Nordic Hydrology, 26, 1995, 21-36

Prediction of Log-Transmissivity 2. Using Lithology and Specific Capacity

Steen Christensen

Department of Earth Sciences,
Aarhus University, Denmark

Two physically based models are derived that can be used to predict the log-transmissivity from the lithology, *e.g.* as logged in boreholes. The first model uses the lithology as the only predictor, whereas the second model uses the lithology and the specific capacity as simultaneous predictors. The model parameters can be estimated by statistical methods from actual observations.

The methods are applied to three case studies. A comparison with Christensen (1995) shows that the lithology in some cases is a better single predictor of log-transmissivity than the specific capacty; in other cases the opposite is true. This suggests that the proposed joint model is preferable to either of the single predictor models. In two cases the joint model improves the prediction considerably when compared with the single predictor models; in the third case the joint model is at least as good as the single predictor models.

Introduction

This is the second of two papers related to the prediction of aquifer log-transmissivity from well and borehole data. In the first paper Christensen (1995) summarizes the theoretical background for the prediction of log-transmissivity from the log-transformed specific capacity, derives the traditional linear statistical model for prediction and demonstrates the application of the method in three case studies of aquifers in different formations. The conclusions from these studies are briefly summarized together with the conclusions of the present paper, which focuses on the prediction of log-transmissivity from the lithological logs of boreholes.

A lithological log is a record of the rocks passed through by a borehole or a well, *i.e.* it is a qualitative description of the rock type and possibly the texture of cores sampled during the drilling of wells. Such logs are traditionally applied by groundwater modellers to set up the initial hydrogeological parameter fields of the model. For instance, the thickness of distinct layers within the formation may be determined quite accurately from the log. Futhermore, with experience the hydraulic conductivity of the layers may be estimated within some orders of magnitude from the rock type and texture. As the transmissivity is a function of the thickness and the conductivity of the layers that form the aquifer, the lithological logs can be used roughly to estimate the transmissivity at the well sites. However, it seems obvious to use statistical methods to estimate the values of the hydraulic conductivity of distinct layers and to quantify the uncertainty of the prediction.

The present paper summarizes the theoretical background for the prediction of log-transmissivity from the lithology, and a nonlinear statistical model for the prediction is proposed. Further, a simple method is proposed that improves the overall prediction by simultaneously using the lithology and the specific capacity as predictors. The application of the methods is demonstated by three case studies from the same aquifers as studied by Christensen (1995).

Theoretical Background

Prediction from Lithology

The transmissivity is the integrated conductivity of a water-bearing formation. For a layered aquifer with horizontal stratification the transmissivity is (Bear 1979)

$$T = \int_0^B K(z)\,dz \tag{1}$$

where we assume essentially horizontal flow. $K(z)$ is the hydraulic conductivity and B is the thickness of the water-bearing formation. B is equal to the thickness of a confined aquifer, whereas B of a phreatic aquifer depends on the water table elevation.

To use Eq. (1) in practice it is necessary to have qualitative *a priori* information about the vertical variability of $K(z)$. Based on this information an approximate model of $K(z)$ is introduced in Eq. (1). *E.g.* in fluvioglacial and limnic aquifers the deposits are approximate horizontally layered, and we may for instance use a model in which $K(z)$ is constant within each of $n$lay distinct homogeneous layers. In this case Eq. (1) reduces to

$$T = \sum_{j=1}^{n\,\mathrm{lay}} K_j B_j \tag{2}$$

where K_j is the conductivity and B_j the thickness of layer j. K_j is mostly unknown, whereas B_j is known or can be estimated accurately in a limited number of points from the lithological logs of wells and boreholes penetrating the aquifer. In the following it is assumed that B_j is a known parameter.

A statistical model for the prediction of $Y = \mathrm{Log}_{10}(T)$ can be

$$y_i = \log_{10}\left(\sum_{j=1}^{nlay} K_j B_{i,j}\right) + e_i \; ; \; i = 1,\ nobs \tag{3}$$

where y_i is the observed log-transmissivity, e_i is the observation error and *n*obs is the number of observations of y. The parameters, K_j may be estimated by the least-squares method from the observations of y, which are observed at different values of the predictor variables, B_j. If the measurement errors, e, are independent and have the same normal distribution then the least-squares estimation is the same as the maximum-likelihood estimation (Draper and Smith 1981).

