

# Vergelijking van het Bayesiaanse en het regressiemodel

Jan Telman, TNO Industrie & Techniek, 17 september 2008

In dit stuk zijn de resultaten vermeld van de vergelijking van twee modellen voor de karakteristieke vermoeiingslijn, gebaseerd op 5 dijkvakken. Eerst worden beide modellen beschreven.

## 1. Beschrijving van de twee modellen

### 1.1. Het Bayesiaanse model

Dit model is beschreven in de notitie “Effecten van correlatie op analyse van breuksterkte en vermoeiingssterkte” van 10 januari 2008.

De basis voor dit model is het volgende:

Voor het duo van de 2 balkjes uit boorkern i veronderstellen we:

- voor de breuksterkte (balkje 1):  
$$\log B_i = \eta + e_{1i} \quad \text{met } e_{1i} \sim N(0, \sigma_B^2) \quad [1]$$
- voor de vermoeiingssterkte (balkje 2):  
$$\log N_i = \mu_i + e_{2i} \quad \text{met } e_{2i} \sim N(0, \sigma_N^2) \quad [2]$$
- voor de curve:  
$$\mu_i = \beta (\eta - \log F_i)^\gamma \quad [3]$$
- verder is er een positieve correlatie  $\rho$  verondersteld tussen de gemeten breuksterkte van het ene balkje uit een kern en de gemeten vermoeiingssterkte van het andere balkje uit dezelfde kern

Bij de Bayesiaanse analyse worden zogenaamde prior verdelingen aangenomen voor de waarde van de modelparameters  $\beta$ ,  $\eta$ ,  $\gamma$ ,  $\rho$ ,  $\sigma_B$  en  $\sigma_N$ . Op basis van het drietal meetwaarden  $\log B$ ,  $\log F$  en  $\log N$  per kern uit een dijkvak worden de posterior verdelingen geschat van die modelparameters. De mediaan uit die verdeling is gehanteerd als beste schatting voor de modelparameters.

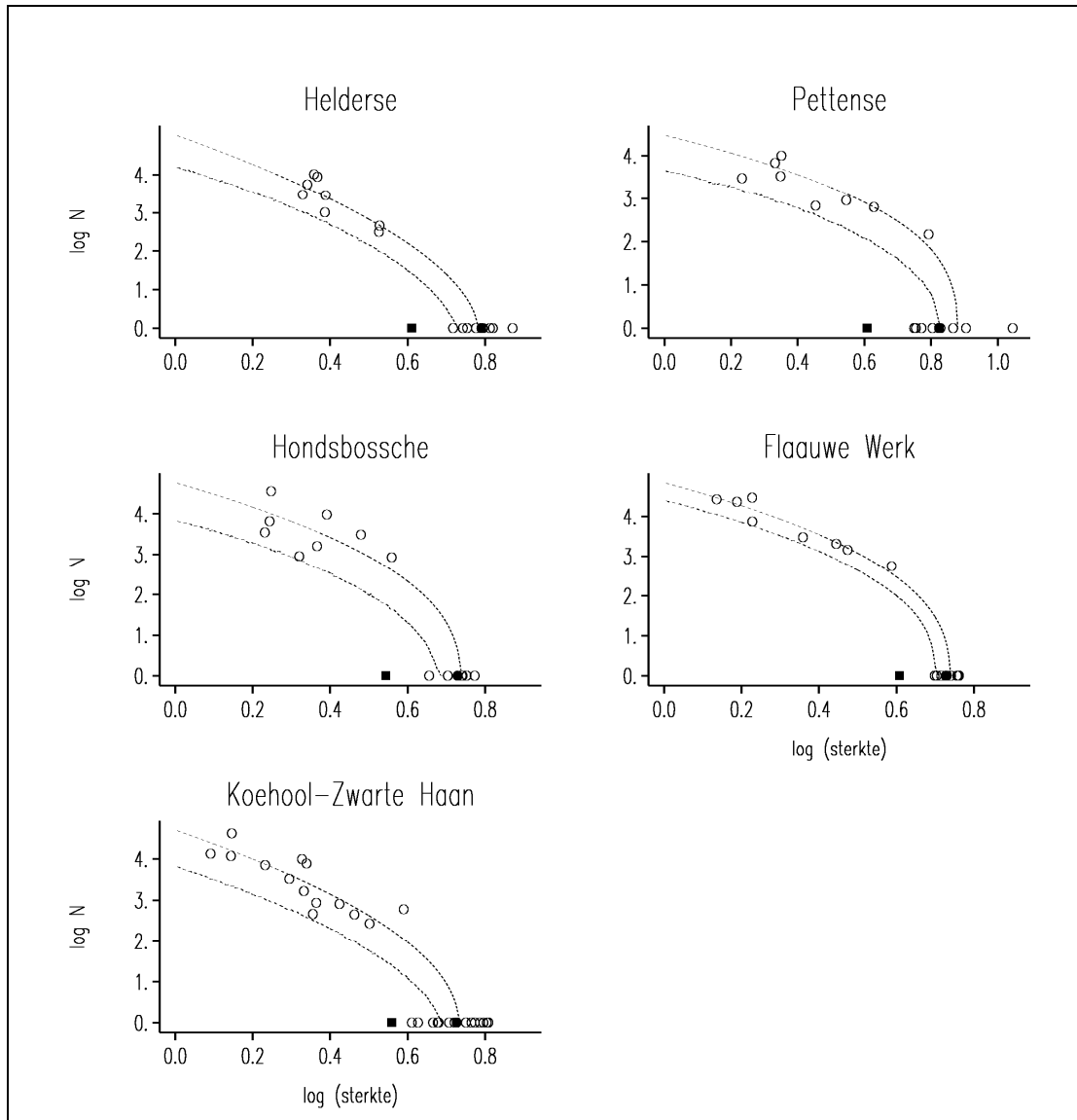
Deze waarden zijn voor de onderzochte dijkvakken vermeld in onderstaande tabel. De waarde voor correlatie  $\rho$  is verondersteld dezelfde te zijn voor alle dijkvakken.

