

Beoordeling van tijdreeksmodellen, theorie? Een pragmatisch geluid!

NHV, De Bilt

27 juni 2013, Martin Knotters



ALTERRA
WAGENINGEN UR

Vaak zijn
theorie en praktijk verenigd

niets klopt
en niemand weet waarom!

ALTERRA
WAGENINGEN UR

Theorie
is
als men alles weet
en niets klopt

ALTERRA
WAGENINGEN UR

Pragmatisme

- filosofische stroming (ontstaan in VS, ca. 1870)
- focus op het verbinden van de praktijk met de theorie, die volgens het pragmatisme niet los van elkaar staan
- praktijk als toetssteen voor de geldigheid van een theorie
- "een opvatting is waar als het in de praktijk werkt"

ALTERRA
WAGENINGEN UR

Praktijk
is
als alles functioneert
en niemand weet waarom

ALTERRA
WAGENINGEN UR

Beoordeling van tijdreeksmodellen

- Hoe en waarom?
 - *Goodness-of-fit*
 - *Diagnostic checks*
 - *Fysische plausibiliteit*
 - *Validatie*
- Enkele stellingen

ALTERRA
WAGENINGEN UR

Goodness-of-fit

Past het model bij de waarnemingen?

- Percentage verklaarde variantie
- RMSE etc.

Diagnostic checking

- Box en Jenkins:

identification – estimation - **diagnostic checking**

Klopt het model in theorie?



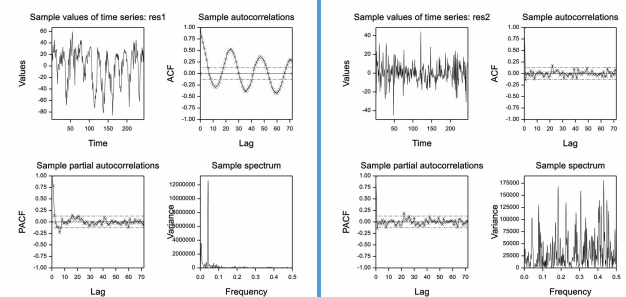
Statistics on the calibration of time series: B02G0367_1

Explained Variance Percentage (EVP): 58.28 %
Root Mean Squared Error (RMSE): 0.107 m
Root mean squared innovation (RMSI): 0.111 m
Akaike's criteria AIC and FPE: -1.543 0.013
 ...

Attention: The explained variance is below 70%.
 Your data or settings are insufficient, or other factors possibly influence the groundwater level.

Diagnostic checks

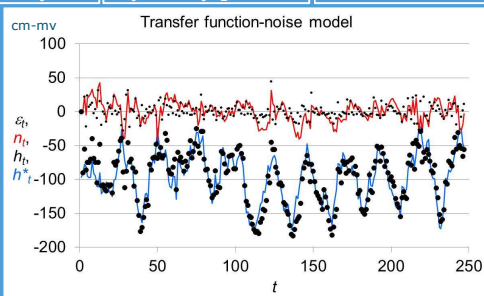
- Voldoet het model aan de veronderstellingen?



Goodness-of-fit

Percentage verklaarde variantie = 84 %, RMSE = 15 cm

$$h_t = h_t^* + n_t \quad h_t^* = \delta h_{t-1}^* + \omega p_{e,t} \quad n_t - c = \delta(n_{t-1} - c) + \epsilon_t$$



Fysische plausibiliteit

Komt het model overeen met hoe ik denk dat de natuur in elkaar zit?

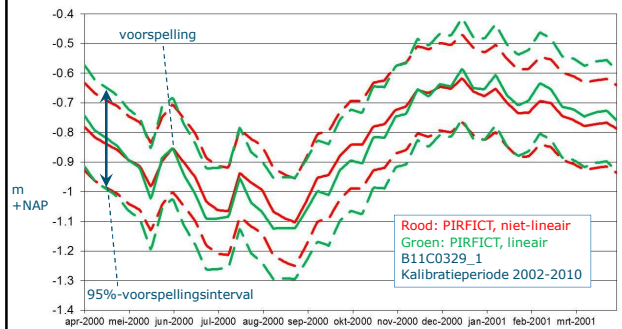
- Samenstelling set kandidaatmodellen
- Beoordeling gekalibreerde parameterwaarden
- Je kennis kan beperkend zijn: data leiden niet tot nieuwe inzichten als je toetst op fysische plausibiliteit

Validatie

The proof of the pudding is in the eating

- Is het model bruikbaar in de praktijk?