Prediction from Lithology and Specific Capacity

Prediction models that join lithology and specific capacity can be derived by inserting Eq. (1) of the present paper in Eqs. (2) and (11) of Christensen (1995 – this issue). However, initial calculations showed that comparable or even better results are obtained by just using a linear combination of the prediction model from lithology (Eq. (3)) and the prediction model from specific capacity (Christensen 1995)

$$y_i = \alpha \log_{10}(SC_i) + \gamma + e_i \tag{4}$$

where SC_i is the observed specific capacity and α and γ are constants. The joint model is

$$y_i = \alpha \log_{10}(SC_i) + \beta \log_{10}\left(\sum_{j=1}^{nlay} K_j B_{i,j}\right) + \gamma + e_i \tag{5}$$

which reduces to

$$y_i = \alpha \log_{10}(SC_i) + \beta \log_{10}\left(\sum_{j=1}^{nlay} k_j B_{i,j}\right) + e_i \tag{6}$$

where ß is a constant and k_j is a scaled hydraulic conductivity (*i.e.* scaled with a factor of $10^{\gamma/\beta}$). The parameters of Eq. (6) (α, ß and k_j; j=1,*n*lay) can be estimated by the least-squares method.

It is noticed that α is the sensitivity of predicted log-transmissivity to a change in log-transformed specific capacity and that ß is the optimal weight given to lithology in order to make the best prediction of log-transmissivity possible. It is also noticed that the hydraulic conductivity cannot be estimated from Eq. (8) unless γ is known *a priori*. Presently our aim is, however, to predict log-transmissivity and not to estimate the hydraulic conductivity.

Data and Data Analysis

In the present study the transmissivity and lithological data are available for three Danish aquifers in differnt formations. A general description of the data and the methods of data analysis are given in the following, whereas specific information is given subsequently in the case study section.

The aquifers consist of consolidated or semi-consolidated layers of sand and gravel. Pumping test data are available for a number of pumping wells within the aquifers. The test results were analyzed by Jacob's straight-line method as described by Christensen (1995) and the estimated transmissivities at the least represent the aquifer transmissivity within som hundred metres distance from the well.

A lithological log and the specific capacity can be obtained for practically every Danish well from the database of the Geological Survey of Denmark (DGU). In Denmark the drilling contractor by statute has to supply DGU with cores from each layer penetrated by the borehole. At DGU the cores are analysed by geologists and the resulting log is recorded in the DGU-database along with the well data. Specific capacity is usually measured by the drilling contractor immediately after the well has been developed. The result is also reported to DGU.

Eq. (3) is used to estimate a model for the prediction of the transmissivity from the lithology of the aquifer and the corresponding residual analysis estimates the parameter and prediction uncertainties. One must expect that the prediction uncertainty depends on how realistic the aquifer lithology is modelled and therefore models are estimated that take into account varying numbers of layers within the aquifer as indicated by the lithological logs.

For each of the studied aquifers Eq. (6) is used to estimate a joint model that uses both lithology and specific capacity as predictors of Y. The prediction statistics of these models are compared to the statistics of the corresponding models, which are based on only one predictor, either lithology or specific capacity.

Case Studies

Data are available for the same three Danish aquifers as in Christensen (1995 – this issue): (A) a fluvioglacial, homogeneous aquifer; (B) fluvioglacial but heterogeneous aquifers; and (C) a homogeneous aquifer of tertiary freshwater deposits. The locations of the aquifers are shown in Fig. 1.

(A) Homogeneous Fluvioglacial Aquifer

In the western part of Zealand (Fig. 1) a 10 to 30 m thick layer of Quarternary fluvioglacial sand and gravel forms an confined aquifer which extends over an area of more than 100 km^2 (see Christensen 1994). The aquifer is overlain by glacial till and underlain by Tertiary clay. Further information on the geology is found in Christensen (1994).

