# Statistical reasoning in diagnostic problem solving

The case of flow-rate measurements

Jeroen de Mast

1

Jeroen de Mast
Professor, University of Waterloo
Academic Director, Jheronimus Academy of Data Science (JADS)
Scientific Director, Holland Innovative

jdemast@uwaterloo.ca





2

## The case of flow-rate measurements

QUALITY ENGINEERING
2019, VOL. 31, NO. 3, 484–498

CASE REPORT

**Statistical reasoning in diagnostic problem-solving—The case of flow-rate measurements**

Jeroen de Mast[a], Stefan H. Steiner[a], Rick Kuijten[b], and Elly Funken-Van den Bliek[c]

Case study
Use of statistics and statistical reasoning in diagnostic problem solving

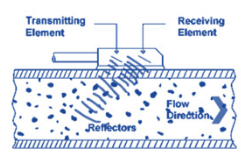
= finding the causes of problems





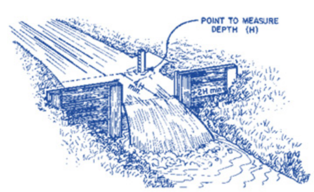
## What's the problem?

- Water boards ("Waterschappen") monitor waterways by measuring flow rates $Q$ in $m^3/s$.




Doppler-type acoustic flowmeter

The *Buulder Aa*



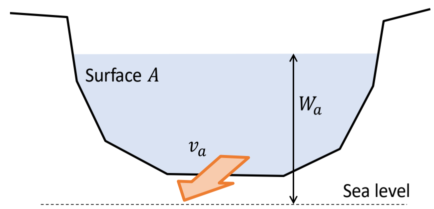

Measuring flow rate using a weir

Problem:

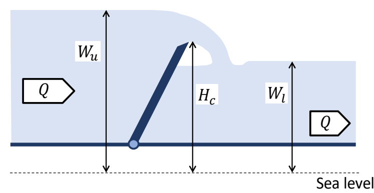
- Two measurement methods for flow rate … wildly different results!

## Measuring flow rate $Q$

- Acoustic flowmeter
  - Measures flow velocity $v_a$ by means of acoustic signals
  - Measures water level $W_a$, and calculates $A$ from there
  - $Q_a = v_a \times A$



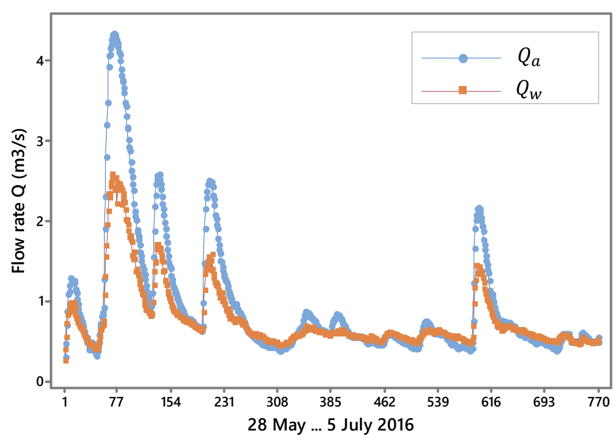

- Weir
  - Computes $Q$ from height of water over weir's crest
  - $Q_w^{\text{raw}} = 3.00 \times 1.86 \times (W_u - H_c)^{1.5}$
  - If water *behind* weir is higher than the weir: "drowned"
    Corrected flow rate:
    $Q_w = Q_w^{\text{raw}} \times (1 + \log(1 - D)\,/2)$








## Large discrepancies between $Q_a$ and $Q_w$




- Root-mean square error $RMSE = \sqrt{\frac{1}{876}\sum_t \left(Q_{w,t} - Q_{a,t}\right)^2} = 0.404.$
- What is the cause of the substantial discrepancies between $Q_a$ and $Q_w$?

## What is the cause of the discrepancies?

"Ask the experts!"

- But … hydrologists couldn't explain the discrepancies
- … and literature didn't give useful clues

"Calibrate both measurement systems!"

- But … impossible to obtain reference values

7

7

## Purpose of this case study

Diagnostic problem solving:

- Finding the causes of a problem by applying smart analytics
- Here: "What is the cause of the discrepancies?"

- Purpose:
  - demonstrate the roles that statistical thinking can play in diagnostic problem-solving, and
  - identify reasoning patterns that make the application of statistical techniques powerful.

