

Lotos-Euros v 1.7: validatierapport voor PM₁₀ en bias-correctie

Martijn de Ruyter de Wildt en Henk Eskes

KNMI, afdeling Chemie en Klimaat
Telefoon +31-30-2206431
e-mail mruijterd@knmi.nl

Conclusies

Bias-correctie:

- De bias-correctie voor PM₁₀ voor v1.7 heeft een sterkere seizoensafhankelijkheid dan de bias-correcties van vorige modelversies. Dit verschil wordt veroorzaakt door het gebruik van de module eqsam in plaats van isorropia. De grotere jaarlijkse gang wordt goed beschreven door de bestaande parametrisatie.
- De resultaten bevestigen dat eqsam vergelijkbare resultaten oplevert als isorropia; eqsam heeft de voorkeur omdat het parallel gedraaid kan worden. Vanwege de meer variabele bias-correctie bij eqsam lijkt het echter nuttig om te onderzoeken of de nieuwe versie van isorropia wel parallel gedraaid kan worden.
- De gemiddelde bias-correctie die moet worden toegepast is iets groter dan bij eerdere modelversies.
- De bias-correctie $y=bx$ zorgt voor een groot dynamisch bereik, waardoor overschrijdingen vrij goed voorspeld worden maar lagere waarden onderschat worden. Dit laatste zorgt voor een negatieve resterende bias en een grotere RMSE. Dit is bij v1.7 en v1.6 meer het geval dan bij v1.3.
- Een bias-correctie met de vorm $y=a+bx$ verkleint de resterende bias en de RMSE, maar zorgt voor een slechtere voorspelling van overschrijdingen.
- Een regressie op basis van $\text{meting}=f(\text{model})$ resulteert in een grotere bias-correctie en (al met al) slechtere validatieresultaten dan een regressie op basis van $\text{model}=f(\text{meting})$. De laatste is tot nu toe steeds gebruikt voor de bias-correctie.

Validatie:

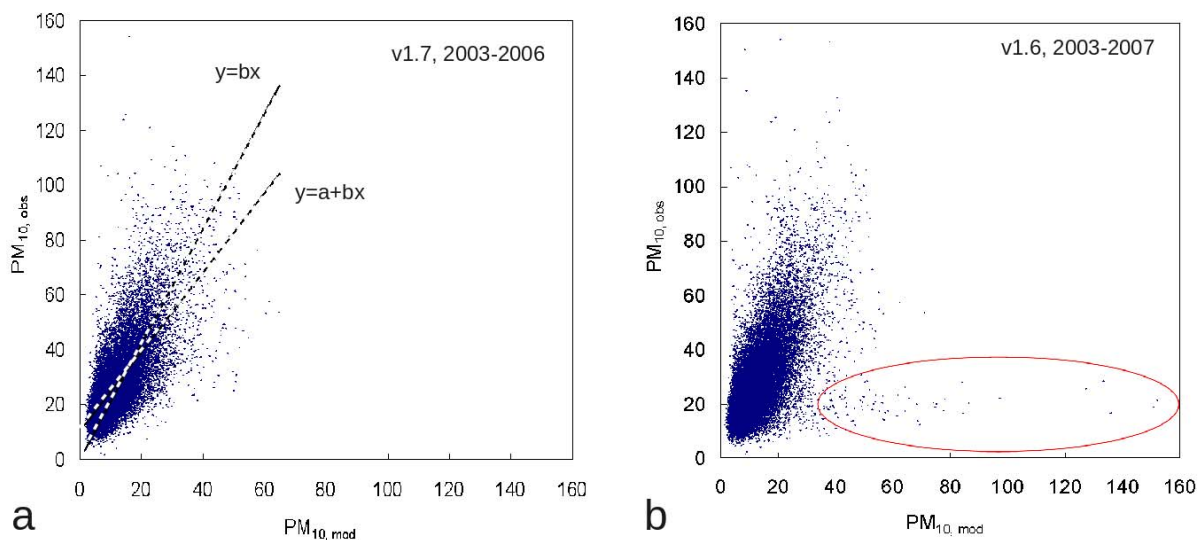
- Met bias-correctie presteert v1.7, vergeleken met v1.6, beter wat betreft de correlatie, vergelijkbaar wat betreft de RMSE en slechter wat betreft de bias. Vergeleken met PROPART zijn de correlatie en de RMSE beter, maar de bias is slechter.
- De overschrijding van twee verschillende drempelwaarden wordt door v1.7 (met de bias-correctie) beter gemodelleerd dan PROPART en de eerdere versies van Lotos-Euros dat doen. Het aantal gemodelleerde overschrijdingen komt redelijk overeen met de waarnemingen en de CSI is relatief hoog.