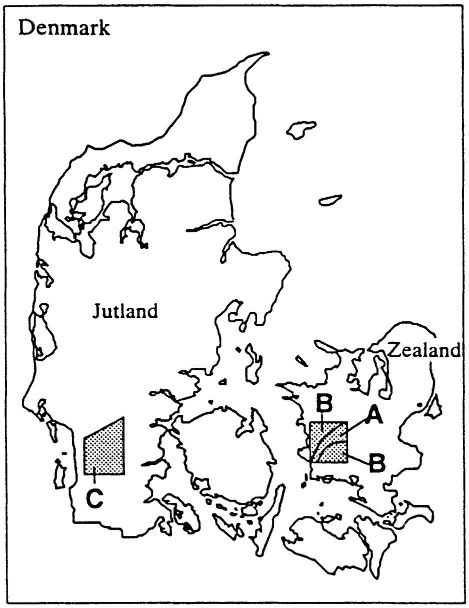

Fig. 1. Location of the studied aquifers.

Table 1 – Statistics for the pumping test results

Aquifer	Aquifer type	Y=Log$_{10}$(T)	
		μ	σ
A	Homogeneous, glaciofluvial	−2.14	0.40
B	Heterogeneous, glaciofluvial	−3.15	0.76
C	Homogeneous, limnic	−1.99	0.42

Transmissivities determined by pumping test analysis and lithological logs from the DGU-database are available for a total of 56 wells within the aquifer. The sample mean and the standard error of Y for these wells are −2,14 and 0,40 (Table 1) which shows that the aquifer is quite high-transmissive and homogeneous.

Approximately half of the wells fully penetrates the aquifer whereas the aquifer lithology below the bottom of the borehole of the other half is estimated from the lithological logs of neighbouring wells.

According to texture description of the lithological logs the aquifer can vertically be divided into layers of three sediment types: fine sand; fine to medium sand; and gravel. For each well the cumulative layer thicknesses of each sediment type are applied in the regression analysis. In 24 logs the texture analayses are missing from the database. At these sites it is therefore not possible to separate layers of fine sand from layers of fine to medium sand. Where no lithological log is available from the immediate vicinity of such a site it is assumed that all sand layers are fine- to medium-grained.

Table 2 – Statistics for regressions between lithology and transmissivity of the homogeneous fluvioglacial aquifer (A)

Case	No. obs.	Layer	K (10^{-4} m/s)	s_K	s_y	r^2
a	56	Fine sand	4.05	1.14	0.35	0.25
		Medium sand	2.98	0.45		
		Gravel	20.7	7.43		
b	56	Sand	3.20	0.39	0.35	0.24
		Gravel	20.1	7.37		
c	28	Sand	3.75	0.67	0.36	0.24
		Gravel	19.1	10.8		
d	56	Sand/Gravel	3.80	0.45	0.39	0.07
*e**)	56	Like *case b*			0.31	0.42

*) Prediction model including specific capacity

The statistics of five regression analyses between lithology and Y are listed in Table 2. K is the estimated hydraulic conductivity of the individual layers; s_k is the standard error of K; s_y is the standard error of the residuals (or the "standard error of prediction"), *i.e.*

$$s_y^2 = \sum_{i=1}^{nobs} \frac{(y_i - y_i^p)^2}{nobs - npar} \tag{7}$$

where y_i^p is the predicted value of y_i and $n$par is the number of estimated model parameters.

r^2 is equal to

$$r^2 = \frac{\sum_{i=1}^{nobs} (y_i^p - y^m)^2}{\sum_{i=1}^{nobs} (y_i - y^m)^2} \tag{8}$$

where y^m is the mean of y. The quantity r^2 is named the coefficient of determination and is the proportion of the variance in Y that is explained by the regression.

To test a null hypothesis that a set of g regression coefficients are all zero (*i.e.* that a set of g predictors has no relation to the response Y) one can use the F-statistic (Wonnacut and Wonnacut 1977)

$$\underset{\sim}{F} = \frac{\Delta r^2/g}{(1-r^2)/(nobs-npar)} \tag{9}$$

where Δr^2 is the increase in r^2 due to the g predictors; and in the denominator r^2 is due to all *n*par predictors. The null hypothesis is rejected at a chosen level of significance, if F is smaller than the critical point of the F-distribution with (g, *n*obs-*n*par-1) degrees of freedom. Two null hypotheses are tested in the following:

The first null hypothesis is that the specific capacity data does not improve the joint prediction model; *i.e.* that model Eq. (3) may be used as well as the joint model Eq. (6). The latter model has *n*par=*n*lay+2 parameters, whereas the former model has *n*lay parameters; *i.e.* g=2.