Challenges:

- Domain knowledge insufficient to identify the cause …
  the hydrologists couldn't explain the discrepancies
- Difficult to collect more data;
  no reference values available, and impossible to do a calibration study

8

8

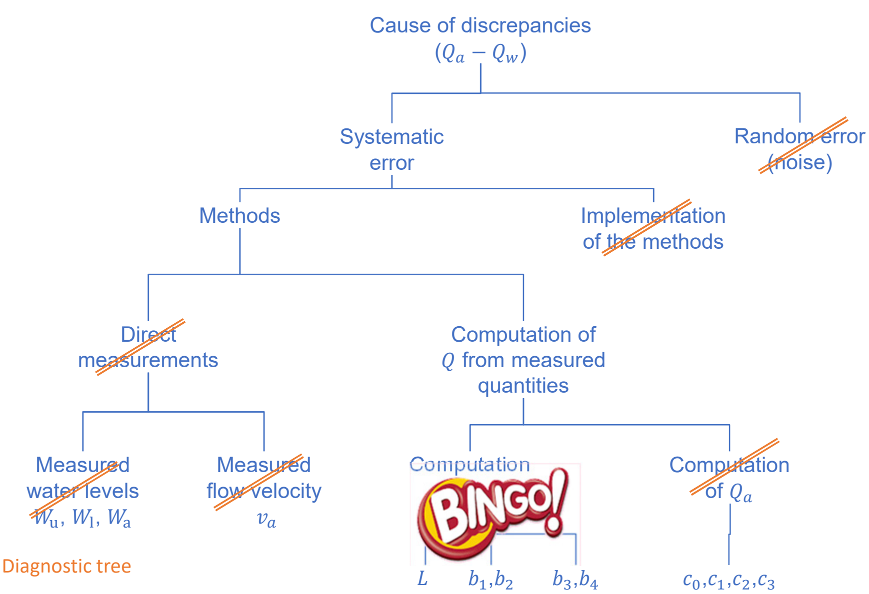
# Diagnostic problem solving

- Literature:
  - Troubleshooting of devices
  - A.I.
  - A.I. for medical diagnosis
  - Practitioners (Dorian Shainin)
- Branch-and-prune strategy
  - Analyze the available data to eliminate as many potential causes as possible
  - Branch-and-prune = hierarchical diagnosis = eliminate & zoom in

9

9

# Search strategy: “Branch and Prune”




Diagnostic tree

10

10

## Discovering the cause of discrepancies

… application of the branch-and-prune strategy

UNIVERSITY OF WATERLOO

JADS Jheronimus Academy of Data Science

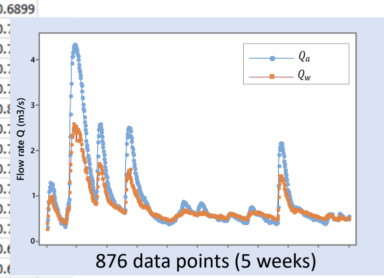
hi holland innovative

11

## Available data

- Data from *weir*: flow rate $Q_w$ and water levels $W_u$, $W_l$
- Data from *acoustic flowmeter*: flow rate $Q_a$, water level $W_a$, velocity $v_a$
- One data point every 1hr, for 876hrs (5 weeks) on a row

