

Vill veta kvaliteten hos våra vattenföringsdata?

Bent Göransson, Bo
Tomas Landelius, FoU

12/2/2009

SMHI

1

Bakgrund

- 2000-2001 genomförde vi en studie för att testa några exempel på noggrannhetskrav på Bo:s Q-data
- En av Bo:s huvuduppgifter är att ta fram kvalitetskontrollerade Q-data
- När Bo (som då hette So) bildades 1998 fanns 1 uttalat tillgänglighetskrav på Q-data, men inga noggrannhetskrav

12/2/2009

SMHI

2

Varför kvalitetskrav?

- För Bo:s resursplanering -

När har vi gjort ett tillräckligt bra jobb och hur visar vi det?

- För att dataanvändarna ska veta vilken kvalitet de har rätt att förvänta sig -

Vilken kvalitet kan de sikta mot i det egna arbetet?

12/2/2009

SMHI

3

Om kvalitetskrav

- Ett kvalitetskrav bör vara så tydligt att man vet om det är uppfyllt eller inte ⇒ kvalitetsmått
- Ett viktigt kvalitetsmått ⇒ noggrannhet i Q-data
- God noggrannhet i Q-data kräver god noggrannhet i avbördningsambanden

12/2/2009

SMHI

4

Projektidé

- Beräkna noggrannheten i Q-data utifrån konfidensintervallen för några avbördningssamband
- Jämföra de beräknade noggrannheterna med potentiella användarkrav

=> Test av kvalitetsmått på Q-data

12/2/2009

SMHI

5

Projektidé

- Ta fram en relevant metod för noggrannhetsberäkningar av Q-data från avbördningskurvor och testa den
- Ta fram exempel på noggrannhetskrav
- Testa metoden och exemplen på noggrannhetskrav på några olika vattenföringsstationer
- Bedöma möjligheterna att operationellt använda beräkningsmetoden och exemplen på noggrannhetskrav

12/2/2009

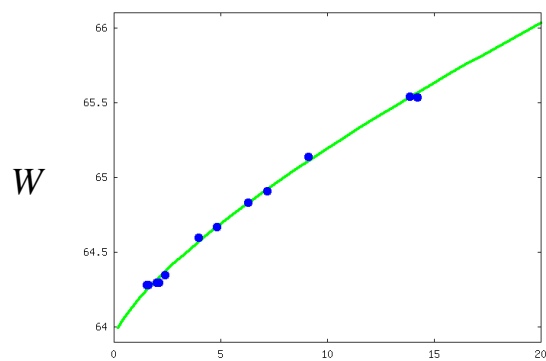
SMHI

6

Noggrannhet hos Q(W)-kurvan

2009-12-02

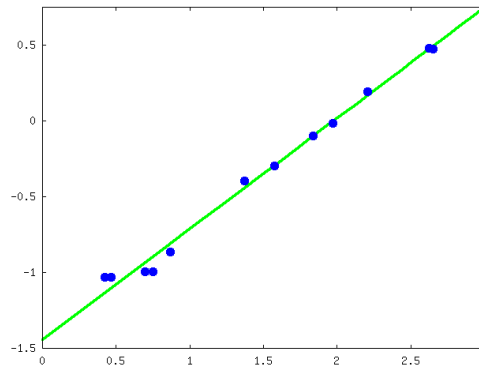
Parametrisering av sambandet



$$Q = c(W - W_0)^b + E$$

Logaritmering av sambandet

$$\begin{aligned}
 q &= \log(Q) \\
 &= \log(c(W - W_0)^b + E) \\
 &= \log(cX^b + E) \\
 &= \log(F(X) + E) \\
 &\approx \log(F(X)) + \frac{E}{F(X)} \\
 &= \log(c) + b \log(X) + e \\
 &= a + bx + e \\
 &= f(x) + e
 \end{aligned}$$

