Developing Regression Models and Control Charts for the Newfoundland Real Time Water Quality Monitoring Network

R. Harvey[1], L. Lye[1], A. Khan[2]

1. Faculty of Engineering and Applied Science, Memorial University
2. Water Resources Management Division, Department of Environment and Conservation

Overview

- The Need for New Research
- Regression Models:
  - The need for developing regression models
  - Methodology
  - Results
- Statistical Process Control
  - The potential for implementing control charts
  - Methodology
  - Results
- Other Research Areas
- Conclusion

The Need for New Research

- In this presentation we'll look at two areas of current research:
- 1. Can regression models be developed for:
  - Water temperature? Dissolved oxygen?
- 2. Can statistical process control tools traditionally used in the manufacturing industry be used for RTWQ data?
- 3. Are there easier ways of analyzing and presenting real-time data?

Regression Models

For Predicting Water Temperature and Dissolved Oxygen

Why Model Water Temperature?

- Water temperature is an important indicator of water quality
- Regression models for water temperature have been extensively studied
- Models usually link air temperature to water temperature
- These models let researchers study:
    - The impact of global warming
    - The impact of nearby industry

Why Model Dissolved Oxygen?

- Dissolved oxygen has a large influence on river health
  - Low levels (0 – 8 mg/L) – fish kills
  - Healthy levels (8 – 12 mg/L)
  - High levels (12 – 20 mg/L) – algal growth
- Dissolved oxygen largely depends on water temperature
- Wide range of models have been proposed in the literature

Methodology

- Step 1 — Get familiar with commonly used models
- Step 2 — Develop datasets
- Step 3 — Develop regression models
- Step 4 — Pick best overall model

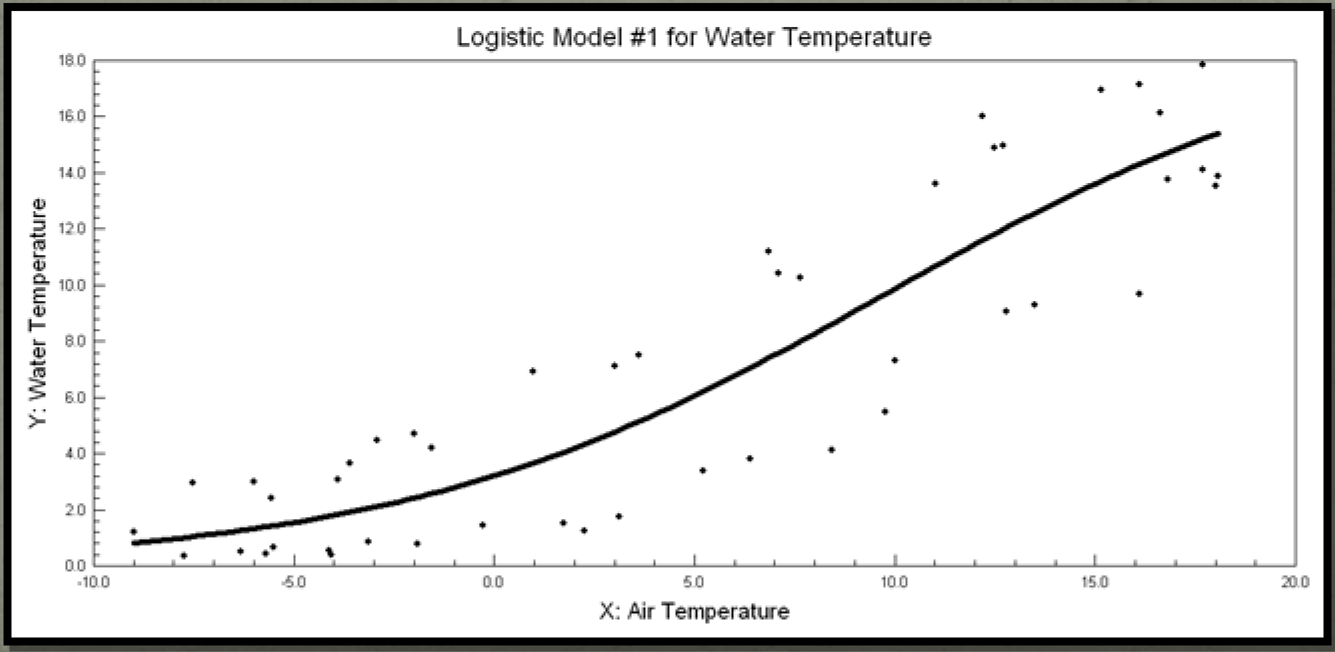
Step 1 – Get Familiar With Models

- Most models use mean monthly and weekly data
- Models developed with daily or hourly data tend to be unreliable
- Three popular options in the literature:
  - 1. Linear regression – using air temperature
  - 2. Multiple regression – using air temperature and streamflow
  - 3. Nonlinear logistic regression – using air temperature

Logistic Regression Model #1

$$Tw = \frac{\alpha}{1 + e^{\gamma(\beta - T_a)}}$$

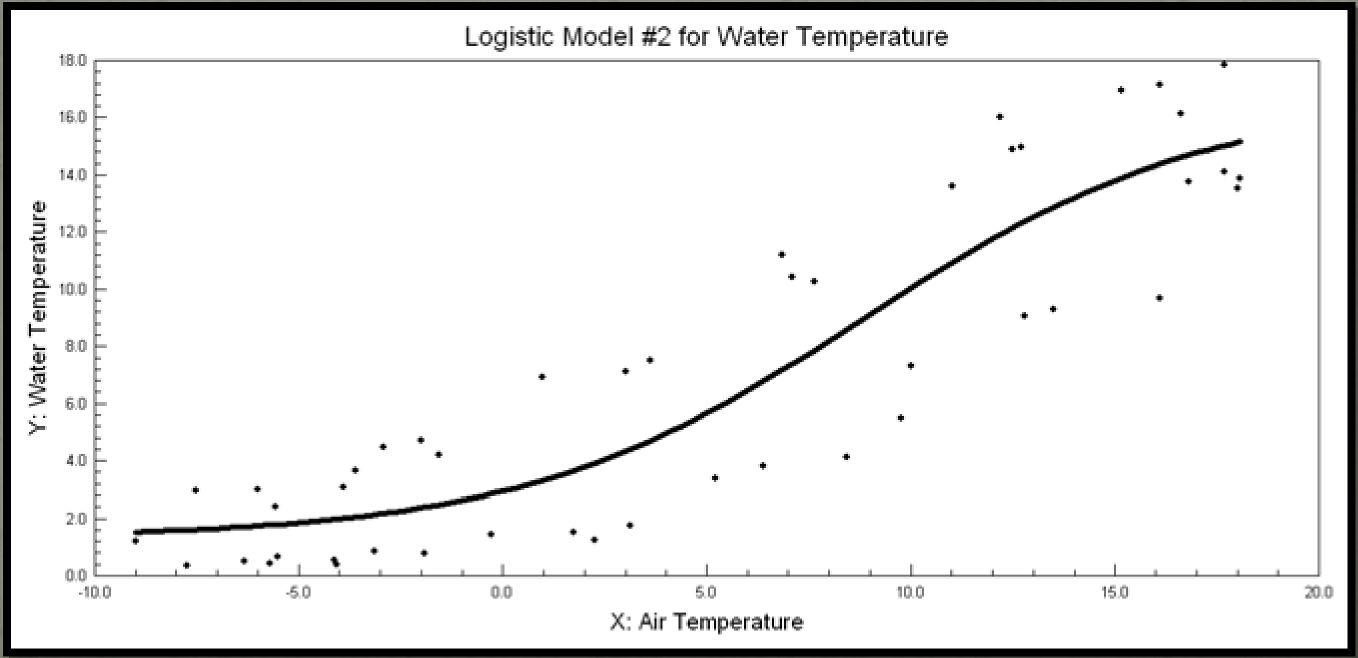
- α – max stream temperature
- γ – steepest slope
- β – air temperature at inflection



Logistic Regression Model #2

$$Tw = \mu + \frac{\alpha - \mu}{1 + e^{\gamma(\beta - T_a)}}$$

- μ – min water temperature



Step 2 – Develop Datasets

- There are a number of sites collecting water quality data
- In order to develop reliable models look for sites with at least 3 years of data
- Three stations selected for preliminary study
  - 1. NF02YL0012 – Humber River
  - 2. NF02YO0121 - Peter's River
  - 3. NF02ZM0178 - Leary's Brook

Available Data

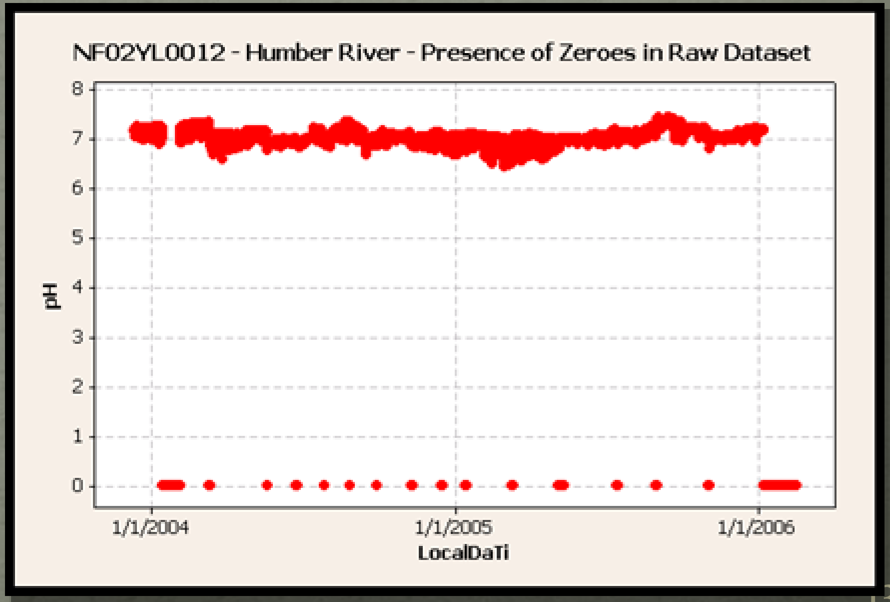
- Historical records available in Excel files
- Water temperature and dissolved oxygen data comes from the sensors
- Streamflow from nearby Environment Canada hydrometric stations
- Air temperature from nearby Environment Canada weather stations
  - Corner Brook for Humber River (<20 km)
  - Badger for Peter's River (<50 km)
  - St. John's for Leary's Brook (same city)

Working With the Historical Records

- Obtaining datasets isn't entirely straightforward as a considerable amount of time is needed for pre and post processing

- Historical records in Excel need to be examined to remove periods of time where the sensor is not working properly - i.e. periods of zero pH

- Remove these periods before any weekly or monthly means are used for regression work



Working With the Historical Records

- Once you have the periods of inaccurate sensor readings removed you're left with a historical record consisting of over 25,000 rows of data

- Manually going through these rows to copy and paste data for the weekly and monthly mean calculations is excessively tedious

- Minitab macros can be written to speed things up by automating the process

Using a Macro to Find Mean Values

NF02YL0012 ***

↓	C1-D	C2	C3	C4	C5
	Date	Water Temperature	pH	Specific Conductance	Dissolved Oxygen
1	12/11/2003 12:50:11 PM	4.9000	7.18920	34.3988	11.0219
2	12/11/2003 1:50:11 PM	4.9001	7.18841	34.3976	11.0738
3	12/11/2003 2:50:11 PM	4.9001	7.18761	34.4964	11.1157
4	12/11/2003 3:50:11 PM	4.9001	7.19681	34.4952	11.0976
5	12/11/2003 4:50:11 PM	4.8002	7.19602	34.4940	11.1495

```
MTB > %CSCEWeekly
Executing from file: C:\Program Files\MINITAB 14\MACROS\CSCEWeekly.MAC
macro is used to find the weekly mean values
Author - Richard Harvey, Memorial University, 2009
c1 contains the date
c2 contains the water temperature
c3 contains the pH
c4 contains the specific conductance
c5 contins the dissolved oxygen
set desired year and weeks (i.e. 2004 1 52)
DATA> 2005 1 52
```

C15	C16	C17	C18	C19	C20
Year	Week	Mean WT	Mean pH	Mean SC	Mean DO
2005	1	2.61565	6.39051	38.1489	14.5635
2005	2	2.13769	6.38834	37.5559	14.8976
2005	3	1.47725	6.50816	38.3904	14.6246
2005	4	0.76453	7.05151	40.4258	13.4390
2005	5	0.30151	6.97476	40.9558	13.74917

- Load Excel historical records into Minitab worksheet
- Write and run a macro that will search through the historical records to find the means you're looking for
- Data ready for analysis

Overview of Datasets

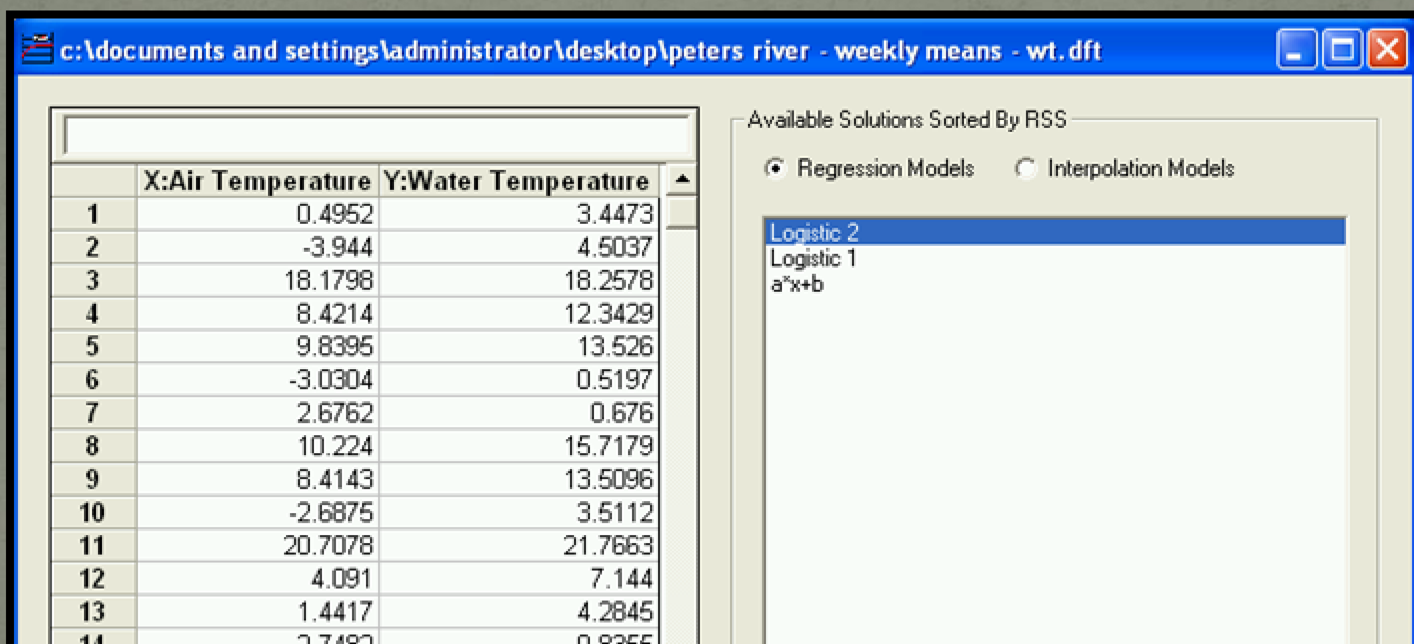
Station Name		Number of Obs.	Mean Water Temperature (°C)	Mean Dissolved Oxygen (mg/L)	Mean Streamflow (m^3/s)	Mean Air Temperature (°C)
Humber River	Weeks Months	194 47	6.85 6.70	12.07 12.09	249.05 248.70	4.67 4.73
Peter's River	Weeks Months	117 30	9.06 9.24	10.83 10.77	4.17 4.08	5.10 5.17
Leary's Brook	Weeks Months	80 24	7.94 8.13	11.33 11.38	1.16 1.12	6.09 6.32