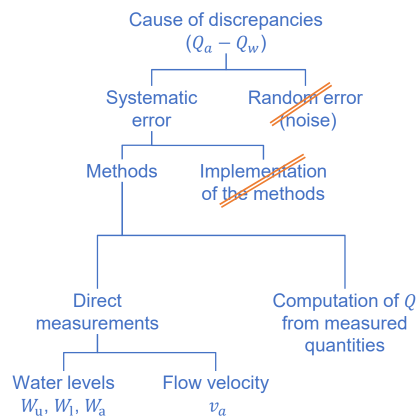
| | W_a | v_a | Q_a | W_u | W_l | H_c | Q_w | D |
|---|---|---|---|---|---|---|---|---|
| 2 | 26.0464 | 0.0784 | 0.3376 | 26.0580 | 25.6520 | 25.8920 | 0.3770 | 0.0000 |
| 3 | 26.0942 | 0.1112 | 0.5244 | 26.1010 | 25.7730 | 25.8920 | 0.5330 | 0.0000 |
| 4 | 26.1400 | 0.1484 | 0.7376 | 26.1450 | 25.8860 | 25.8930 | 0.7060 | 0.0000 |
| 5 | 26.1755 | 0.1812 | 0.9188 | 26.1810 | 25.9550 | 25.8940 | 0.8130 | 0.2125 |
| 6 | 26.1909 | 0.2100 | 1.0812 | 26.1890 | 26.0410 | 25.8590 | 0.8740 | 0.5515 |
| 7 | 26.2077 | 0.2128 | 1.1324 | 26.2020 | 26.0950 | 25.8570 | 0.8430 | 0.6899 |
| 8 | 26.2309 | 0.2228 | 1.2152 | 26.2280 | 26.1230 | 25.8550 | 0.9210 | [illegible] |
| 9 | 26.2499 | 0.2300 | 1.2800 | 26.2470 | 26.1490 | 25.8540 | 0.9600 | [illegible] |
| 10 | 26.2614 | 0.2272 | 1.2716 | 26.2590 | 26.1730 | 25.8560 | 0.9490 | [illegible] |
| 11 | 26.2647 | 0.2200 | 1.2500 | 26.2610 | 26.1720 | 25.8540 | 0.9710 | [illegible] |
| 12 | 26.2623 | 0.2200 | 1.2444 | 26.2540 | 26.1770 | 25.8570 | 0.8990 | [illegible] |
| 13 | 26.2558 | 0.2228 | 1.2328 | 26.2540 | 26.1600 | 25.8540 | 0.9680 | [illegible] |
| 14 | 26.2466 | 0.2272 | 1.2316 | 26.2400 | 26.1630 | 25.8570 | 0.8620 | [illegible] |
| 15 | 26.2341 | 0.2172 | 1.1848 | 26.2320 | 26.1390 | 25.8540 | 0.9020 | [illegible] |
| 16 | 26.2195 | 0.2100 | 1.1256 | 26.2140 | 26.1330 | 25.8570 | 0.8070 | [illegible] |
| 17 | 26.2035 | 0.2100 | 1.1288 | 26.1990 | 26.1160 | 25.8540 | 0.7810 | [illegible] |
| 18 | 26.1872 | 0.2072 | 1.0748 | 26.1810 | 26.0850 | 25.8550 | 0.7630 | [illegible] |
| 19 | 26.1705 | 0.1972 | 0.9960 | 26.1670 | 26.0580 | 25.8540 | 0.7530 | [illegible] |
| 20 | 26.1551 | 0.1872 | 0.9404 | 26.1530 | 26.0410 | 25.8540 | 0.7180 | [illegible] |
| 21 | 26.1414 | 0.1772 | 0.8732 | 26.1390 | 26.0140 | 25.8560 | 0.6910 | 0.5583 |

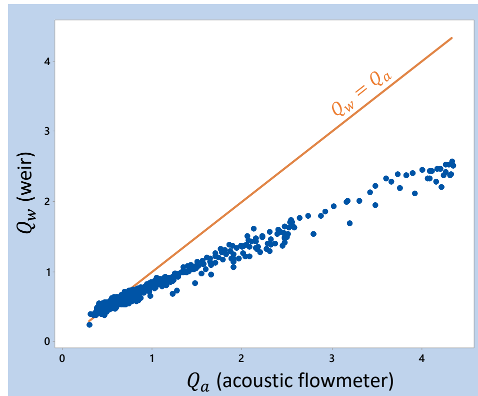





12

## *Systematic* or *random* discrepancies?




Relation between $Q_w$ and $Q_a$ has $R^2 = 98.4\%$:

- 98.4% of discrepancies are systematic.
- Only 1.6% is random noise.

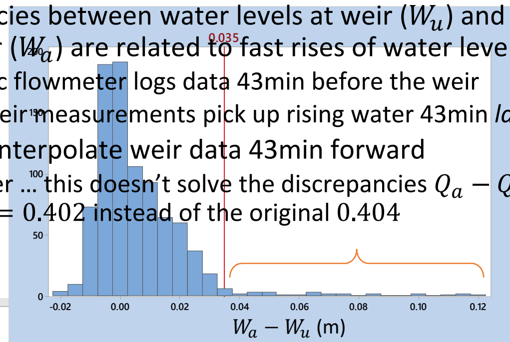
Looking for a cause that correlates systematically with flow rate $Q$.

