Applications of Generalized Least Squares Regression Analysis for Hydrological Trend Detection and Streamflow Projections Under Global Warming Scenarios

Jeannine-Marie St. Jacques[1], Yang Zhao[2], Suzan Lapp[1] and David Sauchyn[1]

[1]Prairie Adaptation Research Collaborative
[2]Mathematics and Statistics Department
University of Regina


Northwest
Territories
Hudson's
Bay
Great Slave Lake
Peace River
Lake Athabasca
Manitoba
Fort McMurray
Nelson River
Churchill River
Alberta
Athabasca River
N. Saskatchewan R.
Saskatchewan
L. Winnipeg
Ontario
Edmonton
Saskatoon
British
Columbia
Calgary
Regina
S. Saskatchewan R.
Winnipeg
Montana
North Dakota

Projected changes in streamflow by the end of the 21st century

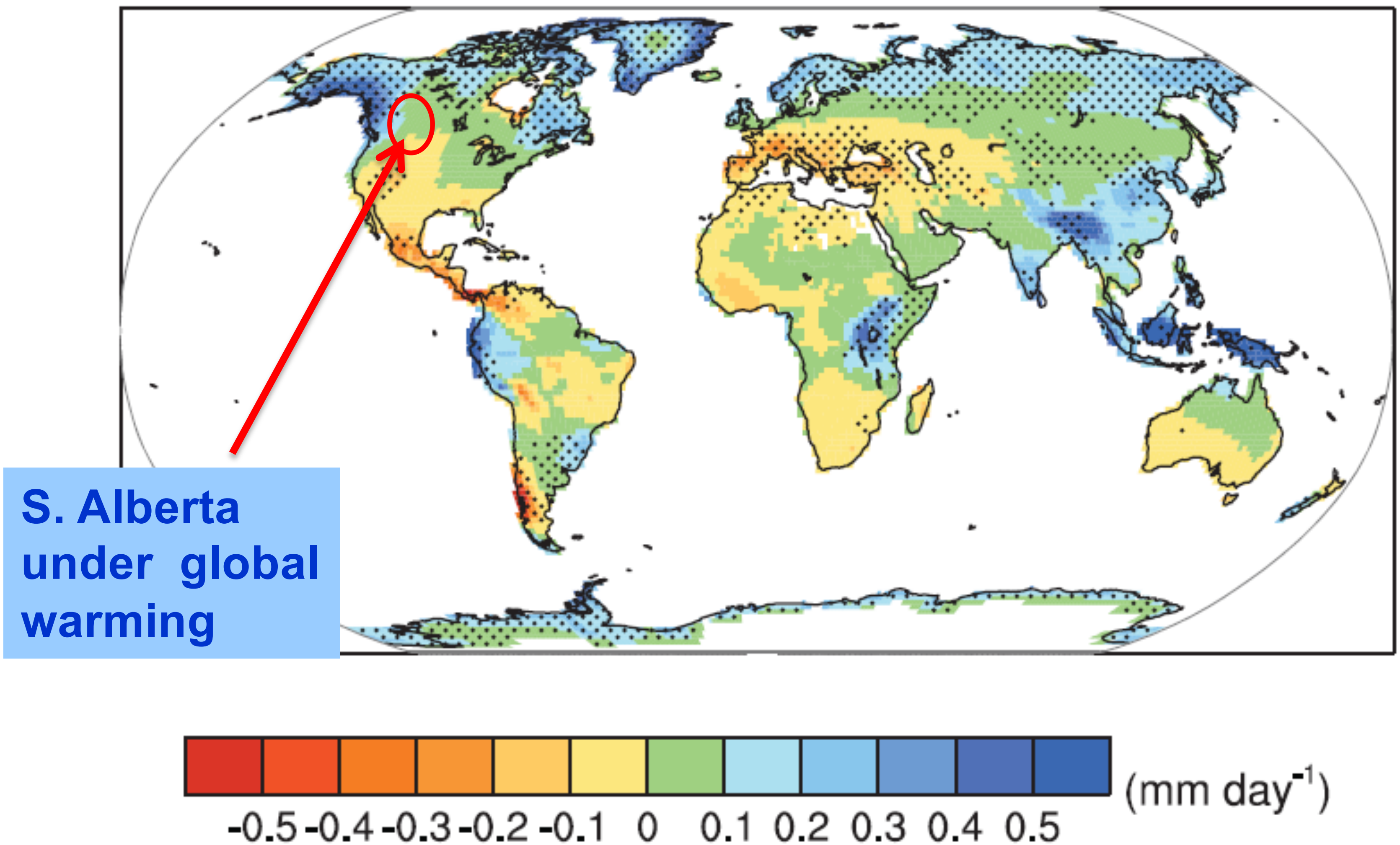


Fig. 10.12 IPCC 4. Multi-model mean changes in streamflow (mm/day). Changes are annual means for the SRES A1B (moderate emissions) scenario for the period 2080 to 2099 relative to 1980 to 1999.

Introduction:

Southern Alberta river basins are located in a **transitional** region of global climate models (GCMs).

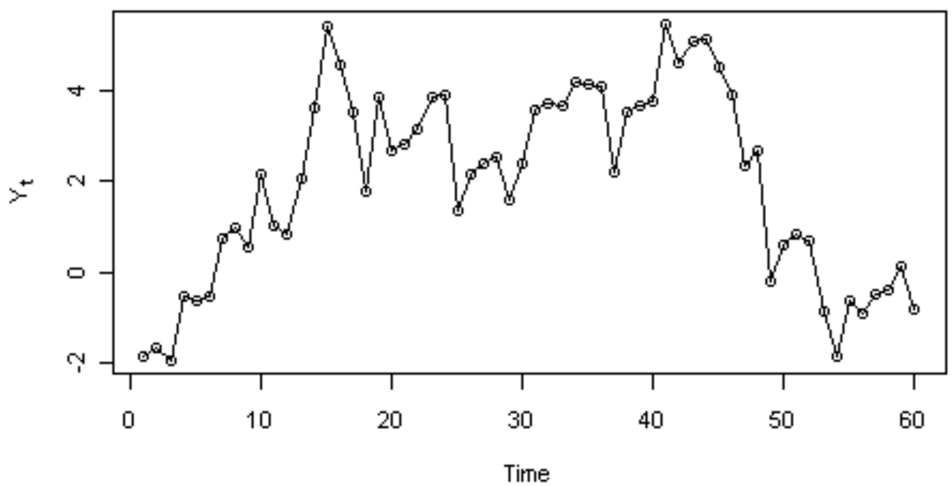
Are there any developing trends in the actual streamflow records?

Recent research showed **declining trends** in S. Alberta streamflow records
(Zhang *et al*., 2001; Rood *et al*., 2005, 2008; Schindler and Donahue, 2006).

However, there are challenging **data analysis issues** in S. Alberta streamflow records that must be **explicitly addressed**
in any trend study:

Problem #1: Autocorrelation in streamflow data

Autocorrelation is the correlation of a time series with its own past and future values.



Geophysical time series are frequently autocorrelated because of ***inertia or carryover processes*** in the physical system.

Example: the slow drainage of groundwater reserves might impart correlation to successive annual flows of a river.