Zeezout:

- De veel te hoge zeezoutpieken die optraden in v1.6 tijdens harde wind boven zee zijn veel kleiner in v1.7.
- Wat in v1.7 resteert van deze pieken verdwijnt helemaal als de bias-correctie niet wordt toegepast op zeezout. Dit bevestigt de conclusie van Vincent Kamphuis dat de bias-correctie niet op zeezout-component van PM₁₀ toegepast hoeft te worden.
- Het wel of niet toepassen van de bias-correctie op zeezout levert al met al ongeveer even goede validatieresultaten op.

Inleiding

In dit document worden de PM₁₀-concentraties geanalyseerd die door v1.7 van Lotos-Euros berekend zijn. Dit gebeurt door, voor de periode 2003-2006, de berekende waarden met daggemiddelde waarnemingen van de regionale LML-stations te vergelijken. Aan de hand van de resultaten is een bias-correctie voor PM₁₀ opgesteld, omdat het model de PM₁₀-concentraties onderschat. De hiermee gecorrigeerde modelwaarden zijn gevalideerd met de LML-waarnemingen. In dit verslag worden de bias-correctie en de resultaten van de validatie gepresenteerd en vergeleken met de uitkomsten voor vorige modelversies. Belangrijke verschillen tussen de hier gebruikte run met v1.7 en de vorige validatieruns met v1.6 (De Ruyter de Wildt et al., 2010a) zijn o.a. een aangepaste beschrijving van zeezout en een hogere resolutie (4 x zoom, oftewel 0.125° x 0.0625°). Ook maakt v1.7 gebruik van de module eqsam, terwijl v1.6 aanvankelijk isorropia gebruikte. De huidige operationele v1.6 gebruikt net als v1.7 ook eqsam (n.a.v. De Ruyter de Wildt et al., 2010b).



Figuur 1. Scatterplot van gemodelleerde en gemeten daggemiddelde PM₁₀-concentraties voor v1.7 (a) en v1.6 (b). Extreem hoge modelwaarden, veroorzaakt door hoge zeezoutconcentraties, zijn omcirkeld in figuur b.

Bias-correctie

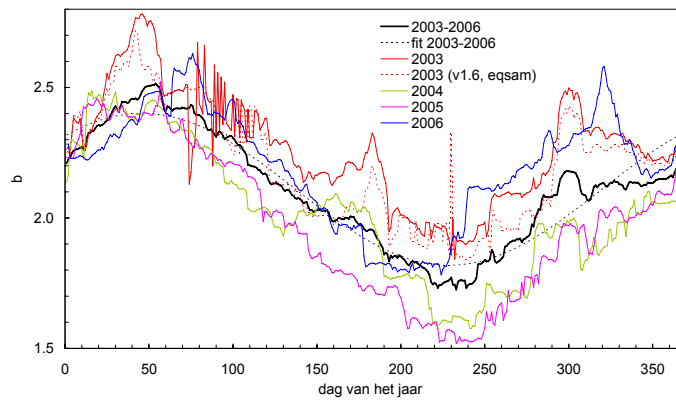
De correctie is bepaald aan de hand van het verschil tussen modelresultaten en metingen. De scatterplot in figuur 1a laat zien dat er eenzelfde systematische afwijking is als bij eerdere versies van Lotos-Euros. Wel zijn er een verschil met v1.6 (zie de scatterplot in figuur 1b): er komen geen zeer hoge waarden meer voor die door abnormaal hoge zeezoutconcentraties veroorzaakt worden. Voor de uitvoer van v1.7 voor de periode 2003-2006 is op dezelfde manier als bij de vorige versies de volgende lineaire regressie bepaald:

$$PM_{10,obs} = b PM_{10,mod} \quad (1)$$

De asafsnede van deze regressie is gelijk gesteld aan nul, zodat een grotere waarde van de correctiefactor b gevonden wordt. Hierdoor neemt de variabiliteit in de gecorrigeerde waarden toe en worden hoge PM₁₀-concentraties minder sterk onderschat. Het dynamisch bereik van het model komt zo dus beter overeen met dat van de metingen. Voor de correctiefactor b geldt weer:

model	periode	γ_0	A	dx	R ²
bias-correctie wel op NaCl					
L-E v1.7 NL	'03-'06	2,11	0,291	319,8	0,913
L-E v1.7 NL	'03	2,27	0,245	317,9	0,682
bias-correctie niet op NaCl:					
L-E v1.7 NL	'03-'06	2,15	0,321	318,3	0,911
L-E v1.7 NL	'03	2,31	0,278	315,5	0,715
L-E v1.6 EU	'03	2,17	0,311	301,4	0,880
(eqsam)					
L-E v1.6 EU	'03-'07	2,03	0,130	19,5	0,814
(isorropia)					

Tabel 1. Coëfficiënten van vergelijking 2 die voor verschillende modelversies gevonden zijn. L-E v1.7 bevat de module eqsam.



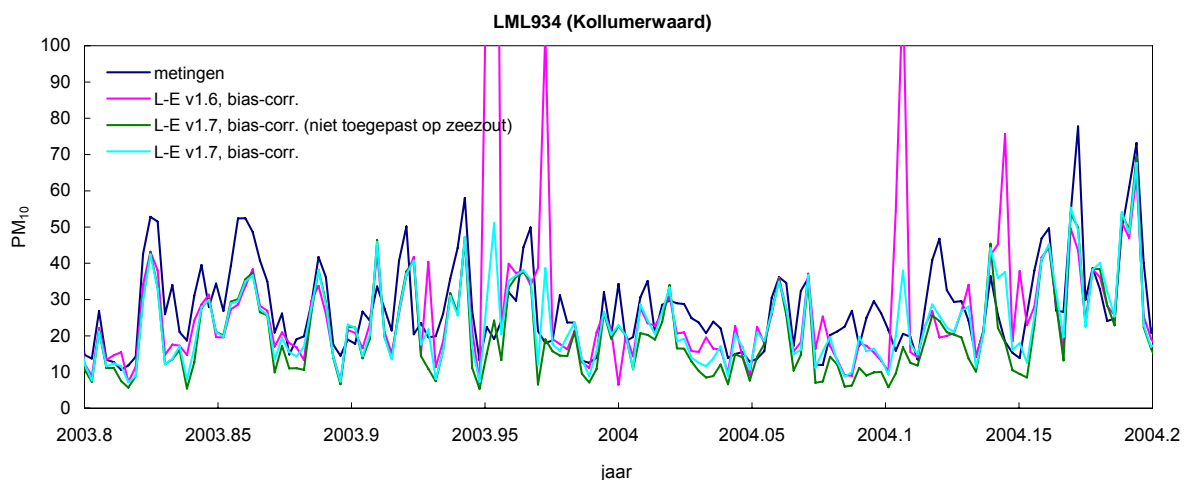
Figuur 2. De bias-correctiefactor b als functie van de dag van het jaar, voor de gehele periode 2003-2006 en voor ieder jaar afzonderlijk. Deze waarden worden gevonden als de bias-correctie ook op zeezout wordt toegepast.

$$b = \gamma_0 + A \sin(2\pi(d - dx) / 365) \quad (2)$$

Voor elke dag van het jaar kan de coëfficiënt b bepaald worden door de lineaire regressie te berekenen voor een bepaalde periode rondom die dag (hier: 60 dagen). Vervolgens kunnen de coëfficiënten γ_0 , A en dx van vergelijking 2 bepaald worden aan de hand van de dagelijkse waarden van b . Tabel 1 bevat de waarden die voor de drie coëfficiënten gevonden worden. In figuur 2 staat de correctiefactor b afgebeeld als functie van de dag van het jaar. Uit de tabel en de figuur blijkt dat de seizoensafhankelijke variatie zeer goed wordt beschreven door de sinusfunctie ($R^2=0,91$). De jaarlijkse gang van de correctiefactor b is behoorlijk groot: voor de gehele periode 2003-2006 ligt b tussen 1,73 en 2,52, met een amplitude (A) van 0,291. Verder blijkt uit de tabel dat de gemiddelde bias-correctie (γ_0) in v1.7 iets groter is geworden ten opzichte van v1.6: 2,27 i.p.v. 2,17 (voor het jaar 2003).

