Hatsi-kD: lognormaal – onzekerheidsmarges doorlatendheid REGIS II

WILLEM JAN ZAADNOORDIJK, JAN HUMMELMAN, ARIS LOURENS EN RONALD VERNES

Doorlatendheden van hydrogeologische eenheden kennen we slechts bij benadering. De onzekerheidsmarge kan groot zijn en ligt asymmetrisch rond het gemiddelde en de mediaan. De verdeling is veelal bij benadering lognormaal. Aan de hand van de hydraulische parametrisatie van het hydrogeologische BRO-model REGIS II versie 2.2.2 worden de relaties tussen gemiddelde, mediaan, standaarddeviatie en onzekerheidsmarges gepresenteerd. Verder wordt een simpele benadering voor onder- en bovengrenzen gegeven.

Artikel

Inleiding

REGIS II is het landelijke hydrogeologische model op regionale schaal dat TNO – Geologische Dienst Nederland onderhoudt (Hummelman e.a., 2019b en Reindersma e.a., 2022). Het is een verfijning van het Digitaal Geologisch Model (DGM, Hummelman, 2019a). Beide modellen zijn opgenomen in de Basisregistratie Ondergrond (BRO, zie bijvoorbeeld https://www.DINOloket.nl). De in DGM onderscheiden geologische eenheden worden in REGIS II op grond van de hydraulische eigenschappen onderverdeeld in hydrogeologische eenheden. De zandige eenheden krijgen een horizontale doorlatendheid en de kleiige eenheden een verticale doorlatendheid. Complexe eenheden hebben beide, terwijl het Holoceen en de gestuwde eenheden momenteel niet hydraulisch geparametriseerd zijn. Voor de geparametriseerde eenheden bevat een download van REGIS II een raster met het gemiddelde en de standaarddeviatie van de doorlatendheid. Bij toepassing is het van belang om rekening te houden met de onzekerheid van de doorlatendheden, die door de standaarddeviatie weergegeven wordt. Complicerende factor hierbij is dat de gemiddelden en standaarddeviaties gebaseerd zijn op een lognormale verdeling.

Hydraulische parametrisatie REGIS II

De doorlatendheid in REGIS II wordt zoveel mogelijk gebaseerd op laboratoriummetingen van de doorlatendheid van ongeroerde monsters uit steekboringen. Deze meetwaarden worden aangevuld met literatuurgegevens voor onbemeten eenheden. Zo wordt voor elke voorkomende combinatie van geologische eenheid en lithoklasse een doorlatendheid bepaald. De gemeten doorlatendheden per combinatie hebben geen normale verdeling. De aantallen zijn vaak niet groot genoeg om de verdeling nauwkeurig te bepalen, maar een lognormale verdeling sluit aan bij de gemeten waarden (Buma e.a., 2021). Een lognormale verdeling is een verdeling waarbij de logaritme van de variabele normaal verdeeld is.

Met de doorlatendheden en diktes per lithoklasse in de boorbeschrijvingen wordt een totale kD- of c-waarde met standaarddeviatie bepaald voor elke hydrogeologische eenheid die voorkomt in een boring. Deze totale kD- of c-waarde wordt met de dikte van die eenheid omgerekend naar een corresponderende uniforme horizontale of verticale doorlatendheid voor die eenheid ter plaatse van de boring. Deze doorlatendheden worden met hun onzekerheid per eenheid ruimtelijk geïnterpoleerd tussen de gebruikte boringen (Hummelman e.a., 2019b). Zo wordt per gridcel van 100 m x 100 m een gemiddelde en standaarddeviatie bepaald. De standaarddeviatie geeft dus de onzekerheid weer die ontstaat door de ruimtelijke variatie van de doorlatendheid binnen lithoklassen en van de verdeling van de lithoklassen binnen een hydrogeologische eenheid, en door de opschaling. Hierbij wordt aangenomen dat deze geïnterpoleerde doorlatendheden ook een lognormale verdeling hebben (Hummelman e.a., 2019b).
Met deze aanname voor de verdeling van de doorlatendheid ligt de vertaling vast van gemiddelde en standaarddeviatie naar een betrouwbaarheidsinterval.
De exacte formules om de bandbreedte te bepalen bij een gegeven onzekerheid zijn ingewikkelder dan voor de normale verdeling waar de bandbreedte een factor maal de standaarddeviatie rond het gemiddelde is.
De standaarddeviaties van hydraulische parameters zijn over het algemeen vrij groot door de heterogeniteit van de ondergrond, soms zelfs groter dan de gemiddelden.