Streamflow data has frequent positive serial correlation in the residuals **therefore classical linear regression and Mann-Kendall non-parametric methods will disproportionately detect trend.**

(Kulkarni and von Storch, 1995; Zheng *et al.,* 1997; Zheng and Basher, 1999; Zhang *et al.,* 2000, 2001; Burn and Hag Elnur, 2002; Yue *et al.,* 2002)

How autocorrelation messes up OLS

$$Y_t = \beta_0 + \beta_1 t + e_t$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sum_{t=1}^{n} (Y_t - \hat{\beta}_0 + \hat{\beta}_1 t)^2}{(n-2)\sum_{t=1}^{n}(t - \bar{t})^2}$$

No autocorrelation in residuals case

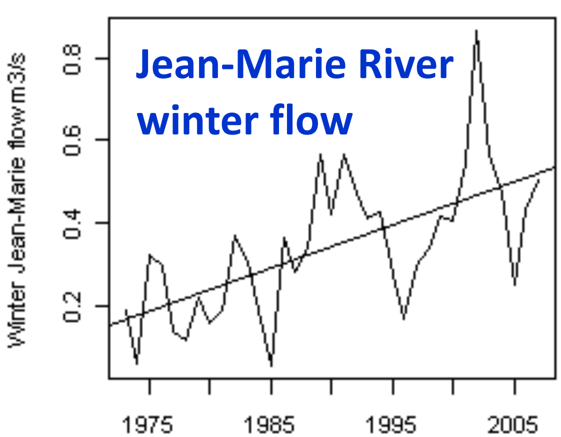
$$\hat{\beta}_1 / \sqrt{\text{Var}(\hat{\beta}_1)} \sim t_{(\alpha/2,\, n-2)}$$

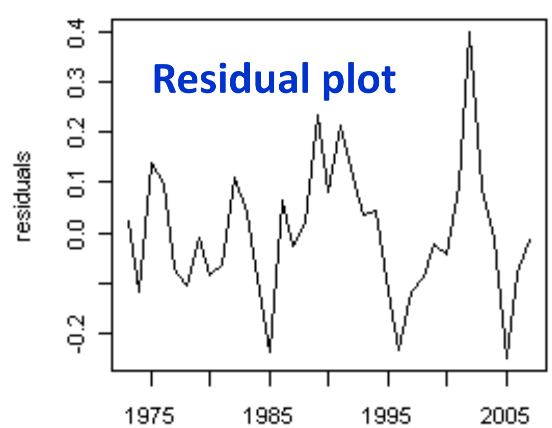
$$\text{Var}(\hat{\beta}_1) = \frac{12\gamma_0}{n(n^2-1)} \left\{ 1 + \frac{24}{n(n^2-1)} \sum_{s=2}^{n} \sum_{t=1}^{s-1} (t - \bar{t})(s - \bar{t})\, \rho_{s-t} \right\}$$

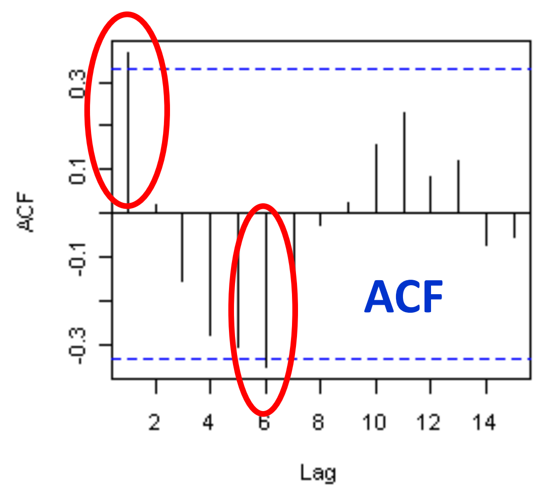
Residual autocorrelation

positive residual autocorrelation underestimate $\text{Var}(\hat{\beta}_1)$

Autocorrelated residuals AR(1)?







Regression $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}$

for **OLS**,

$\boldsymbol{\beta}_{OLS} = (\mathbf{X'X})^{-1}\,\mathbf{X'Y}$

Variance-covariance matrix $\mathbf{cov}(\boldsymbol{\beta}_{OLS}) = (\mathbf{X'X})^{-1}\,\mathbf{X'}\,\boldsymbol{\Sigma}_n\,\mathbf{X}\,(\mathbf{X'X})^{-1}$
where $\boldsymbol{\Sigma}_n = \mathbf{cov}(\mathbf{WW'})$

$$\boldsymbol{\Sigma}_n = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ . & . & . & & . \\ . & . & . & & . \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix}$$

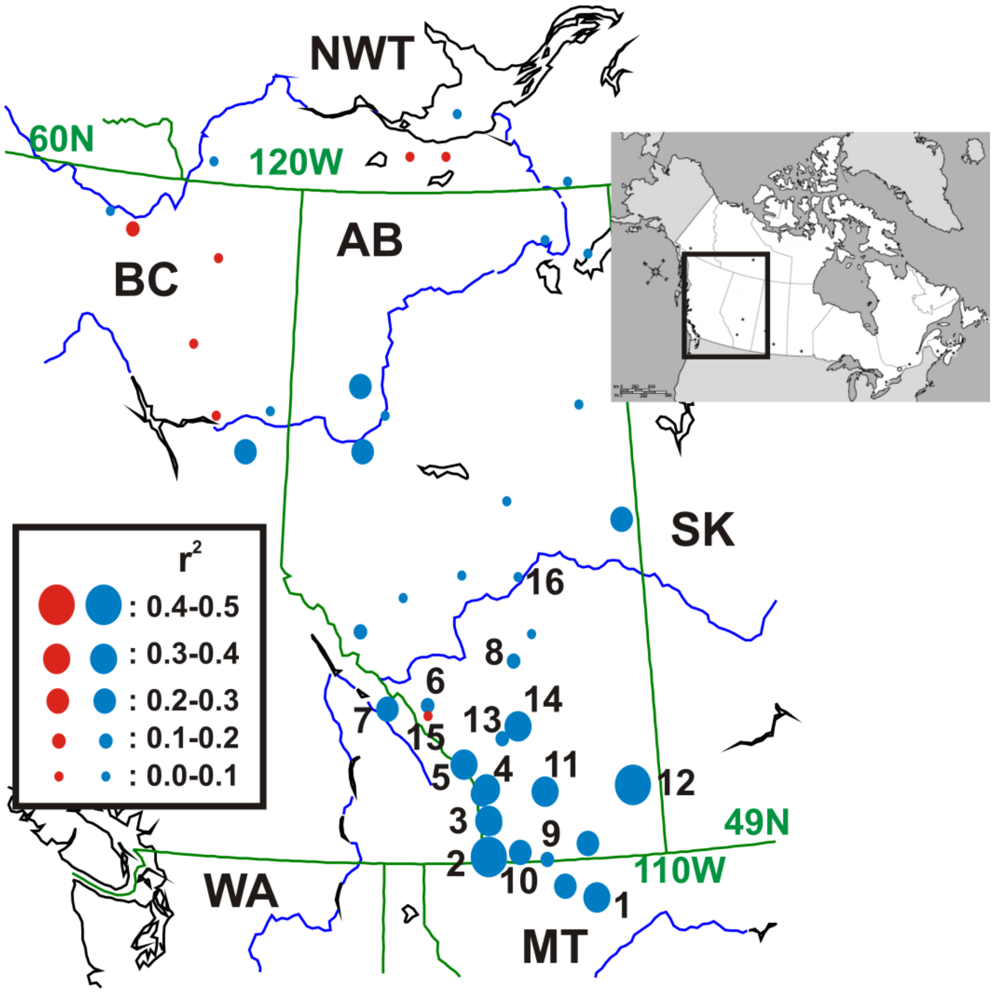
If residuals are normal i.i.d., $\mathbf{cov}(\boldsymbol{\beta}_{OLS}) = \sigma^2(\mathbf{X'X})^{-1}$

Since $$\boldsymbol{\Sigma}_n = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2\,\mathbf{I}$$

