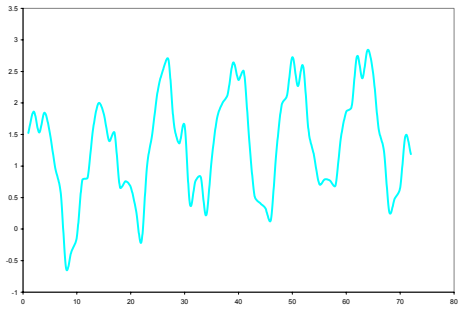
Trends and Trend Assessment Tools

March 15, 2005
Bio 0843

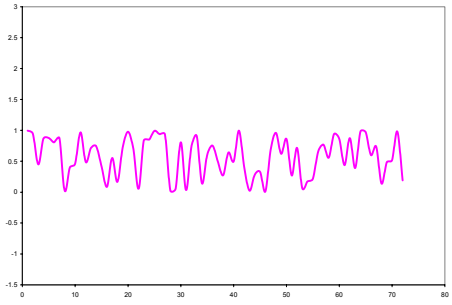
Time Series Components

- Trend
- Seasonality
- Random Variation
- Period of Observation
- Frequency of Observation

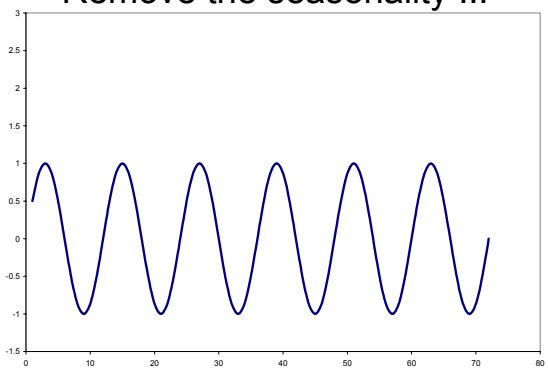
Given a series ...



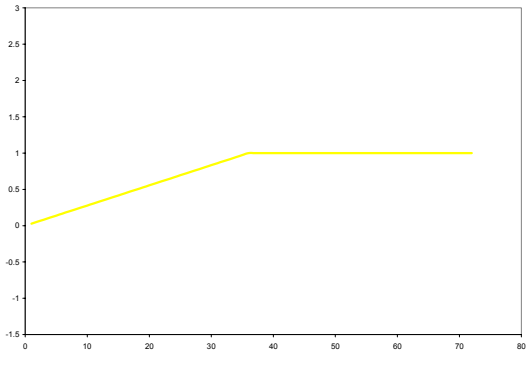
Take out the random variation ...



Remove the seasonality ...



and the trend is left



What I want to cover

- What is a trend ?
- What is trend detection ?
- Hypotheses
  - Narrative and statistical
  - Explicit and implicit
- Critical evaluation of trend detection methodologies
- Critical evaluation of trends reported in the literature

Trend

- "A monotonic change in the level of a series, usually overtime but sometimes in space"
- "A general increase or decrease in observed values of some random variable over time"
- *i.e.*
  - *1980, there has been an increasing trend in pH in Sudbury lakes (time)*
  - *The further one gets from the smelter at Sudbury, the lower the sulfate levels in lake water (space)*

What is trend detection

- "it is not generally possible for statistics to detect trends that are not apparent by inspection, especially for data records of short to moderate length – say 20 years or less. It is preferable to think of trend tests as a quantitative basis for deciding whether apparent trends are real"
  - Loftis et al. 1989

Narrative vs. Statistical Hypotheses

- Narrative hypothesis is an informal statement, i.e., *there are fewer spring peepers in recent years*
- Statistical hypothesis is a formal (mathematical) statement
  - contains explicit and implicit components
  - implicit components can trip you up

Explicit and Implicit Hypotheses

- Explicit: *there is a trend ..."*
- Implicit
  - type of trend, (linear, monotonic, etc.)
  - population distribution (normal, etc.)
  - power of the test
  - appropriate sampling strategy
- Implicit hypotheses can trip you up and muddy the interpretation of results

Interpreting trends in the literature

- What was the period of observation ?
  - Older results may be less relevant
  - Shorter series are more prone to "drift"
- What test was used ?
  - Different tests make different assumptions about the data
  - Some tests are more optimistic than others
- Are the reported results relevant to your situation ?