dijkvak	beta	eta	gamma	sigmaB	sigmaN	rho
Helderse	5.783	0.781	0.562	0.090	0.351	0.716
Pettense	4.681	0.880	0.382	0.154	0.424	0.716
Hondsbossche	5.434	0.737	0.427	0.107	0.475	0.716
Flaauwe Werk	5.462	0.738	0.400	0.071	0.220	0.716
Koehool-Zwarte Haan	5.505	0.732	0.508	0.098	0.470	0.716

Tabel 1. Modelparameters uit het Bayesiaanse model

In feite geeft de relatie [3] de curve weer van  $\log N$  als functie van de vermoeiingssterke  $\log F$  voor een gemiddelde kern uit het dijkvak. Daarnaast is op basis van de geschatte spreiding rond deze curve de 95% ondergrens berekend voor  $\log N$  bij gegeven  $\log F$  voor individuele kernen. Beide curves zijn weergegeven in

onderstaande figuur. De bovenste lijn is steeds de gemiddelde lijn en de onderste lijn de 95% ondergrens voor de logN van individuele kernen bij een gegeven logF.



**Figuur 1. Verwachte waarde en 95% ondergrens voor individuele kernen volgens het Bayesiaanse model**

Met een open rondje zijn de meetpunten aangegeven. De bovenste puntenwolk geeft de gemeten combinaties van  $\log(\sigma_V)$  en  $\log(N)$  weer. De horizontale rij punten zijn gemeten waarden voor de breuksterkte  $\log(\sigma_B)$ . Het dichte rondje in die rij is een schatting voor de gemiddelde breuksterkte. Het dichte vierkantje geeft een schatting van de 95% ondergrens voor de breuksterkte van individuele kernen in het dijkvak. De waarden voor deze ondergrens zijn vermeld in onderstaande tabel.

dijkvak	breuksterkte
Helderse	0.610
Pettense	0.608
Hondsbossche	0.544
Flauwe Werk	0.608
Koehool-Zwarte Haan	0.559

**Tabel 2. 95% ondergrens voor de breuksterkte**

## 1.2. Het regressiemodel

Voor de karakteristieke vermoeiingslijn volgens het regressiemodel geldt:

$$\log(N) = \beta \times (\log(\sigma_B) - \log(\sigma_V))^\alpha \quad [4]$$

De notatie kan verwarring opleveren met de voorgaande paragraaf. Hier wordt met  $\sigma_B$  de breuksterkte bedoeld en met  $\sigma_V$  de vermoeiingsterkte van de betreffende kern.