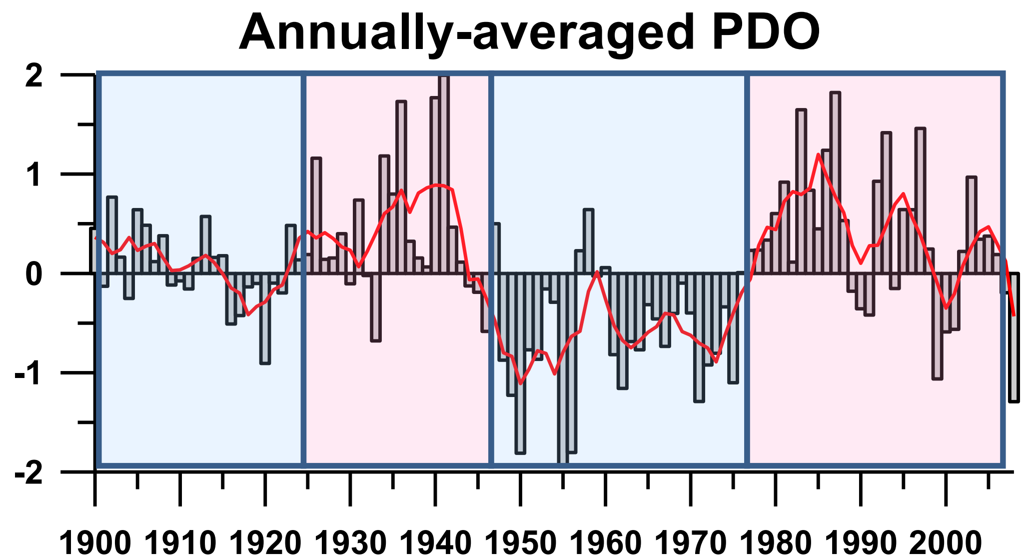
Therefore $\mathbf{cov}(\boldsymbol{\beta}_{OLS}) = (\mathbf{X'X})^{-1}\,\mathbf{X'}\,\sigma^2\,\mathbf{I}\,\mathbf{X}\,(\mathbf{X'X})^{-1} = \sigma^2(\mathbf{X'X})^{-1}$ if have normal i.i.d. residuals

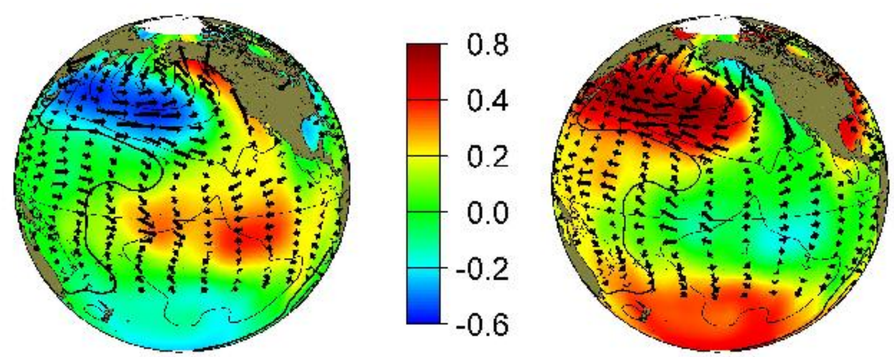
Problem #2: The *Pacific Decadal Oscillation (PDO)* is a major factor controlling streamflow in Alberta.

A strong negative relationship exists between the two



Correlations between same yr PDO and rivers
Both filtered by 5-yr binomial smoother



