

Trendanalyse op maat voor een meetnet waterkwaliteit

NHV - dinsdag 6 maart 2012

drs. Paul K. Baggelaar
Icastat

ir. Eit C.J. van der Meulen
AMO

Hoofddoelstellingen milieumeetnetten

Beschrijven en beoordelen van:

1. de toestand
objectiveren met behulp van normen
2. de verandering van de toestand
objectiveren met behulp van statistische methoden
vergt veel inspanning bij grootschalig meetnet (duizenden reeksen)

Trend

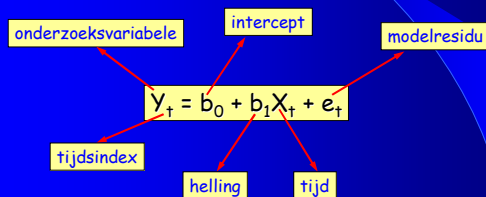
- Vorm van niet-stationariteit
- Verandering in het *centrum* van de kansverdeling van meetwaarden over tenminste enkele jaren
- We richten ons op de *monotone* trend

Onderdelen van trendanalyse

1. Trend*detectie*: objectieve uitspraak over wél of géén trend
2. Trend*kwantificering*: schatting van de grootte van de trend

Toetsen op monotone trend

Bekendste: toets op lineaire regressiehelling



Trendtoetsen met lineaire regressie

Toetshypothesen:

$$H_0 : \beta_1 = 0 \text{ en } H_a : \beta_1 \neq 0$$

Toetsingsgroottheid:

$$T = \frac{|b_1|}{s[b_1]}$$

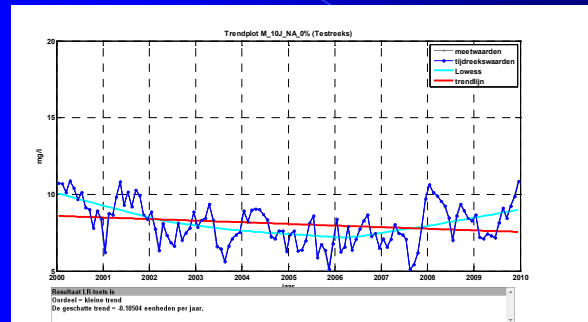
Student-t-waarde

Verwerp H_0 als $T > t_{(0,975; r-2)}$

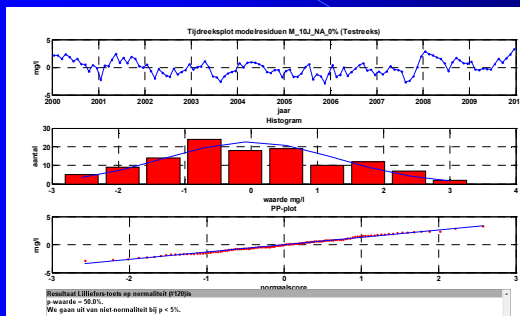
Voorwaarden trendtoetsen met lineaire regressie

1. Modelresiduën zijn afkomstig uit (dezelfde) normale kansverdeling
2. Modelresiduën vertonen geen autocorrelatie

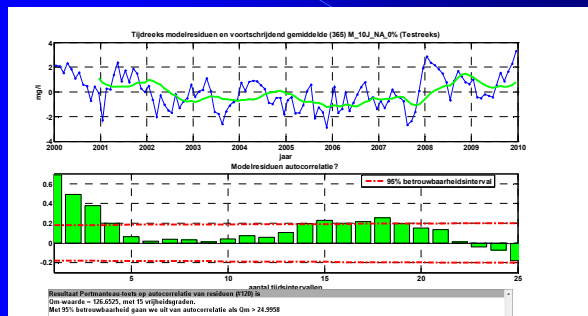
Voorbeeld lineaire regressie



Modelresiduën normaal verdeeld?



Modelresiduën geen autocorrelatie?