- Randomize data to remove any autocorrelation between observations
- At this point we now have the data so its possible to go ahead and start developing the models

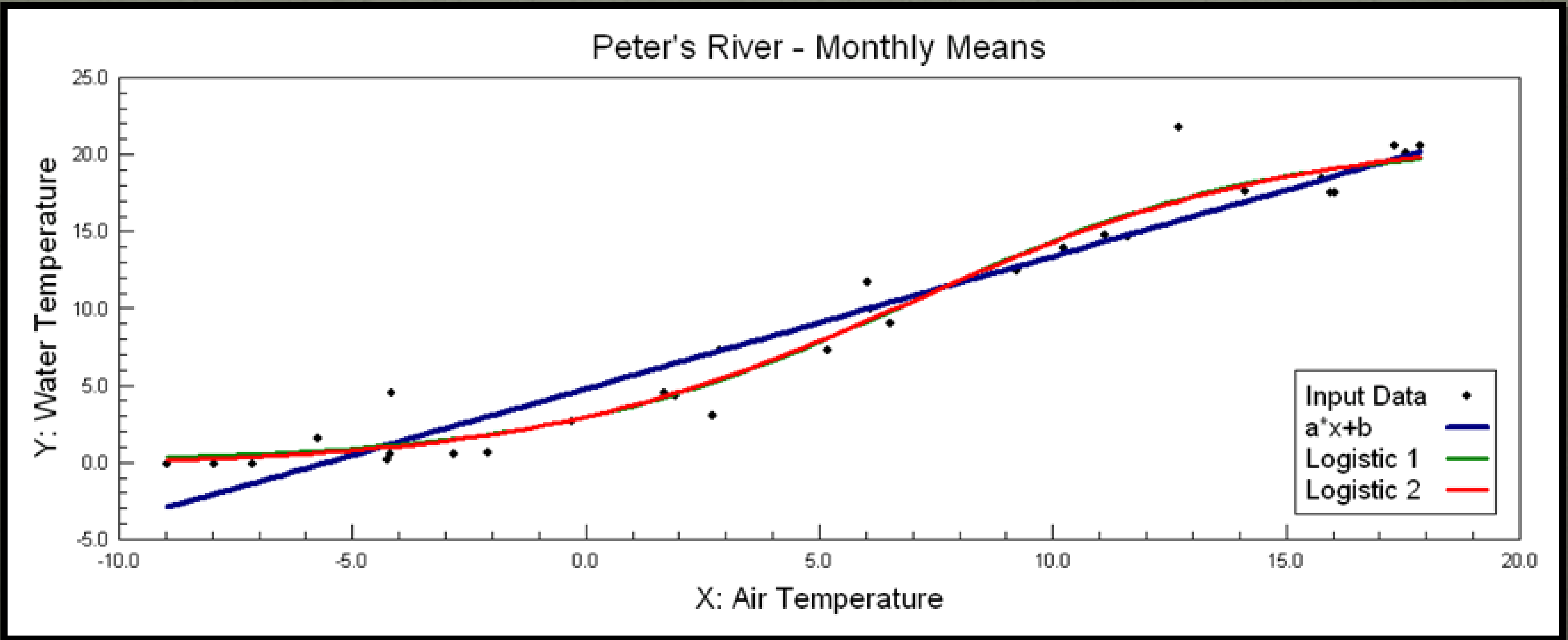
Step 3 – Develop Regression Models

- Can use Minitab or Datafit

- Datafit is useful in that you can enter the models you want to solve or solve a variety of models all at once



Sample Monthly Mean Models



Linear

$$Tw = 0.86Ta + 4.79$$

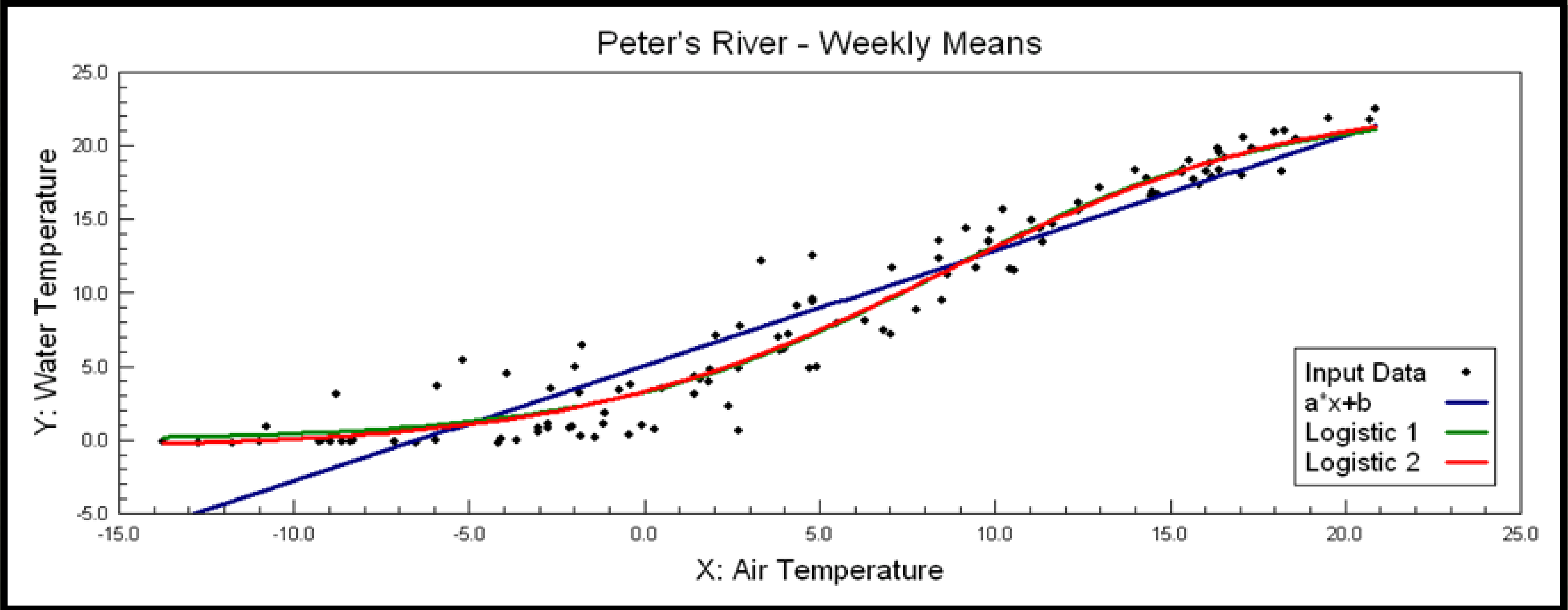
$$R^2 adjusted = 0.93$$

Logistic #1

$$Tw = \frac{20.92}{\left(1 + \exp\left(0.26 * \left(6.97 - Ta\right)\right)\right)}$$

$$R^2 adjusted = 0.96$$

Sample Weekly Mean Models



Linear

$$Tw = 0.78Ta + 5.06$$

$$R^2 adjusted = 0.90$$

Logistic #1

$$Tw = \frac{22.63}{(1 + \exp(0.21 * (8.42 - Ta)))}$$

$$R^2 adjusted = 0.95$$

Step 4 – Pick the Best Overall Model

- Visual inspection - looks like the data follows S-shape
- Take a look at the adjusted R-squared values as well:

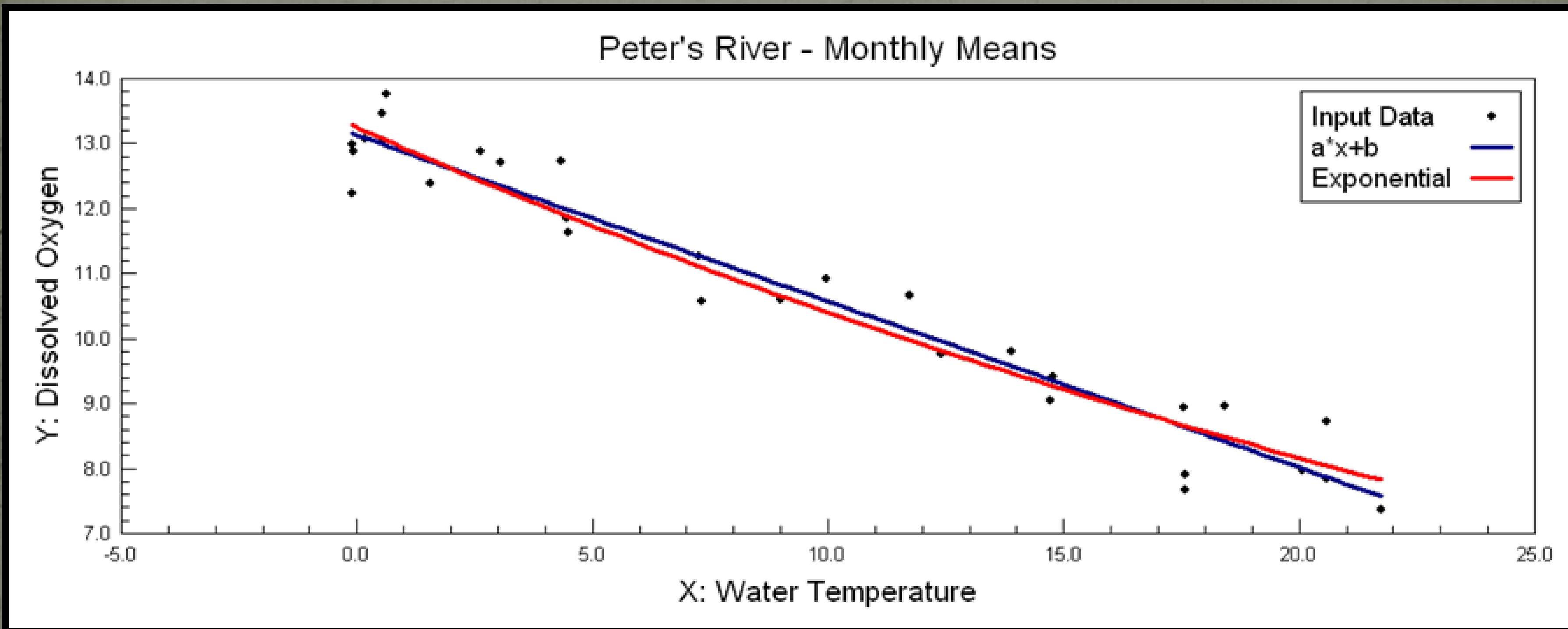
Station name		Number of Data Points	Linear	Logistic 1	Logistic 2	Multiple Regression
Humber River	Weekly Monthly	194 47	0.726 0.771	0.764 0.794	0.765 0.792	0.759 0.801
Leary's Brook	Weekly Monthly	117 30	0.946 0.937	0.953 0.950	0.954 0.948	0.947 Not significant
Peter's River	Weekly Monthly	80 24	0.901 0.931	0.946 0.958	0.946 0.957	0.911 0.945

- Looks like Logistic #1 is the best choice here

Modeling Dissolved Oxygen

- So things worked out pretty well for water temperature
- Can models be developed for dissolved oxygen in the same way?
- Looking through some of the literature there are three popular options:
  - 1. Linear regression – using water temperature
  - 2. Multiple regression – using water temperature and streamflow
  - 3. Nonlinear exponential regression – using water temperature
- Can use the same monthly and weekly datasets as before

Sample Monthly Mean Models



Linear

$$DO = -0.26Tw + 13.13$$

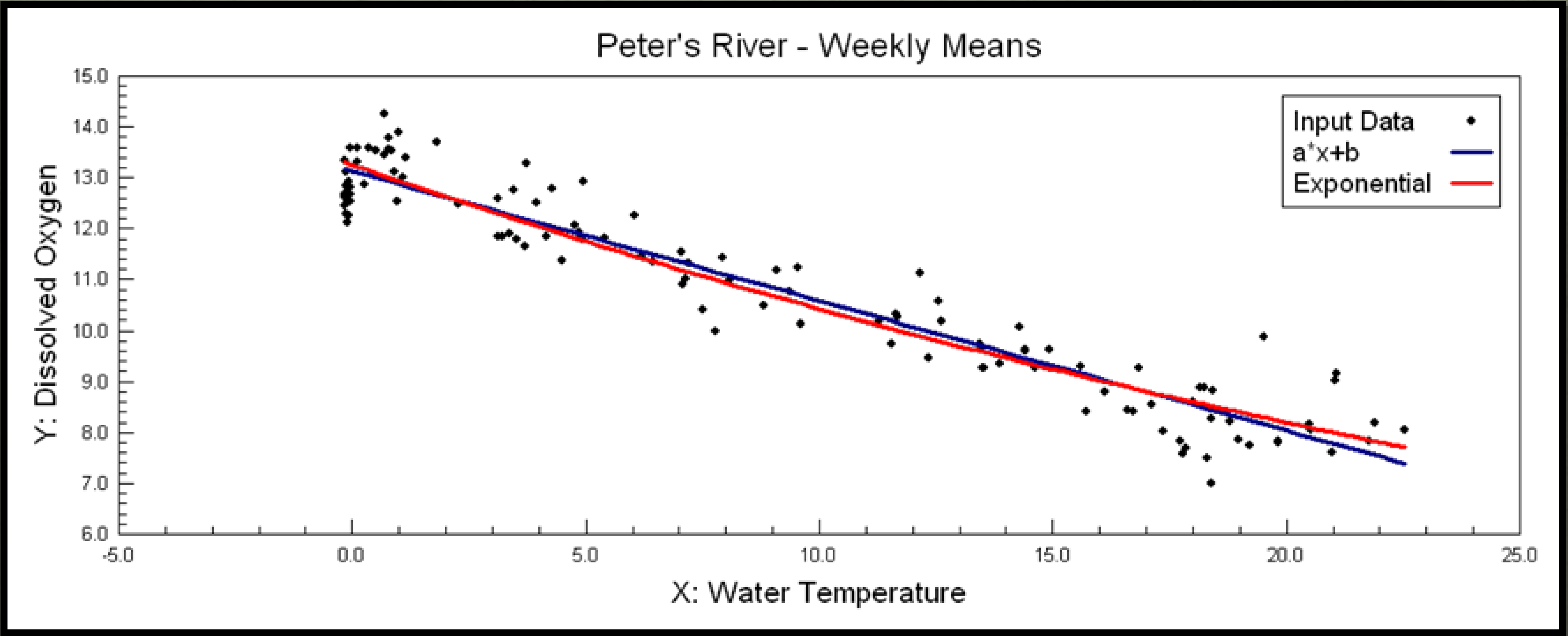
$$R^2 adjusted = 0.94$$

Exponential

$$DO = \exp(2.58 - 0.02Tw)$$

$$R^2 adjusted = 0.94$$

Sample Weekly Mean Models



Linear

$$DO = -0.25Tw + 13.13$$

$$R^2 adjusted = 0.92$$

Exponential

$$DO = \exp(2.58 - 0.02Tw)$$

$$R^2 adjusted = 0.91$$

Picking the Best Overall Model

- Use visual inspection and review adjusted R-squared values:

Station name		Number of Data Points	Linear	Multiple Regression	Exponential Regression
Humber River	Weekly Monthly	194 47	0.686 0.677	Streamflow not significant	0.698 0.698
Leary's Brook	Weekly Monthly	117 30	0.790 0.830	0.867 0.891	0.758 0.801
Peter's River	Weekly Monthly	80 24	0.911 0.941	Streamflow not significant	0.911 0.940

- Both linear and exponential models work well

- Streamflow for Humber River and Peter's River is not significant

Interpretation of Results

- The relationship between air temperature and water temperature for these three stations is not linear but is in fact S-shaped
- The relationship between water temperature and dissolved oxygen can be described by either a linear or exponential model
- Streamflow is not always a significant explanatory variable for dissolved oxygen when looking at monthly and weekly mean values
- For some stations the relationship is not always perfect
- Does size of drainage area matter? Any regional behaviour?