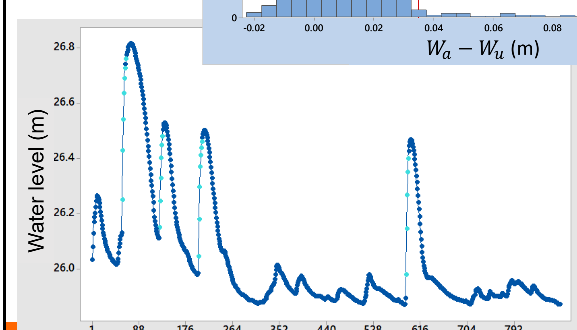
13

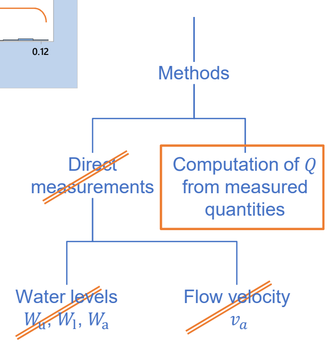
13

## Water levels $W_a$ (acoustic) and $W_u$ (weir)

- Discrepancies between water levels at weir ($W_u$) and acoustic flowmeter ($W_a$) are related to fast rises of water level
  - Acoustic flowmeter logs data 43min before the weir
  - Thus, weir measurements pick up rising water 43min *later*
- Solution: interpolate weir data 43min forward
  - However … this doesn't solve the discrepancies $Q_a - Q_w$ $RMSE = 0.402$ instead of the original 0.404



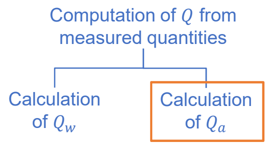



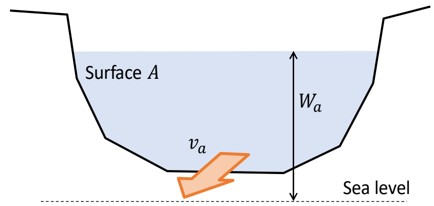




14

14

# Calculation of $Q_a$ by acoustic flowmeter




$$
\begin{aligned}
Q_a &= v_a \times A \\
&= v_a \times (c_0 + c_1 W_a + c_2 W_a^2 + c_3 W_a^3) \\
&= v_a \times (14478 - 1673 W_a + 64 W_a^2 - 0.82 W_a^3)
\end{aligned}
$$

- Are the calibration constants $c_0$, $c_1$, $c_2$ and $c_3$ correct?
  - Determined new values $\tilde{c_0}$, $\tilde{c_1}$, $\tilde{c_2}$ and $\tilde{c_3}$ such that discrepancies disappear (that is, $Q_a(\tilde{c_0}, \tilde{c_1}, \tilde{c_2}, \tilde{c_3}) \approx Q_w$).

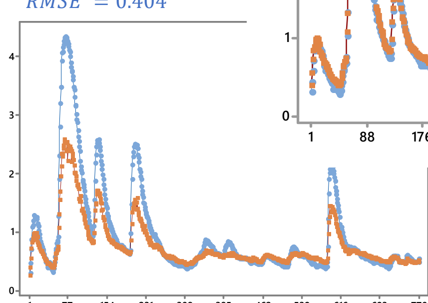
    $\frac{Q_w}{v_a} = \tilde{c_0} + \tilde{c_1} W_a + \tilde{c_2} W_a^2 + \tilde{c_3} W_a^3 + \epsilon$

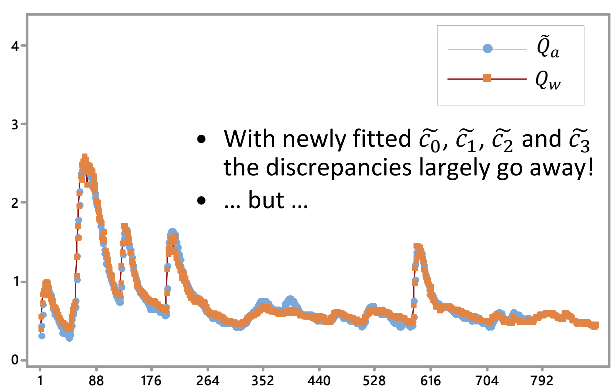




# Calculation of $Q_a$ by acoustic flowmeter

$RMSE = 0.404$

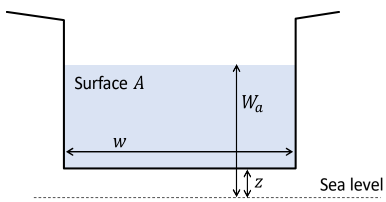



- With newly fitted $\tilde{c_0}$, $\tilde{c_1}$, $\tilde{c_2}$ and $\tilde{c_3}$ the discrepancies largely go away!
- … but …

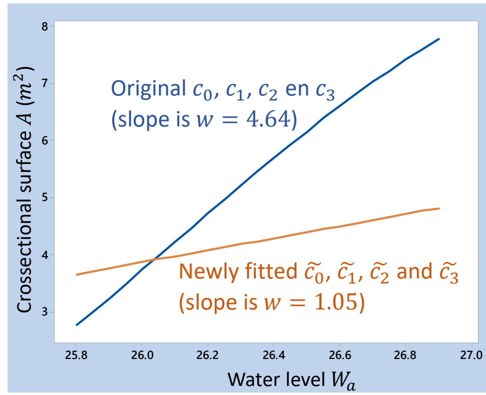
$$\text{RMSE} = \sqrt{\frac{1}{876}\Sigma(\tilde{Q}_a - Q_w)^2} = 0.056$$