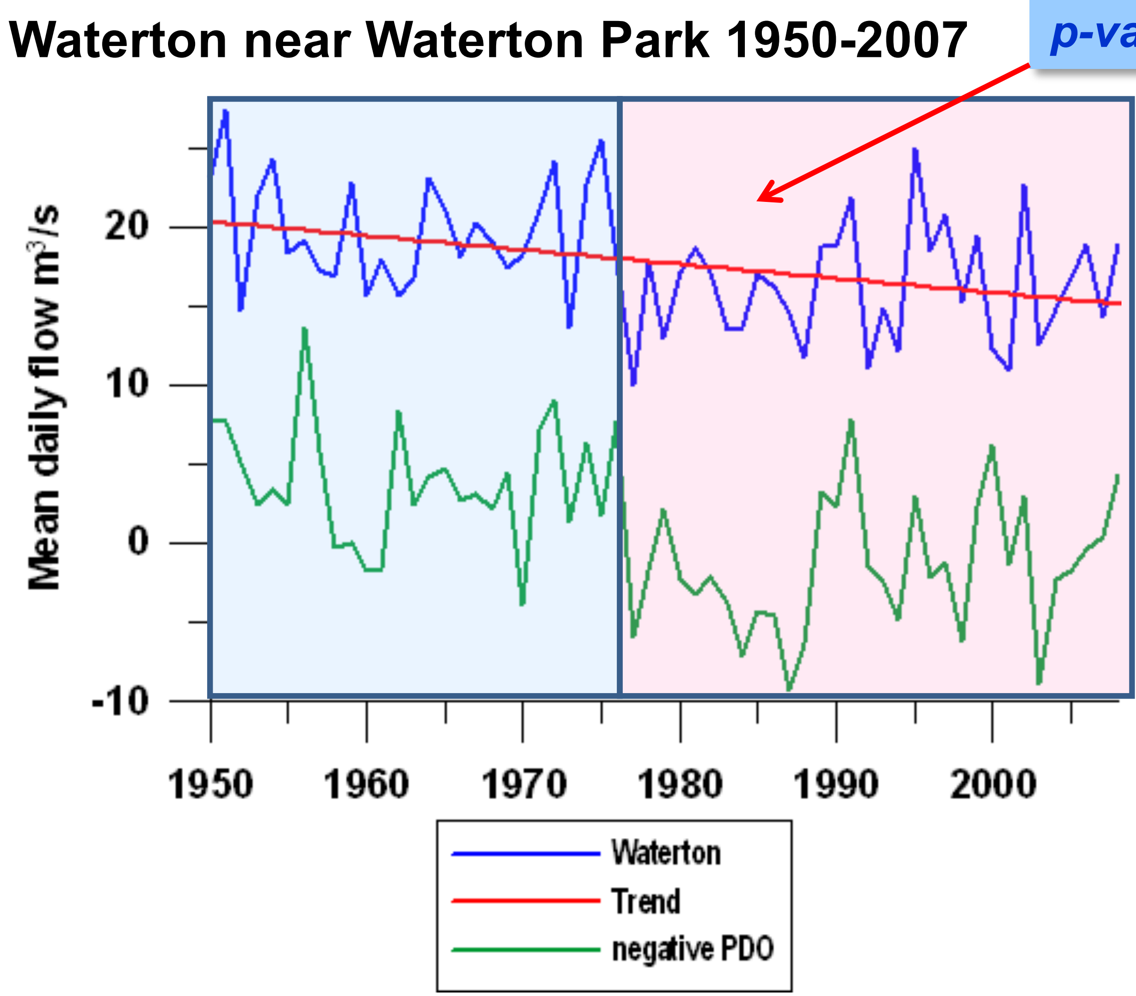


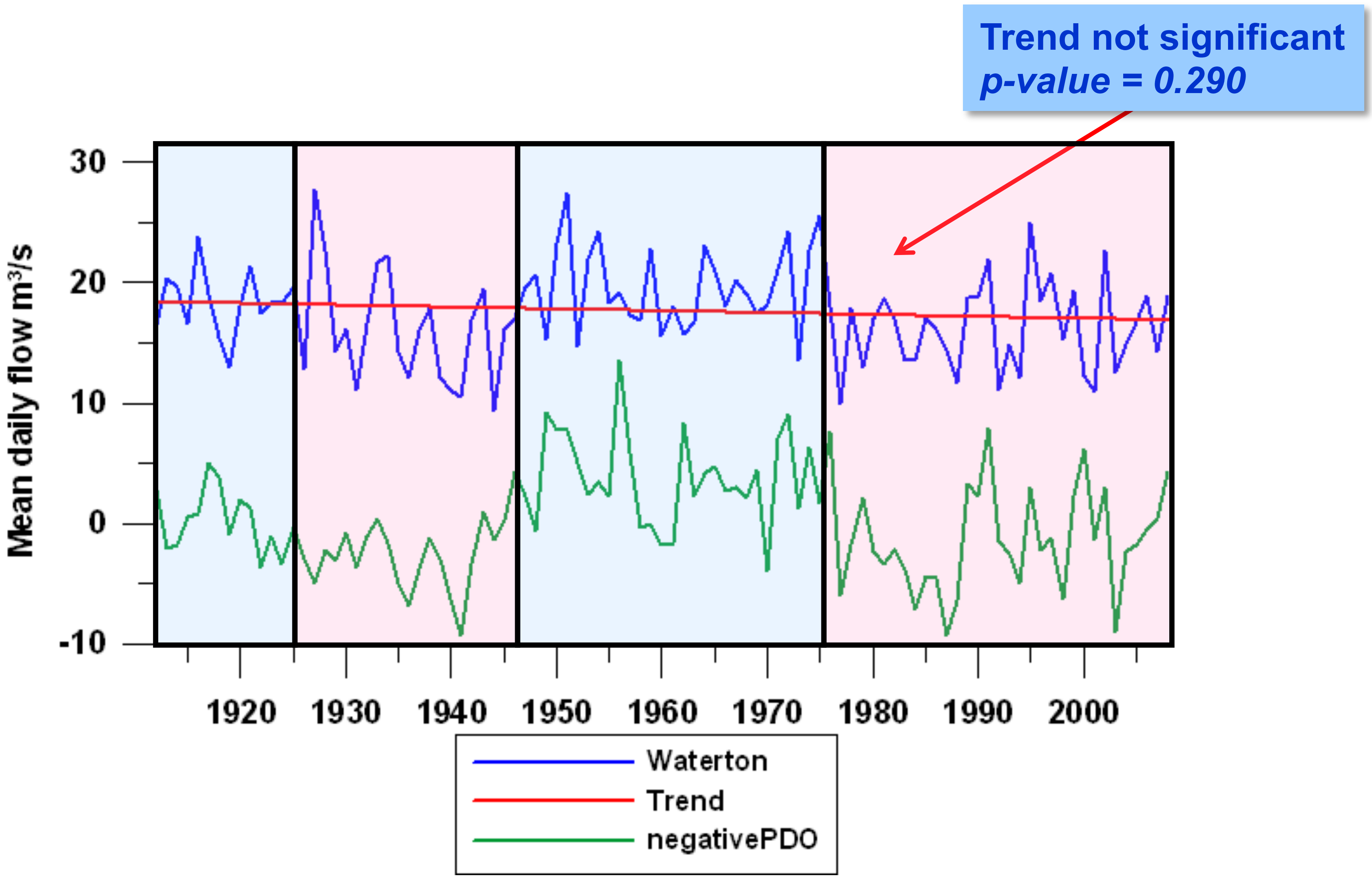
http://jisao.washington.edu/pdo/

Problem: the phase of the low frequency PDO (~60 yr) and sampling period can induce false global warming trends

Significant trend

p-value = 0.004





Many Alberta instrumental records begin in the 1950s, or omit the 1930s and 1940s (periods of high positive PDO, hence low AB streamflow).

If PDO not taken into account, could produce false global warming declines.

Three further problems with Southern Alberta streamflow data:

- ***Short*** typically ~40-50 years in N. Alberta and at most ~95 years in S. Alberta.

- ***Gappy*** especially in 1930s (economic collapse) and the 1940s (war).

- ***Heavy human impact*** from irrigation, dams, cities, tar sands, especially in S. Alberta, obscuring natural hydrology.

Solutions

Serial correlation in residuals: use **Generalized Least Squares regression (GLS)** which fits ARMA models to the residuals. Use **R** programming language. Data is mean daily flow (m^3/s) annualized over the year, so Central Limit Theorem, essentially normally distributed.

PDO: explicitly include its effect in **model**. Also include **El Niño** or **Southern Oscillation Index (SOI)** and **North Atlantic Oscillation (NAO)** to improve signal-to-noise ratio.

Short, gappy data: use **longest** (80-90 years), **most complete** records with modest infilling.

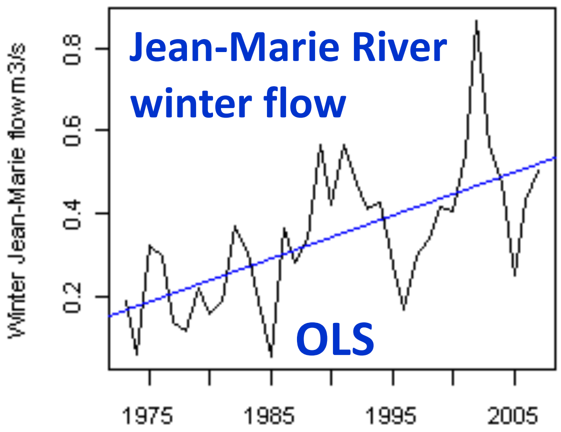
Heavy human impact:

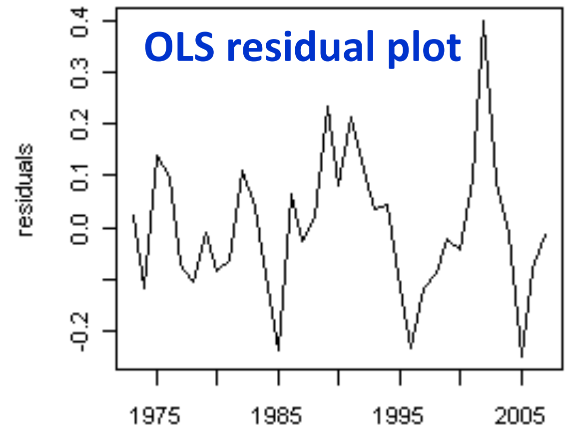
(1) examine **unregulated rivers,** and

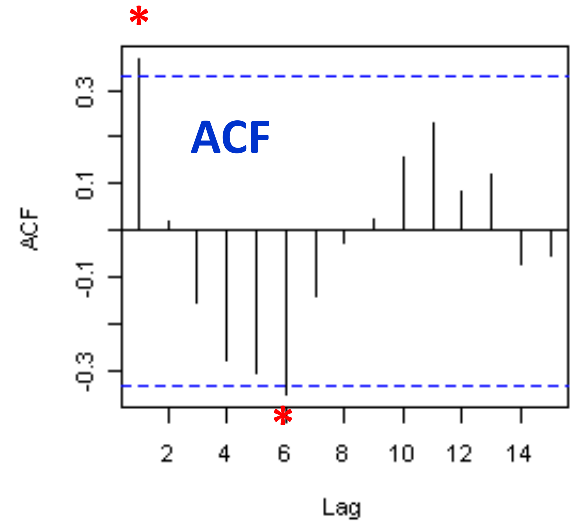
(2) compare actual flows to their corresponding **naturalized** flows from Alberta Environment.

Definition: ***Naturalized flow*** is an estimate of what the flow should have been if we hadn't removed the water.

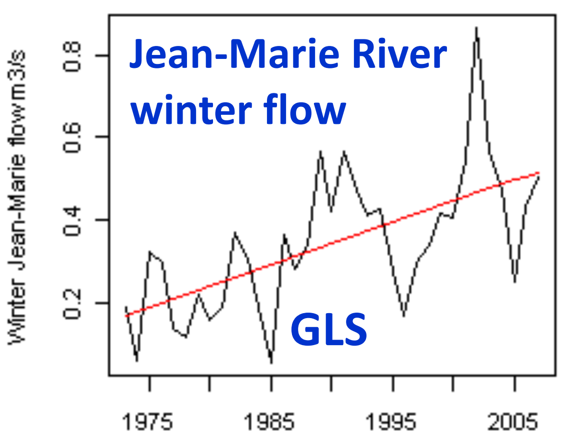
Generalized Least Squares Regression

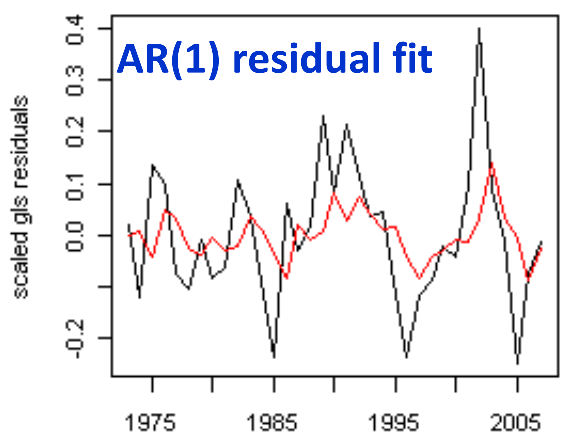


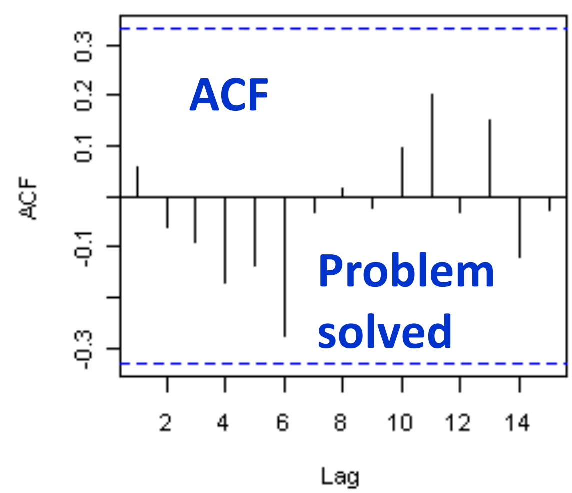




Have residual autocorrelation? Model it with ARMA(*p,q*) process and throw it into the fit!