Comparing Regression Models

Seeing how the different stations are related

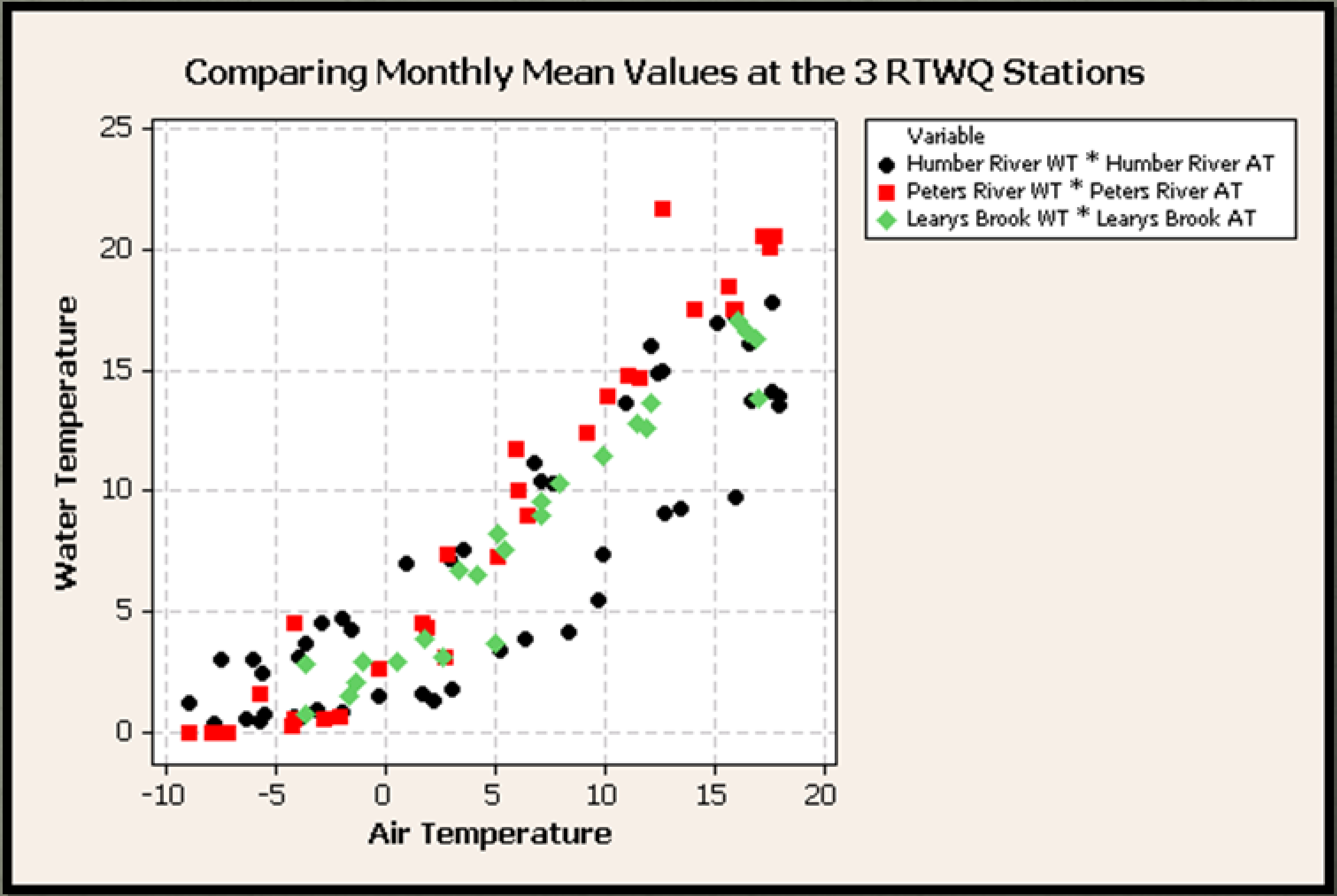
Overview

- Regression equations developed for WT and DO using the three different stations
- Is it possible that the relationships are the same at each site (i.e. one general relationship that applies for all regions on the island)?
- Or is each site unique and there is no way to relate them to each other?

Water Temperature Models

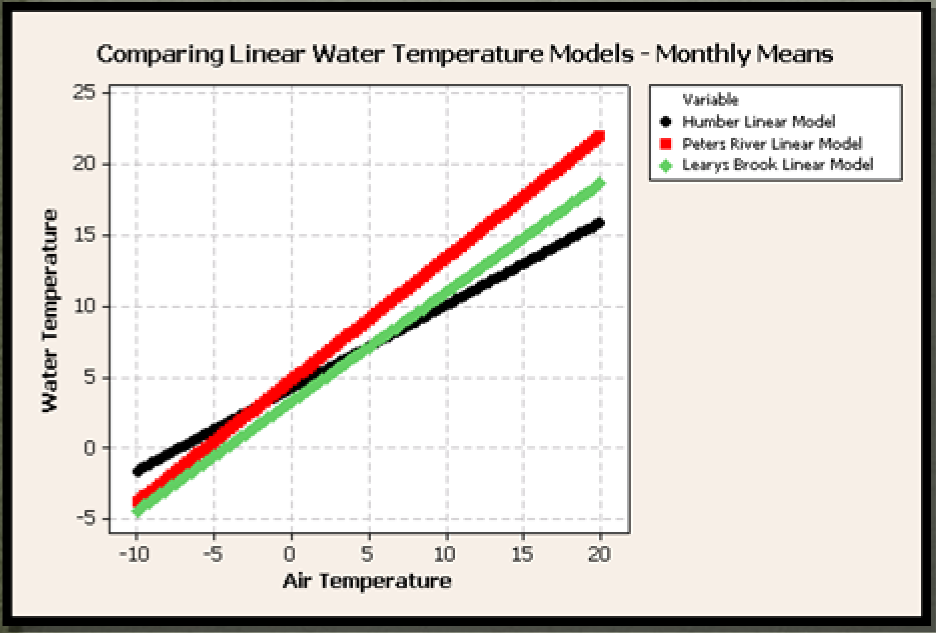
How do the models developed for each site compare to each other?

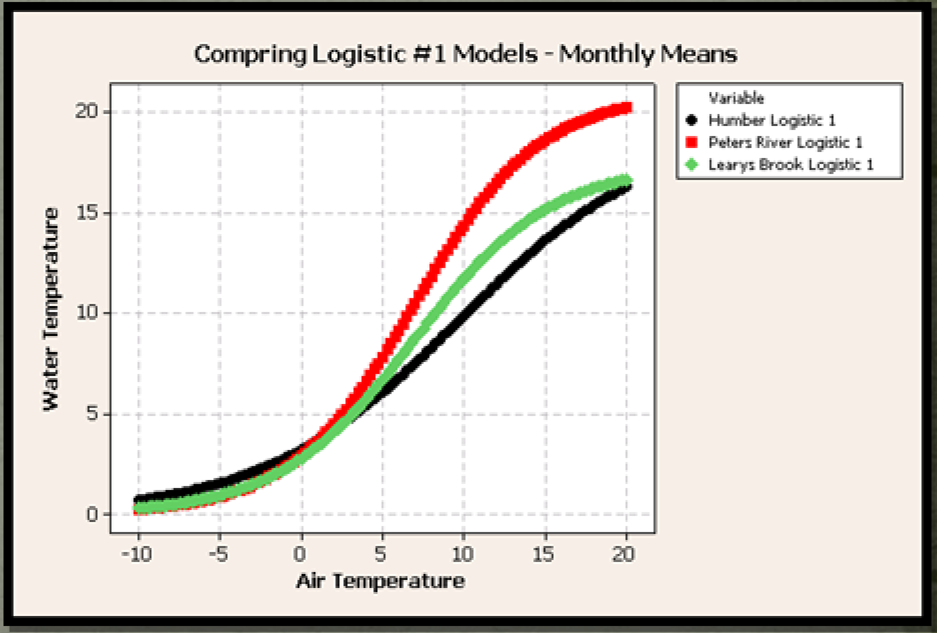
Looking at Monthly Mean WT Models



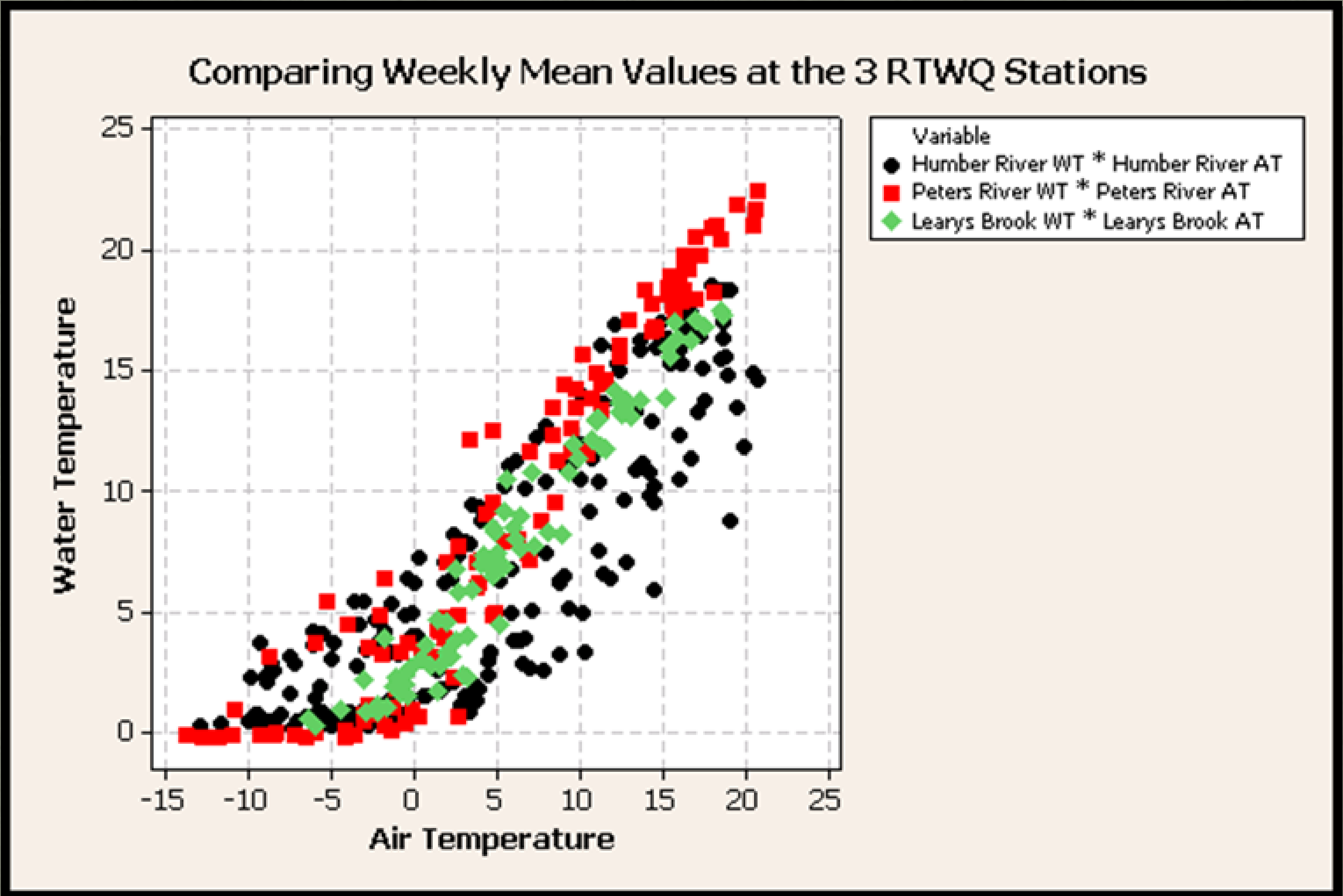
Looking at Monthly Mean WT Models

- Just define a set of air temperatures in Minitab (x-axis)
- Use the linear models and plot on the same graph
- Use logistic #1 models and plot on the same graph



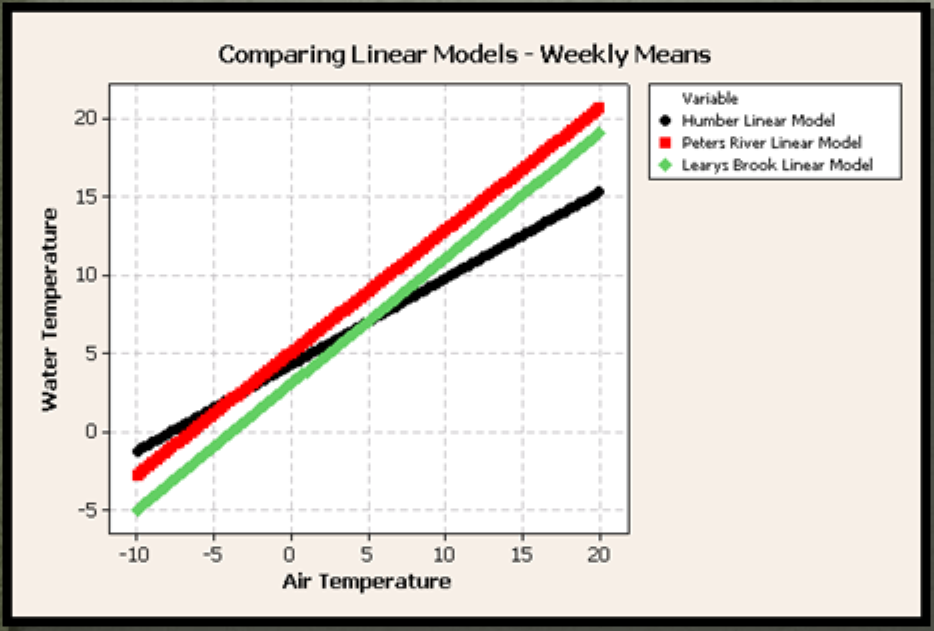


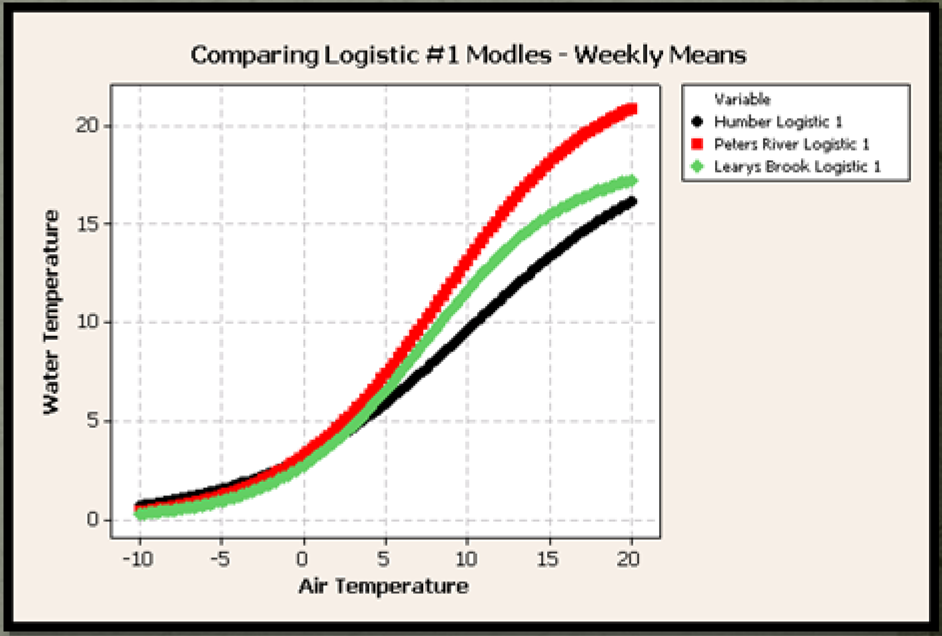
Looking at Weekly Mean WT Models



Looking at Weekly Mean WT Models

- Just define a set of air temperatures in Minitab (x-axis)
- Use the linear models and plot on the same graph