Trend Detection Methodologies

- Time Series Analysis
- Environmental statistics
  - Linear regression
  - Mann Kendall test
  - Seasonal Kendall test
  - Sen Slope test
  - Wilcoxon test

Time Series Analysis Requirements

- Long data series (typically 10's – 100's of records)
- Equally spaced observations
- No missing data
- Known distribution of parameters
- Statisticians

Environmental Statistics Requirements

- Can work with short series (5 or more observations)
- Handle unequally spaced data
- Not necessarily as sensitive to distributional assumptions
- Readily available and easily applied
- Not as necessary to seek professional help

Time Series vs. Environmental trend statistics

- Forecasting vs. prediction
- Time series analysis can be used to make predictions about the future state of a variable; environmental trend statistics cannot
- Very important when evaluating the literature; environmental trend statistics are valid only for the period of observation

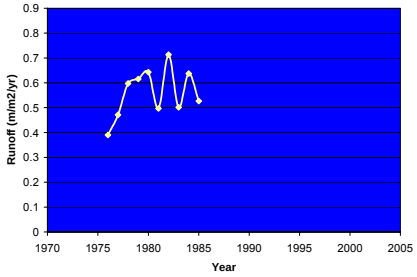
What to believe ?

- Thermocline depths are getting deeper in boreal shield lakes (Smith et al. 2000)
- Linear regressions were performed of maximum thermocline depth against year for 12 lakes in the Boreal Shield with data from 1990-1998. In all cases, a downward trend was observed; 7 of these slopes were significant at the $p<0.05$ level

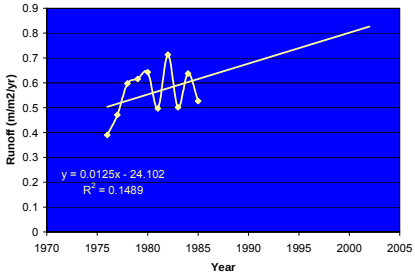
Trend is dependent on time period

- It is expected that climate change will affect stream flow in Canada. Over the past 15 years, published reports suggest that stream flows are both increasing and decreasing. What should one believe ?

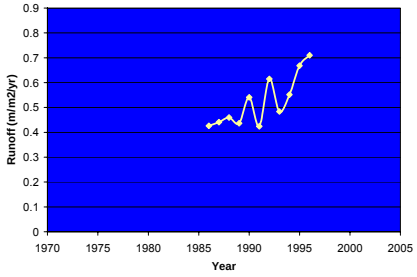
Is there a trend ?



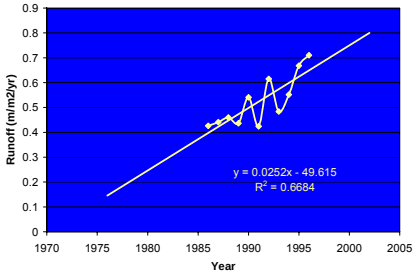
Is there a trend ?



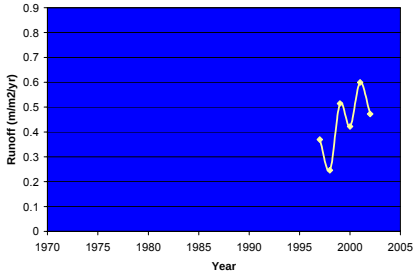
Is there a trend ?



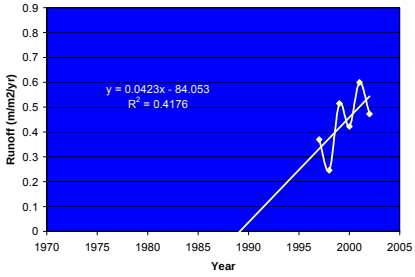
Is there a trend ?



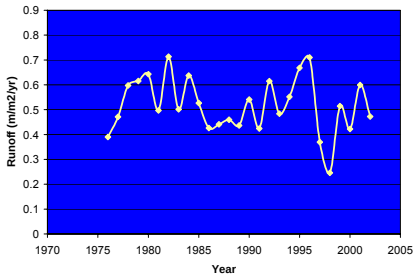
Is there a trend ?



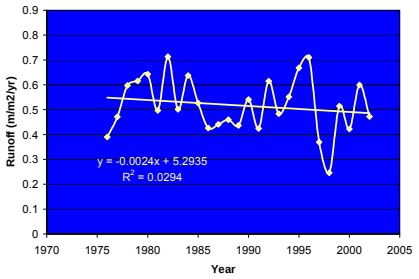
Is there a trend ?



Is there a trend ?



Is there a trend ?