Eqsam vs. isorropia

De grote jaarlijkse gang in de bias-correctie wordt vooral veroorzaakt door het gebruik van de module eqsam voor het berekenen van de samenstelling van aërosolen in de troposfeer. De versies 1.3 en 1.4 gebruikten



Figuur 3. Tijdreeksen rond het einde van 2003 voor LML station 934 (Kollumerwaard).

nog de module isorropia, maar eqsam kan met OpenMP parallel gedraaid worden en isorropia niet. In juni 2010 is al onderzocht wat de invloed van de twee modules op de resultaten is (De Ruyter de Wildt et al., 2010a), waarna voor de operationele v1.6 voor eqsam gekozen is. De laatste 2 regels in tabel 1 laten zien dat, voor v1.6, het gebruik van eqsam de amplitude A van de bias-correctie meer dan verdubbeld; het verschil in periode, 2003 vs. 2003-2007, heeft veel minder invloed zoals de eerste 2 regels in de tabel laten zien. Er bleek toen verder dat eqsam niet systematisch slechter presteert dan isorropia: eqsam levert een iets slechtere RMSE maar ook een iets betere correlatie en CSI (zie ook tabel 2).

Toepassing biascorrectie op zeezout

Uit onderzoek van Vincent Kamphuis (TNO) is gebleken dat de bias-correctie niet op de zeezout-component van PM₁₀ toegepast hoeft te worden. Wanneer dit niet gedaan wordt, maar de correctie wel op alle andere PM₁₀-componenten wordt toegepast, worden iets andere waarden voor de drie coëfficiënten gevonden (tabel 1). De gemiddelde bias-correctie is - logischerwijs - iets hoger, evenals de jaarlijkse gang. De tijdreeksen voor één van de kuststations (Kollumerwaard, figuur 3) laten het effect op de berekende concentraties zien van het wel of niet toepassen van de correctie op zeezout. Versie 1.6 van L-E produceerde tijdens de getoonde periode een paar extreem hoge pieken; deze pieken komen overeen met de uitschieters in de scatterplot in figuur 1b. In de uitvoer van v1.7 zijn deze pieken niet meer aanwezig, maar op deze dagen is de concentratie nog steeds te hoog indien de bias-correctie op zeezout wordt toegepast (lichtblauwe lijn). Als dit niet gedaan wordt (groene lijn), zijn ook de resterende pieken verdwenen en komen op deze dagen de berekeningen goed overeen met de metingen. Dit bevestigt dat de bias-correctie niet op de zeezout-component van PM₁₀ toegepast hoeft te worden.

model	periode	gemid- delde ($\mu\text{g}/\text{m}^3$)	bias ($\mu\text{g}/\text{m}^3$)	corre- latie	RMSE ($\mu\text{g}/\text{m}^3$)	# data- punten	# over- schrijdingen obs. mod.	% raak (R)	% mis (M)	% loos alarm	hit rate / detectie- waarsch.	false alarm ratio	critical success index	
drempelwaarde=40 $\mu\text{g}/\text{m}^3$														
L-E v1.7 NL	'03-'06	24,0	-3,8	0,75	11,1	15133	2387	2157	8,9	6,9	5,3	0,57	0,37	0,42
idem, met correctie NaCl	'03-'06	25,3	-2,6	0,74	10,6	15133	2387	2154	8,8	7,0	5,4	0,56	0,38	0,42
PROPART	'03-'06	28,3	0,4	0,65	11,3	15133	2387	2339	8,5	7,3	7,0	0,54	0,45	0,37
L-E v1.6 EU (isor.)	'03-'06	25,5	-2,5	0,69	11,1	14987	2387	2034	8,3	7,6	5,2	0,52	0,39	0,39
L-E v1.3 NL (isor.) (Manders et al., 2009)	'04-'06	25,2	-1,3	0,70	9,6									
<i>eqsam vs. isorropia</i>														
L-E v1.6 EU (eqsam)	'03	29,6	-2,6	0,77	12,3	4116	963	856	14,7	8,7	6,1	0,63	0,29	0,50
L-E v1.6 EU (isor.)	'03	28,7	-3,6	0,76	11,9	4164	963	1027	16,5	6,7	8,2	0,71	0,33	0,53
<i>biascorrectie $y=a+bx$</i>														
L-E v1.7 NL	'03-'06	27,7	-0,2	0,75	9,2	15133	2387	1767	7,7	8,0	3,9	0,49	0,34	0,39
idem, met correctie NaCl	'03-'06	27,6	-0,2	0,74	9,4	15133	2387	1669	7,3	8,4	3,7	0,46	0,34	0,38
L-E v1.6 EU (isor.)	'03-'06	27,8	-0,2	0,69	10,1	14987	2387	1577	7,0	8,9	3,5	0,44	0,33	0,36
L-E v1.3 NL (isor.) (Manders et al., 2009)	'04-'06	26,4	-0,0	0,70	9,1									
<i>regressie model=$f(\text{meting})$</i>														
L-E v1.7 NL	'03-'06	27,6	-0,2	0,75	11,9	15133	2387	3181	11,3	4,5	9,8	0,71	0,46	0,44
L-E v1.7 NL ($y=a+bx$)	'03-'06	27,6	-0,2	0,75	11,8	15133	2387	3169	11,2	4,6	9,7	0,71	0,46	0,44
drempelwaarde=50 $\mu\text{g}/\text{m}^3$														
L-E v1.7 NL	'03-'06					15133	1114	966	3,4	4,0	3,0	0,46	0,47	0,33
idem, met correctie NaCl	'03-'06					15133	1114	920	3,3	4,1	2,8	0,45	0,46	0,32
PROPART	'03-'06					15133	1114	1123	3,3	4,0	4,1	0,45	0,55	0,29
L-E v1.6 EU (isor.)	'03-'06					14987	1114	854	3,1	4,3	2,6	0,42	0,45	0,31