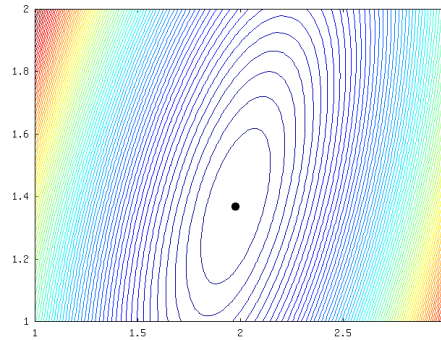


Estimering av parametrar

$$\begin{aligned}
 q &= f(\log(W - W_0)) + e \\
 &= a + b \log(W - W_0) + e, \quad e \sim N(0, \sigma)
 \end{aligned}$$

$$\begin{aligned}
 \hat{q} &= \hat{f}(\log(W - \hat{W}_0)) \\
 &= \hat{a} + \hat{b} \log(W - \hat{W}_0)
 \end{aligned}$$

Minimering av kvadratisk fel



$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (q_i - (\hat{a} + \hat{b} \log(W_i - \hat{W}_0)))^2$$

11

Least squares - Maximum likelihood

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2} (q_i - (a + b \log(W_i - W_0)))^2\right)$$

$$l = \log(L)$$

$$= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (q_i - (a + b \log(W_i - W_0)))^2$$

$$\frac{\partial l}{\partial a} = 0, \quad \frac{\partial l}{\partial b} = 0, \quad \frac{\partial l}{\partial W_0} = 0, \quad \frac{\partial l}{\partial \sigma} = 0 \quad \text{ger } \hat{a}, \hat{b}, \hat{W}_0, \hat{\sigma}$$

12

Uppskattning av fel

$$\begin{aligned}\hat{q} &= \hat{a} + \hat{b} \log(\hat{W} - \hat{W}_0) \\ &\approx \hat{a} + \hat{b} \log(W - W_0) + \hat{b} \frac{\delta W - \delta W_0}{W - W_0} \\ &= \hat{a} + \hat{b} \log(W - W_0) - \hat{b} \frac{\hat{W}_0 - W_0}{W - W_0} + \hat{b} \frac{\delta W}{W - W_0}\end{aligned}$$

$$\begin{aligned}d &= \hat{q} - q \\ &= \hat{\alpha} + \hat{\beta} \log(W - W_0) - \delta W_0 \frac{b + \delta b}{W - W_0} + (b + \delta b) \frac{\delta W}{W - W_0} + e \\ &\approx (\hat{f} - f) + b \frac{\delta W}{W - W_0} + e\end{aligned}$$

13

Beräkning av feluppskattning

Kovarianser mellan parametrarna, $p = (a \ b \ W_0 \ \sigma)$:

$$C_{ij} = \text{cov}(p_i, p_j)$$

$$C^{-1}_{ij} = -\frac{\partial^2 l}{\partial p_i \partial p_j}$$

Ex. Inget fel i W eller W_0 :

$$\begin{aligned}\text{var}(d) &= \text{var}(a) + \log(W - W_0)^2 \text{var}(b) + \\ &\quad 2 \log(W - W_0) \text{cov}(a, b) + \text{var}(e) \\ &= g(\{Q_i, W_i\}, W - W_0) \cdot \sigma^2\end{aligned}$$

14

Beräkning av feluppskattning, forts.

$$d \text{ är } N(0, \sqrt{g\sigma^2 + \frac{b^2}{(W - W_0)^2} \sigma_w^2})$$

$$P\left(\frac{|\hat{q} - q|}{\sqrt{\frac{ng\hat{\sigma}^2 + \frac{b^2}{(W - W_0)^2} \delta W^2}{n - m}}} < t_{1-\eta/2}\right) = 1 - \eta$$

m är antalet linjära beroenden, $1-\eta$ är konfidensnivån.

