

A critique of some aspects of statistical usage in hydroclimate research.

Robin Clarke

Instituto de Pesquisas Hidraulicas

Porto Alegre – RS, Brazil

Area of concern: 1.

**Tests of hypotheses often applied
incorrectly.**

Area of concern 1: use of data both to suggest a hypothesis to be tested, and to test it.

■ **Examples:**

Testing for a “break” in record, at a particular time determined by looking at it.

Area of concern 1 contd:

- **Testing whether an observation at a particular time is an outlier, when the time is determined by looking at the record.**
- **Testing whether a “significant” trend exists in a record, after visual inspection suggested that trend exists.**

Area of concern 1 contd:

- In agricultural/biological research, where data suggest a hypothesis, an experiment can be designed (e.g., at an experimental station, or in a laboratory) to test it by collecting new data, explicitly obtained for that purpose.
- In hydroclimate research, such experiments are rarely if ever possible: we can only locate unused data sets, not designed for the purpose of testing hypotheses, and hope to use them for that purpose.

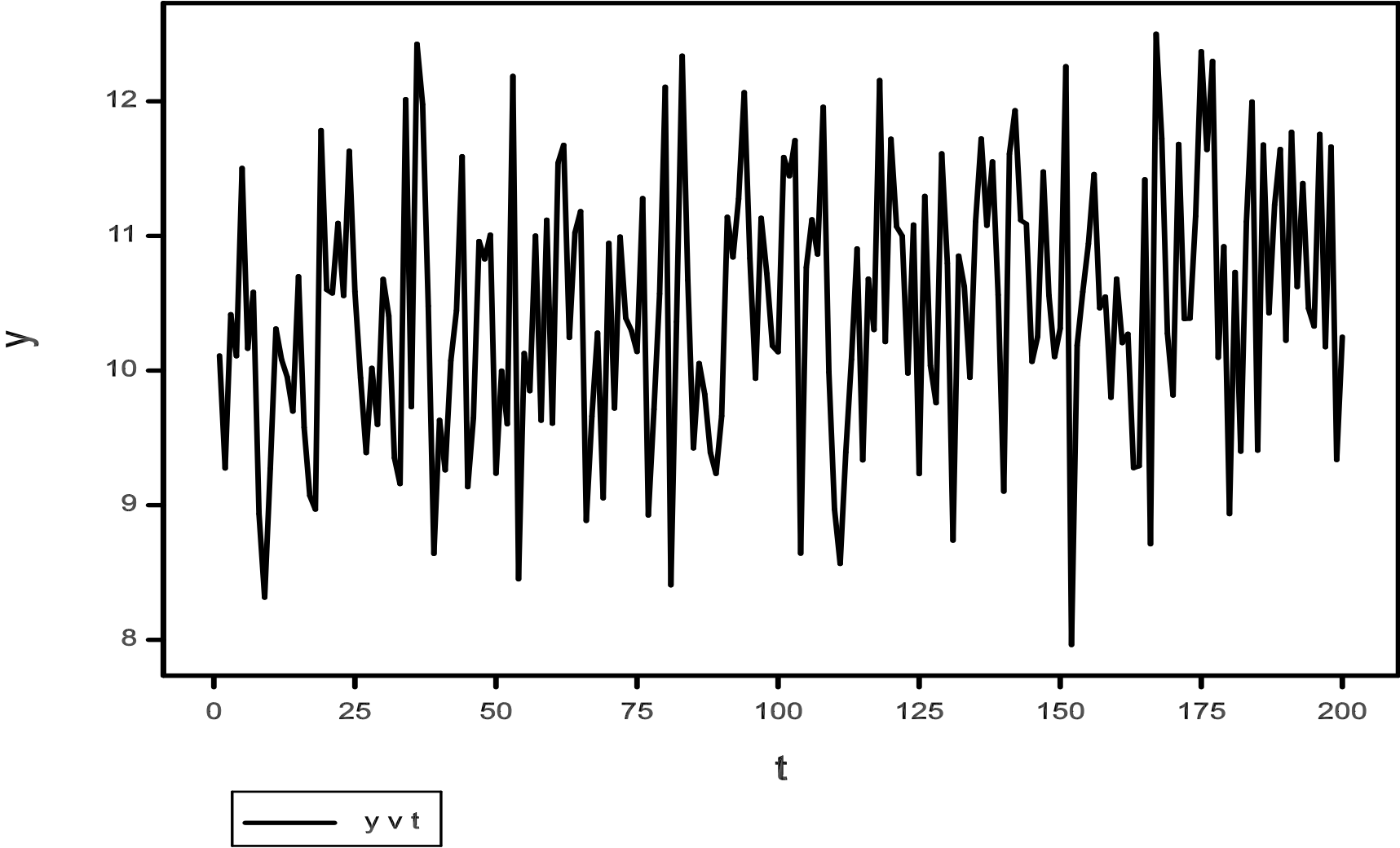
Area of concern: 2.