The second null hypothesis is that the lithology data does not improve the joint model; *i.e.* that model Eq. (4) may be used as well as the joint model Eq. (6). In this case the latter model has *n*lay more parameters than the former; *i.e.* g=*n*lay.

Case a – In this case the aquifer is divided into the three types of layers mentioned above. The estimated conductivity of the fine sand is comparable to that of the fine to medium sand, namely about 0.0003 m/s, whereas the standard error of the estimated conductivity of the former layer is more than twice that of the latter. One may therefore conclude that the regression analysis cannot reveal a significant difference between the conductivity of the layers that are recorded as fine sand and of those that are recorded as fine to medium sand. The statistics of the three-layer regression are thus similar to the statistics of the two-layer regression of *case b,* Table 2.

Case b – This case differs from *case a* in that the aquifer is divided into two types of layers only: layers of sand, and layers of gravel. The estimated conductivity of the sand at 0.0003 m/s is within the range of conductivity for clean sand (Freeze and Cherry 1979) and the estimated conductivity of the gravel a 0.0020 m/s is at the lower end of the conductivity range for gravel. The standard errors of the estimated conductivities are rather large, *i.e.* up to about 35% of the estimated mean.

The prediction error of the two-layer model is 0.35 and the proportion of variance explained by the regression, r^2, is 0.24. This is somewhat better than r^2=0.17 of the linear model between Y and log-transformed specific capacity estimated by Christensen (1995).

Case c – The regression analysis of *case b* is based on 56 data sets of which only 15 have complete lithological logs, texture data is lacking in 13 logs, 17 logs are from boreholes that only partially penetrate the aquifer and 11 logs from partially penetrating boreholes also lack the texture analysis. In practice the layer thicknesses thus have to be estimated, which contradicts the assumption made in the development of Eq. (3). The imperfection of lithological data introduces some unknown amount of uncertainty into the regression analysis.

In the present case the regression analysis is based on the 28 logs from the fully penetrating boreholes both with (15) and without (13) complete texture analyses. That is, the imperfection of the lithological data is reduced as compared to the data used in *case b*. According to Table 2 the proportion of the variance explained by the estimated model is comparable to that of *case b*. The lack of improvement of r^2, suggests that the model estimation in the previous case is not seriously influenced by the fact that the bottoms of the aquifers were estimated at half of the well sites.

Case d –In this case the aquifer is only represented by its entire thickness at the 56 well sites. The statistics of Table 2 show that the single-layer model yields the worst fit, *i.e.* the model explains practically none of the variance of *Y*.

Case e – In this case the estimated model uses both the lithology of the borehole and the specific capacity of the well as predictors of *Y*. A two-layer lithological model like that of *case b* is assumed, and the parameters of the model Eq. (6) are estimated on the basis of the data of all 56 wells.

The statistics of Table 2 show that the prediction error in the present case is 0.31 and the proportion of the variance explained by the estimated model is 0.42, whereas the corresponding values of *case b* were 0.35 and 0.24 respectively. This shows that the simultaneous use of lithology and specific capacity improves the prediction considerably.

The first null hypothesis, that the specific capacity data does not improve the prediction, is rejected at a 0.1% level of significance. The second null hypothesis, that lithology does not improve the prediction, is also rejected at a 0.1% level of significance (r^2 of the estimated model Eq. (4) is 0.19).

Summary – For the present homogeneous fluvioglacial aquifer the fit of the two-layer model, in which sand layers are separated from gravel layers, is significantly better than the fit of the one-layer model, in which the entire aquifer thickness is applied as the predictor of *Y*. The proportion of the total variance that is explained