15

Feluppskattning – två fall

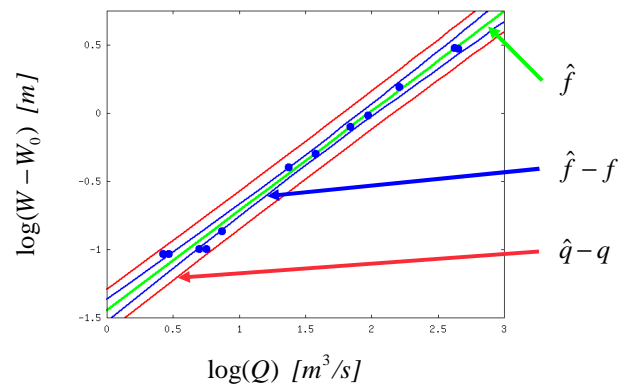
- W_0 är känd. Inget fel i W
- W_0 okänd. Fel i mätningen av W .
 δW är $N(0, \sigma_w)$ och oberoende av övriga fel:

Herschky i "Hydrometry": $\sigma_w \approx 3$ mm.

16

Feluppskattning – enkla fallet @ 95%

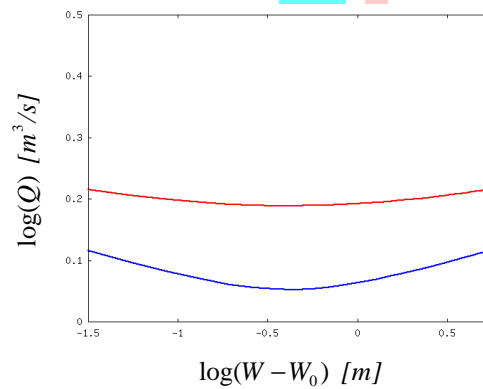
$$\hat{q} - q = (\hat{a} - a) + (\hat{b} - b) \log(W - W_0) + e$$



17

Feluppskattning – enkla fallet, forts.

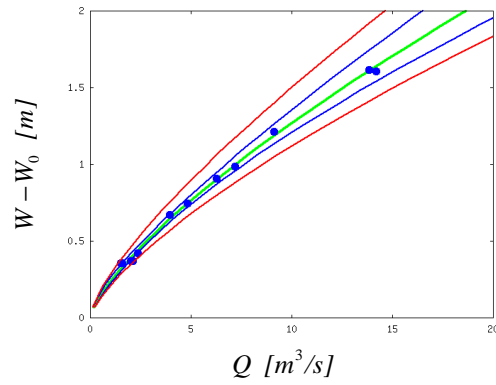
$$\hat{q} - q = \hat{f} - f + e$$



18

Feluppskattning – enkla fallet, forts.

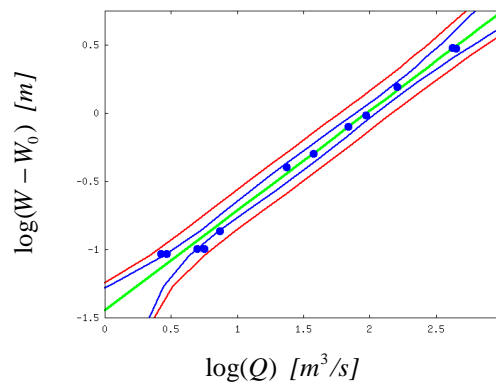
$$\hat{Q} - Q \approx (\hat{q} - q)Q$$



19

Feluppskattning – svårare fallet

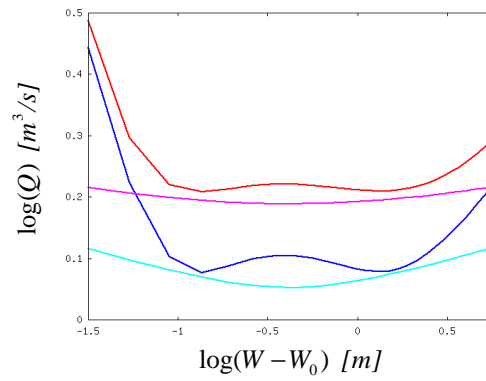
$$\hat{q} - q = (\hat{a} - a) + (\hat{b} - b) \log(W - W_0) - (\hat{W}_0 - W_0) \frac{b}{W - W_0} + b \frac{\partial W}{\partial W_0} + e$$



20

Feluppskattning – svårare fallet, forts.