De gemodelleerde doorlatendheden in REGIS II v2.2.2 variëren van 1×10^{-6} m/d tot 246 m/d (Zaadnoordijk, 2023). De onzekerheid hierbij varieert sterk. Het totale bereik van de variatiecoëfficiënt (de standaarddeviatie gedeeld door het gemiddelde) loopt van 0 tot 29091. De extreem grote waarden komen echter weinig voor. 75% van de hydrogeologische eenheden heeft voor alle rastercellen een variatiecoëfficiënt kleiner dan 1, 45% kleiner dan 0.75. Van alle rastercellen van alle eenheden tezamen heeft 98% een variatiecoëfficiënt kleiner dan 1 en 45% kleiner dan 0.5.

Hatsi-kD – Benadering onzekerheidsmarge

Als Hatsi-kD willen we graag een praktische benadering. Die presenteren we in deze paragraaf en in de volgende kijken we hoe goed die is.
Bij een lognormale verdeling is de logaritme van de doorlatendheid normaal verdeeld. Dit betekent dat de symmetrische marge van het betrouwbaarheidsinterval rond het gemiddelde van de normale verdeling (waarbij dat gemiddelde gelijk is aan de mediaan en de modus) getransformeerd wordt naar een gelijke factor ten opzichte van de mediaan voor de doorlatendheid (die nu niet meer gelijk is aan het gemiddelde en niet aan de modus). Een voorbeeld: als de mediaan de helft van de bovengrens van het 95%-betrouwbaarheidsinterval is, is de ondergrens ervan gelijk aan de helft van de mediaan. Het 95%-betrouwbaarheidsinterval ligt dus asymmetrisch om de mediaan. We willen toch graag uitgaan van het gemiddelde omdat dat wel opgenomen is in REGIS II en de mediaan niet.

We kiezen een benadering met de gemiddelde k-waarde m voor de middenwaarde en het gemiddelde plus twee keer de standaarddeviatie s voor de bovenwaarde. Geven we de bovengrens weer met:

$$k_b = m + 2s$$

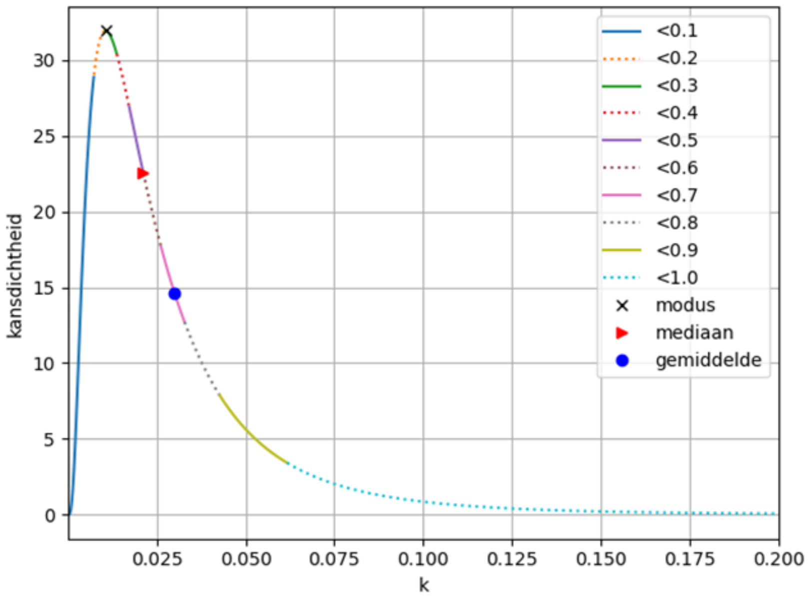
dan krijgen we voor de ondergrens:

$$\mathrm{k_o} \approx \frac{m}{\left(\frac{k_b}{m}\right)} = \frac{\mathrm{m}^2}{\mathrm{m} + 2\mathrm{s}}$$

De waarde van deze benadering bekijken we in de volgende paragrafen met de formules voor de lognormale verdeling.

Lognormale verdeling

Een lognormale verdeling bevat alleen positieve waarden en is scheef, waardoor de meest waarschijnlijke waarde (modus) kleiner is dan de mediaan die op zijn beurt kleiner is dan het gemiddelde (Afbeelding 1).