Tabel 2. Evaluatie van de resultaten van de regionale LML-stations, voor Lotos-Euros en PROPART. De overschrijdingsstatistiek (rechter helft) is berekend voor de drempelwaarden 50 $\mu\text{g}/\text{m}^3$ en 40 $\mu\text{g}/\text{m}^3$. De eerste is de wettelijke grenswaarde, de tweede waarde levert meer overschrijdingen op, wat de berekende statistiek een bredere basis geeft. R: een overschrijding wordt zowel gemeten als gemodelleerd; M: een overschrijding wordt wel gemeten maar niet gemodelleerd; LA: een overschrijding wordt niet gemeten maar wel gemodelleerd; hit rate= $R/(R+M)$; false alarm ratio= $LA/(R+LA)$; critical success index= $R/(R+M+LA)$.

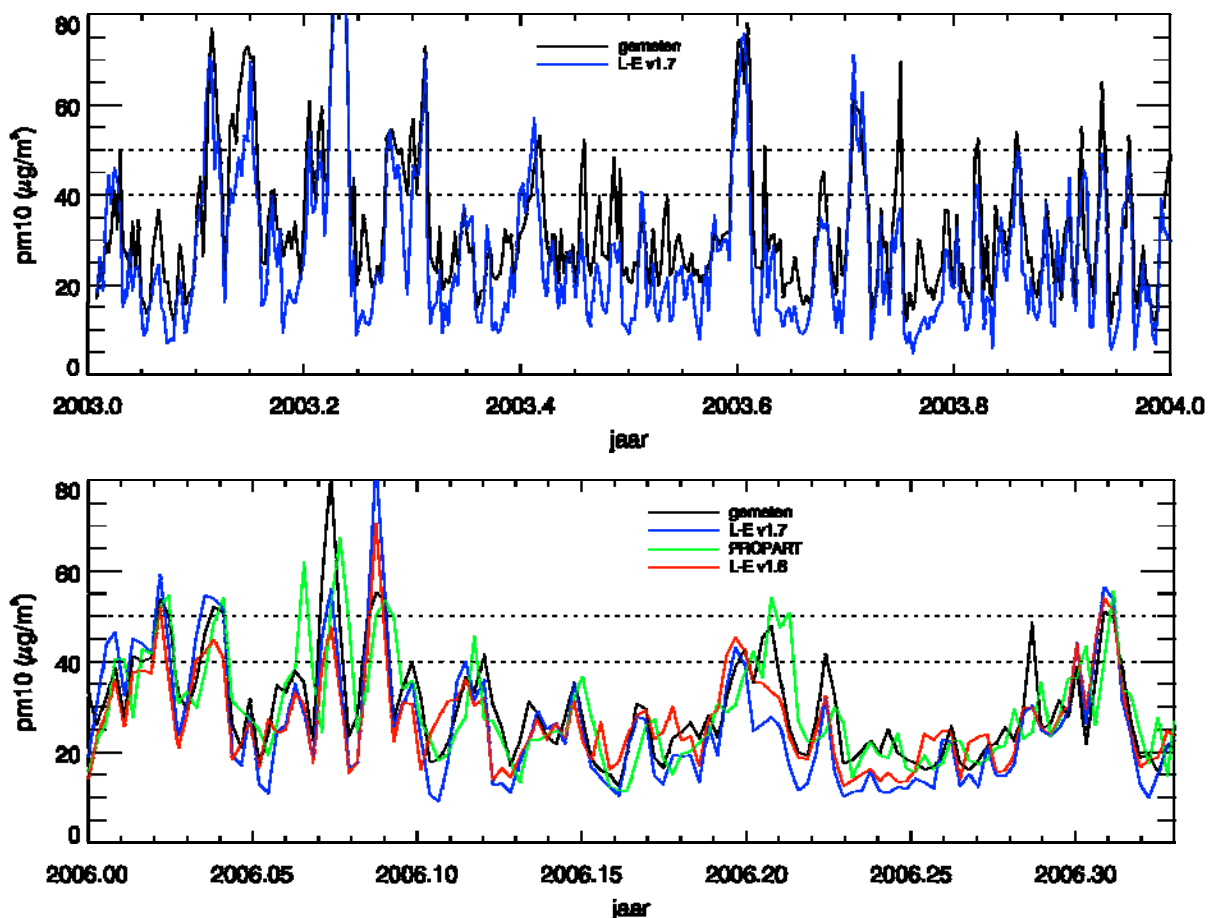
Validatie

Een validatie van de PM₁₀-waarden die door v1.7 berekend worden met metingen van alle regionale LML-stations laat het volgende zien (zie tabel 2). Uitgaande van een bias-correctie die niet op zeezout wordt toegepast, vinden we dat de correlatie hoger is t.o.v. v1.6, de RMSE hetzelfde en de resterende bias negatiever. Uit de tabel blijkt ook dat overschrijdingen door v1.7 beter voorspeld worden dan door v1.6. Het aantal gemodelleerde en gemeten overschrijdingen komen nu beter met elkaar overeen. Ook de Critical Success Index (CSI), die een algemeen beeld geeft van de overschrijdingsstatistiek, is nu hoger. Verder laat de tabel zien dat L-E vergeleken met PROPART veel beter presteert wat betreft de correlatie, vergelijkbaar wat betreft de RMSE en slechter wat betreft de bias. Vergeleken met PROPART is ook de CSI duidelijk hoger.