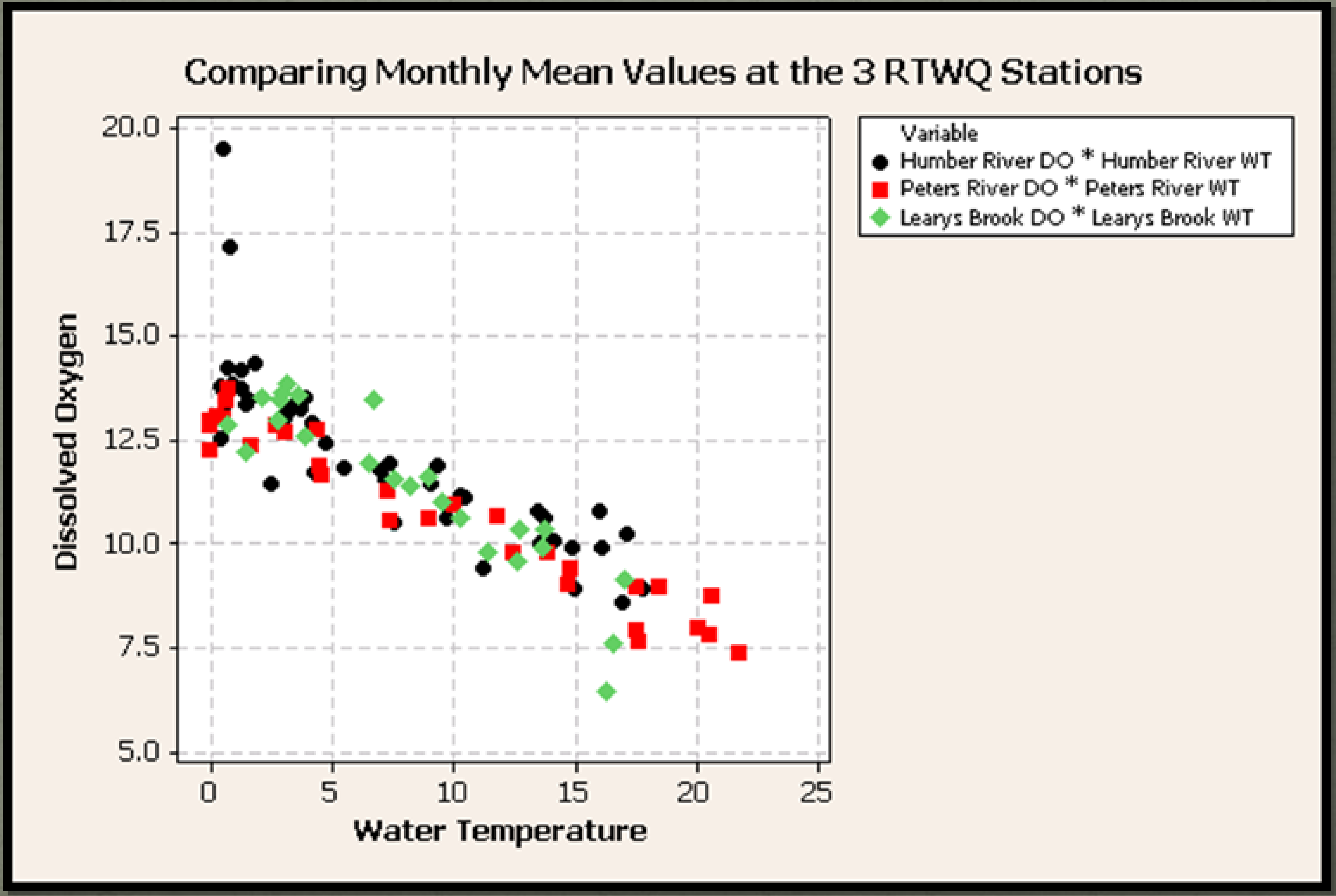




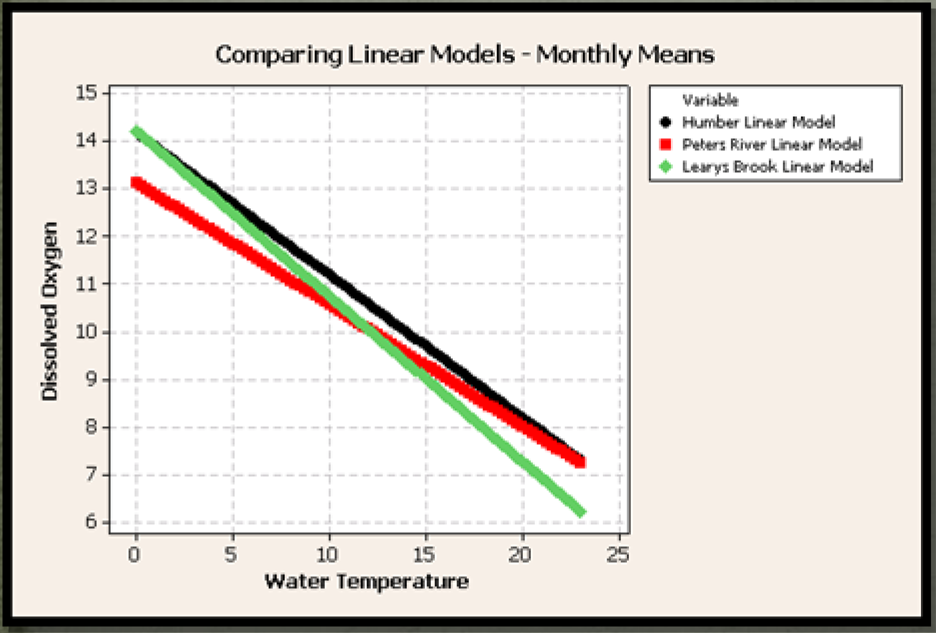
Dissolved Oxygen Models

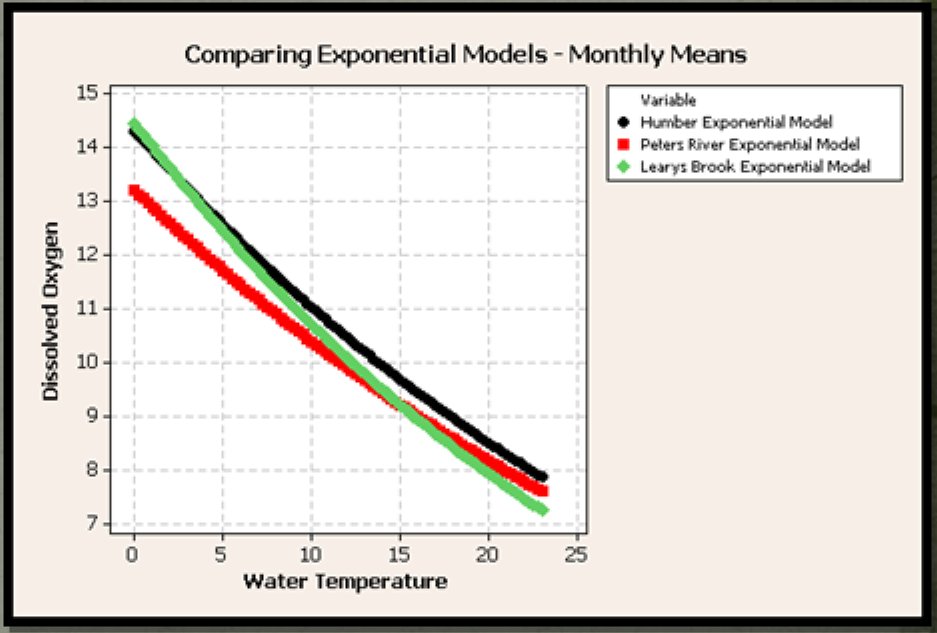
How do the models developed for each site compare to each other?

Looking at Monthly Mean DO Models

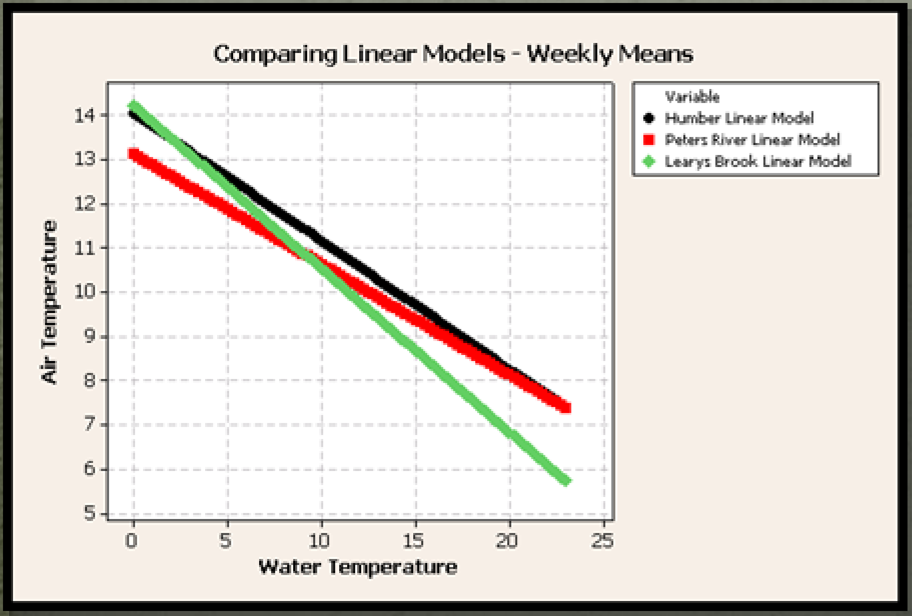


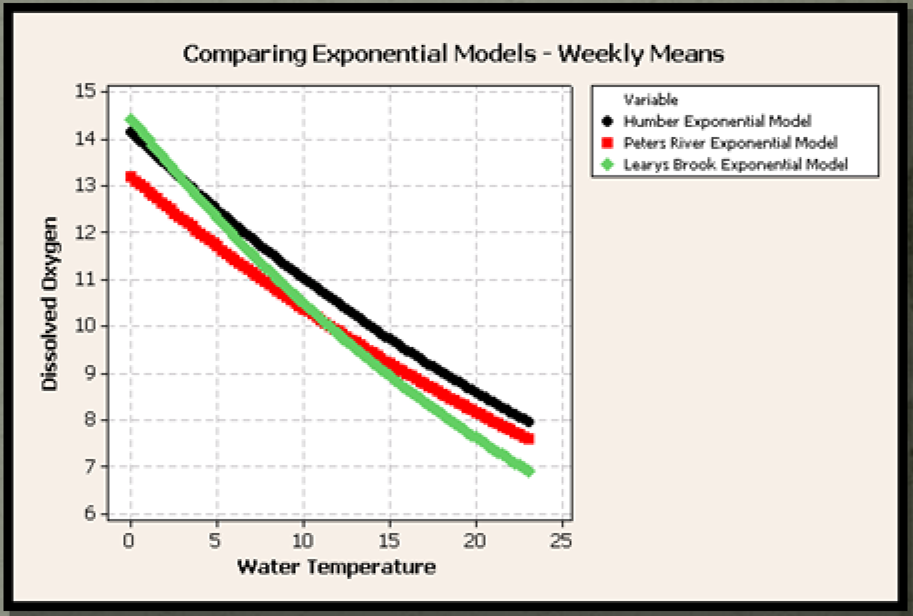
Looking at Monthly Mean DO Models





Looking at Weekly Mean DO Models





Making A 3-Way Plot

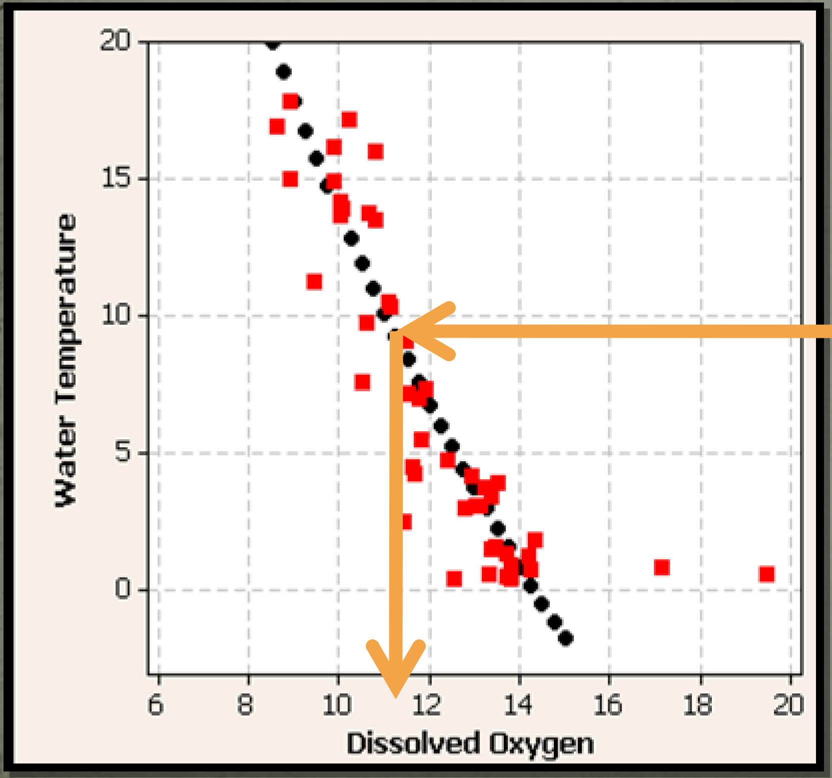
One That Shows AT, WT and DO

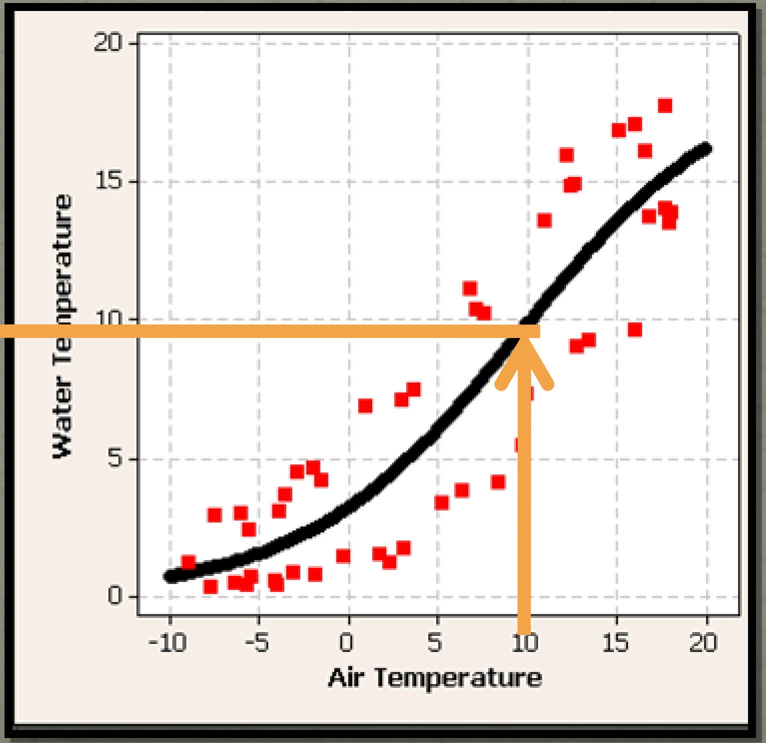
Overview

- It would be great if a plot could be developed that would let the user start with an air temperature, draw a line up to the water temperature, then draw a line from water temperature to dissolved oxygen

Using Humber River – Monthly Means

- It would be better if the WT and DO plot could be mirrored.