Afbeelding 1 Kansdichtheid voor een lognormale verdeling met een gemiddelde en standaarddeviatie van 0.03 met intervallen 10% en modus, mediaan en gemiddelde

Om de bandbreedte voor een bepaalde betrouwbaarheid te berekenen, gaan we nader kijken naar de lognormale verdeling. Hierbij gebruiken we de symbolen m en s voor het gemiddelde en de standaarddeviatie van een lognormale doorlatendheid k.

De logaritme hiervan $Y = \log k$ (met "log" de natuurlijke logaritme) heeft dan een normale verdeling waarvan we het gemiddelde aanduiden met μ en de standaarddeviatie met σ. Deze kunnen we berekenen met de formules:

$$\mu = \log\left(\frac{m}{\sqrt{1+\frac{s^2}{m^2}}}\right) = \log\left(\frac{m^2}{\sqrt{m^2+s^2}}\right)$$

$$\sigma = \sqrt{\log\left(1+\frac{s^2}{m^2}\right)}$$

De modus T is de waarde bij de top van de kansverdeling. Deze kan berekend worden uit het gemiddelde m en de standaarddeviatie s van REGIS II:

$$T = e^{\mu-\sigma^2} = e^{\frac{m^4}{(m^2+s^2)^{3/2}}}$$

De mediaan M van de verdeling van k is gelijk aan de $\exp$-functie (inverse van de "log" -functie) van de mediaan van Y (die ook gelijk is aan het gemiddelde en de modus van Y): $M = e^{\mu}$. De mediaan van k kan ook berekend worden uit het gemiddelde m en de standaarddeviatie s:

$$M = \frac{m}{\sqrt{1+\frac{s^2}{m^2}}} = \frac{m^2}{\sqrt{m^2+s^2}}$$

Percentielen k_p en $k_{1\text{-}p}$ zijn factorsymmetrisch ten opzichte van de mediaan M: ze voldoen exact aan relatie $k_p / M = M / k_{1\text{-}p}$ ofwel $k_{1\text{-}p} = M^2 / k_p$.

Onzekerheidsmarges bij een lognormale verdeling

De functies voor de lognormale verdeling zijn beschikbaar in verschillende reken-omgevingen, bijvoorbeeld Excel en het Pythonpakket SciPy (Zaadnoordijk, 2023). Met deze functies kunnen de betrouwbaarheidsmarges en percentielen voor een lognormale verdeling berekend worden op basis van het gemiddelde m en de standaarddeviatie s. Zo geeft Tabel 1 voor verschillende percentielen het verschil met het gemiddelde relatief ten opzichte van de standaarddeviatie ((p-m)/s).

Tabel 1 Overzicht van standaarddeviatiefactoren voor bovengrens-percentielen voor lognormale verdelingen met een gemiddelde m en standaarddeviatie s

m	s	p0.90	(p-m)/s	p0.95	(p-m)/s	p0.99	(p-m)/s	k_b=m+2s
0.020	0.002	0.0226	1.3	0.0234	1.7	0.0251	2.5	0.024
0.020	0.005	0.0266	1.3	0.0291	1.8	0.0344	2.9	0.030
0.020	0.010	0.0328	1.3	0.0389	1.9	0.0537	3.4	0.040
0.020	0.020	0.0411	1.1	0.0556	1.8	0.0981	3.9	0.060
40.000	5.000	46.5581	1.3	48.7126	1.7	53.0266	2.6	50.000
40.000	10.000	53.2028	1.3	58.1813	1.8	68.8109	2.9	60.000
40.000	20.000	65.5418	1.3	77.8127	1.9	107.3645	3.4	80.000

Voor de voorbeelden in Tabel 1 ligt voor de voor de benadering gekozen bovengrens van $k_b = m + 2s$ in de buurt van de bovengrens van een symmetrische 90%-marge met 5% onderschrijding (=p0.05) en 5% overschrijding (=p0.95).