# Calculation of $Q_a$ by acoustic flowmeter




Rectangular cross-section:

- $A = w \times (W_a - z)$

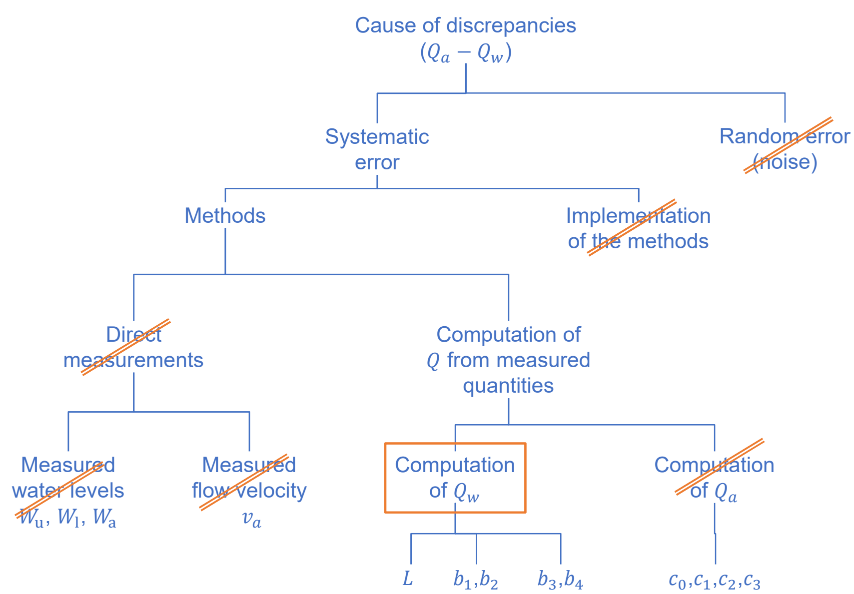



- To eliminate the discrepancies $Q_a - Q_w$ …
  - We need to use the newly fitted $\tilde{c_0}$, $\tilde{c_1}$, $\tilde{c_2}$ and $\tilde{c_3}$
  - Which imply that the width of the waterway is $w = 1.05$m
  - Whereas in fact the Buulder Aa has a width of $w = 4.50$m

17

17

# Branching and pruning …




18

18

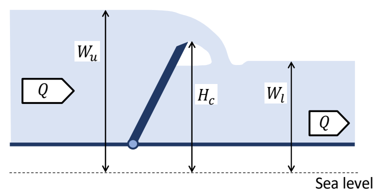
## Computation of $Q_w$ by weir

Weir computes $Q$ from height of water over weir's crest

- $Q_w = 3.00 \times 1.86 \times (W_u - H_c)^{1.5} \times (1 + \log(1-D)/2)$
  $= \underbrace{L \times b_1 \times (W_u - H_c)^{b_2}}_{Q_w^{\text{raw}}} \times \underbrace{(b_3 + b_4 \log(1-D))}_{\text{Correction for "drowned weir"}}$

  $D = \dfrac{W_l - H_c}{W_u - H_c}$

- Is there an error in one of the calibration constants $L$, $b_1$, $b_2$, $b_3$ or $b_4$?





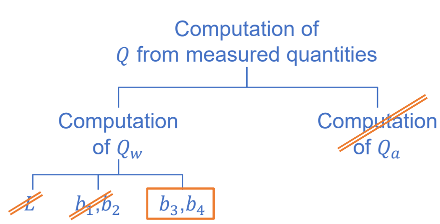



## Computation of $Q_w$ by weir

Is there an error in one of the calibration constants $L$, $b_1$, $b_2$, $b_3$ or $b_4$?

- $L$ is the width of the weir
  - The value $L = 3.00$ is the correct value
  - And there is no other value for $L$ that makes the discrepancies go away.

- The values of $b_1 = 1.86$ and $b_2 = 1.5$ are based on theory in hydrology
  - Depend on viscosity of the fluid, shape of the weir, and other properties
  - The alternative values $\tilde{b}_1 = 3.44$ and $\tilde{b}_2 = 1.9$ eliminate the discrepancies
  - … but are totally out of the range of comparable weirs in waterways!