Als de bias-correctie ook op zeezout wordt toegepast, net als in de vorige operationele versies, zijn de RMSE en bias iets kleiner (regel 2 in tabel 2). Daar tegenover staat een lichte verbetering van de correlatie, de hit rate en de false alarm ratio. Al met al levert het wel of niet toepassen van de bias-correctie op zeezout ongeveer even goede resultaten op.

Versterking dynamisch bereik: bias-correctie $y=bx$ vs. $y=a+bx$

Uit de bovenste 5 regels in tabel 2 blijkt dat de RMSE bij v1.7 en v1.6 hoger is dan bij v1.3. De relatief hoge RMSE heeft te maken met de keuze van de bias-correctie: deze is gericht op een juiste modellering van hoge PM₁₀-concentraties, wat ten koste gaat van een iets hogere RMSE (Manders et al., 2009). Doordat de



Figuur 4. Gemeten en gemodelleerde daggemiddelde PM₁₀-concentraties, gemiddeld over alle regionale LML-stations, voor het jaar 2003 (boven) en voor de eerste vier maanden van 2006 (onder).

lineaire bias-correctie alleen een richtingscoëfficiënt (b) heeft en geen asafsnede (a), is de correctiefactor zeer klein voor lage concentraties. Er is daarom ook nog eens gekeken naar een bias-correctie met de vorm $y=a+bx$. Hiervoor zijn de volgende correcties gevonden voor v1.7 (respectievelijk zonder en met toepassing van de correctie op zeezout):

$$PM_{10,corr} = a + b PM_{10,mod} \\ = 1,14 + \{1,44 + 0,318 \sin(2\pi(d - 314,2) / 365)\} PM_{10,mod} \quad (3)$$

en

$$PM_{10,corr} = 10,4 + \{1,44 + 0,285 \sin(2\pi(d - 316,1) / 365)\} PM_{10,mod} \quad (4)$$