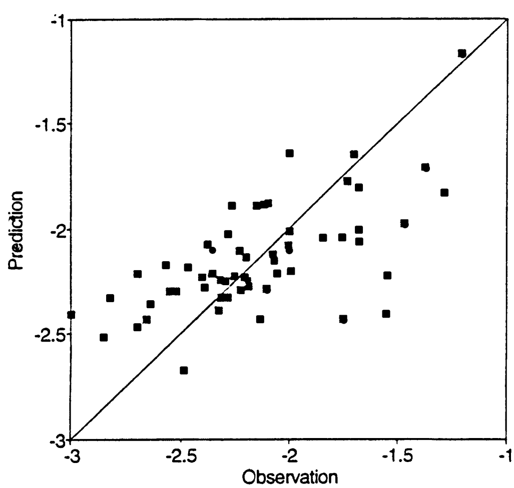

Fig. 2.
Y observations and predictions from the homogeneous fluvioglacial aquifer.

by the estimated two-layer model is 0.25 which is somewhat larger than that of the model of Christensen (1995) who used the specific capacity as a single predictor of Y (r^2=0.17).

For the estimated joint model r^2 is 0.42. That is, if the lithology and the specific capacity are used simultaneously as predictors of Y then the estimated model is significantly better than a model that uses either lithology or specific capacity as a single predictor of Y. However, Fig. 2 shows that the predictions of the joint model obviously tend to be too small in the high-transmissive zones and too large in the low-transmissive zones of the aquifer respectively. This tendency is even clearer for the single predictor models, especially the model that is based on specific capacity.

(B) Heterogeneous Fluvioglacial Aquifers

Within the catchment east and west of the previous aquifer (Fig. 1) the fluvioglacial layer is heterogeneous and embedded in the glacial till. The layer is present within most of the catchment but the thickness varies greatly within short distances. It is likely that the hydraulic conductivity is also highly variable. The fluvioglacial deposits form confined aquifers from which groundwater is hardly recoverable in some areas and recoverable in other areas. More details about the hydrogeology are given by Christensen (1994).

The transmissivity of 24 wells was estimated by pumping test analysis. The average Y for these wells is −3.15 with a standard error of 0.76 (Table 1) showing that the transmissivity of the aquifers is generally smaller and more variable than that of the homogenous aquifer (A).

The aquifer is fully penetrated by the boreholes of 15 wells for which the lithological logs are recorded in the DGU-database. However, the texture of 7 of these logs are missing from the database. Nine wells penetrate the aquifer only partially and 5 of these lack texture data. The lithological information is thus complete at only 8 well sites, whereas the lithology at the remaining 16 sites has to be partially estimated from nearby complete logs. If such nearby logs are also lacking it is assumed that all sand layers are fine- to medium-grained.

The statistics of the regression analysis between Y and lithology are listed in Table 3.

Case f – In this analysis the cumulative layer thicknesses of each of the three sediment types are applied as predictors. The estimated hydraulic conductivities (Table 3) are up to one order of magnitude smaller than those estimated for the homogeneous fluvioglacial aquifer (A). The estimated conductivities of the fine sand and the fine to medium sand are within the expected ranges, while the conductivity of gravel is just below the expected range, Freeze and Cherry (1979). The standard errors of the conductivities are large, *i.e.* almost up to 100% of the estimated mean.

Table 3 – Statistics for regressions between the lithology and transmissivity of the heterogeneous fluvioglacial aquifer (B)

Case	No. obs.	Layer	K (10^{-4} m/s)	s_K	s_y	r^2
f	24	Fine sand	0.15	0.14	0.60	0.44
		Medium sand	1.09	0.37		
		Gravel	3.86	2.95		
g	12	Fine sand	0.11	0.09	0.48	0.67
		Medium sand	1.78	0.73		
		Gravel	24.0	22.3		
h	24	Sand	0.75	0.24	0.62	0.36
		Gravel	4.36	3.29		
i	24	Sand/Gravel	0.98	0.30	0.65	0.28
j *)	24	Like *case h*			0.53	0.58

*) Prediction model including specific capacity

The standard error of prediction is 0.60 and the proportion of variance explained by the regression is 0.44.

Case g – As mentioned above the lithological logs recorded in the DGU-database are only complete at 8 well sites. Texture data are also available at 4 other sites but these boreholes do not penetrate the aquifer completely. At the remaining 12 sites there is no on-the-spot texture information. In the present case the analysis is based on the 12 borehole logs of which the texture is described in the DGU-database.