## Final suspect: *drowning correction*

$$Q_w = \underbrace{L \times b_1 \times (W_u - H_c)^{b_2}}_{Q_w^{\mathrm{raw}}} \times \underbrace{(b_3 + b_4 \log(1 - D))}_{\text{Correction for “drowned weir”}}$$

- Original values: $b_3 = 1$ and $b_4 = 0.5$
- Can we find values $\tilde{b}_3$ and $\tilde{b}_4$ that turn $Q_w(\tilde{b}_3, \tilde{b}_4)$ into a good predictor for $Q_a$?
  $\rightarrow Q_a \approx Q_w(\tilde{b}_3, \tilde{b}_4)$

or

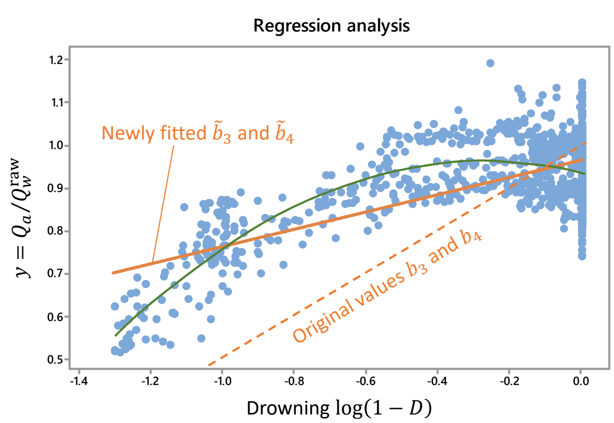
$$\frac{Q_a}{Q_w^{\mathrm{raw}}} \approx \frac{Q_w(\tilde{b}_3, \tilde{b}_4)}{Q_w^{\mathrm{raw}}} = \tilde{b}_3 + \tilde{b}_4 \log(1 - D) \quad \text{(linear in } \log(1 - D)\text{)}$$

- Find better values $\tilde{b}_3$ and $\tilde{b}_4$ by least squares.

21

21

## Final suspect: *drowning correction*




Drowning correction should be:

$$\frac{Q_w}{Q_w^{\mathrm{raw}}} = 0.935 - 0.212 \log(1 - D) - 0.387 \log^2(1 - D)$$

22

22

## Error in drowning correction!

Original computation method applied by weir setup:

$$Q_w = \underbrace{3.00 \times 1.86 \times (W_u - H_c)^{1.5}}_{Q_w^{\text{raw}}} \times \underbrace{(1 + \log(1 - D)/2)}_{\text{Correction for "drowned weir"}}$$

But this should be:

$$Q_w = 3.00 \times 1.86 \times (W_u - H_c)^{1.5} \times (0.935 - 0.212 \log(1 - D) - 0.387 \log^2(1 - D))$$



## Conclusion … What's the cause of discrepancies?

- Systematically considered what could go wrong in both measurement systems
- Most potential causes could be eliminated
  - Some cannot explain the discrepancies
  - Others *can* explain the discrepancies, but cannot be true for other reasons
- One potential explanation remains:
  *Correction for drowning in the weir setup*
  - Better correction formula makes discrepancies go away
  - Drowning correction formula lacks a solid theoretical foundation, and has emerged in practice

## Statistical reasoning in diagnostic problem solving

Lessons learned

UNIVERSITY OF WATERLOO

JADS Jheronimus Academy of Data Science

hi holland innovative

25

## Basic reasoning pattern

**Hypotheses bring focus**

*Not*: blindly look for correlations or fit curves to data,
*But*: systematically generate hypotheses about candidate causes from known theory.
*Because*: brings focus to the study and results in more interesting hypotheses.