Uitgebreide lineaire regressie

$$Z_t = b_0 + b_1 \cdot X_t + N_t$$

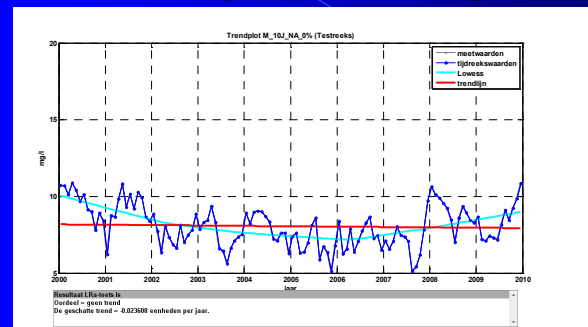
$$N_t = \phi_1 \cdot N_{t-1} + e_t$$

modelruis

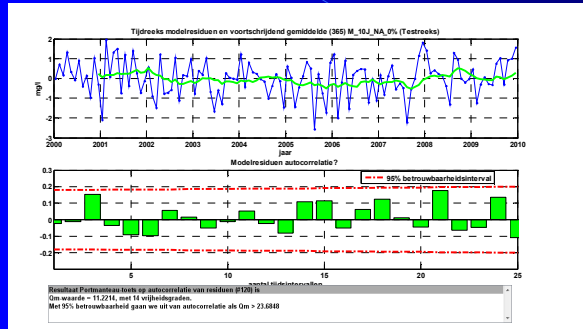
autoregressieve modelparameter

modelresidu

Voorbeeld uitgebreide lineaire regressie



Modelresiduën geen autocorrelatie?

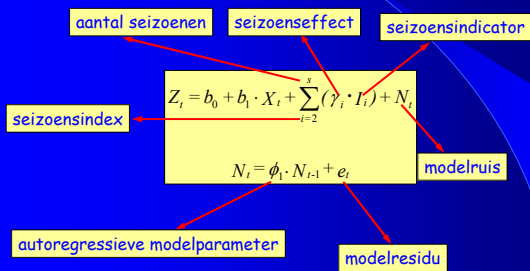


Mogelijke kenmerken milieugegevens

- Ondergrens van nul
- Gecensureerde meetwaarden (bv. $< 1 \mu\text{g/l}$)
- Uitschieters, meestal naar boven
- Scheve kansverdeling (naar rechts)
- Seizoenspatroon
- Correlatie in tijd of ruimte

Rekening mee houden bij statistische analyse

Nóg uitgebreidere lineaire regressie



Voorbeeld verdelingsvrije toets: Mann-Kendall-toets

$$S = \sum_{i=1}^{n-1} \sum_{k=i+1}^n \text{sgn}[Z_k - Z_i]$$

$$\begin{aligned} \text{sgn}[\theta] &= 1 & \text{als } \theta > 0 \\ \text{sgn}[\theta] &= 0 & \text{als } \theta = 0 \\ \text{sgn}[\theta] &= -1 & \text{als } \theta < 0 \end{aligned}$$

$$E[S] = 0 \text{ en } \text{Var}[S] = \frac{n \cdot (n-1) \cdot (2n+5)}{18}$$

$$\begin{aligned} T &= \frac{S-1}{\sqrt{\text{Var}[S]}} & \text{als } S > 0 \\ T &= 0 & \text{als } S = 0 \\ T &= \frac{S+1}{\sqrt{\text{Var}[S]}} & \text{als } S < 0 \end{aligned}$$

Statistisch significante trend als:
 $|T| > U_{(1-\alpha/2)}$ bij tweezijdig toetsen

Uitbreidingen Mann-Kendall-toets

Toetsingsgrootheid per seizoen

$$S_g = \sum_{i=1}^{n_g-1} \sum_{k=i+1}^{n_g} \text{sgn}[Z_{kg} - Z_{ig}]$$

$$S^* = \sum_{g=1}^s S_g$$

$$E[S^*] = \sum_{g=1}^s E[S_g] = 0$$

$$\text{Var}[S^*] = \sum_{g=1}^s \text{Var}[S_g] + \sum_{g=1}^s \sum_{h \neq g} \text{Cov}[S_g, S_h]$$