De modelparameters  $\beta$  en  $\alpha$  kunnen worden geschat met het excelprogramma “grafiekenmaker karakteristieke vermoeiingslijn\_sjabloon.XLT”. Voor bovengenoemde 5 dijkvakken zijn deze parameters berekend. Bij de rekenprocedure moet worden aangegeven over welke range van waarden voor  $X = \log(\log(\sigma_B) - \log(\sigma_V))$  de marge voor de 95% ondergrens moet worden bepaald. Wij kozen hier voor een range tussen 95% van het minimum en 105% van het maximum voor de in de proef gemeten waarden.

De gekozen range en de resulterende modelparameters per dijkvak zijn weergegeven in onderstaande tabel. In de kolom “alfa” en “beta\_kar” staan de parameters voor de karakteristieke 95%-grens volgens formule [4]. Voor de gemiddelde lijn geldt dezelfde uitdrukking, met dezelfde waarde voor  $\alpha$  maar de waarde voor  $\beta$  zoals aangegeven in de kolom “beta\_gem”.

dijkvak	ondergrens X	bovengrens X	alfa	beta_kar	beta_gem
Helderse	-0.605	-0.251	0.627	5.010	6.093
Pettense	-0.929	-0.156	0.261	3.296	4.275
Hondsbossche	-0.717	-0.225	0.438	4.952	5.529
Flaauwe Werk	-0.715	-0.211	0.392	4.972	5.457
Koehool-Zwarte Haan	-0.761	-0.203	0.423	4.022	5.078

Tabel 3. Gekozen range en modelparameters regressiemodel per dijkvak

## 2. Vergelijking van de twee modellen

Om vergelijking van beide modellen mogelijk te maken is als eerste stap voor het Bayesiaanse model een relatie van de vorm als in formule [3] gefit. Vervolgens is een nadere vergelijking uitgevoerd.

### 2.1. Parametrisering van de karakteristieke lijn volgens het Bayesiaanse model

Voor het Bayesiaanse model was eerder nog geen relatie bepaald voor de 95% ondergrens voor individuele punten. De lijnen in Figuur 1 suggereren dat voor die ondergrens, in feite de karakteristieke lijn, een zelfde relatie geldt als die voor de gemiddelde lijn volgens formule [3].

Daarom zijn de in de Bayesiaanse analyse berekende waarden gebruikt om per dijkvak een model te fitten van de vorm:

$$\log(N) = \beta \times (\eta - \log(\sigma_v))^\alpha \quad [5]$$

Merk op dat we vanaf hier in dit model de notatie gelijk hebben gekozen als die in het regressiemodel [4].

Het model [5] bleek goed te passen bij de berekende waarden voor de 95% ondergrens uit de Bayesiaanse analyse. De geschatte parameters in formule [5] per dijkvak zijn weergegeven in onderstaande tabel.

dijkvak	beta	eta	alfa
Helderse	5.147	0.720	0.576
Pettense	4.023	0.824	0.442
Hondsbossche	4.735	0.679	0.503
Flaauwe Werk	5.126	0.705	0.421
Koehool-Zwarte Haan	4.816	0.679	0.582

Tabel 4. Parameters voor de karakteristieke lijn volgens het Bayesiaanse model

## 2.2. Nadere vergelijking van beide modellen

Uit vergelijking van de formules [4] en [5] blijkt meteen dat beide modellen niet gelijk zijn.

Het regressiemodel [4] geeft  $\log(N)$  als functie van twee inputs, namelijk de vermoeiingssterkte  $\log(\sigma_v)$  en de breuksterkte  $\log(\sigma_B)$ .

Het Bayesiaanse model [5] geeft  $\log(N)$  als functie van alleen de vermoeiingssterkte. In plaats van de breuksterkte heeft dit model een parameter  $\eta$ , die kan worden gezien als een grenswaarde voor de vermoeiingssterkte, berekend uit de verdeling van de op het betreffende dijkvak gemeten breuksterktes.