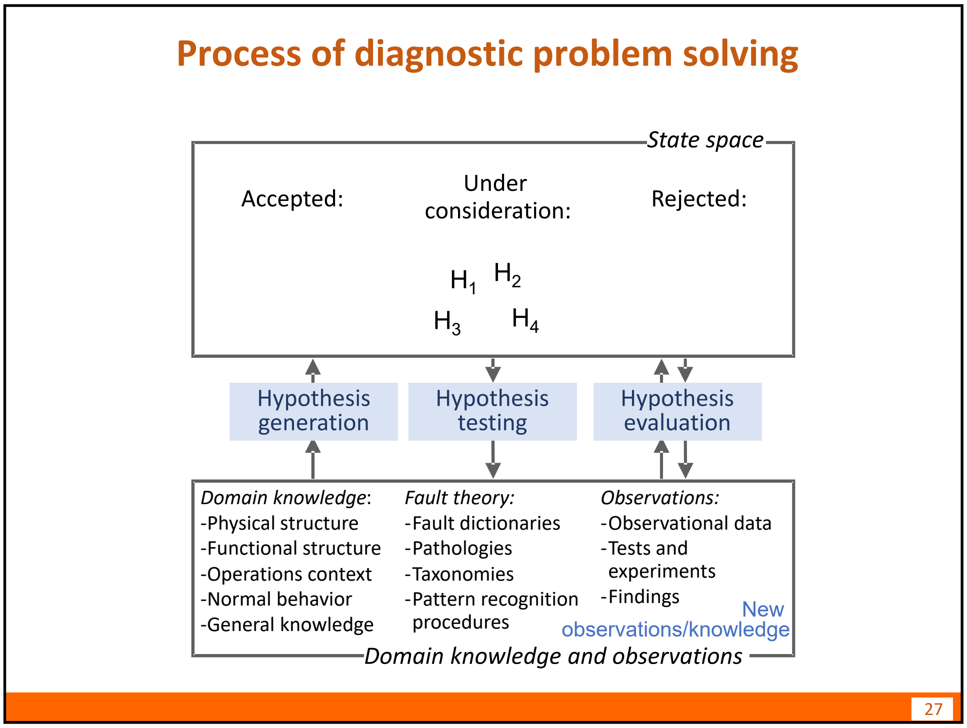
$$\frac{Q_a}{Q_w^{\text{raw}}} = \tilde{b}_3 + \tilde{b}_4 \log(1 - D) + \epsilon$$

**Branch-and-prune strategy**

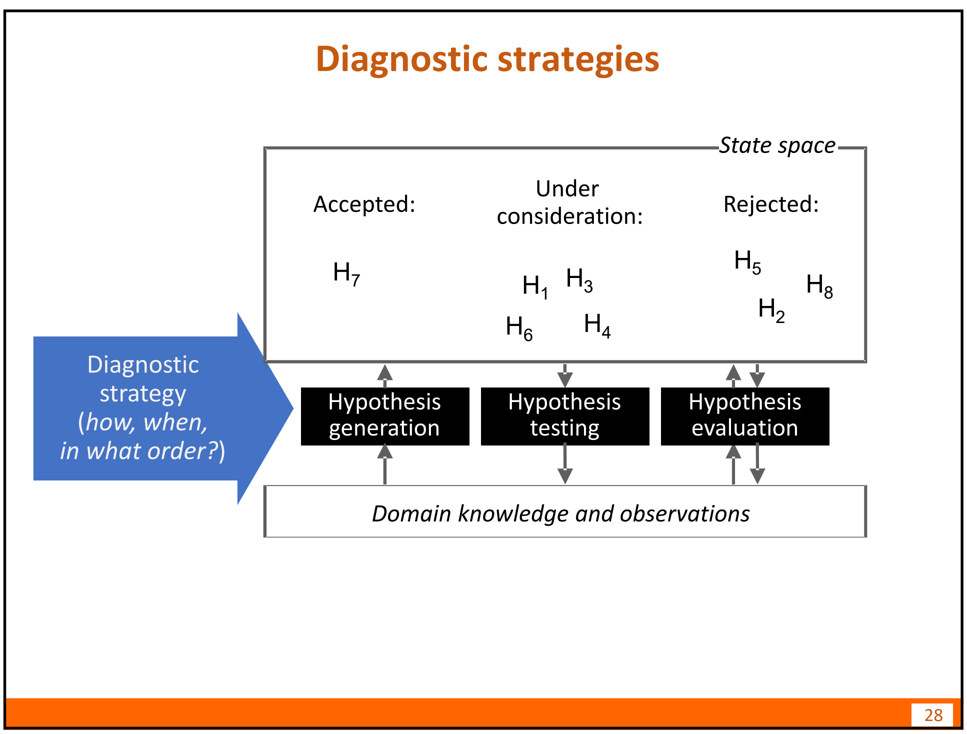
*Hierarchy*: rule out broad hypotheses early in the study,
only elaborate retained hypotheses into more detailed sub-hypotheses.
*Because*: makes the search efficient

26

26


Process of diagnostic problem solving
State space
Accepted:
Under consideration:
Rejected:
H1 H2
H3 H4
Hypothesis generation
Hypothesis testing
Hypothesis evaluation
Domain knowledge:
-Physical structure
-Functional structure
-Operations context
-Normal behavior
-General knowledge
Fault theory:
-Fault dictionaries
-Pathologies
-Taxonomies
-Pattern recognition procedures
Observations:
-Observational data
-Tests and experiments
-Findings
New observations/knowledge
Domain knowledge and observations
27