The estimated conductivities of the fine sand and the fine to medium sand are comparable to those of *case f,* whereas the gravel conductivity estimated in the present case is almost one order of magnitude larger than that of the previous case. The estimated conductivities are within the expected ranges but the standard errors of the estimates are still large. The prediction error of the present case is smaller and the proportion of variance explained by the regression is larger than those of *case f.* One may therefore conclude that the estimated model parameters and the error of prediction depend on the quality of the data which is used for estimation and prediction respectively.

Case h – In this case only two predictors are applied: layers of sand; and layers of gravel. The estimated conductivities and the statistics of this case are quite comparable to those of *case f.*

Case i – In *case i* the entire aquifer thickness is the only predictor of Y. The prediction error is larger and the proportion of variance explained is smaller than those of the previous cases. The one-layer model therefore seems to be less reliable for the prediction of Y than the multi-layer models.

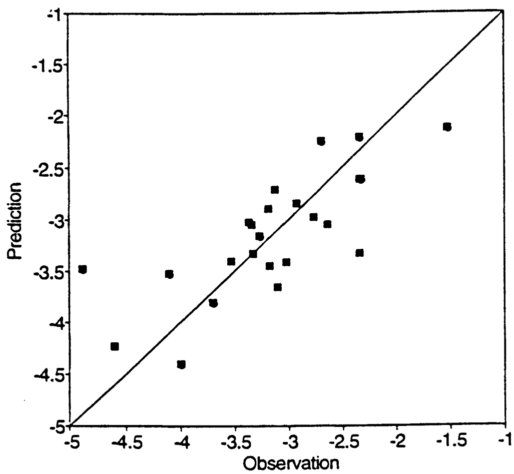

Fig. 3.
Y observations and predictions from the heterogeneous fluvio-glacial aquifer.

Case j – In this case a model is estimated that uses lithology and specific capacity as joint predictors of *Y*. Similar results are achieved whether a two-layer or a three-layer model of the lithology is assumed. In the following the joint model is compared to the two-layer model of *case h*.

Compared to the model of *case h* the joint model reduces the prediction error from 0.62 to 0.53, whereas the proportion of explained variance is increased from 0.36 to 0.58, Table 3.

The first null hypothesis, that the specific capacity data does not improve the prediction, is rejected at a 2.5% level of significance. The second null hypothesis, that lithology does not improve the prediction, can not be rejected at a 25% level of significance (r^2 of the estimated model Eq. (4) is 0.53).

Fig. 3 shows that the predictions made with the joint model are equally good within the entire range of observations.

Summary – For the present heterogeneous fluvioglacial aquifers the lithological logs recorded in the DGU-database can be applied as rather poor predictors of *Y*. One can use either the two-layer model, which distinguishes between sand and gravel layers, or the three-layer model, which distinguishes between layers of fine sand, fine to medium sand and gravel. In contrast, the one-layer model, in which the aquifer thickness is the only predictor, seems to be a less reliable model for the prediction of *Y*.

The estimated conductivities and the prediction errors are sensitive to the quality of the data applied in the regression analysis. When possible one should therefore only apply the lithological logs which include texture data.

If linear regression is made between *Y* and the log-transformed specific capacity of the present 24 wells (*i.e.* lithology is not used) then the prediction error of this model is 0.53 and the proportion of variance explained is 0.53. These values are practically the same as when both lithology and specific capacity are used as predictors. That is, the lithology does not improve the prediction significant, and one may just as well use the specific capacity as single predictor of Y.

(C) Homogeneous Tertiary Aquifer

The third case study is from the southwestern part of Jutland (Fig. 1) where aquifers are found in the Quarternary and Miocene formations. The present aquifer is in a 15 to 40 m thick Miocene formation of limnic sand extending over an area of some thousands of square kilometres. This formation, which is named the Ribe formation, is overlain by 150-200 m of deposits of Miocene clay and sand (containing substantial amounts of mica) and Quartenary sand and till, whereas the underlying deposits are marine clay and sand from the Miocene or Oligocene.

Transmissivity estimated from pumping test analyses and lithological logs of the aquifer are available at 17 wells within a 500 km^2 sub-area. The estimated values of Y are relatively high, with a mean at -1.99, and a moderate standard error at 0.42 (Table 1) which indicates that the aquifer is quite homogeneous.