Linear Regression

- Readily available, easy to use
- Slope gives an estimate of magnitude of change over time
- Regression statistics provide an estimate of degree of association, potentially spurious significance
- Over-applied

Assumptions of Linear Regression

- Trend is linear; rate of increase or decrease is constant
- Errors are normally distributed
- Response variable (y) is drawn from a normal distribution

Linear Regression Hypotheses

- Explicit: There is a trend in the value of a series of interest over time
- Implicit
  - Change is the same over each time period
  - Data are normally distributed
  - Error is homogeneous
  - Variance is constant

Other problems with linear regression

- Very sensitive to outliers; abnormally high or low values at the start or end of a series will have a disproportionate influence on the estimate of slope
- Has difficulties with data below detection limit; no clear guidelines as to what numbers should be used

Non-parametric tests: an alternative to linear regression

- Do not assume the data follow a specific distribution
- Less sensitive to extreme values
- Less sensitive to missing values
- Tests are not as powerful when assumptions of parametric tests are met

Mann-Kendall test

- Are the values of a series going up or down in a manner not expected by chance alone ?
- Compares all pairs of observations, counts the number where values are increasing, subtracts the number decreasing and calculates a probability

Mann-Kendall Calculation (1)

$x_1(t_1)$	$x_2(t_2)$	$x_3(t_3)$	$x_4(t_4)$	…	$x_{n-1}(t_{n-1})$	$x_n(t_n)$
	x_2 x_1	x_3 x_1	x_4 x_1	…	x_{n-1} x_1	x_n x_1
		x_3 x_2	x_4 x_2	…	x_{n-1} x_2	x_n x_2
			x_4 x_3	…	x_{n-1} x_3	x_n x_2
					…	…
					x_{n-1} x_{n-2}	x_n x_{n-2}
						x_n x_{n-1}

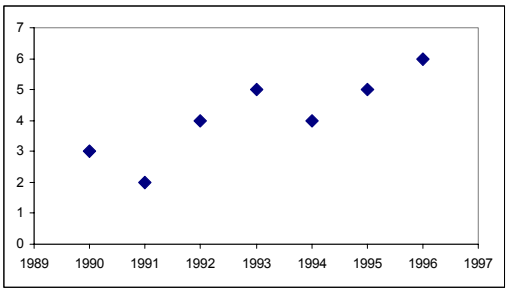
Mann-Kendall Calculation (2)

$x_1(t_1)$	$x_2(t_2)$	$x_3(t_3)$	$x_4(t_4)$	…	$x_{n-1}(t_{n-1})$	$x_n(t_n)$
	$y_{2,1}$	$y_{3,1}$	$y_{4,1}$	…	$y_{n-1,1}$	$y_{n,1}$
		$y_{3,2}$	$y_{4,2}$	…	$y_{n-1,2}$	$y_{n,2}$
			$y_{4,3}$	…	$y_{n-2,3}$	$y_{n,3}$
$x_k - x_j > 0$- $>y_{kj} = 1$ $x_k - x_j < 0$- $>y_{kj} = -1$ $x_k - x_j = 0$- $>y_{kj} = 0$					…	…
					$y_{n-1,n-2}$	$y_{n,n-2}$
						$y_{n,n-1}$

Mann-Kendall Calculation (3)

- $S = \Sigma_{k=1,n} \Sigma_{j=k,n} y_{k,j}$
- $V(S) = n(n-1)(2n+5)/18$
- $Z = (S-1)/V(S)^{1/2}$ {S > 0}
- $Z = (S+1)/V(S)^{1/2}$ {S < 0}
- $p = z_{(1-\alpha)}$
- If abs(Z) > $z_{(1-\alpha)}$ reject H_0: no trend, sign of Z tells direction of trend

Mann-Kendall Example



Mann-Kendall Example (Δx)

	1990	1991	1992	1993	1994	1995	1996
	3	2	4	5	4	5	6
3		-1	1	2	1	2	3
2			2	3	2	3	4
4				1	0	1	2
5					-1	0	1
4						1	2
5							1

Mann-Kendall Example (y)

	1990	1991	1992	1993	1994	1995	1996
	3	2	4	5	4	5	6
3		-1	1	1	1	1	1
2			1	1	1	1	1
4				1	0	1	1
5					-1	0	1
4						1	1

Mann-Kendall Example (stats)

- $S = \Sigma_{k=1,n} \Sigma_{j=k,n} y_{k,j} = 15$
- $V(S) = n(n-1)(2n+5)/18 = 53.667$
- $Z = (S-1)/V(S)^{1/2}$ {S > 0} = 1.91
- $p = z_{(1-\alpha)} = 0.027$
- If abs(Z) > $z_{(1-\alpha)}$ reject H_0: no trend, accept H_1: trend
- S > 0, therefore positive trend

So there's a trend, now what

- Magnitude can be estimated using Sen's slope
- median (50th percentile) inter-annual difference between values
- can be calculated from Mann-Kendall tableau
- more robust than regression slope
- somewhat subject to outliers

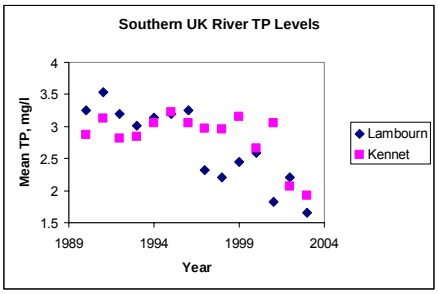
That's not all ...