Bruikbaarheid van voorspellingen én van informatie over onzekerheid

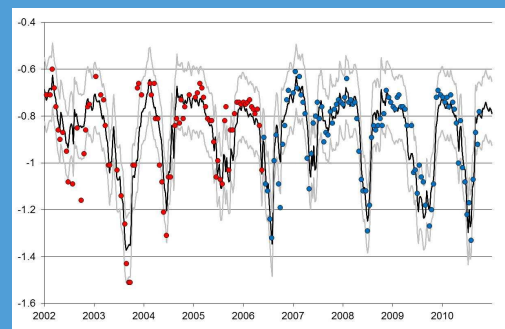


Validatie

- Bohlin (1991):
 - een testprocedure, waarvan de uitkomst aangeeft of een model voldoet aan zijn doel

Bohlin, T. (1991) *Interactive system identification: prospects and pitfalls*; Springer, Berlijn

Validatieset/Kalibratieset (reeks splitsen)



Validatie

- Chatfield (1995):
 - Validation: model checking
 - Benadrukt bestaan van 'model uncertainty': onzekerheid wordt vaak onderschat
 - Adviseert model te valideren op een *volledig nieuwe reeks*, 'true replication'
 - *Data splitting, hold-out samples*: onvermijdelijk bij validatie van tijdreeksmodellen

Chatfield, C. (1995) *Model uncertainty, data mining and statistical inference*. *Journal of the Royal Statistical Society A*; 158(3), pag 419-466

Validatiecriteria

Fouten: $e_i = h_i - \hat{h}_i, i = 1 \dots n$

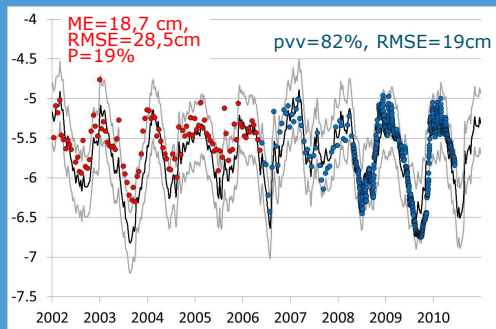
Systematische fout: $ME = \frac{1}{n} \sum_{i=1}^n e_i$

Toevallige fout: $SDE = \sqrt{\frac{\sum_{i=1}^n (e_i - ME)^2}{n-1}}$

(On)nauwkeurigheid: $RMSE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} = \sqrt{ME^2 + \frac{(n-1)}{n} SDE^2}$

P: Percentage waarnemingen buiten het 95%-voorspellingsinterval

Validatieresultaten 31A0103_1



Conclusies/stellingen

- Praktische bruikbaarheid van modellen kan niet worden beoordeeld op basis van 'goodness-of-fit', diagnostic checks en fysische plausibiliteit
- Praktische bruikbaarheid van statistische informatie (overschrijdingsrisico's etc.) kan alleen door validatie met onafhankelijke waarnemingen worden getoetst

Validatie van overschrijdingsduren

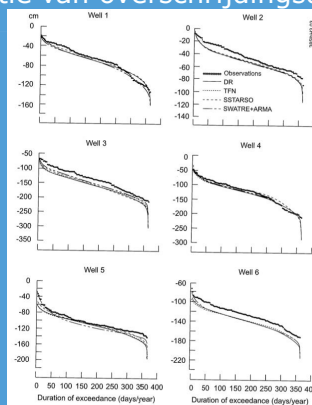


Figure 3.3: Observed and simulated durations of exceedance in the validation period.

DE BËTACANON VAN FOKKE & SUKKE



REID, GELEIJNSE & VAN TOL
VOORWOORD ROBERT DIJKRAAF

Validatie van overschrijdingsduren

RMSE in cm

Well	Model			
	DR	TFN	TARSO	Swatire+ARMA
1	5.9	6.0	4.5	5.6
2	8.1	7.8	7.3	8.7
3	20.7	19.9	18.8	13.6
4	8.0	7.4	6.9	6.2
5	11.5	11.5	7.6	9.0
6	13.5	13.1	13.4	13.0