**Excessive reliance on the
testing of hypotheses.**

Area of concern 2: excessive use of hypothesis-testing procedures.

- **Given enough data, any null hypothesis can be falsified.**
- **Even where result of a significance test is “very highly significant ($P < 0.001$)”, this conclusion may be of little practical use.**

Simulated data, $y=10 + 0.004 t + e$ with $e \sim N(0,1)$



Trivial example:-

- Generate samples of increasing size N from $y=10+0.004 t + \varepsilon$, where $\varepsilon \sim N(0,1)$. (Seed =170934).

Linear trend: 4 units per 1000 "years".

N	$b \times 10^3$	$se[b] \times 10^3$	F	R^2
50	8.21	± 9.21	0.79	~ 0
100	4.22	± 3.25	1.69	0.7
150	4.34	± 1.76	6.10 **	3.3
200	3.63	± 1.17	9.68 ***	4.2

Area of concern 2 contd: excessive use of hypothesis-testing procedures.

- **Opinion: Less emphasis should be given to the testing of hypotheses, greater emphasis on**
- **(i) identifying sources of uncertainty in model parameters;**
- **(ii) seeking to reduce such uncertainties, by identifying “components of variance” in residuals ε .**
- **See T A Cohn and H F Lins (2005) “Nature’s style: naturally trendy”, GRL Vol.32, L23402 for excellent discussion of the value of significance testing.**

Area of concern: 3.

Failure to take account of spatial correlation in variables being analysed;
failure to recognise effects of spatial correlation on significance tests for
trend.

Area of concern 3: failure to take account of spatial correlation.

- **Cohn & Lins discussed “significance” of trend in a single series. In practice, regional trends are sometimes more important.**
- **Example: are there trends in Amazon rainfall? (In Europe, STarDEX Project)**
- **Common approach: analyse record from each station separately, ignoring spatial correlation between sites.**

Area of concern 3: failure to take account of spatial correlation.

From StarDEX Final Report:

“Table 2: Number of precipitation stations with positive/negative trends and the corresponding no. of significant ($p < 5\%$) trends for precipitation indices across Europe (481 stations)”

[No mention of trend magnitudes; no mention of trend standard errors; no allowance for spatial correlation;...]

Area of concern 3 contd: failure to take account of spatial correlation.

- **Allowance for spatial correlation necessary for:**
- **Establishing reality of regional trends;**
- **Comparing e.g. satellite rainfall fields with raingauge-network observations.**
- **Example: comparison of TRMM with raingauge network in Brazilian Amazon:-**

238 Sites in Brazilian Amazon: means 1998-2005.

Mean annual rainfall (mm):-

Gauges:	TRMM:	Diff.	SE(Unc).	SE(Corr).
2101	2113	-12.0	± 21.0	± 84.4

Mean value, 95% quantile of daily rainfall (mm):-

31.5	29.0	+2.49	± 0.32	± 2.37
------	------	-------	------------	------------

Mean value, number of days in year with > 2mm:-

117	139	- 22	± 1.4	± 10.2
-----	-----	------	-----------	------------

Area of concern: 4:

- Use of sequences of values generated by GCMs (e.g., up to year 2100) as if they were data, for the purpose of looking at how/whether frequency of extreme events will change.