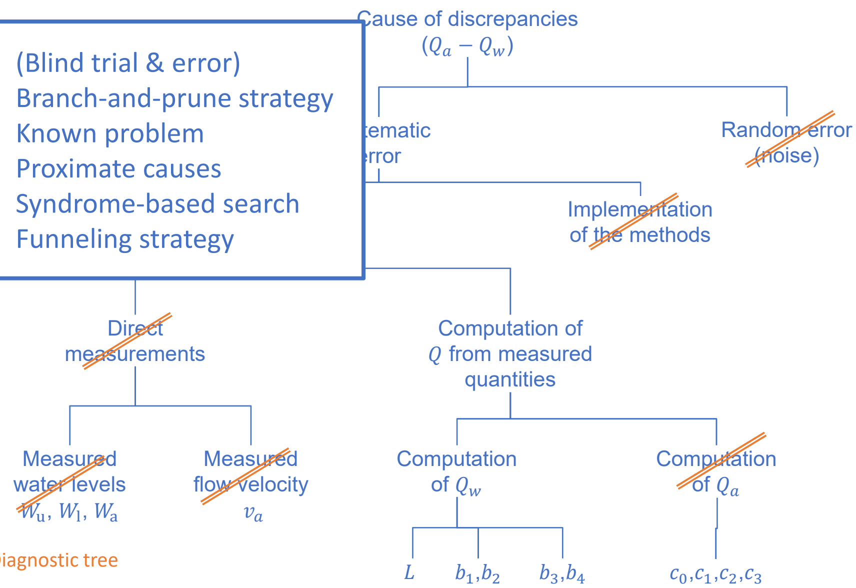
27


Diagnostic strategies
State space
Accepted:
Under consideration:
Rejected:
H7
H1 H3
H6 H4
H5
H8
H2
Diagnostic strategy (how, when, in what order?)
Hypothesis generation
Hypothesis testing
Hypothesis evaluation
Domain knowledge and observations
28

28

## Diagnostic strategies

- (Blind trial & error)
- Branch-and-prune strategy
- Known problem
- Proximate causes
- Syndrome-based search
- Funneling strategy




Diagnostic tree

29

29

## Useful statistics to guide the search

*Statistics for inference*

*Not so useful*: statistical significance.
*More useful*: statistics expressing a cause's contribution to the total problem.
*Because*: aim is not to find *all* causes, but to find the few dominant causes (Pareto).

Here: $RMSE = \left(\frac{1}{876}\sum_t \left(Q_{w,t} - Q_{a,t}\right)^2\right)^{1/2}$

- Initial $RMSE = 0.404$
- Asynchronous logging of weir and acoustic flowmeter
  Reduces $RMSE$ to 0.402
- Drowning correction
  Reduces $RMSE$ to 0.103

30

30

## Criterion for rejecting candidate explanations

Candidate explanations:

- Calibration constants $c_0$, $c_1$, $c_2$ and $c_3$ of acoustic flowmeter
  Reduces $RMSE$ to 0.056
- Changing theoretical constants $b_1$, $b_2$ applied by weir
  Reduces $RMSE$ to 0.077
- Changing drowning correction applied by weir
  Reduces $RMSE$ to 0.103
- Fitting more complex equations (or trying more contrived explanations)
  Probably could reduce $RMSE$ to close to 0.000

Goal is not simply to minimize $RMSE$.





## Criterion for rejecting candidate explanations

*Explanatory coherence*

Hypotheses are retained or rejected based on:

- Explain much of the problem
- Are parsimonious
- Agree with accepted knowledge

- Changing calibration constants $c_0$, $c_1$, $c_2$ and $c_3$ of acoustic flowmeter
  *does* resolve the discrepancies
  but *contradicts* the accepted fact that the waterway's width is 4.50m
  $\rightarrow$ Candidate explanation is rejected
- Changing drowning correction applied by weir
  *does* resolve the discrepancies
  *and* agrees with fact that this correction formula has no solid basis
  $\rightarrow$ Candidate explanation is accepted
- More complex and contrived hypotheses could explain the discrepancies as well …
  $\rightarrow$ But weren't even considered as they aren't parsimonious

## Final thought …

"Uninformed" stats. approach:
- Brainstorming
- List of candidate causes
- Experiment

Machine-learning approaches:
- Brainstorming
- Data sources
- Fit predictive algorithms / train AE or neural network

Both are *one-shot* approaches!

Sequential studies:
- Sequence of smaller experiments / analyses
- Where one experiment builds on the findings in previous experiments
- Early experiments: focus on the relevant part of the problem space
- Later experiments: efficient testing of detailed hypotheses