- probabilities must be adjusted for small sample sizes (n <10) and serial correlation
- adjustments can be made for tied data
- multiple series can be compared for trend homogeneity
- http://www.epa.gov/quality1/dqa.html, *Guidance for Data Quality Assessment: Practical Methods for Data Analysis (G-9) QA00 Version*

Sen Slopes (Δx/ Δy)

	1990	1991	1992	1993	1994	1995	1996
	3	2	4	5	4	5	6
3		-1	0.5	0.67	0.25	0.4	0.5
2			2	1.5	0.67	0.75	0.8
4				1	0	0.33	0.5
5					-1	0	0.33
4	Sen = 0.67					1	1
5							1

Does choice of test matter?



Background

- Lambourn and Kennet are two tributaries of the Thames. The Thames supplies drinking water for 2/3 of the 8 million people in greater London. Eutrophication is seriously impacting the water quality of these rivers. Agricultural incentives were initiated in 1990 to encourage best management practices of fertilizer application. Are they working ?

What is trend detection

- "it is not generally possible for statistics to detect trends that are not apparent by inspection, especially for data records of short to moderate length – say 20 years or less. It is preferable to think of trend tests as a quantitative basis for deciding whether apparent trends are real"
  - Loftis et al. 1989

SUMMARY OUTPUT Lambourn

Regression Statistics	
Multiple R	0.89
R Square	0.80
Adjusted R Square	0.78
Standard Error	0.28
Observations	14

ANOVA

	df	SS	MS	F	F-prob
Regression	1	3.69	3.69	46.90	0.000018
Residual	12	0.94	0.08		
Total	13	4.64			

	Coefficients	s.e.	t Stat	P-value	L 95%	U 95%
Intercept	257.08	37.14	6.92	1.6E-05	176.15	338.00
time	-0.13	0.02	-6.85	1.78E-05	-0.17	-0.09

SUMMARY OUTPUT Kennet

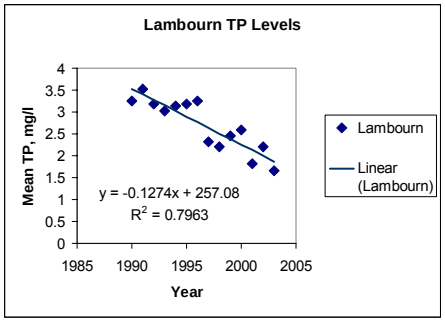
Regression Statistics	
Multiple R	0.55
R Square	0.30
Adjusted R Square	0.25
Standard Error	0.34
Observations	14

ANOVA

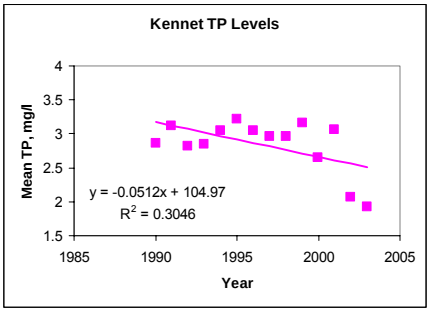
	df	SS	MS	F	F-prob
Regression	1	0.60	0.60	5.26	0.04
Residual	12	1.36	0.11		
Total	13	1.95			

	Coefficients	s.e.	t Stat	P-value	L 95%	U 95%
Intercept	104.97	44.55	2.36	0.04	7.90	202.03
time	-0.05	0.02	-2.29	0.04	-0.10	-0.003

Lambourn Regression



Kennet Regression



Trend Summaries

	Lambourn	Kennet
Regression p	1e 05	0.04
M Kp	5e 4	0.16
M KZ	- 328	- 095
Regression slope	- 013	- 005
Sen Slope	- 012	- 001
Overall	Decline	No Trend

Conclusions

- Non-parametric tests (such as Mann-Kendall and Sen's slope) are preferable when implicit assumptions about the data are not met
- Non-parametric statistical tests give similar (but slightly less powerful) results when the implicit assumptions about data are met