The aquifer is penetrated by 11 boreholes but the texture of the lithological log of one of these holes is lacking. For the 6 boreholes that do not penetrate the aquifer the lithology below the bottom of the hole is estimated roughly from the nearest wells that do penetrate the aquifer. The texture has to be estimated for the entire log of two of these boreholes.

According to the texture of the lithological logs the limnic formation mainly consists of layers of quartz sand, but a thin layer of gravel appears at three sites. The sand layers are mostly medium- to coarse-grained but layers of fine sand also appear in more than half of the profiles. (It is mentioned that due to the few observations of gravel no reliable estimate could be made of a three-layer model).

The statistics of the regresion analysis are listed in Table 4.

Table 4 – Statistics for regressions between the lithology and transmissivity of the limnic aquifer (B)

Case	No. obs.	Layer	K (10^{-4} m/s)	s_K	s_y	r^2
k	17	Fine sand	0.26	0.41	0.31	0.51
		Medium sand & Gravel	5.34	0.97		
l	17	Medium sand & gravel	5.59	0.95	0.31	0.48
m	17	Sand/gravel	4.03	0.92	0.41	0.08
*n**)	17	Like *case l*			0.25	0.70

*) Prediction model including specific capacity

Case k – Two predictors are applied in this case: the cumulative thickness of fine sand; and the cumulative thickness of the coarse grained deposits (sand and gravel). The hydraulic conductivities (Table 4) are estimated within the expected ranges (Freeze and Cherry 1979). The standard error of the estimated conductivity of the medium sand/gravel is less that 20% of the estimated mean, whereas the standard error of the estimated fine sand conductivity is large.

The proportion of variance explained by the regression is 0.51 which is relatively large compared to the two previous aquifers (A) and (B).

Cases l and m – For the above mentioned reasons only the cumulative thickness of the coarse deposits is used as predictor in *case m*. As expected, the results of *case k* and *case l* are practically identical (Table 4).

In *case m* the entire aquifer thickness (the cumalative thickness of all sand and gravel layers) is applied as a predictor of Y. Table 4 shows that in this case the estimated model explains practically none of the sample variance of Y.

Case n – In this case the specific capacity of the well and the cumulative thickness of coarse deposits are used as simultaneous predictors of Y. Table 4 shows that when the present case is compared to *case l* the prediction is significantly improved: the prediction error is reduced from 0.31 to 0.25; and the proportion of explained variance is increased from 0.48 to 0.70.

The first null hypothesis, that the specific capacity data does not improve the prediction, is rejected at a 5.0% level of significance. The second null hypothesis, that lithology does not improve the prediction, is rejected at a 0.5% level of significance (r^2 of the estimated model Eq. (4) is 0.35).

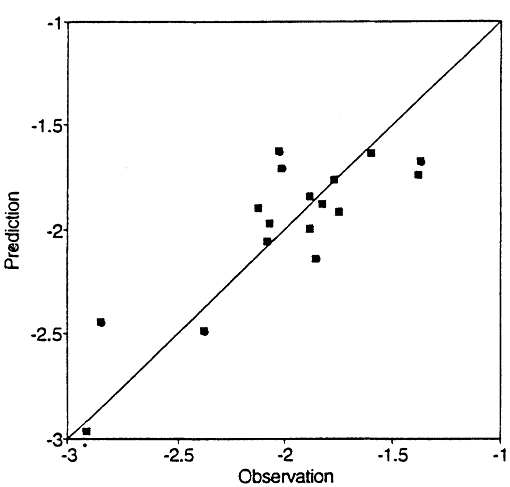

Fig. 4.
Y observations and predictions from the limnic aquifer.

Summary – The lithological data recorded in the DGU-database is a better predictor of Y for the limnic aquifer than for the former fluvioglacial aquifers (A) and (B). The prediction model is very sensitive to whether one distinguishes between layers of fine sand and the layers of more coarse deposits. If no distinction is made the model explains practically none of the sample variance, whereas the proportion of variance explained is approximately 0.5 if the distinction is made. In the latter case the standard error of prediction is made. In the latter case the standard error of prediction is approximately 0.3. This result applies even if the cumulative thickness of medium-coarse sand and gravel is applied as a single predictor of Y.