Area of concern 4:

- **What is done:** suppose Y is a variable of interest, determined by forcing variables X and parameters φ ; $Y = F(X, \varphi)$ in the real world.
- **Model (e.g., GCM) is postulated,** $Y = f(X^*, \theta) + \varepsilon$.
- **After estimating θ , as θ_{est} say, a future scenario of X^* is used to calculate Y_{est} , the corresponding "Y",**
 $Y_{est} = f(X^*, \theta_{est})$.
- **Y_{est} which is really a deterministic series, is treated as "data".**

Area of concern 4 contd:

- The “data sequence” Y_{est} is used to examine whether (e.g.) the frequency of extreme values of Y , changes, under the future scenario of X^* .

Area of concern 4: A trivial example (contd)

- Generate 200 values from $y=10+0.004 t + \varepsilon$, where $\varepsilon \sim N(0,1)$. (Seed =170934, so sequence is “deterministic” in the sense that a GCM-generated sequence is deterministic).
- Use first 50 pairs (y, t) to fit linear model. Result is $y=9.981 + 0.00821 t$.
- Use fitted model (note that f has the same form as F in this example!) to calculate y_{est} for $t=51\dots 200$.
- Treating this sequence y_{est} as “data”, mean of y_{est} is 11.01 (compared with true value 10.54) BUT $\text{s.d.}(y_{\text{est}}) = \pm 0.357$, whilst $\text{s.d. of } y_{51\dots 200}$ is ± 0.971 .

Area of concern 4 contd:

- Hence we see that:
 - the deterministically-generated sequence y_{est} will have (some) characteristics different from those of observed data-sequences.
 - In particular, variability in Y_{est} is under-estimated.
 - Therefore, values of Y with given return period (e.g., 20 years) will be under-estimated.

Why state what is so obvious?

- Because the use of GCM-generated sequences to study extreme events in the future appears to fall into the same trap.**

Example from a recent paper:-

- **Aim**: to answer question “when will changes in extreme precipitation (P) due to climate change be detectable?”
- **Method**: Use of 160-year long GCM-generated sequences of daily P, (a) with CO₂ forcing, (b) without CO₂ forcing (Control).
- **Non-stationary** GEV model fitted to sequence (a) (non-stationary in position parameter); **stationary** GEV model fitted to sequence (b).
- Time of detectable change defined as time at which hypothesis “daily rainfall with 20-year return period is equal for (a) and (b)” is rejected.

Example from a recent paper (contd):-

- Paper says:
- “Climate model output is deterministic in that, if given the same initial conditions, the model will output the exact same measurements. However, because of the chaotic nature of the models and their sensitivity to initial conditions, *statistical models can be used to analyze the output in much the same way as they are used to analyze observed weather data.*” [my italics]
- QUESTION: IS THIS TRUE?

Example from a recent paper (contd):-

Doubtful, because:-

(a) real world $F(X,\varphi) \neq$ model $f(X^*,\theta)$.

**(b) the model output $Y_{est}=f(X^*,\theta_{est})$,
i.e., without the component " ε ".**

Example from a recent paper (contd):-

■ **SO: WHAT IS THE ALTERNATIVE?**

- **Recall the original question:-** "when will changes in extreme precipitation (P) due to climate change be detectable?"
- **Analyst needs answers to questions:**
- **Why is the question important? What major changes in policy/investment will occur if (say) annual maximum daily rainfall changes after 2050? Or after 2040?.....**
- **[Need to balance cost of a major analytical study against the cost of not knowing the answer to the question posed].**

Conclusions (opinions only):

- **Tests of hypotheses often applied incorrectly;**
- **Reliance on hypothesis testing is excessive;**
- **Failure to recognise importance of spatial correlation and its consequences;**
- **Use of GCM output to explore future changes in extremes is questionable.**

Winston Churchill:

**“My Right Honourable friend uses
statistics as a drunkard
uses a lamp-post -
for support, not for illumination.”**

**That is all –
thank you for your
attention!**