Discussie

Het hydrogeologische model REGIS II geeft informatie over doorlatendheden van de hydrogeologische eenheden met een inschatting van de gemiddelde waarde (m) en de standaarddeviatie (s). De standaarddeviatie is een universele maat voor de spreiding. Om deze te kunnen vertalen in een bandbreedte of een betrouwbaarheidsmarge, moet ook de kansverdeling bekend zijn. De aanname in REGIS II is dat de doorlatendheden een lognormale verdeling hebben. Deze verdeling wijkt sterk af van de veel voorkomende normale verdeling als de standaarddeviatie groot is ten opzichte van het gemiddelde. De kansverdeling is dan sterk asymmetrisch. De kansen voor over- en onderschrijding zijn gelijk bij een betrouwbaarheidsinterval begrensd door de mediaan maal een factor en de mediaan gedeeld door die factor. Hiervoor moet de mediaan M berekend worden ($M = m^2 / \sqrt{m^2 + s^2}$) en voor het kiezen van een bovengrens moet de cumulatieve kansdichtheidsverdeling gebruikt worden met de te accepteren overschrijdingskans.

Als een dergelijke precisie (nog) niet vereist is, kan een indruk van de bandbreedte verkregen worden door de bovengrens twee keer de standaarddeviatie groter dan het gemiddelde te kiezen ($m + 2s$) en de ondergrens de overeenkomstige factor kleiner dan het gemiddelde ($m^2 / \{m + 2s\}$). Tabel 2 geeft de bijbehorende percentielen van de onder- en bovengrens voor een aantal lognormale verdelingen.

Tabel 2 Overzicht van percentielen voor lognormale verdelingen met een gemiddelde m en standaarddeviatie s bij een bovengrens gelijk aan m+2s en een ondergrens gelijk aan $m^2/(m+2s)$

m	s	onder	boven	%onder	%boven	%tussen
0.020	0.002	0.017	0.024	4	97	93
0.020	0.005	0.013	0.030	6	96	90
0.020	0.010	0.010	0.040	11	96	85
0.020	0.020	0.007	0.060	18	96	78
40.000	5.000	32.000	50.000	4	97	93
40.000	10.000	26.667	60.000	6	96	90
40.000	20.000	20.000	80.000	11	96	85

De voorbeelden in Tabel 2 hebben een betrouwbaarheidsinterval dat veelal meer dan 80% van de waarden bevat. De verdeling van de resterende waarden is wel scheef: de kans op overschrijding is zo'n 3 tot 4 procent terwijl de kans op onderschrijding groter wordt dan 10% als de standaarddeviatie de helft van het gemiddelde is of meer.

Dit geeft een goed beeld voor de 45% van de cellen in REGIS II v2.2.2 die een variatiecoëfficiënt hebben die kleiner is dan 0.5. Maar voor de 2% van de cellen met een standaarddeviatie groter dan het gemiddelde kan de vuistregel beter niet gebruikt worden omdat gemiddelde plus twee keer de standaarddeviatie overeenkomt met een percentiel van (afgerond) 100%, terwijl het percentiel van ondergrens boven de 40% uitkomt. Er kan een betere ondergrens verkregen worden door in plaats van het gemiddelde een centrale waarde c te kiezen gelijk aan $m/2$ bij een standaarddeviatie van twee keer het gemiddelde of $m/3$ bij een standaarddeviatie van drie keer het gemiddelde, $m/5$ bij vijf keer het gemiddelde. De ondergrens is dan: $c^2/(m+2s)$, maar in plaats van dergelijke ingewikkelder benaderingen is het aan te bevelen om de exacte waarden te berekenen (wat eenvoudig kan in bijvoorbeeld Excel of met Python, zie Zaadnoordijk, 2023).

De gepresenteerde benadering sluit aan op de veel gebruikte aanpak van gevoeligheids-analyse, waarbij parameters gevarieerd worden door ze te vermenigvuldigen met een factor 0.5 en 2. Deze factoren komen overeen met de onder- en bovengrens van een lognormale doorlatendheid, als de standaarddeviatie de helft is van het gemiddelde. Voor een standaarddeviatie die beduidend groter is dan de helft van het gemiddelde geeft deze aanpak dus een onderschatting van de invloed van het bereik van de parameter.

De voorgaande beschouwing ging over de verdeling van de doorlatendheid per gridcel in REGIS II op basis van het gemiddelde en de standaarddeviatie voor k per gridcel. Bij REGIS II wordt momenteel geen informatie over de correlatie van de doorlatendheid tussen naburige gridcellen geleverd. Daardoor geven de percentielwaarden per cel niet de percentielen van de doorlatendheid op een andere schaal. In hoeverre de bandbreedte verandert bij verschaling hangt af van de ruimtelijke correlatie van de doorlatendheid.