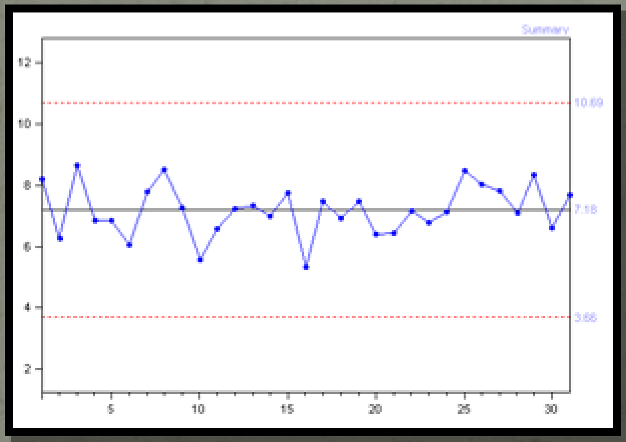
Control Charts

For Statistical Process Control of RTWQ Data

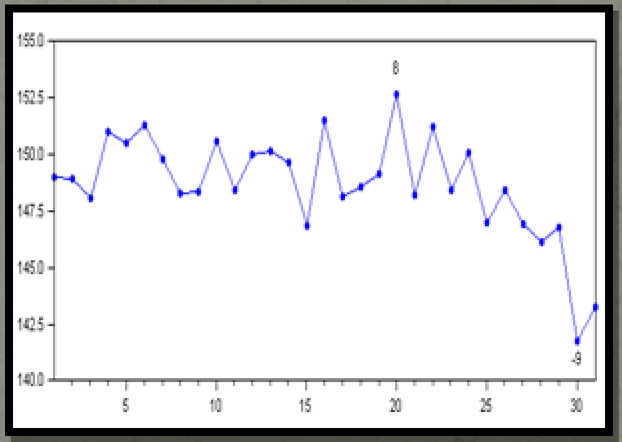
What is a Statistical Process Control Chart?

- An effective method of using graphical plots known as control charts for monitoring a process over time
- Traditionally used for monitoring industrial processes - manufacturing
- Control charts test the hypothesis that there are only common causes of variability in a process versus the alternative that there are special causes of variability
- Variety of charts have been developed for SPC over the years
  - Shewhart chart
  - CUSUM chart
  - Many others and modified versions of the first two…

Types of Control Charts



Shewhart Chart



CUSUM Chart

- Control charts were developed for SPC of the manufacturing industry where it can be assumed that the data being studied is independent (i.e. no autocorrelation)
- Some authors have looked at using control charts for water quality data but control charts for RTWQ data has not been studied before

The Problem …

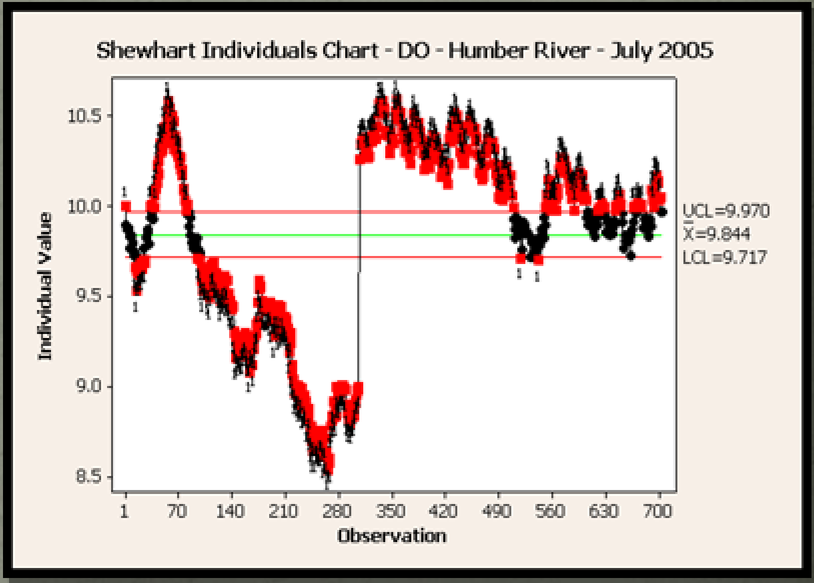
- We can rarely assume water quality data is not autocorrelated

- If we try and use statistical software like Minitab to develop control charts for the RTWQ data (without modifying the charts in some way) we end up with control charts that are useless

  - Develop control chart for historical DO

  - Control chart would tell us our process is out of control (red – Red – RED!)

  - We already know that something like DO will vary over time (i.e. it won't always be 10 mg/L)



- So is there a way to modify the control charts to handle our RTWQ data?

Control Charts for Correlated Data

- The problem of developing control charts for data that is highly correlated has been studied before (but not seasonal data):

- Different approaches:
  - 1. If correlation isn't too high just use the standard Shewhart Chart – but it may be necessary to expand the control limit lines which are traditionally set at 4 standard deviations from the center line

  - 2. If correlation is high, model the data using the sophisticated Box – Jenkins approach (ARIMA) then use control charts on the residuals

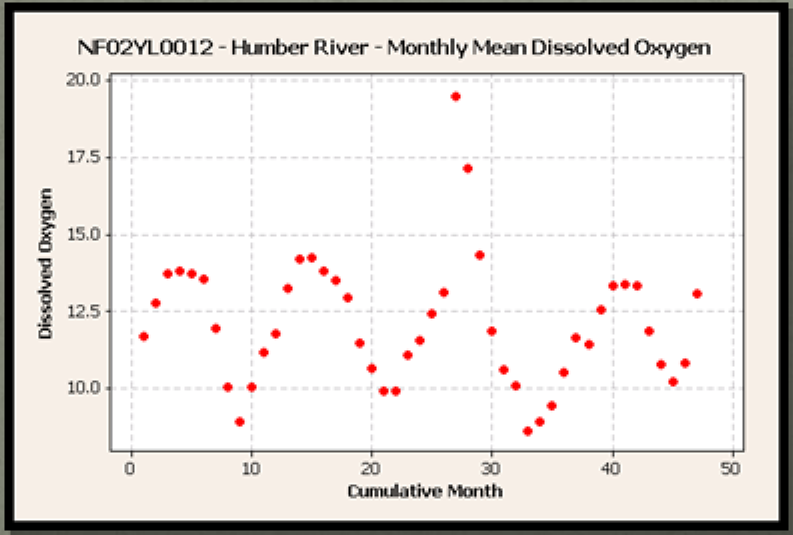
  - 3. Use Manly and MacKenzie CUSUM approach to compare measurements at different rivers. Data can be correlated in both space and time

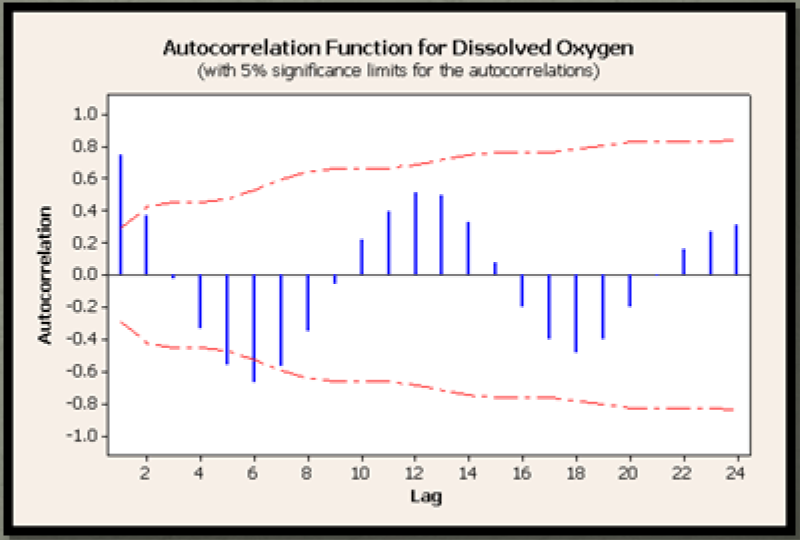
Approach 1

If Autocorrelation is Low – Use the Standard Control Chart

Shewhart Chart – Low Correlation

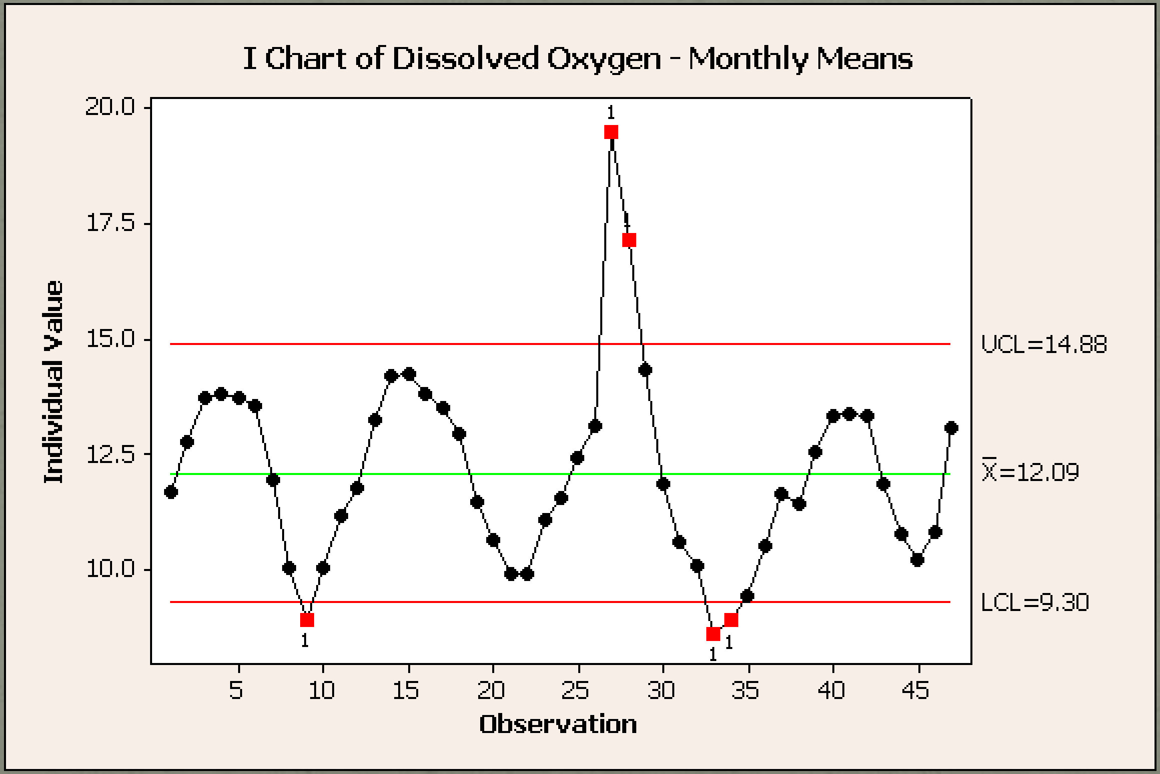
- If we are dealing with monthly mean dissolved oxygen data correlation between the data isn't overly high – but still seasonal
- NF02YL0012 – Humber River – 47 monthly observations of DO





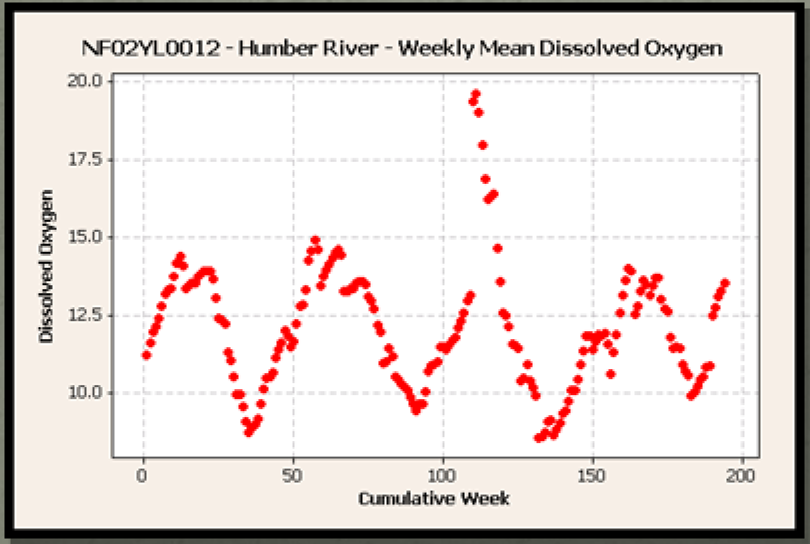
Shewhart Chart – Low Autocorrelation

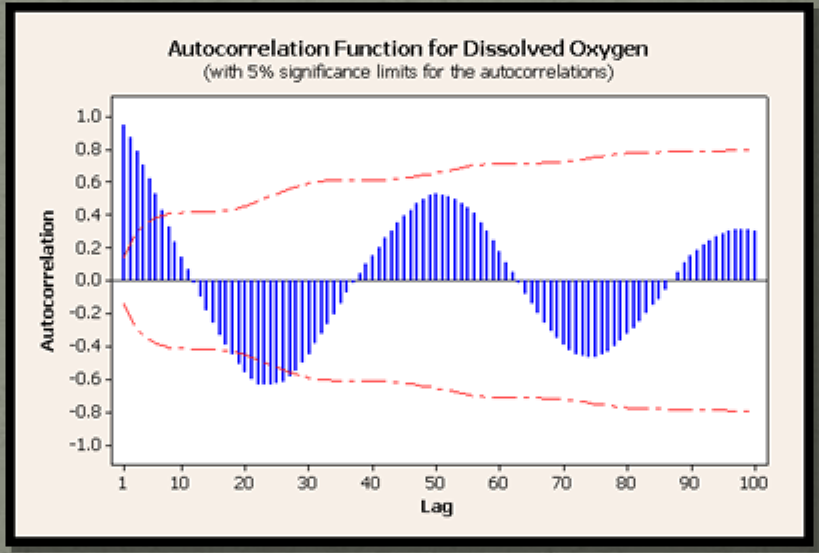
- Because of the lower autocorrelation we could try the standard Shewhart Individuals Chart to identify out of control points. But the limits are too narrow because of the autocorrelation and seasonality effects.