Christensen (1995) shows that the proportion of variance that can be explained by a linear model for the prediction of Y from specific capacity is 0.35 and the standard error of prediction is 0.35. The prediction made by the present model from the lithological data is thus considerably better than that from applying specific capacity as a predictor.

However, the prediction is even better if both lithology and specific capacity are used as simultaneous predictors of Y. In this case the estimated prediction model explains as much as 70% of the observed variation in Y.

Fig. 4 shows that the predictions are equally good within the entire range of observations.

Conclusion

Hydrogeologists ought to use statistical methods to predict the hydrogeological (physical) parameters from the log of lithology, which is available for most boreholes. In the present paper two models are proposed that can be used to predict the log-transmissivity, Y, from the lithology. The model parameters are physical and they may be estimated by statistical methods from observations of Y and of the predictor variables.

The first model predicts Y from the integrated hydraulic conductivity of a water-bearing formation which is composed of distinct homogeneous layers. The lithology is introduced into the model as the cumulative thickness of layers of equal deposits, whereas the conductivity of the individual deposits can be estimated, for example, by non-linear least-squares estimation.

The second model is a joint model which uses lithology and specific capacity as simultaneous predictors of Y. It is derived as a linear combination of the above-mentioned model with the model described by Christensen (1995), which uses the specific capacity as the predictor of Y. Again, the parameters of the joint model can be estimated by non-linear least-squares estimation.

The proposed methods are shown to be applicable in case studies from three Danish aquifers in different formations. In these studies the lithological logs of boreholes and the specific capacities of wells are taken from the database of the

Geological Survey of Denmark, and the transmissivites are determined by pumping test analysis . These studies yield the following conclusions:

Only the layers that are preferential to groundwater flow need to be included in the model that uses the lithology as the single predictor. In the present cases this means that one- or two-layer models are adequate. The estimated hydraulic conductivities are within the expected ranges.

In some cases the lithology may be a better predictor than the specific capacity; in other cases the opposite applies. In two of the present cases the lithology is better because the specific capacity depends much more on the highly variable efficiency of the wells (two orders of magnitude, Christensen 1995) than on the transmissivity, which varies moderately (one order of magnitude). In the third case the natural variation of transmissivity is large (three orders of magnitude) and it thus has a large influence on the specific capacity. In this case the specific capacity is a better single predictor than the lithology.

If the lithology and the specific capacity are used simultaneously as predictors of Y this improves the prediction considerably. In each of the studied aquifers the proportion of variance explained by the estimated prediction models is increased by almost 0.2 if the joint model is used instead of the model that only uses lithology.

In two cases the joint model is better than the model that uses specific capacity as single predictor; in the third case the predictions are almost indentical.

Acknowledgements

Aarhus University Research Foundation supported this study by a research fellowship and the Danish Natural Science Research Council funded the field work in South Jutland. I am grateful to the following who have made their pumping tests available for this study: Water Works of Slagelse, Kemp & Lauritzen Ltd., H. Brøker Brøndboring, Bjarne Christiansen Brøndboring, Næstved Brøndboring, Verner Nielsen Brøndboring and Keld Rømer Rasmussen. I also wish to thank Conrad Aub-Robinson for improving the English text.

References

Bear, J. (1979) *Hydraulics of groundwater*, McGraw-Hill, New York, 569 pp.

Draper, N.R., and Smith, H. (1981) *Applied regression analysis,* Second edition, John Wiley & Sons, New York, 709 pp.

Christenen, S. (1994) Hydrological Model for the Tude å Catchment, *Nordic Hydrology,* Vol.25, pp. 145-166.

Christensen, S. (1995) Prediction of Log-Transmissivity, 1. Using Specific Capacity, *Nordic Hydrology, Vol. 26,* pp. 1-20.

Freeze, R.A., and Cherry, J.A. (1979) *Groundwater*, Prentice-Hall, New Jersey, 604 pp.

Wonnacott, T.H., and Wonnacott, R.J. (1977) *Introductory Statistics,* Third ed. John Wiley & Sons, New York. 650 pp.

First received: 7 March, 1994
Revised version received: 13 June, 1994
Accepted: 20 June, 1994

Address:
Department of Earth Sciences,
Ny Munkegade, build. 520,
Aarhus University,
DK 8000 Aarhus C,
Denmark