Verder is het goed om te beseffen dat de doorlatendheden in REGIS II gebaseerd zijn op uniforme horizontale of verticale doorstroming van de hydrogeologische eenheden en dat de doorlatendheden voor andere stromingspatronen (bijvoorbeeld lokaal sterk variërende stroming) nog extra kunnen afwijken.

Conclusies

Doorlatendheden, zoals bijvoorbeeld toegekend aan de gridcellen van het hydrogeologische BRO-model REGIS II, hebben vaak een grote onzekerheid, die bovendien asymmetrisch verdeeld is. Een lognormale verdeling sluit hierbij aan. Uitgaande van een gemiddelde m en standaarddeviatie s kan dan de betrouwbaarheidsmarge waarbinnen 90% van de doorlatendheden liggen benaderd worden als $[m^2 / (m + 2s);\ m + 2s]$. Hierbij is de kans op overschrijding ongeveer 3 tot 4%. De kans op onderschrijding is rond 4 tot 11 %, maar kan oplopen boven de 20 % als de standaarddeviatie groter is dan het gemiddelde.

Dankwoord

De auteurs bedanken Martin Knotters voor zijn opmerkingen bij een eerdere versie van dit artikel en de daaruit voortvloeiende discussie.

Literatuur

Buma, J.T., S.A.R. Bus, R. Harting, W.J. Zaadnoordijk (2021) Karakterisering van de doorlatendheid van de ondiepe ondergrond van Noord-Nederland. TNO 2021 R10310, TNO Geologische Dienst Nederland. Beschikbaar via https://www.grondwatertools.nl: https://www.grondwatertools.nl/sites/default/files/2021-08/TNO-GDN2021_KarakteriseringDoorlatendheidOndiepeOndergrondNoord-Nederland_TNO-R10310.pdf.

Hummelman, H.J., D. Maljers, A. Menkovic, R. Reindersma, J. Stafleu en R. Vernes (2019a) Totstandkomingsrapport Digitaal Geologisch Model (DGM). TNO 2019 R11653, TNO Geologische Dienst Nederland, Utrecht. Beschikbaar via https://www.DINOloket.nl: https://www.dinoloket.nl/sites/default/files/Totstandkomingsrapport-DGM.pdf.

Hummelman, H.J., D. Maljers, A. Menkovic, R. Reindersma, R. Vernes en J. Stafleu (2019b) Totstandkomingsrapport Hydrogeologisch Model (REGIS II). TNO 2019 R11654, TNO Geologische Dienst Nederland, Utrecht. Beschikbaar via https://www.DINOloket.nl: https://www.dinoloket.nl/sites/default/files/Totstandkomingsrapport-REGIS-II.pdf.

Reindersma, R., H.J. Hummelman, W. Dabekaussen en J. Stafleu (2022) Totstandkomingsrapport Kleine Release REGIS II onzekerheden geometrie. TNO 2022 R12013, TNO Geologische Dienst Nederland, Utrecht. Beschikbaar via https://www.DINOloket.nl: https://www.dinoloket.nl/sites/default/files/2023-06/R12013 Totstandkomingsrapport Kleine Release REGIS II onzekerheden geometrie.pdf.

Zaadnoordijk, W.J. (2023) Overzicht doorlatendheden REGIS II v2.2, memo, TNO Geologische Dienst Nederland, Utrecht. Beschikbaar via: https://www.grondwatertools.nl/doorlatendheden.

Summary Confidence Intervals for the Log-normal Hydraulic Conductivities in the Dutch National Hydrogeological Model REGIS II

An approximation as well as exact formulas are presented to calculate confidence intervals of the hydraulic conductivities in REGIS II, using the average values and standard deviations that are present in this hydrogeological model of the Netherlands. The hydraulic conductivities have a log-normal distribution.

Auteurs

WILLEM JAN ZAADNOORDIJK
TNO – Geologische Dienst Nederland / TU Delft - CiTG - Watermanagement
willem_jan.zaadnoordijk@tno.nl

JAN HUMMELMAN
TNO – Geologische Dienst Nederland
jan.hummelman@tno.nl

ARIS LOURENS
TNO – Geologische Dienst Nederland
aris.lourens@tno.nl

RONALD VERNES
TNO – Geologische Dienst Nederland
ronald.vernes@tno.nl

TNO – Geologische Dienst Nederland
Princetonlaan 6
3584 CB Utrecht
support@geologischedienst.nl
088 866 4300