Om beide modellen te vergelijken zijn twee benaderingen gevolgd. Allereerst zijn per dijkvak beide curves grafisch vergeleken, waarbij in het regressiemodel [4] voor  $\log(\sigma_B)$  een vaste waarde is gehanteerd ter grootte van de  $\eta$  uit het Bayesiaanse model. Vervolgens is nagegaan welk percentiel die  $\eta$  representeert in de verdeling van de gemeten waarden voor  $\log(\sigma_B)$ .

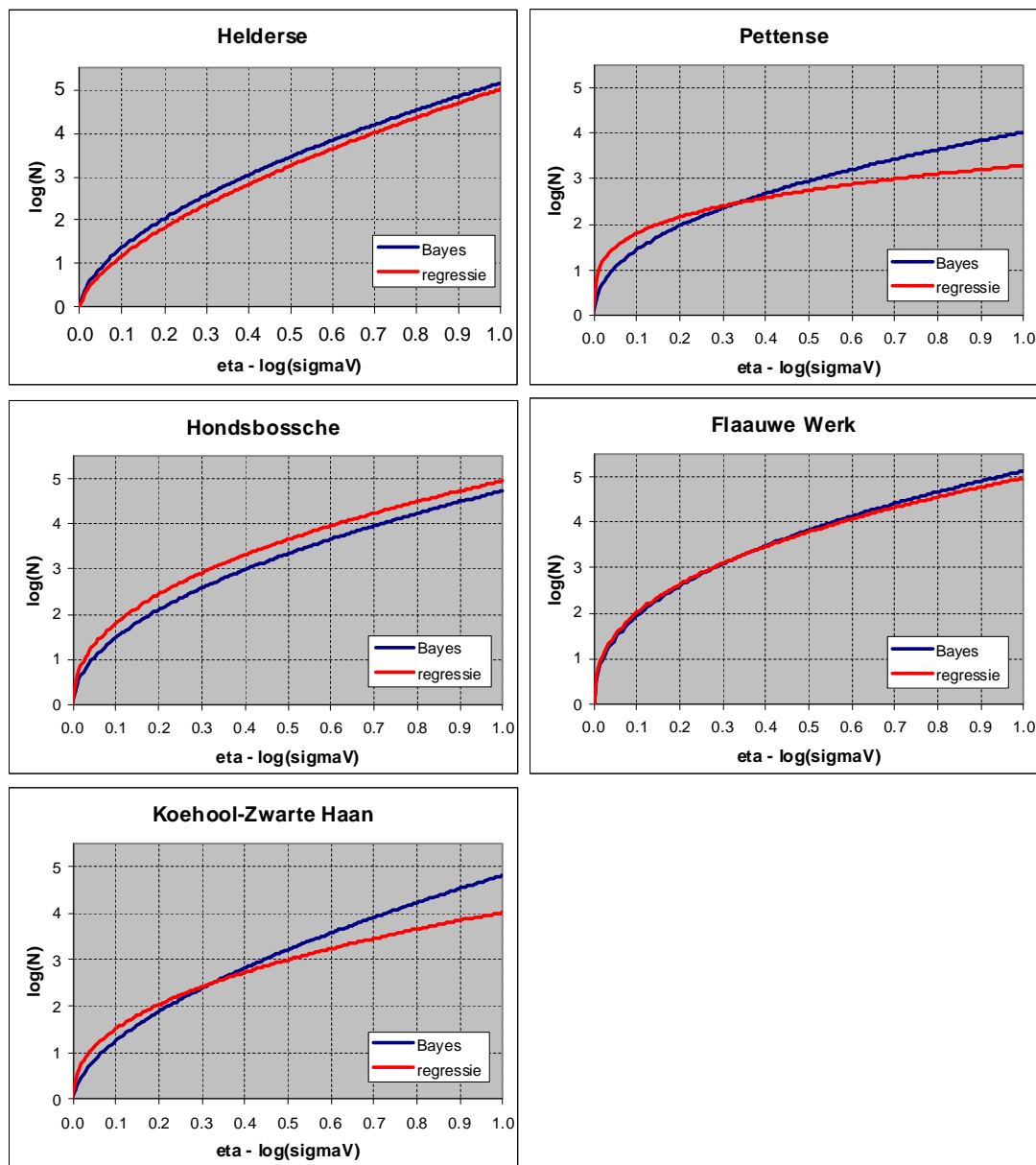
## 2.3. Grafische vergelijking karakteristieke lijn van beide modellen