Shewhart Chart – High Correlation

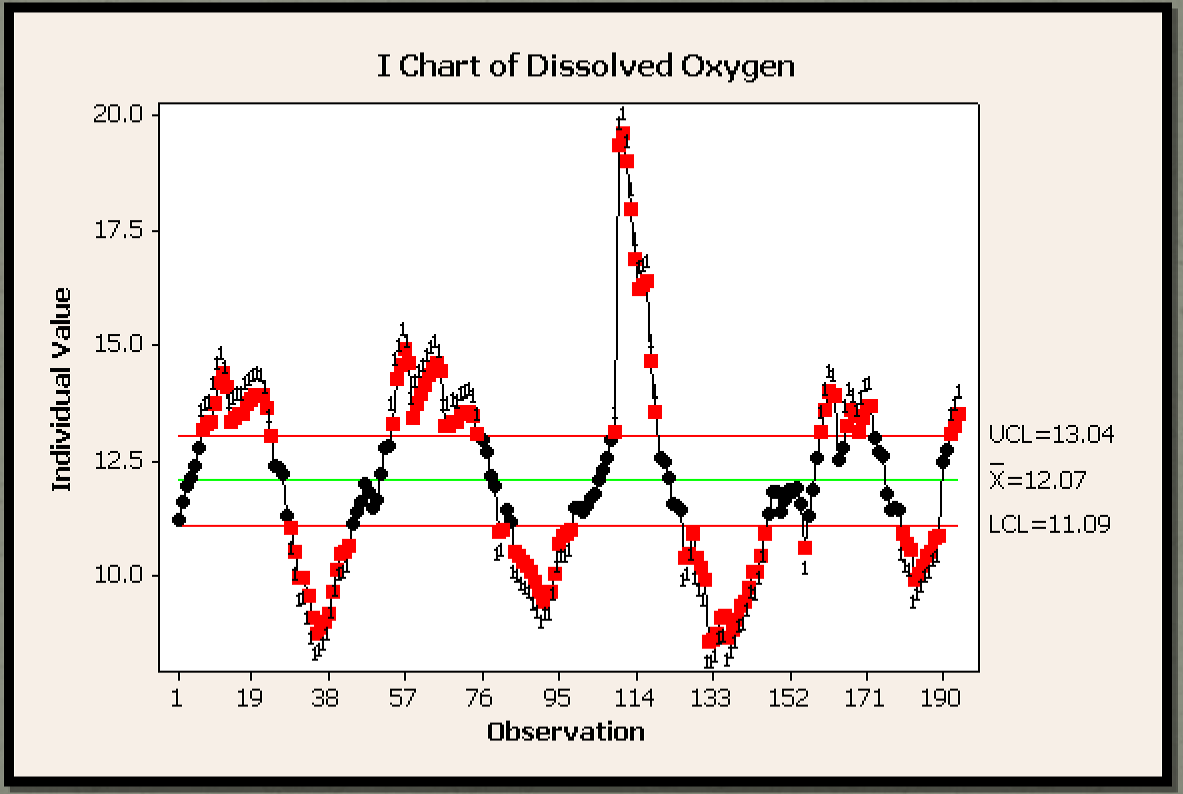
- If we take a look at weekly means there will be higher autocorrelation
- NF02YL0012 – 194 weekly observations of DO





Shewhart Chart – High Correlation

- Chart shows a lot of points as being out of control



Approach 2

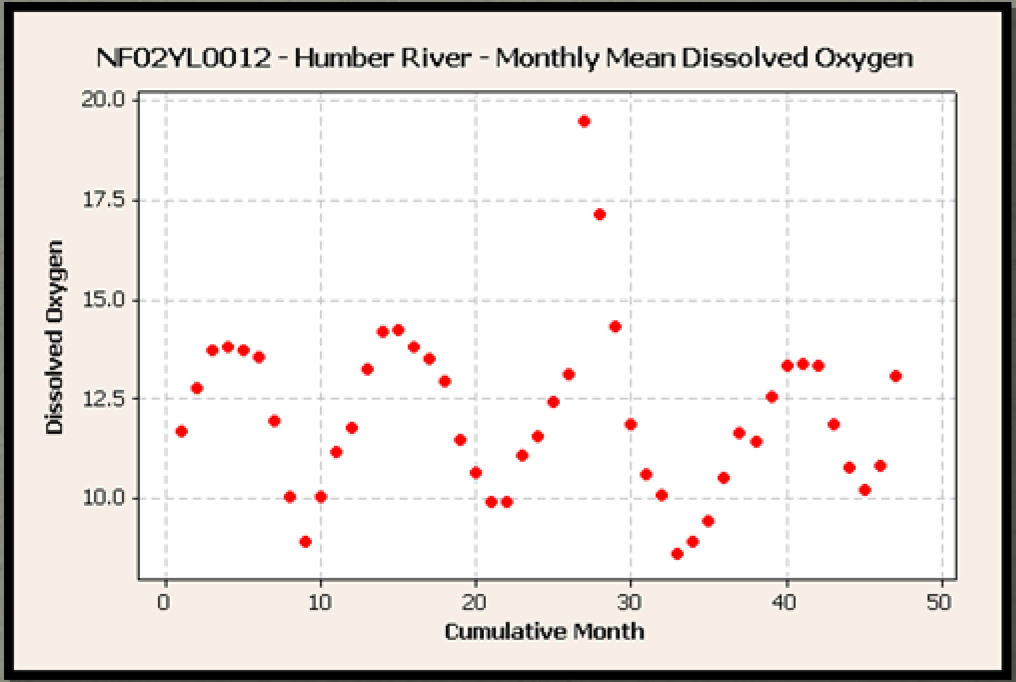
If Correlation is High – Model with ARIMA

Control Charts for ARIMA Residuals

- A number of authors have fit ARIMA models to water quality data
- Fitting to hourly and daily values is pretty difficult
- More potential for modeling monthly and weekly means
- Methodology
  - 1. Obtain dataset
  - 2. Fit ARIMA model to the data
  - 3. Use control charts to study the residuals
  - 4. use control chart findings to investigate original data

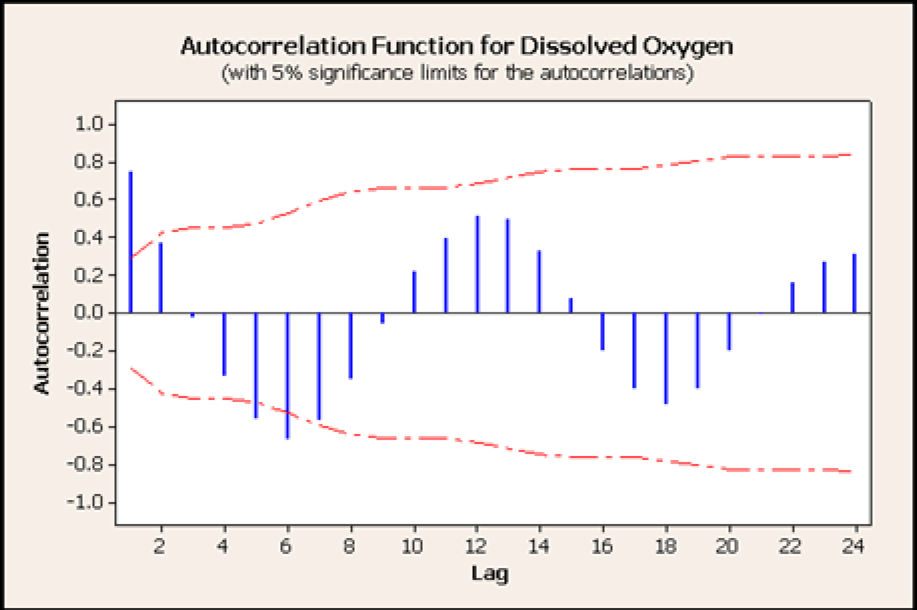
Monthly Mean Dissolved Oxygen Example

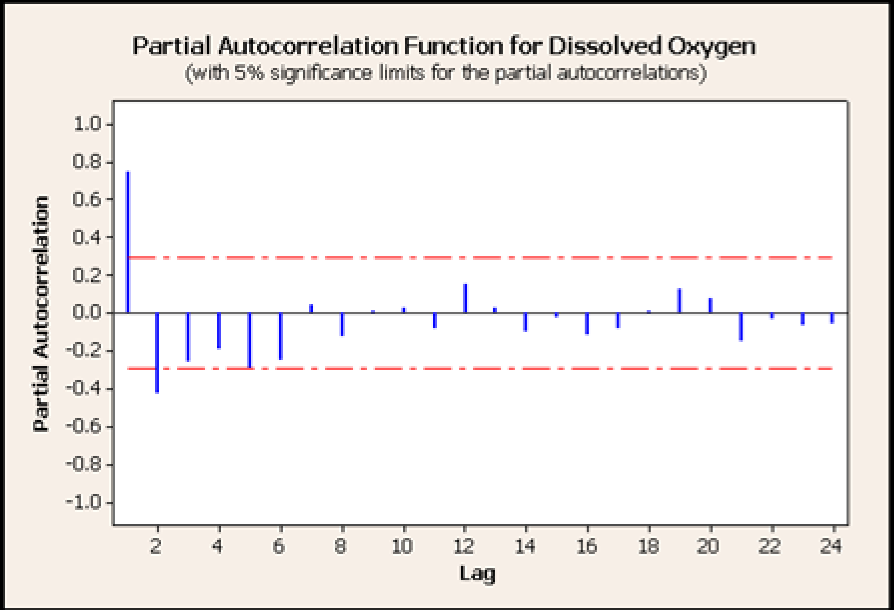
- If we come back again to the 47 monthly DO observations recorded at the Humber River station



- Appears to have seasonal variation – need to investigate that first …

Seasonal Variation in Dissolved Oxygen





- Autocorrelation plot shows DO follows a 12 month cycle at the site
- With these plots in mind we can look for the best ARIMA model

ARIMA Model for Mean Monthly DO

- Start with an AR(1) model with a Seasonal Component (1)
- Residuals aren't very normal but might not be a problem

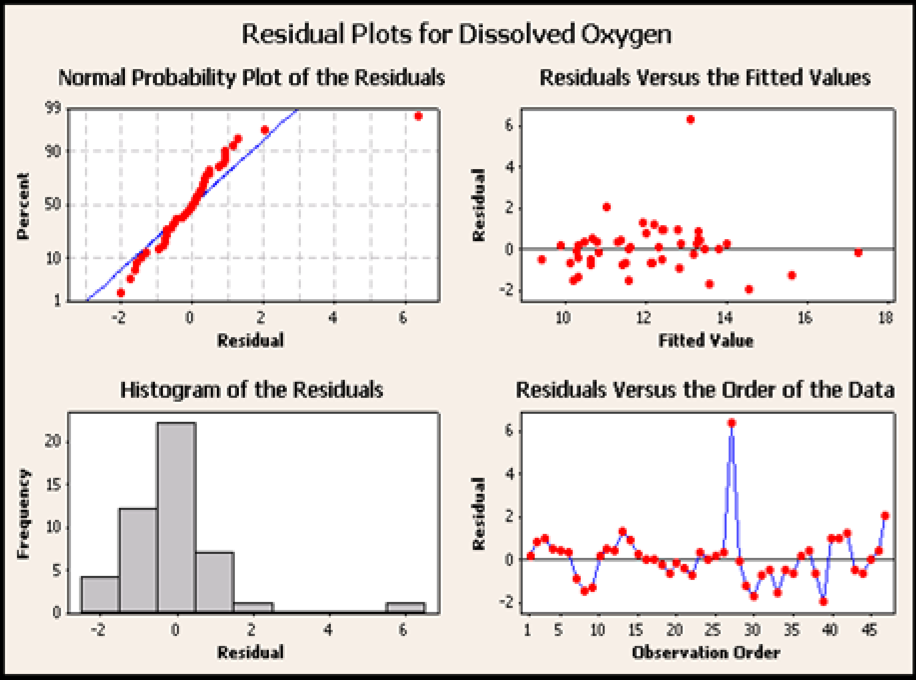
Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	0.6860	0.1144	6.00	0.000
SAR	12	0.4359	0.1438	3.03	0.004
Constant		2.1448	0.1891	11.34	0.000
Mean		12.108	1.067		

Number of observations: 47
Residuals: SS = 73.7307 (backforecasts excluded)
MS = 1.6757 DF = 44

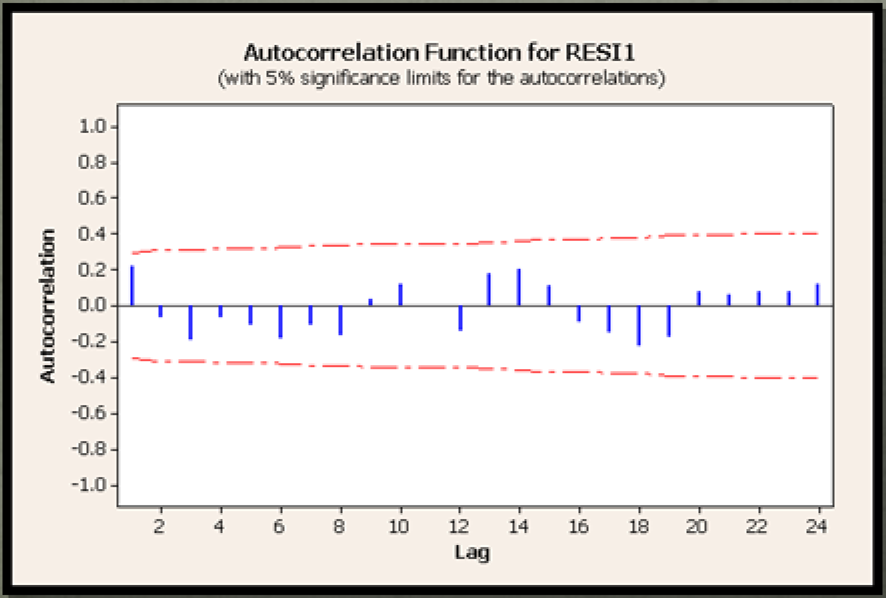
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

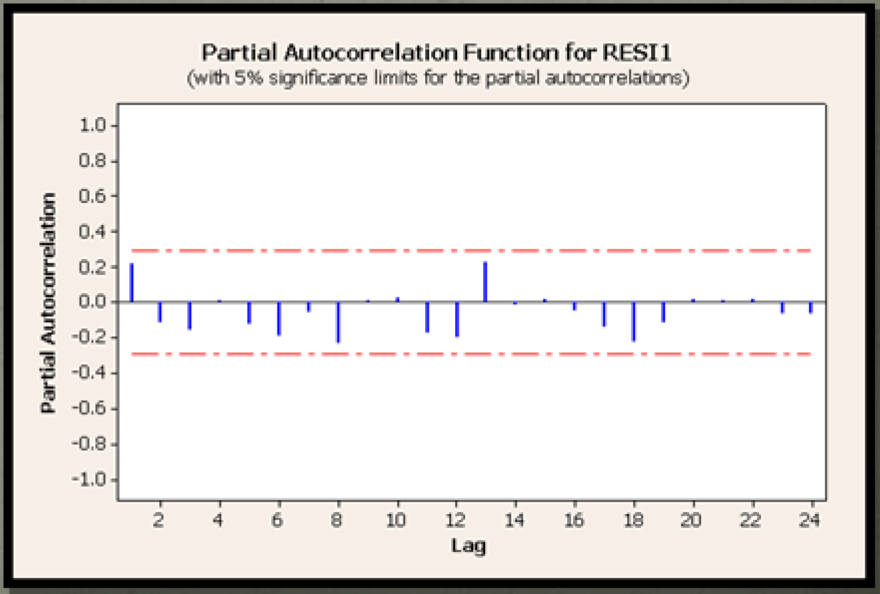
Lag	12	24	36	48
Chi-Square	11.8	29.6	30.7	*
DF	9	21	33	*
P-Value	0.223	0.100	0.580	*