Toetsen op monotone trend

Parametrisch

Lineaire regressie | +s | +a | +sa |

Verdelingsvrij

Mann-Kendall | +s | +sa |

Spearman | +s |

Lettenmaier | +a | +sa |

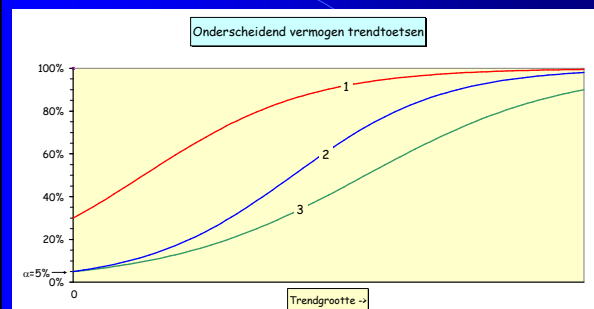
Farrell | +s |

Wanneer welke toets gebruiken?

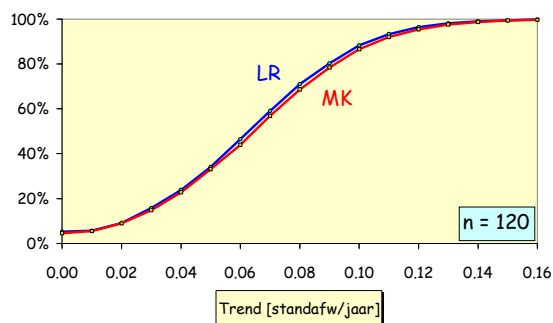
Criteria bij keuze trendtoets

1. Empirisch significantieniveau niet hoger dan gehanteerd significantieniveau (α)
verleent objectiviteit en zeggingskracht aan signaleringsfunctie van het meetnet
2. Hoogste onderscheidend vermogen ($1-\beta$) van alle trendtoetsen die aan 1. voldoen
er wordt dan zo efficiënt mogelijk informatie gefilterd uit de duur betaalde meetgegevens

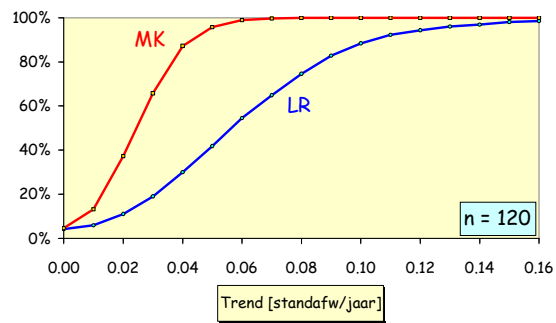
Voorbeeld keuze trendtoets



Percentage trenddetectie bij normale kansverdeling



Percentage trenddetectie bij lognormale kansverdeling



Kenmerken verdelingsvrij toetsen

1. Doet bij normale kansverdeling niet veel onder voor parametrisch toetsen
2. Is bij niet-normale kansverdeling krachtiger dan parametrisch toetsen
3. Geen last van uitschieters !

En transformeren?

- Lukt zelden volledig en toepassen van een parametrische toets/schatter is dan niet optimaal
- Geeft 'kromme' trends in de meetschaal