De grafische vergelijking van de beide modellen is gebaseerd op de curves, waarvan de parameters zijn weergegeven in onderstaande tabel.

dijkvak	beta		eta		alfa	
	Bayes	regressie	Bayes	regressie	Bayes	regressie
Helderse	5.147	5.010	0.720	0.720	0.576	0.627
Pettense	4.023	3.296	0.824	0.824	0.442	0.261
Hondsbossche	4.735	4.952	0.679	0.679	0.503	0.438
Flaauwe Werk	5.126	4.972	0.705	0.705	0.421	0.392
Koehool-Zwarte Haan	4.816	4.022	0.679	0.679	0.582	0.423

Tabel 5. Parameters voor de karakteristieke lijn volgens het Bayesiaanse en het regressiemodel

De curves zijn per dijkvak grafisch weergegeven in de volgende figuur. Om aan te sluiten bij de grafieken op enkele logschaal uit het excelprogramma “grafiekenmaker karakteristieke vermoeiingslijn\_sjabloon.XLT” is steeds  $\log(N)$  uitgezet tegen  $\eta - \log(\sigma_V)$ . Als range is gehanteerd dat  $\eta - \log(\sigma_V)$ , en voor het regressie model dus  $\log(\sigma_B) - \log(\sigma_V)$ , ligt tussen 0 en 1. Dit komt overeen met vermoeiingssterktes  $\sigma_V$  tussen 0.1 en 1 maal de breuksterkte  $\eta$  of  $\sigma_B$ .



Figuur 2. Karakteristieke lijnen per dijkvak, volgens Bayesiaans en regressiemodel

Tevens is voor beide lijnen het gemiddelde voor  $\log(N)$  berekend over de range van  $\eta - \log(\sigma_V)$  tussen 0 en 1. Deze waarden zijn, evenals het verschil regressie minus Bayes, weergegeven in onderstaande tabel.

dijkvak	Bayes	regressie	verschil
Helderse	3.262	3.076	-0.19
Pettense	2.796	2.619	-0.18
Hondsbossche	3.157	3.451	0.29
Flauwe Werk	3.601	3.564	-0.04
Koehool-Zwarte Haan	3.051	2.833	-0.22

**Tabel 6. Gemiddelde waarden van de karakteristieke curve**

Uit Figuur 2 en Tabel 6 blijkt:

- voor het Flauwe Werk leveren beide modellen vrijwel dezelfde karakteristieke lijn op
- voor het Helderse dijkvak ligt het regressiemodel overal lager dan het Bayesiaanse model, gemiddeld is  $\log(N)$  in het regressiemodel 0.19 lager (= een factor 0.65 op  $N$ ); dit is deels te verklaren uit de lagere waarde van  $\beta$  voor het regressiemodel en deels door de flauwere kromming (hogere  $\alpha$ )
- voor de Hondsbossche zeevering ligt het regressiemodel overal hoger dan het Bayesiaanse model, gemiddeld is  $\log(N)$  0.29 hoger (= een factor 1.97 op  $N$ ); dit is deels te verklaren uit de hogere waarde van  $\beta$  voor het Bayesiaanse model en deels door de scherpere kromming (lagere  $\alpha$ )
- voor het Pettense dijkvak en voor Koehool-Zwarte Haan kruisen beide lijnen; het regressiemodel heeft een scherpere kromming, wat veroorzaakt wordt door een lagere waarde van  $\alpha$ ; gemiddeld ligt  $\log(N)$  in het regressiemodel voor deze vakken resp. 0.18 en 0.22 lager (= factor 0.67 en 0.60 op  $N$ ) dan het Bayesiaanse model