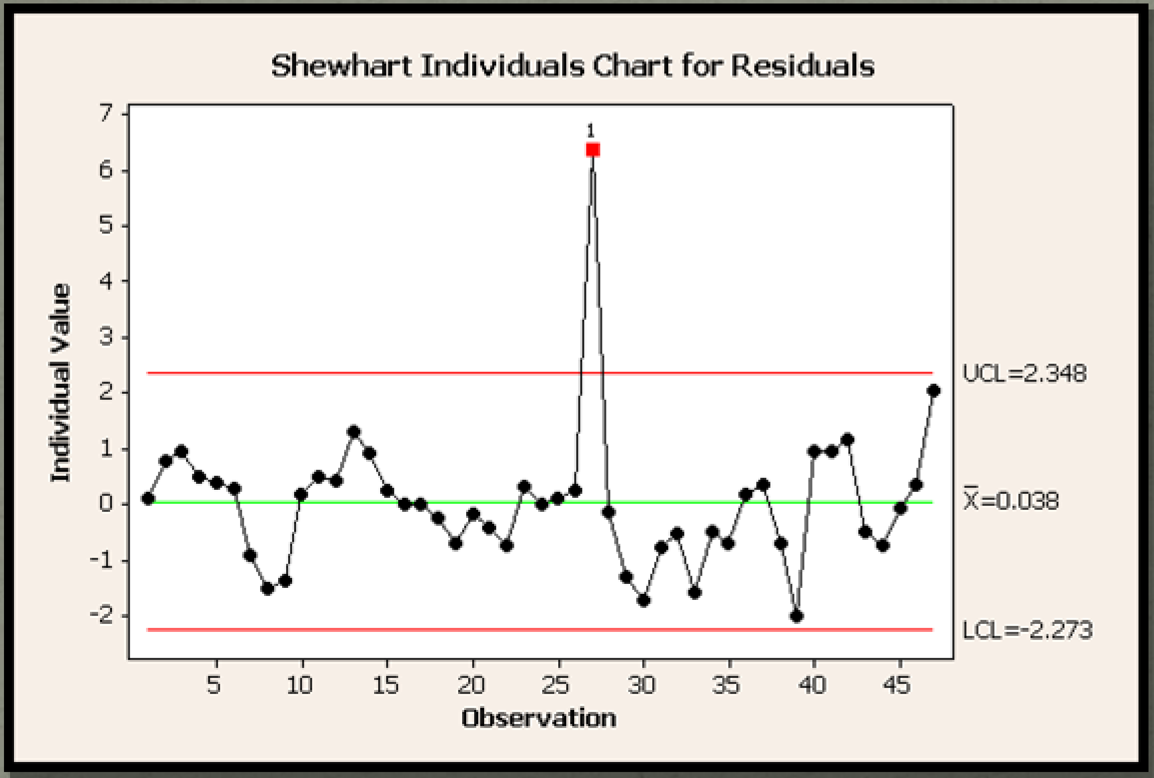
ARIMA Model for Mean Monthly DO

- Start with an AR(1) model with a Seasonal Component (1)
- No longer need to worry about autocorrelation – so its now possible to use a control chart to examine the residuals





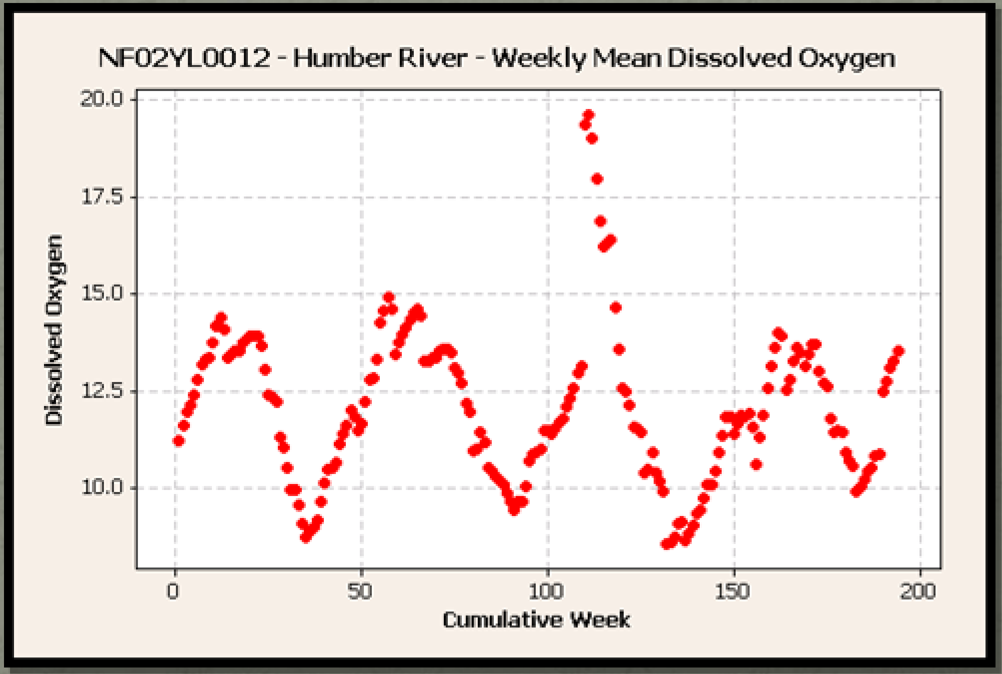
Shewhart Chart for Residuals



- Red point corresponds to out of control observation 27
- Dissolved Oxygen of 19.48 mg/L for February 2006 could be investigated further

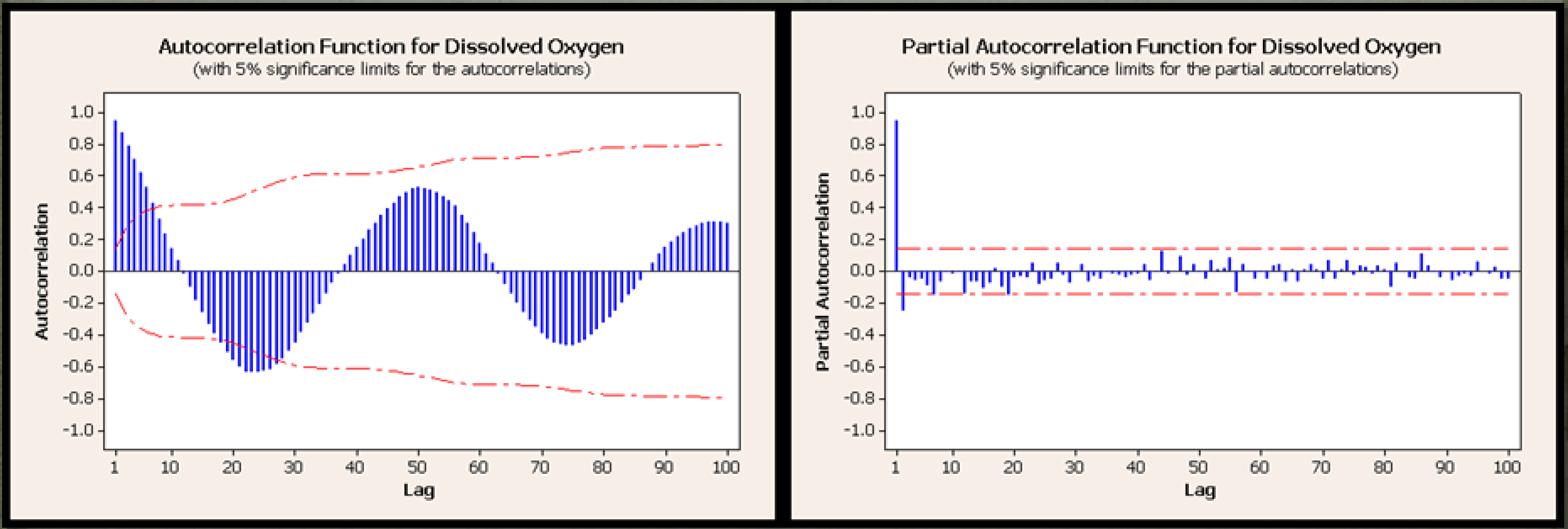
Weekly Mean Dissolved Oxygen Example

- What about the weekly mean DO set with higher autocorrelation



- Once gain there appears to be seasonal variation – need to investigate that first …

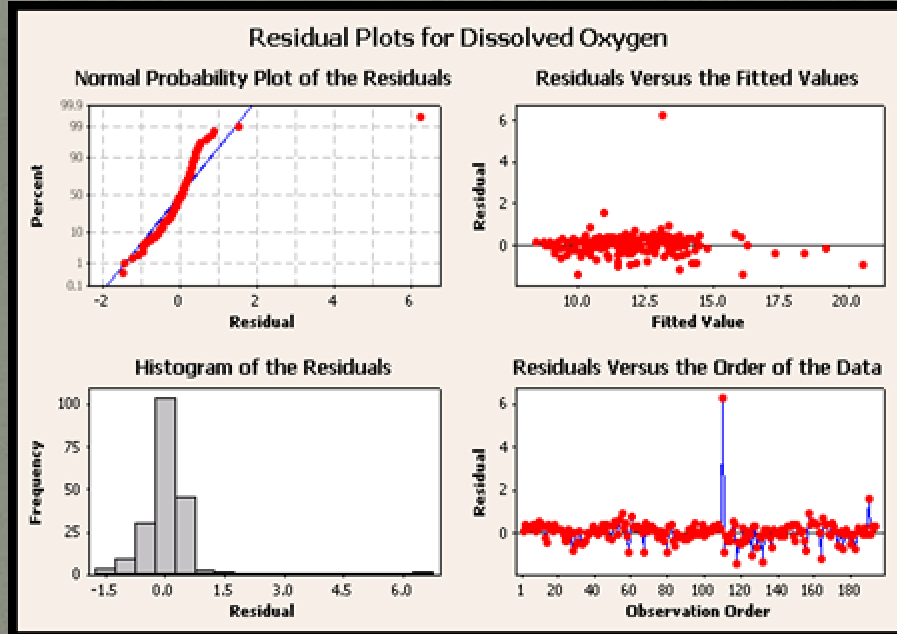
Seaonality in Weekly Mean DO



- May not have to introduce seasonality into the model

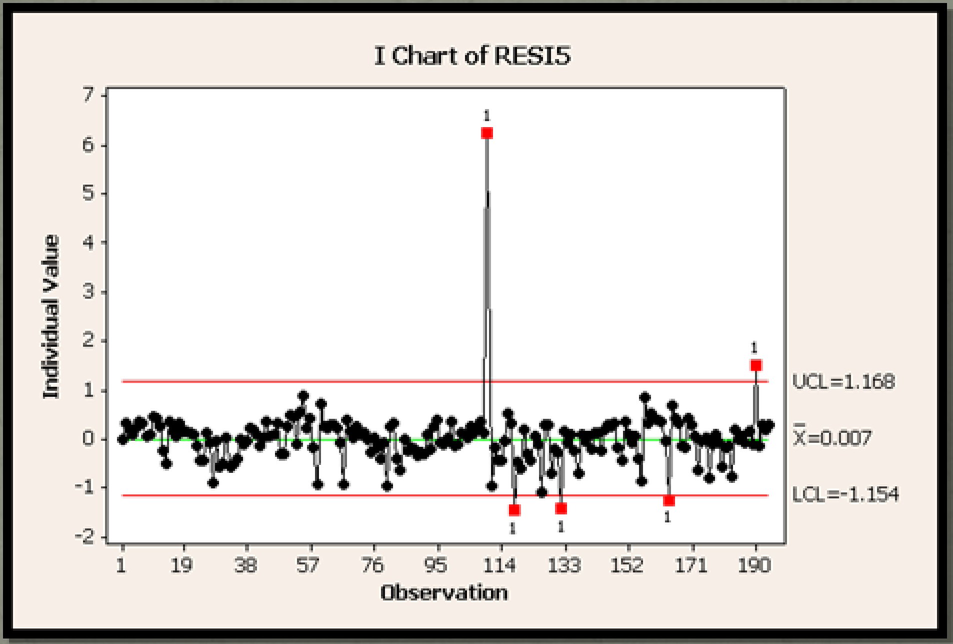
Fitting AR(2) Model to Weekly Mean DO

```
Final Estimates of Parameters

Type          Coef   SE Coef      T      P
AR   1      1.2014    0.0698  17.21  0.000
AR   2     -0.2660    0.0699  -3.81  0.000
Constant   0.78113   0.04400  17.75  0.000
Mean       12.0943    0.6813


Number of observations:  194
Residuals:    SS =  71.6906 (backforecasts excluded)
              MS =  0.3753  DF = 191


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            12      24      36      48
Chi-Square    7.1    26.3    36.6    45.9
DF              9      21      33      45
P-Value     0.624   0.196   0.305   0.436
```



- An AR(2) model without a seasonality term works well
- Now use control chart to examine the residuals

Shewhart Chart for AR(2) Residuals



- Observation 110 corresponds to dissolved oxygen over 19 mg/L in the 16th week of 2006
- All of the data before Observation 110 were in the 10 – 12 range (i.e. out of statistical control)

Approach 3

Use Manly and MacKenzie CUSUM Approach to Compare Data from Different Rivers in the Network

The Manly and MacKenzie Approach

- Manly and MacKenzie modified the CUSUM approach to detect changes in a monitored variable at more than one site
  - They compared mean January values of pH for 25 rivers in New Zealand from 1989 to 1996
  - Developed a piece of software known as the CUSUM Analysis Tool (CAT)
  - CAT v2.2 can handle data that is correlated in both space and time

- We can use the modified Manly and MacKenzie approach and the CAT software to take a look at the RTWQ Data.

Manly and MacKenzie for RTWQ Data

- For illustration purposes consider the monthly mean DO values for September for a number of sites in the network. Enter this data in CAT

Worksheet 1 ***

↓	C1-T	C2	C3	C4	C5	C6	C7	C8
	Site	Month	2003	2004	2005	2006	2007	2008
1	Humber River at Humber Village Bridge - NF02YL0012	9	*	9.6253	9.7609	8.2244	9.6246	7.1426
2	Peters River - NF02YO0121	9	8.4266	7.0945	9.5664	8.0345	9.1063	*
3	Waterford River - NF02ZM0009	9	*	*	8.2713	7.6796	7.5582	10.0820
4	Learys Brook - NF02ZM0178	9	*	7.3109	*	*	9.7650	9.8070
5	Lower Reid Brook Below Tributary - NF03NE011	9	*	9.4790	10.5914	12.1391	10.3773	11.3040
6	Reid Brook at Outlet of Reid Pond - NF03NE0009	9	10.0695	10.8427	9.7320	9.7322	12.0802	10.4958
7	Camp Pond Brook Below Camp Pond - NF03NE0010	9	8.9256	9.6137	10.3524	9.7635	10.2338	10.5384
8	Main River at Paradise Pool - NF02YG0009	9	*	*	*	*	10.0798	9.3288
9	Tributary to Gills Pond Brook - NF02YO0019	9	*	*	*	9.9634	9.7735	9.2865
10	East Pond Brook - NF02YO0192	9	*	*	*	10.0553	10.0103	9.7125
11	Southwest Brook below Southwest Pond - NF02ZE0033	9	*	*	*	*	4.4881	9.5437
12	Come by Chance River Near Goobies - NF02ZH0009	9	*	*	*	*	9.7488	9.3717
13	Rattling Brook Below Bridge - NF02ZK0023	9	*	*	*	*	3.9666	8.0677
14	Tributary to Lower Reid Brook - NF03NE0012	9	*	*	*	9.3611	12.2537	11.9293
15	Minipi River below Minipi Lake - NF03OE0030	9	*	*	*	10.7491	10.5997	9.6141

Dissolved Oxygen | Variable 2 | Variable 3 | Variable 4 | Variable 5 | Variable 6 | Variable 7 | Variable 8

	X Coord.	Y Coord.	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Station 1			*	9.6253	9.7609	8.2244	9.6246	7.1426
Station 2			8.4266	7.0945	9.5664	8.0345	9.1063	*
Station 3			*	*	8.2713	7.6796	7.5582	10.0820
Station 4			*	7.3109	*	*	9.7650	9.8070
Station 5			*	9.4790	10.5914	12.1391	10.3773	11.3040
Station 6			10.0695	10.8427	9.7320	9.7322	12.0802	10.4958
Station 7			8.9256	9.6137	10.3524	9.7635	10.2338	10.5384
Station 8			*	*	*	*	10.0798	9.3288
Station 9			*	*	*	9.9634	9.7735	9.2865
Station 10			*	*	*	10.0553	10.0103	9.7125
Station 11			*	*	*	*	4.4881	9.5437
Station 12			*	*	*	*	9.7488	9.3717
Station 13			*	*	*	*	3.9666	8.0677
Station 14			*	*	*	9.3611	12.2537	11.9293
Station 15			*	*	*	10.7491	10.5997	9.6141
Station 16								