Bij niet-normaliteit geven wij de voorkeur aan verdelingsvrije methoden

Toetsen geselecteerd voor de procedure

Parametrisch

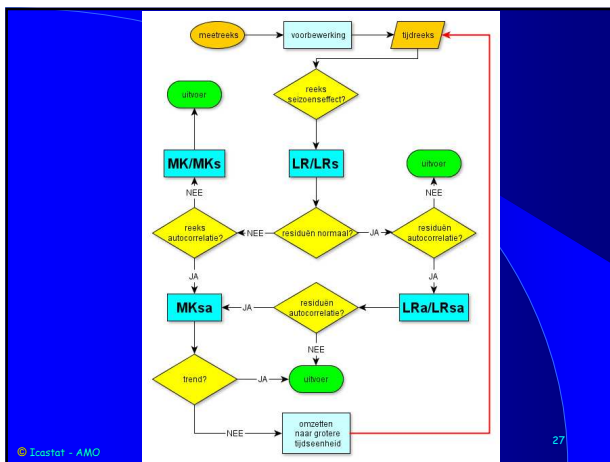
Lineaire regressie en uitbreidingen

Verdelingsvrij

Mann-Kendall en uitbreidingen

Preferentieprocessen van de toetsen

| Toets | Statistische kenmerken van proces | | | | | |
|-------|-----------------------------------|---|-----|----------------------------|---|-----|
| | Normale kansverdeling | | | Geen normale kansverdeling | | |
| | S | A | S+A | S | A | S+A |
| LR | | | | | | |
| LRs | | | | | | |
| LRa | | | | | | |
| LRsa | | | | | | |
| MK | | | | | | |
| MKs | | | | | | |
| MKsa | | | | | | |



Trendkwantificering

| Trendtoets | Trendschatter |
|--|---------------------------|
| (Uitgebreide) lineaire regressie | Lineaire regressiehelling |
| Mann-Kendall | Theilhelling |
| Seizoenale Mann-Kendall | Kendall-seizoenshelling |
| Seizoenale Mann-Kendall met verdiscontering autocorrelatie | |

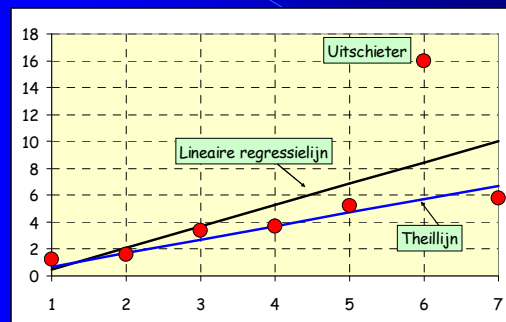
Theilhelling en Kendall-seizoenshelling zijn beide **zuivere en robuuste** schatters, met **grotere nauwkeurigheid** dan lineaire regressiehelling bij scheve kansverdelingen

Principe Theilhelling

| Tijd | Tijd | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------|------------|------|------|------|------|-------|--------|-----|
| Tijd | Meetwaarde | 1.2 | 1.6 | 3.4 | 3.7 | 5.2 | 16.0 | 5.8 |
| 1 | | | | | | | | |
| 2 | | 0.40 | | | | | | |
| 3 | | 1.10 | 1.80 | | | | | |
| 4 | | 0.83 | 1.05 | 0.30 | | | | |
| 5 | | 1.00 | 1.20 | 0.90 | 1.50 | | | |
| 6 | | 2.96 | 3.60 | 4.20 | 6.15 | 10.80 | | |
| 7 | | 0.77 | 0.84 | 0.60 | 0.70 | 0.30 | -10.20 | |

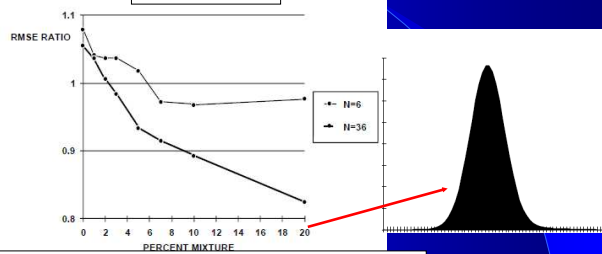
Theilhelling is mediaan van de hellingen = 1.00

Robuustheid Theilhelling



Nauwkeurigheid Theilhelling

[Hirsch et al., 1991]



RMSE Ratio = (RMSE Theilhelling) / (RMSE lin reghelling)

Toepassingen procedure

Grootschalige meetnetten waterkwaliteit

- RIWA
- RWS
- 10 waterschappen
- drinkwaterbedrijf
- Provincie
- VMM

Structureren presentatie uitvoer is uitdaging!

Conclusies

1. Maatwerk bij trendanalyse loont: meer onderscheidend vermogen bij trendtoetsen en grotere nauwkeurigheid bij trendschatten
2. Selectie toets/schatter obv soort kansverdeling en al of geen seizoenseffecten en/of autocorrelatie
3. Bij niet-normaliteit verdelingsvrije methoden
4. Selectieprocedure en trendanalyse zijn zodanig geobjectiveerd dat ze automatisch uitgevoerd kunnen worden

Vragen?