95% C.I. for β_1 = 0.1050 ± 0.0624

	ar1	β_0	β_1
est.	0.3555	0.0001	0.1050
s.e.	0.1546	0.0317	0.0312

Regression $Y = X\beta + W$

for **OLS**,

$$\beta_{OLS} = (X'X)^{-1} X'Y$$

Variance-covariance matrix $\mathbf{cov}(\beta_{OLS}) = (X'X)^{-1} X' \Sigma_n X (X'X)^{-1}$
where $\Sigma_n = \mathbf{cov}(WW')$

$$\Sigma_n = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ . & . & . & & . \\ . & . & . & & . \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix}$$

for **GLS**,

$$\beta_{GLS} = (X' \Sigma_n^{-1} X)^{-1} X' \Sigma_n^{-1} Y$$

Variance-covariance matrix $\mathbf{cov}(\beta_{GLS}) = (X' \Sigma_n^{-1} X)^{-1}$

GLS is the **best linear unbiased estimator** of β

Statistical Methodology

Use **low-pass filtered mean daily streamflow** (5-year binomial smoother).

Use as predictors: **trend**, **PDO**, **SOI**, **NAO**.
Climate variables also low-pass filtered and leading streamflow by **-1, 0, +1, +2** years.

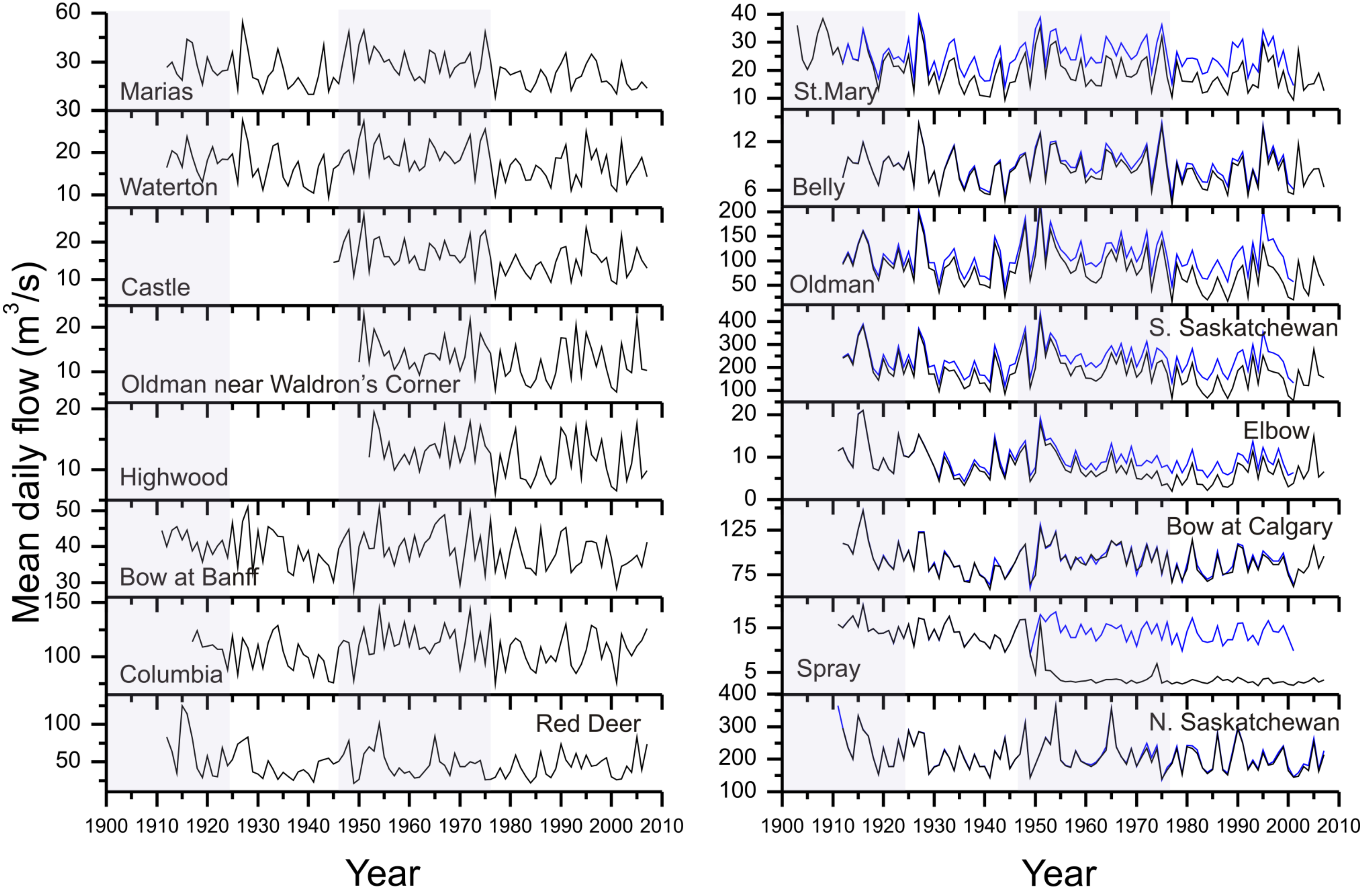
For each river
Loop { for all |{predictor subsets}| ≤ *6*, for all *p,q* such that *p ≤ 8, q ≤ 5*
fit GLS model predicting river flow, using subset of predictors and ARMA(*p,q*) residuals
(arima(river,order=c(*p*,0,*q*), xreg=predsubset, method=c("ML"))
} end Loop

Choose model with least **corrected Akaike Information Criterion (AIC_c)** goodness-of-fit statistic.

Assess significance of trend with **Neyman-Pearson** statistic (RP).

following Zheng *et al.* (1997) *Journal of Climate*

24 Southern Alberta streamflow records analyzed so far…



Grey shading of negative phase of PDO

Results

Flow Record	Actual flow record			Naturalized flow record			Human impact /yr
	Record period	Significant linear Trend?	Change %/yr	Record period	Significant linear trend?	Change %/yr	
Marias R. near Shelby, MT	1912-2007	decreasing	**-0.26**	n.a.			
Waterton R. near Waterton Park	1912-2007	none	-0.05	n.a.			
Castle R. near Beaver Mines	1945-2007	none	-0.04	n.a.			
Oldman R. near Waldron's Corner	1950-2007	increasing	**0.43**	n.a.			
Highwood R. at Diebel's Ranch	1952-2007	none	0.11	n.a.			
Bow R. at Banff	1911-2007	decreasing	**-0.12**	n.a.			
Columbia R. at Nicholson, BC	1917-2007	none	-0.001	n.a.			
Red Deer R. at Red Deer	1912-2007	decreasing	**-0.22**	n.a.			
St. Mary R. at International Boundary	1903-2007	decreasing	**-0.46**	1912-2001	none	0.006	-0.47
Belly R. near Mountain View	1912-2007	none	0.02	1912-2001	none	0.02	-0.002
Oldman R. near Lethbridge	1912-2007	decreasing	**-0.76**	1912-2001	decreasing	**-0.18**	-0.58
S. Saskatchewan R. at Medicine Hat	1912-2007	decreasing	**-0.36**	1912-2001	increasing	**0.05**	-0.41
Elbow R. below Glenmore Dam	1911-2007	decreasing	**-0.70**	1912-2001	decreasing	**-0.35**	-0.35
Bow R. at Calgary	1912-2007	decreasing	**-0.16**	1912-2001	decreasing	**-0.16**	-0.01
Spray R. at Banff	1911-2007	decreasing	**-2.20**	1912-2001	decreasing	**-0.11**	-2.09
N. Saskatchewan R. at Edmonton	1912-2007	decreasing	**-0.14**	1911-2007	decreasing	**-0.10**	-0.04