En, ter vergelijking, voor v1.6 (isorropia, EU, NaCl ook gecorrigeerd):

$$PM_{10,corr} = 10,6 + \{1,37 + 0,157 \sin(2\pi(d - 336,5) / 365)\} PM_{10,mod} \quad (5)$$

Wat opvalt is de lagere RMSE voor deze bias-correcties (zie tabel 2). De RMSE voor v1.7 is nu vergelijkbaar met wat eerder voor v1.3 met de biascorrectie $y=a+bx$ gevonden is (Manders et al., 2009). Ook is de resterende bias nu vrijwel nul; gemiddeld is er dus geen onderschatting meer. Aan de andere kant is het aantal overschrijdingen nu wel behoorlijk lager dan in de metingen, wat resulteert in een lagere CSI's dan bij gebruik van de bias-correctie $y=bx$. Concluderend kunnen we zeggen dat met de correctie $y=a+bx$ het dynamische bereik van de modelwaarden kleiner is dan met de correctie $y=bx$. Met de correctie $y=bx$ worden zowel meer lage als meer hoge waarden gemodelleerd; het eerste leidt tot een wat hogere RMSE en een negatieve bias, het tweede tot een meer correcte beschrijving van overschrijdingen. Bij v1.6 en v1.7 is het negatieve effect op RMSE en bias van de correctie $y=bx$ groter dan bij v1.3.

Afhankelijkheid bias-correctie van keuze x- en y-as

Tot nu toe is de bepaling van de bias-correctie voor PM₁₀ steeds gebaseerd op een lineaire regressie met de waarnemingen als functie van de modelwaarden (dus met de waarnemingen op de y-as en de modelwaarden op de x-as, zie figuur 1). Omdat hierbij impliciet aangenomen wordt dat de grootste fout in de waarnemingen aanwezig is en niet in de modelwaarden, wat het regressieresultaat beïnvloedt, is er nu ook gekeken naar de omgekeerde situatie. Als de afhankelijkheid omgedraaid wordt, dus met de modelwaarden als functie van de waarnemingen,

$$PM_{10,mod} = b PM_{10,obs} \quad (6)$$

wordt bij de bepaling van de regressie de grootste onzekerheid aan de modelwaarden toegekend. De regressiecoëfficiënt b die dan gevonden wordt is 0,405, gemiddeld over het jaar (bij $y=bx$, geen biascorrectie van NaCl). De correctiefactor waarmee de modelwaarden nu vermenigvuldigd moet worden is $1/0,405=2,47$. Deze waarde is hoger dan de eerder gevonden waarde van 2,15 (bij $meting=f(model)$); doordat nu impliciet wordt aangenomen dat de fout in de modelwaarden zit, worden deze nog sterker gecorrigeerd. Door de hogere correctiefactor is de negatieve restbias vrijwel verdwenen (zie tabel 2). Verder is de correlatie hetzelfde gebleven en de RMSE is hoger geworden. Als we naar de overschrijdingen kijken, zien we dat er nu te veel overschrijdingen voorspeld worden. Hierdoor zijn zowel de hit rate, de false alarm ratio als de CSI hoger. Al met al zijn de validatieresultaten dus eerder verslechterd dan verbeterd. Dit is ook het geval als bij $model=f(meting)$ de biascorrectie de vorm $y=a+bx$ heeft.

Referenties

- Manders, A. M. M., M. Schaap and R. Hoogerbrugge (2009). Testing the capability of the chemistry transport model LOTOS-EUROS to forecast PM₁₀ levels in the Netherlands. *Atmospheric Environ.*, 42(26), 4050-4059.
- Ruyter de Wildt, M. S. de en H. Eskes (2010a). Lotos-Euros v1.6: bias-correctie voor PM₁₀. KNMI rapport.
- Ruyter de Wildt, M. S. de en H. Eskes (2010b). Lotos-Euros v1.6: invloed van EQSAM en ozonassimilatie op de PM₁₀-concentratie. KNMI rapport.