- There are missing points in the data set – fortunately CAT is setup to handle this situation

Manly and MacKenzie for RTWQ Data

- CAT output shows us that there is little evidence of Dissolved Oxygen changing from 2003 to 2008 (p-value = 0.3680)

- Any unusual years in the dataset would have an individual p-value < 0.05

- CAT also develops CUSUM plots for each year

Univariate CUSUM Analysis
Variable Analysed:
Dissolved Oxygen
Significance of the CUSUM values were determined with 4999 randomizations

Stations in Worksheet Order:

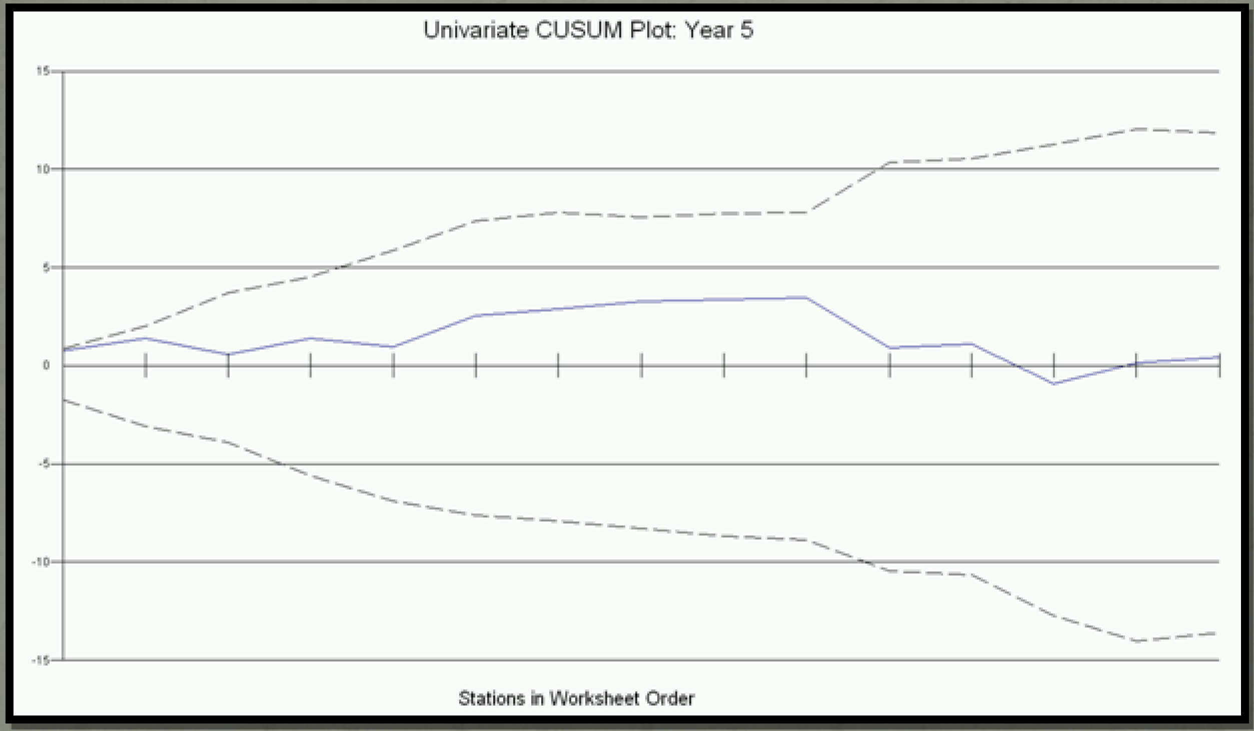
1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	

Test for Serial Correlation
Initial Correlation Estimate = -0.4563
Test Statistic = 0.0120
Probability of a value being more extreme than the observed = 0.8774

Standard CUSUM Method
Test Statistics:
Z Total = 7.7900, p = 0.3680
Year 1 max Z = 0.5826, p = 0.5302
Year 2 max Z = 1.5900, p = 0.2900
Year 3 max Z = 1.4934, p = 0.3576
Year 4 max Z = 1.0598, p = 0.6016
Year 5 max Z = 1.3864, p = 0.6466
Year 6 max Z = 1.6778, p = 0.3040

Sample CUSUM Plot – Year 5 - 2007

- The CUSUM plot shows that the dissolved oxygen levels were about the same as all other years (blue line is within the boundary limits)



Overall Thoughts on Control Charts

- Determining the best ARIMA model can be a bit challenging
- Usually takes some familiarity with the data and the process
- There is certainly potential for using the charts for monthly and weekly means but this requires advanced statistical knowledge
- But likely that it will not be efficient to try and implement this approach for daily or hourly data

Other Research Areas

Other Regression Models

Other Control Chart Applications

Other Research Areas

- Some of the USGS real time stations have had success developing regression models for linking real time data to chemical constituent data collected through manual grab sampling
  - Currently being investigated for Newfoundland RTWQ network
  - Some initial difficulties establishing reliable relationships
  - Grab sample collection methods may need to be reviewed

Modifying the Control Chart

A Better Way of Identifying Unsafe Water Quality

Overview

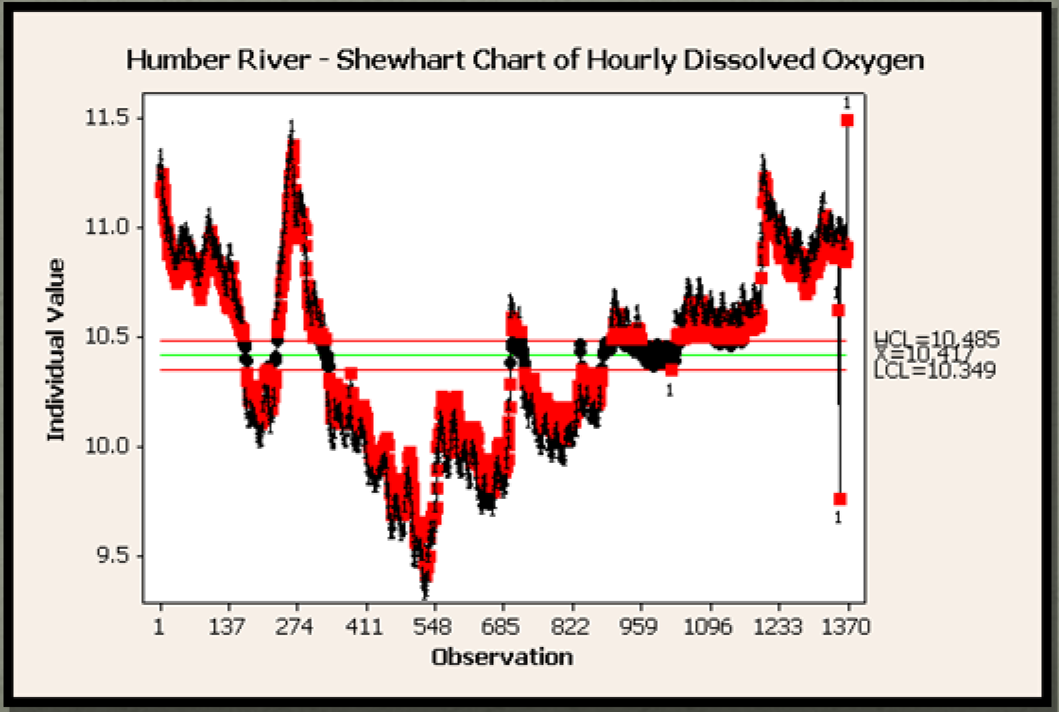
- Control charts were developed to investigate a process for statistical control and these charts are great for looking at manufacturing or industrial processes
- Water quality data isn’t the same as manufacturing data
- Instead of the standard control chart that compares individual values to the overall mean of the data set, a more useful plot for RTWQ data would flag values when they go outside of defined water quality guidelines.

The Plan…

- Develop a chart that shows each value
- Define regions for safe and unsafe values
- For example, when dealing with DO:
  - Draw line at 6 mg/L
    - Points below this level will be flagged in Red (unsafe)
  - Draw an additional line to flag points in the 6 to 6.5 mg/L range
    - Any points in this region are flagged in Green (close to being unsafe)
- Lets try it out with some hourly values from the RTWQ Network station NF02YL0012 – Humber River at Humber Village Bridge

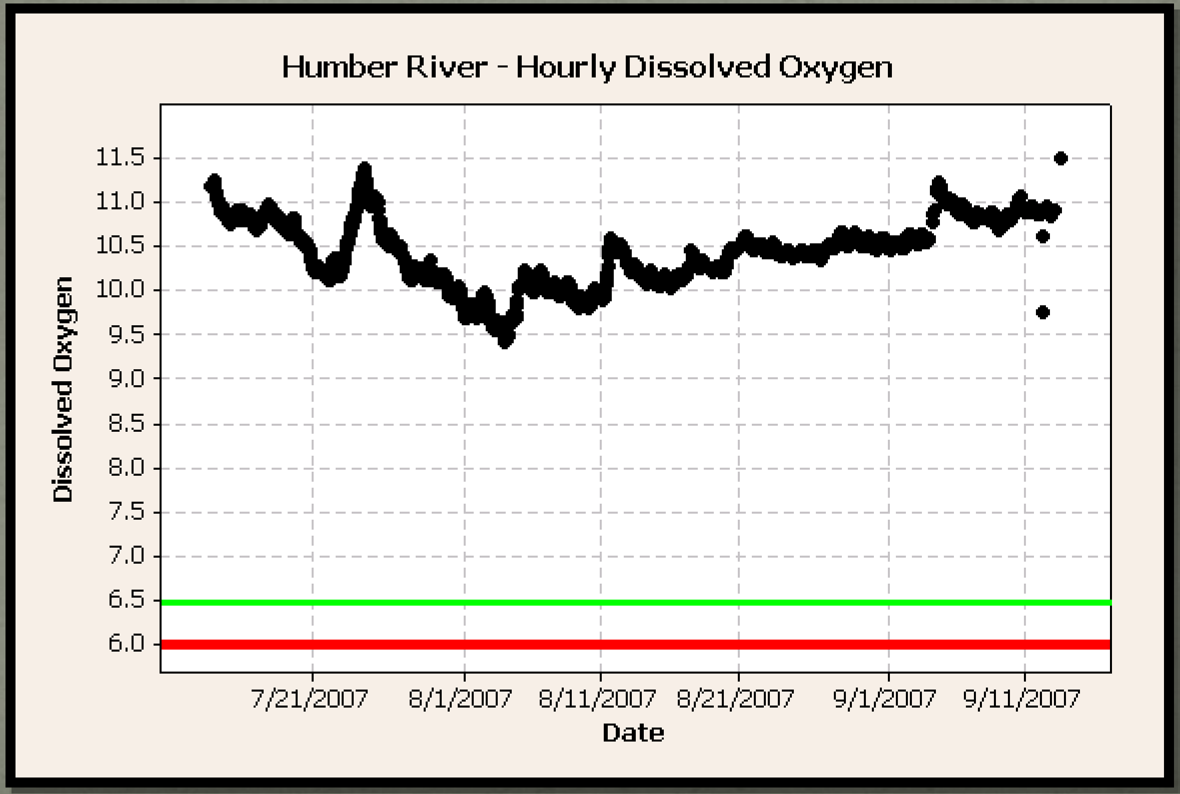
Standard Control Chart – RTWQ Data

- If we were to look at a section of all the available RTWQ data (i.e. one month long deployment period of the sensor) we would see that parameters like DO vary over the month. The standard control chart would show the process as being entirely out of control.



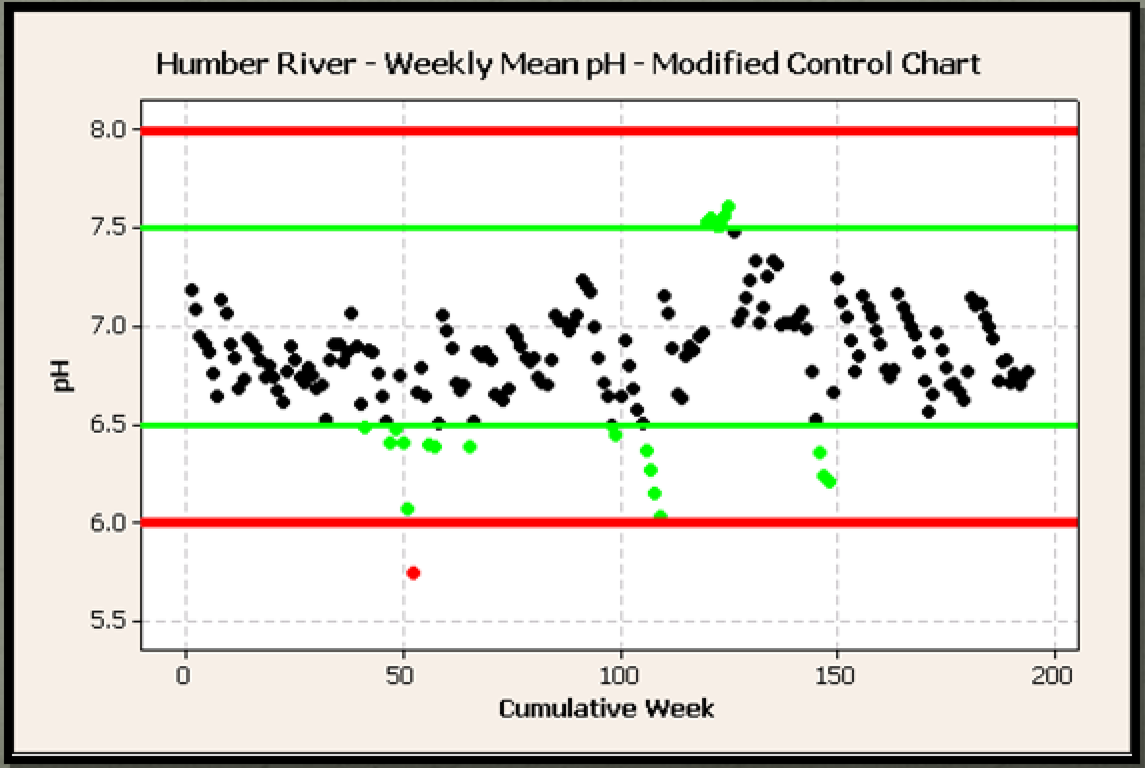
Modified Control Chart – RTWQ Data

- A modified control chart with safe levels of DO defined as being above 6 mg/L shows that we really have nothing to worry about.
- DO never goes below the unsafe limit



Modified Control Chart – Weekly Means

- We could even develop one of these modified control charts for the weekly mean data used previously in regression.
- Looking at say the pH define the safe regions and flag points



Conclusions

- This research shows the potential for developing regression models for water temperature and dissolved oxygen from using air temperature.
- Regional analysis of these relationships requires more data from more sites.
- Relating grab sample data to sensor data requires good quality data from more sites.
- There is potential for implementing SPC control charts for the RTWQ data but this is not easy for those with no advanced statistical training due to autocorrelation and seasonality.
- Simpler and more meaningful control chart methodologies are currently being developed.

Thank You

- Natural Sciences and Engineering Council of Canada
- Faculty of Engineering and Applied Science, Memorial University
- Institute for Biodiversity, Ecosystem Science and Sustainability (IBES)
- Water Resources Management Division, Department of Environment and Conservation