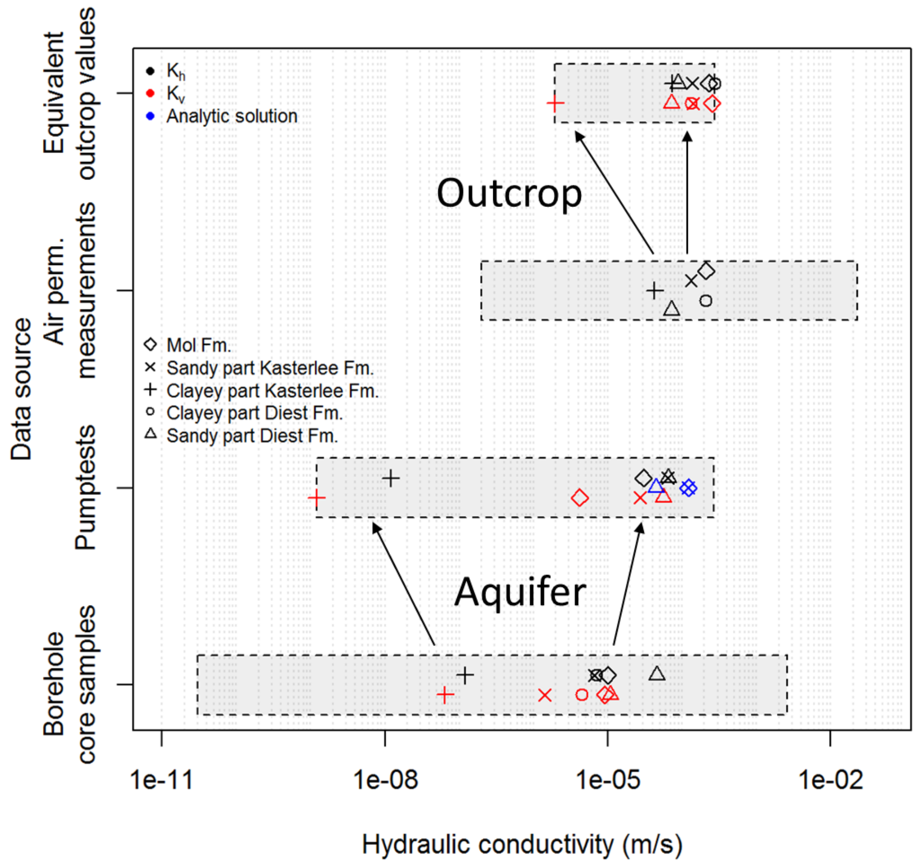

Fig. 6. Comparison of geometric mean *K* values obtained from borehole core samples, pump tests, outcrop air permeameter measurements and calculated equivalent values. The gray boxes represent the data limits, and the arrows indicate the contrasting effects of upscaling for the aquifer and aquitard units.

HESSD
10, 9689–9720, 2013

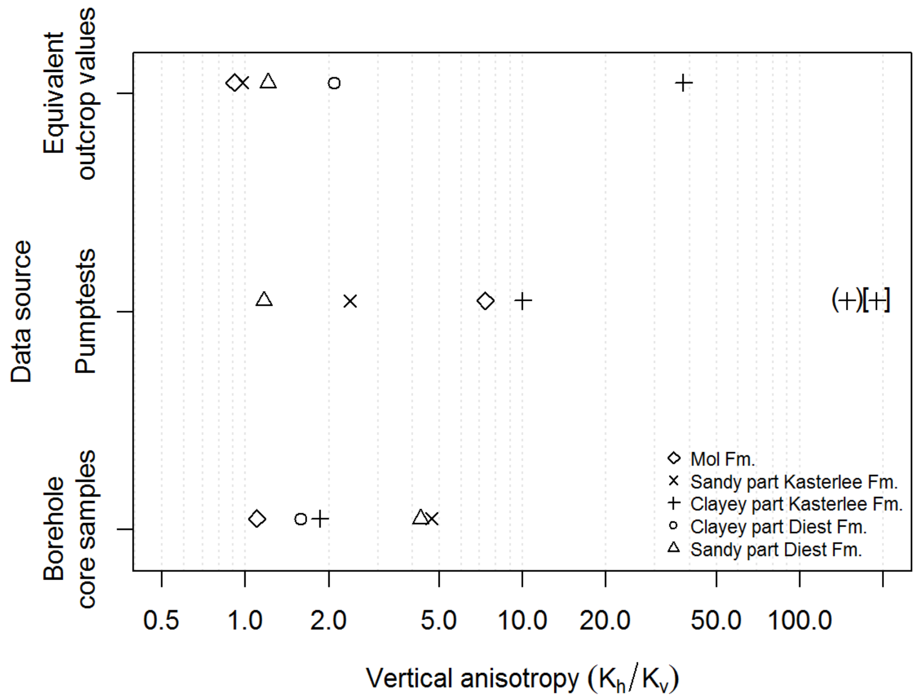

Fig. 7. Comparison of the vertical anisotropy factors derived from the geometric mean *K* values from Fig. 6. The plusses between round and square brackets represent respectively the parameter value obtained by Gedeon and Mallants (2012) using regional inverse modelling and the value representing a part of the aquitard in the original Dessel 2 pump test interpretation by Lebbe (2002).

HESSD
10, 9689–9720, 2013

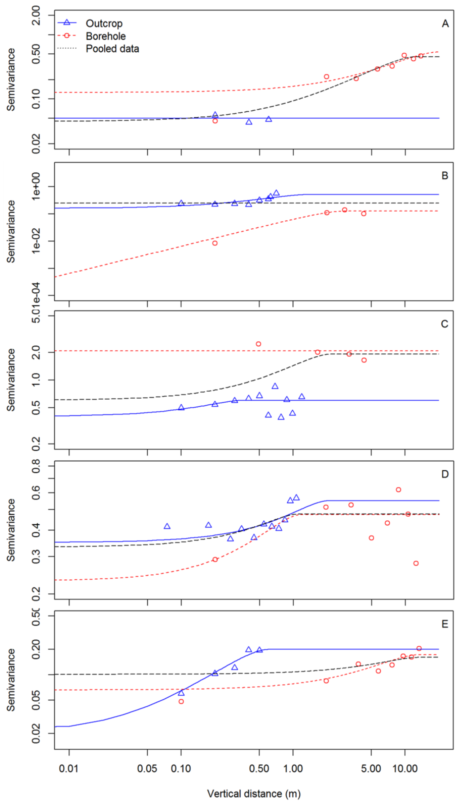

Fig. 8. Comparison between vertical experimental and modelled semivariograms (fitted using a least squares approach) for outcrop and borehole data. **(A)** Mol Fm, **(B)** sandy Kasterlee Fm, **(C)** clayey Kasterlee Fm, **(D)** clayey Diest Fm, and **(E)** sandy Diest Fm.

HESSD

10, 9689–9720, 2013

Testing outcrop hydrogeological parameters

B. Rogiers et al.