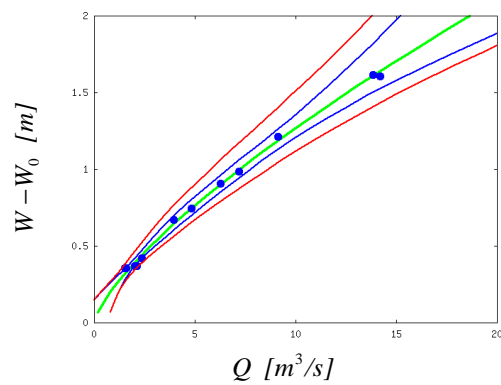
$$\hat{q} - q = \hat{f} - f + b \frac{\delta W}{W - W_0} + e$$



21

Feluppskattning – svårare fallet, forts.

$$\hat{Q} - Q \approx (\hat{q} - q) Q$$



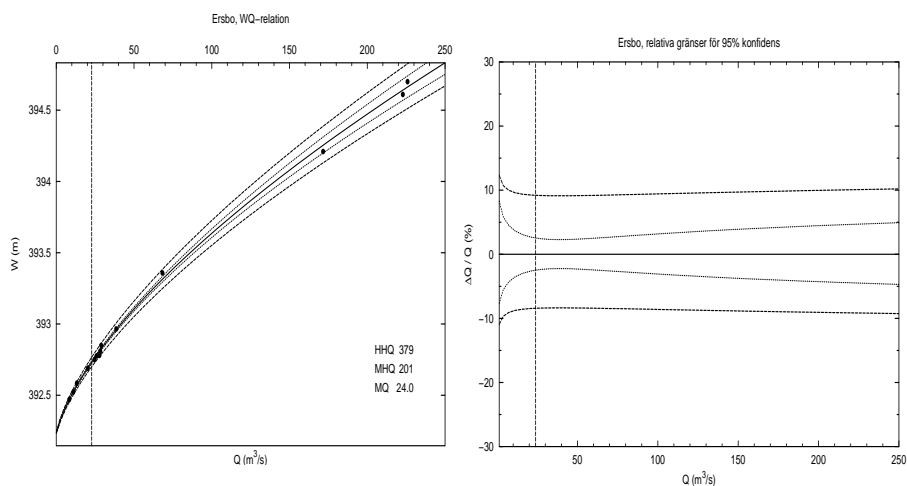
22

Funderingar

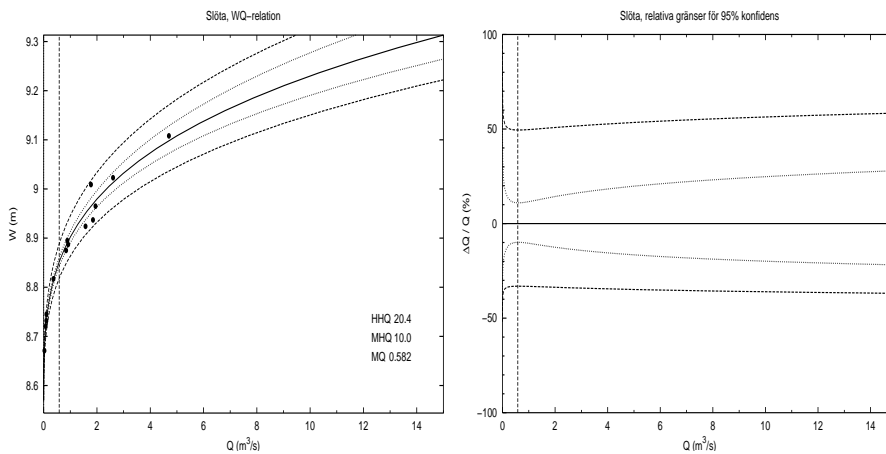
- Bättre att estimerar parametrar genom att minimera absoluta fel istället för relativa?
- Fel i W saknar betydelse i detta exempel (3 mm). Fel i W_0 spelar dock roll.
- Enklare och exaktare feluppskattning med alternativ parameterisering av $Q(W)$:

$$Q = \sum p_i f_i(W) + E$$

Konfidensintervall för avbördningskurva och Q-data



Konfidensintervall för avbördningskurva och Q-data

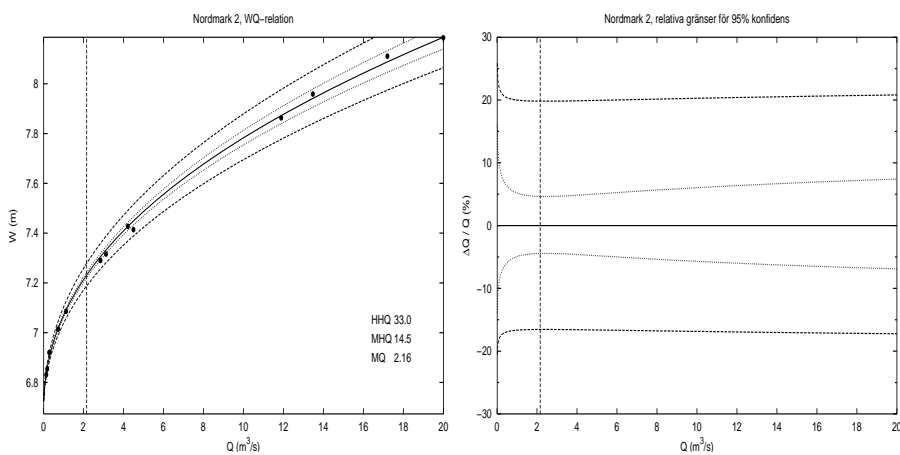


12/2/2009

SMHI

25

Konfidensintervall för avbördningskurva och Q-data



12/2/2009

SMHI

26

FoU:s exempel på noggrannhetskrav

- Medelfel i alla Q < 5 %

Ersbo	6 %
Slöta	43 %
Nordmark	12 %

- Medelfel i alla låga Q (i de 20% lägsta värdena) < 10 %

Ersbo	4 %
Slöta	25 %
Nordmark	10 %

forts.

12/2/2009

SMHI

27

FoU:s exempel på noggrannhetskrav

forts.

- Medelfel i årsmax (HQ) < 5 %

Ersbo	4 %
Slöta	20 %
Nordmark	9 %

- Fel i medelvärdet av alla Q (MQ) < 1 %

Ersbo	1,1 %
Slöta	6 %
Nordmark	2 %

12/2/2009

SMHI

28

Slutsatser om beräkningsmetoderna och noggrannhetsmåten

- Det är *synnerligen* viktigt att veta vad beräkningarna kan förmedla, liksom vad man vill ha förmedlat
- För SMHIs stationer är det *idag* svårt att kunna använda något riktigt bra noggrannhetsmått på Q med klassisk statistik (bl.a. bör manuella justeringar ibland göras på de framräknade sambanden)

forts.

12/2/2009

SMHI

29

Slutsatser om beräkningsmetoderna och noggrannhetsmåten

forts.

- Även om vi skulle räkna om våra avbördnings-samband så skulle det *idag* bli svårt att garantera en statistisk noggrannhet, som är i nivå med FoU:s exempel på krav
- I *framtiden* vore det möjligt nå FoU:s exempel på krav, men då krävs betydande resurstillskott, eller att ett begränsat antal stationer prioriteras på bekostnad av andra

12/2/2009

SMHI

30