15 declines, 7 no trends and only 2 increases

From analyzing both actual and corresponding naturalized flows, infer direct human impacts:

Change%/yr

Metric for global warming versus human impact

$Q_t = \mu + \lambda T_t + \beta_1 x_{1,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t$, $\qquad t = 1, \ldots, L,$

$Q_t = \mu + \lambda T_t$

Change%/yr = $100\ \lambda\ / mean(Q_t)$

Naturalized record Change%/yr reflects only **global warming**

Actual record Change%/yr reflects **global warming and human impact**

human impact = difference between Change%/yr for actual flow record and its corresponding naturalized flow

Results

Flow Record	Actual flow record: Record period	Actual flow record: Significant linear Trend?	Actual flow record: Change %/yr	Naturalized flow record: Record period	Naturalized flow record: Significant linear trend?	Naturalized flow record: Change %/yr	Human Impact /yr
Marias R. near Shelby, MT	1912-2007	**decreasing**	**-0.26**	n.a.			
Waterton R. near Waterton Park	1912-2007	none	-0.05	n.a.			
Castle R. near Beaver Mines	1945-2007	none	-0.04	n.a.			
Oldman R. near Waldron's Corner	1950-2007	**increasing**	**0.43**	n.a.			
Highwood R. at Diebel's Ranch	1952-2007	none	0.11	n.a.			
Bow R. at Banff	1911-2007	**decreasing**	**-0.12**	n.a.			
Columbia R. at Nicholson, BC	1917-2007	none	-0.001	n.a.			
Red Deer R. at Red Deer	1912-2007	**decreasing**	**-0.22**	n.a.			
St. Mary R. at International Boundary	1903-2007	**decreasing**	**-0.46**	1912-2001	none	0.006	-0.47
Belly R. near Mountain View	1912-2007	none	0.02	1912-2001	none	0.02	-0.002
Oldman R. near Lethbridge	1912-2007	**decreasing**	**-0.76**	1912-2001	**decreasing**	**-0.18**	-0.58
S. Saskatchewan R. at Medicine Hat	1912-2007	**decreasing**	**-0.36**	1912-2001	**increasing**	**0.05**	-0.41
Elbow R. below Glenmore Dam	1911-2007	**decreasing**	**-0.70**	1912-2001	**decreasing**	**-0.35**	-0.35
Bow R. at Calgary	1912-2007	**decreasing**	**-0.16**	1912-2001	**decreasing**	**-0.16**	-0.01
Spray R. at Banff	1911-2007	**decreasing**	**-2.20**	1912-2001	**decreasing**	**-0.11**	-2.09
N. Saskatchewan R. at Edmonton	1912-2007	**decreasing**	**-0.14**	1911-2007	**decreasing**	**-0.10**	-0.04

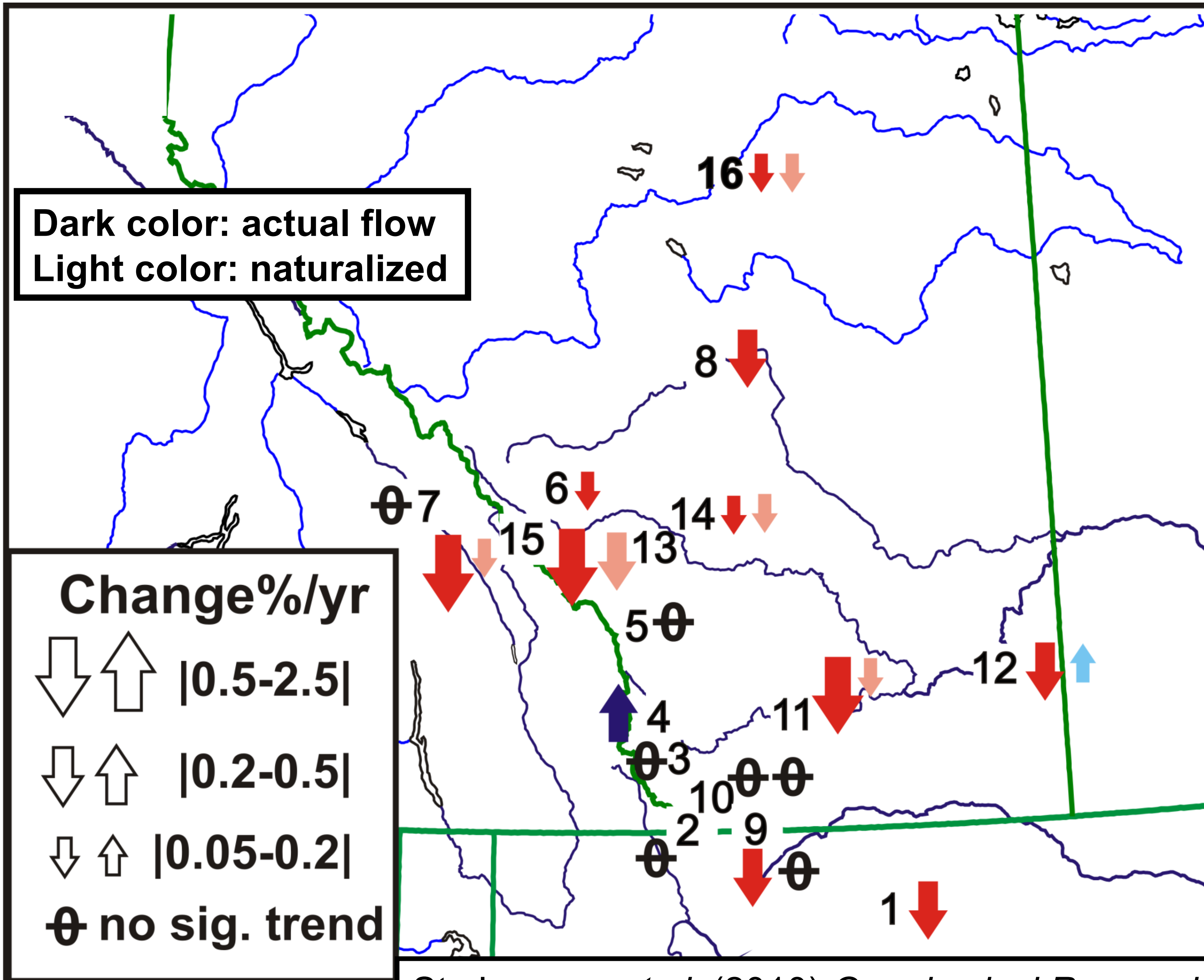
AGW **Human impacts**

15 declines, 7 no trends and only 2 increases

From analyzing both actual and corresponding naturalized flows, infer direct human impacts:

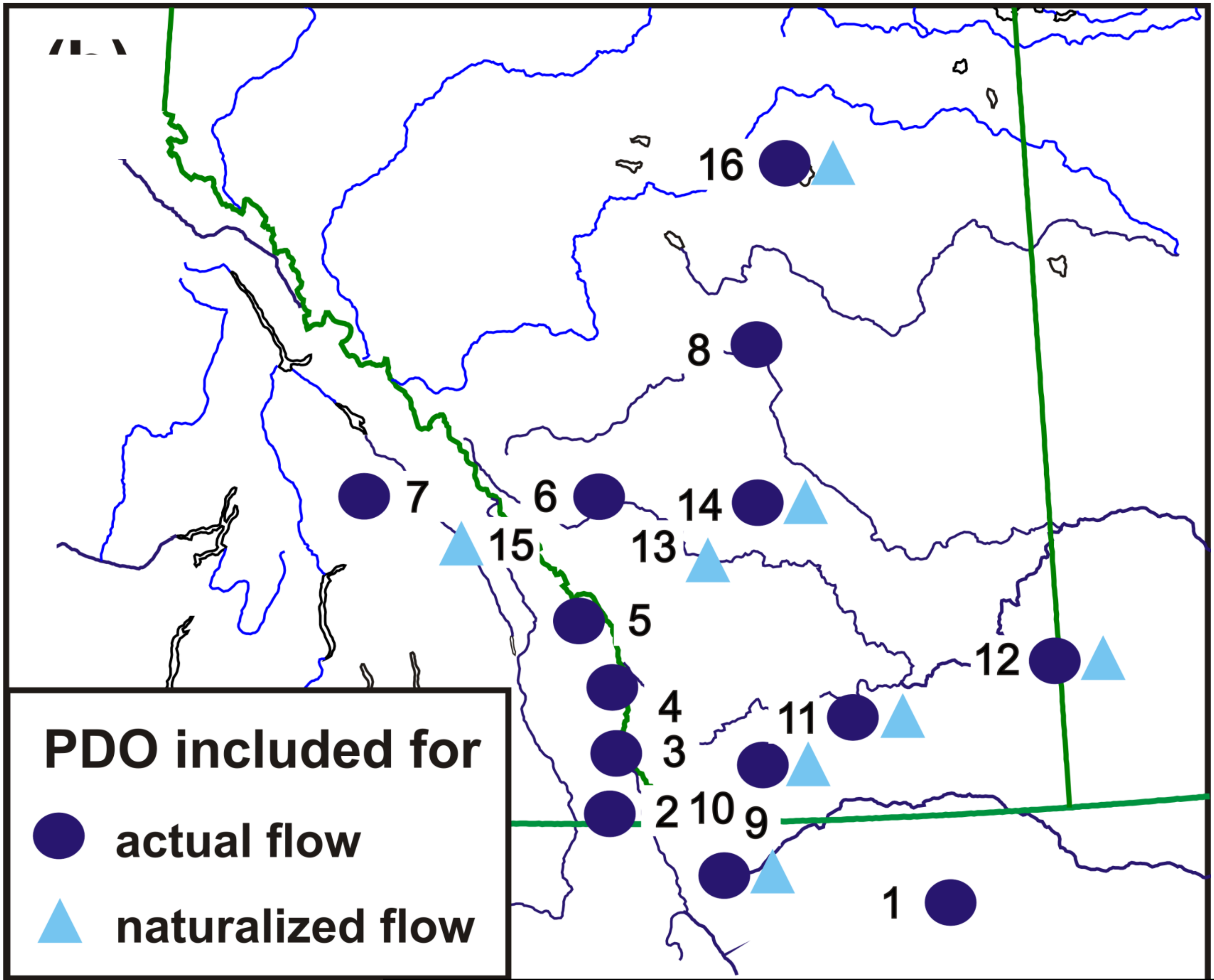
Human impacts ≥ global warming (AGW) effects

Geographical pattern: Bow River Valley worst?



St. Jacques *et al*. (2010) *Geophysical Research Letters*

PDO in optimum predictor subset in all but 2 records:



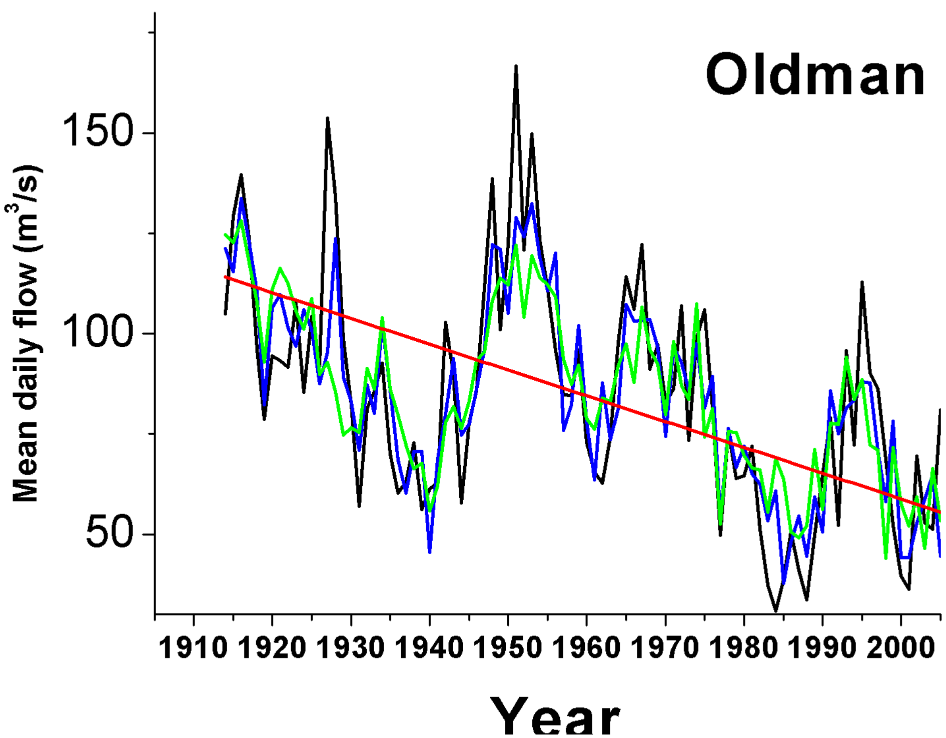
St. Jacques *et al.* (2010) *Geophysical Research Letters*

GLS regression equation projection

$$Oldman(Q_t) = 0.11 - 17.17 * trend - 9.25 * PDO - 9.52 * PDO_{P2} - 9.75 * SOI_{P2}$$