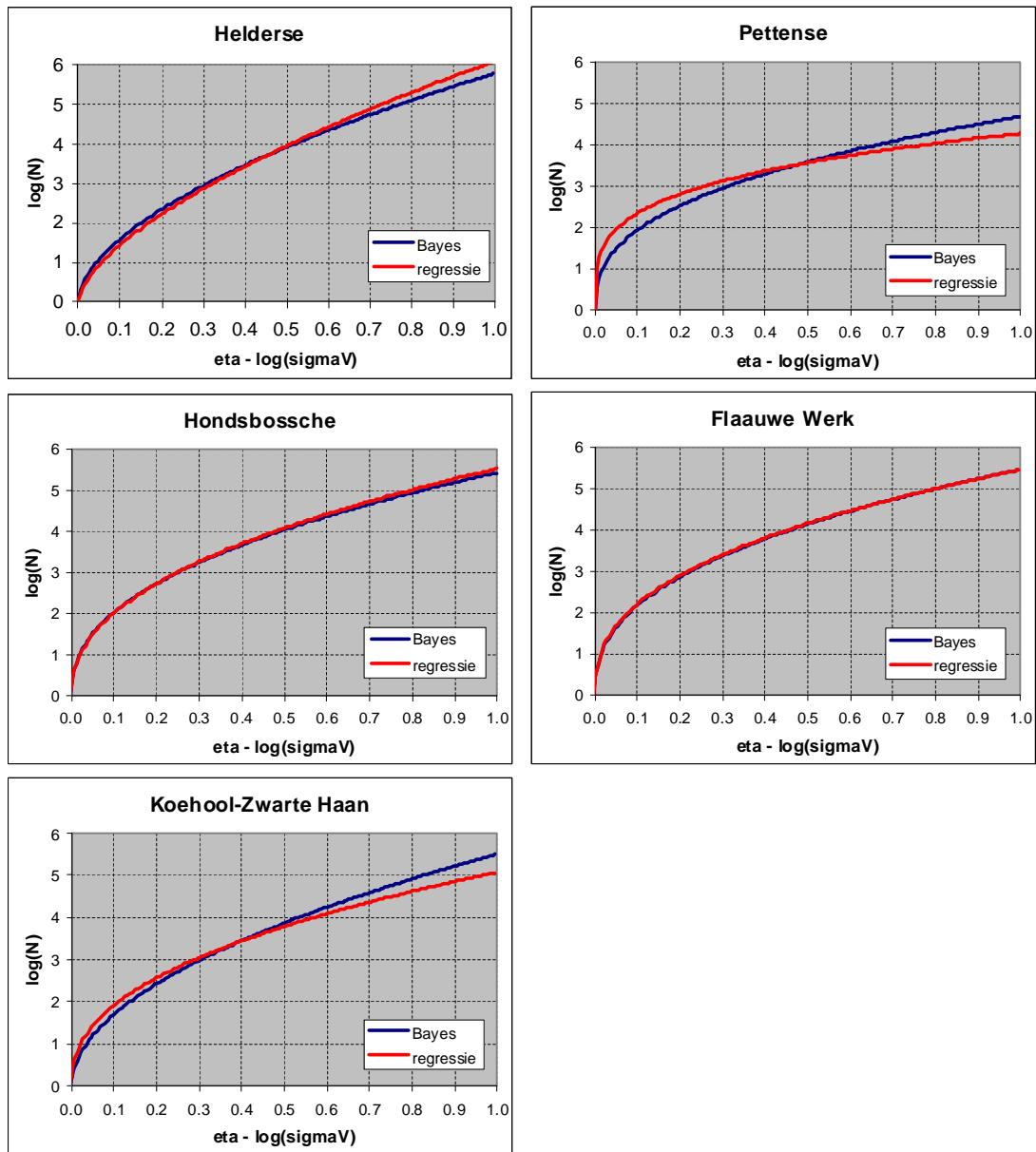
## 2.4. Grafische vergelijking gemiddelde lijn van beide modellen

De vergelijking van de beide modellen is eveneens uitgevoerd voor de curves van de gemiddelde lijn, waarvan de parameters zijn weergegeven in onderstaande tabel.

dijkvak	beta		eta		alfa	
	Bayes	regressie	Bayes	regressie	Bayes	regressie
Helderse	5.783	6.093	0.781	0.781	0.562	0.627
Pettense	4.681	4.275	0.880	0.880	0.382	0.261
Hondsbossche	5.434	5.529	0.737	0.737	0.427	0.438
Flauwe Werk	5.462	5.457	0.738	0.738	0.400	0.392
Koehool-Zwarte Haan	5.505	5.078	0.732	0.732	0.508	0.423

**Tabel 7. Parameters voor de gemiddelde lijn volgens het Bayesiaanse en het regressiemodel**

De curves zijn per dijkvak grafisch weergegeven in de volgende figuur.