$$+ ARMA(2,3) \text{ error term } \varepsilon_t$$

$R^2_{(regular)} = 0.62$
$R^2_{(innovations)} = 0.73$

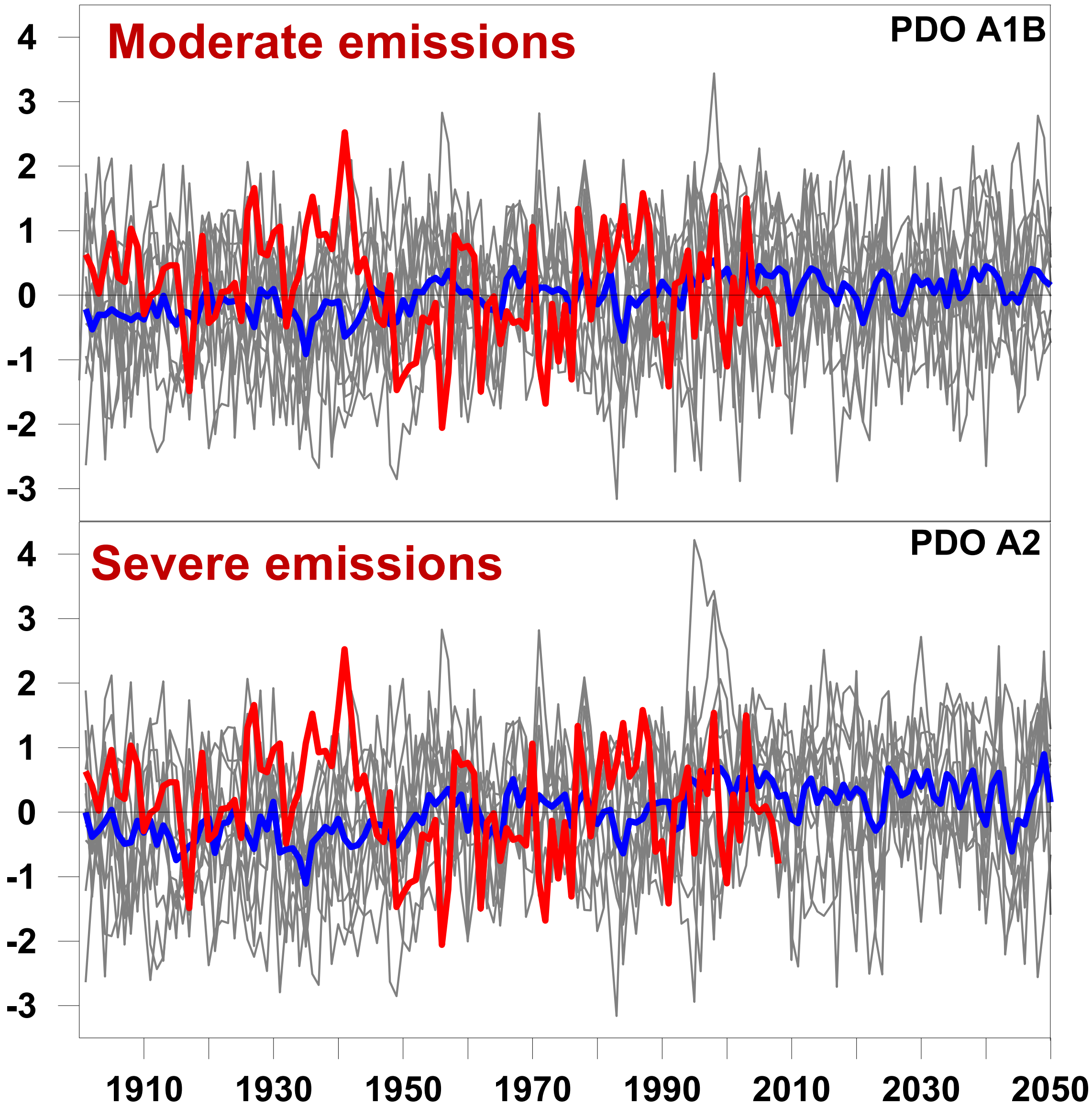


Idea: use archived GCM data project PDO, SOI, and NAO.

If have projected PDO, SOI and NAO, can project out streamflow regression equation ~45 yrs.

Black line : observed streamflow
Red line: trend
Blue line: fitted GLS model with error term
Green line: fitted GLS model without error term

PDO projections: 2010-2050



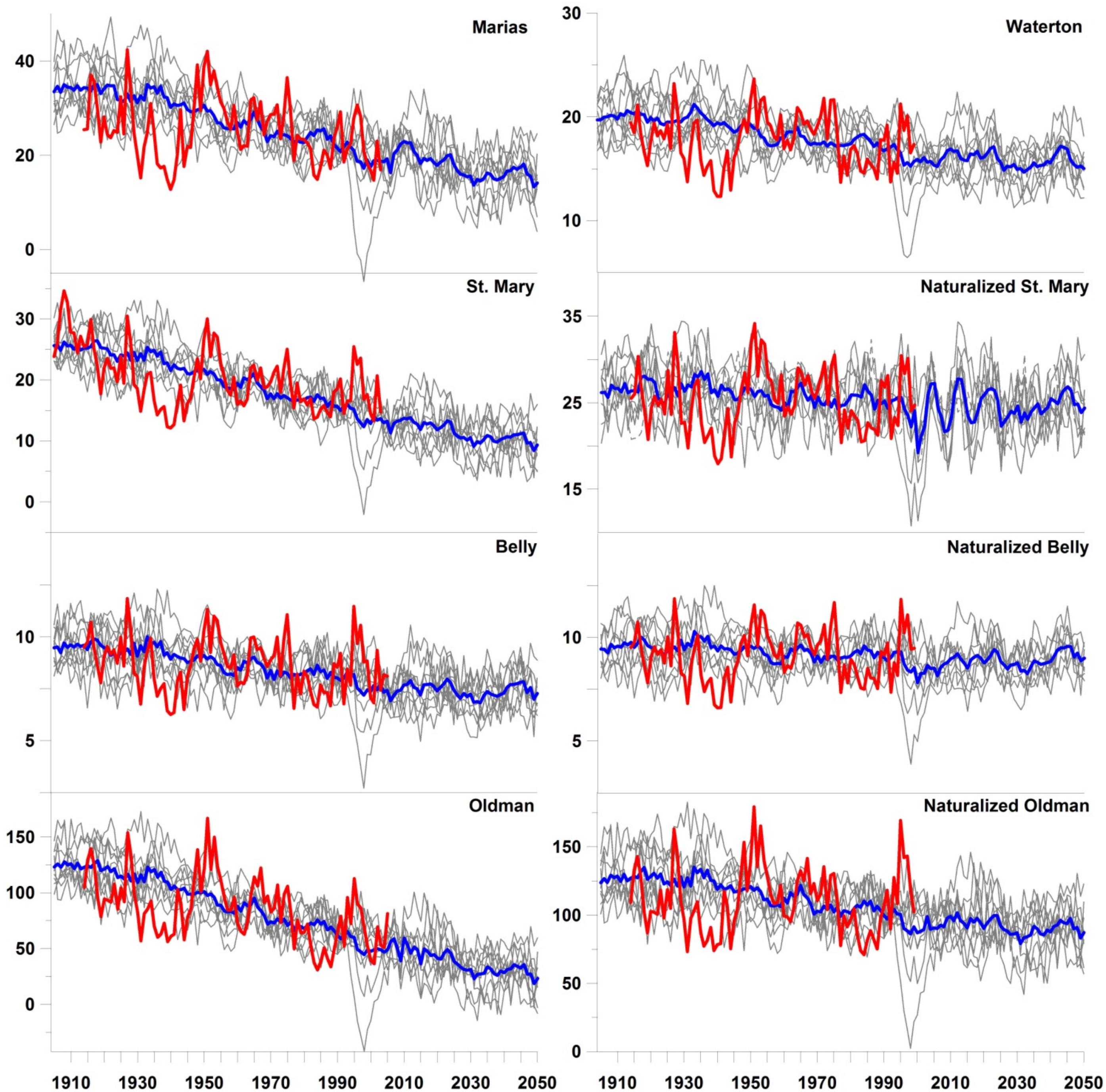
All-model means show shift towards more positive PDO-like conditions.

Also have SOI and NAO projections.

Lapp *et al.* (*in prep. a*)
International J. of Climatology

Red line: observed PDO
Grey lines: individual GCM runs PDO
Blue line: all-model mean PDO

Southern Alberta streamflow projections



Idea: using the best 8 streamflow GLS equations (*R^2* > 0.64) project for 2010-2050

A2 emissions scenario: 6 of 8 all-model means show declines, no increases.

A1B same.

Lapp *et al*. (*in prep. b*)

Red line: observed streamflow
Grey lines: individual GCM runs
Blue line: all-model mean streamflow

Conclusions

- **GLS is very useful** for modeling certain types of streamflow data (*i.e*., daily mean flow), allowing correct computation of trend tests in presence of autocorrelated data.
- **PDO** has a large effect on Southern Alberta streamflow.
- There are **15 decreasing trends**, **7 no trends**, and **2 increasing trends** detected in the 24 S. Alberta streamflow records.
- Most streamflows are **declining** due to hydroclimatic changes (from global warming) and severe human impacts, which are of the same order of magnitude as the global warming changes, if not greater.
- Our GCM projections show a shift towards **more positive-phase PDO mean state**. GLS streamflow projections show mainly **declines** (**6 out of 8**) and no increases.

Alberta Environment

EPCOR

Manitoba Hydro

Thanks to Mike Seneka, Xiaogu Zheng, Chris Ray, Greg MacCulloch and our sponsors:

NSERC CRSNG

PARC

DRI
DROUGHT RESEARCH INITIATIVE
RÉSEAU DE RECHERCHE SUR LA SÉCHERESSE