**Figuur 3. Gemiddelde lijnen per dijkvak, volgens Bayesiaans en regressiemodel**

Tevens is voor beide lijnen het gemiddelde voor  $\log(N)$  berekend over de range van  $\eta - \log(\sigma_V)$  tussen 0 en 1. Deze waarden zijn, evenals het verschil regressie minus Bayes, weergegeven in onderstaande tabel.

dijkvak	Bayes	regressie	verschil
Helderse	3.696	3.736	0.04
Pettense	3.396	3.400	0.00
Hondsbossche	3.802	3.840	0.04
Flauwe Werk	3.902	3.919	0.02
Koehool-Zwarte Haan	3.648	3.566	-0.08

**Tabel 8. Gemiddelde waarden van de gemiddelde curve**

Uit Figuur 3 en Tabel 8 blijkt dat de gemiddelde curves voor de Hondsbossche zeekering en ook voor het Flauwe werk gelijk zijn voor beide modellen. Voor de overige drie dijkvakken zien we kruisende lijnen, waarbij voor het Pettense dijkvak het regressiemodel gemiddeld dezelfde  $\log(N)$  oplevert als het Bayesiaanse model.

Voor het Helderse dijkvak en Koehool-Zwarte Haan ligt het regressiemodel respectievelijk 0.04 hoger en 0.08 lager dan het Bayesiaanse model. De verschillen tussen het regressiemodel en het Bayesiaanse model zijn voor de gemiddelde curves kleiner dan voor de hierboven beschouwde karakterstieke curves.

## 2.5. Percentielwaarde voor $\log(\sigma_B)$ in relatie tot $\eta$

In bovenstaande vergelijking is in de karakteristieke lijn volgens het regressiemodel [4] voor  $\log(\sigma_B)$  een vaste waarde gehanteerd ter grootte van de  $\eta$  uit het Bayesiaanse model [5]. Deze  $\eta$  is een grenswaarde voor  $\log(\sigma_V)$ , in de Bayesiaanse analyse geschat uit de variaties in de gemeten waarden van  $\log(\sigma_B)$ .

Nagegaan is welk percentiel die  $\eta$  representeert in de verdeling van de gemeten waarden voor  $\log(\sigma_B)$ . Daarbij is uitgegaan van een normale verdeling van deze meetwaarden. Bovendien is rekening gehouden met het aantal meetwaarden. Technisch betekent dit dat de percentielwaarde van de t-verdeling is berekend.

De resultaten staan in onderstaande tabel. Daarbij is eveneens aangegeven hoeveel de  $\log(N)$  van de regressiecurve gemiddeld hoger of lager ligt dan bij de Bayesiaanse curve.

dijkvak	gemiddelde van $\log(\sigma_B)$	stdev van $\log(\sigma_B)$	n	eta	(eta-gemiddelde $\sigma_B$ ) / stdev $\sigma_B$	%punt	gemiddeld verschil $\log(N)$
Helderse	0.790	0.0706	8	0.720	-1.00	18%	-0.19
Pettense	0.824	0.1136	8	0.824	0.00	50%	-0.18
Hondsbossche	0.729	0.1128	8	0.679	-0.44	33%	0.29
Flauwe Werk	0.727	0.0553	8	0.705	-0.41	35%	-0.04
Koehool-Zwarte Haan	0.727	0.0800	14	0.679	-0.59	28%	-0.22

**Tabel 9. Parameter  $\eta$  in de karakteristieke lijn in relatie tot de verdeling van  $\log(\sigma_B)$**

Uit deze tabel blijkt dat het niet goed mogelijk is om aan te geven welke waarde voor  $\log(\sigma_B)$  gekozen moet worden om de regressiecurve te laten aansluiten op de Bayesiaanse curve. Voor het Flauwe Werk passen beide curves goed bij elkaar en zou geconcludeerd kunnen worden dat voor de parameter  $\eta$  of voor  $\log(\sigma_B)$  het 35% punt uit de verdeling van meetwaarden voor  $\log(\sigma_B)$  gekozen moet worden, overeenkomend met een waarde gelijk aan het gemiddelde minus 0.41 maal de standaardafwijking. Maar voor de Hondsbossche geeft nagenoeg dezelfde keuze een groot verschil te zien tussen de regressiecurve en de Bayesiaanse curve.