Contents

List of Figures

List of Tables

1 Survey of Quantitative Methods

The bread and butter of quantitative consulting is the ability to apply quantitative methods to business problems in order to obtain actionable insight. Clearly, it is impossible (and perhaps inadvisable, in a more general sense) for any given individual to have expertise in every field of mathematics, statistics, and computer science.

We believe that the best consulting framework is reached when a small team of consultants possesses expertise in 2 or 3 areas, as well as a decent understanding of related disciplines, and a passing knowledge in a variety of other domains: this includes keeping up with trends, implementing knowledge redundancies on the team, being conversant in non-expertise areas, and knowing where to find detailed information (online, in books, or through external resources).

In this section, we present an introduction for 9 "domains" of quantitative analysis:

- survey sampling and data collection;
- data processing;
- data visualisation;
- statistical methods;
- queueing models;
- data science and machine learning;
- simulations;
- optimisation, and
- trend extraction and forecasting;

Strictly speaking, the domains are not free of overlaps. Large swaths of data science and time series analysis methods are quite simply statistical in nature, and it's not unusual to view optimisation methods and queueing models as sub-disciplines of operations research. Other topics could also have been included (such as Bayesian data analysis or signal processing, to name but two), and might find their way into a second edition of this book.

Our treatment of these topics, by design, is brief and incomplete. Each module is directed at students who have a background in other quantitative methods, but not necessarily in the topic under consideration. Our goal is to provide a quick "reference map" of the field, together with a general idea of its challenges and common traps, in order to highlight opportunities for application in a consulting context. These subsections are emphatically NOT meant as comprehensive surveys: they focus on the basics and talking points; perhaps more importantly, a copious number of references are also provided.

We will start by introducing a number of motivating problems, which, for the most part, we have encountered in our own practices. Some of these examples are reported on in more details in subsequent sections, accompanied with (partial) deliverables in the form of charts, case study write-ups, report extract, etc.).

As a final note, we would like to stress the following: it is **IMPERATIVE** that quantitative consultants remember that acceptable business solutions are not always optimal theoretical solutions. Rigour, while encouraged, often must take a backseat to applicability. This lesson can be difficult to accept, and has been the downfall of many a promising candidate.

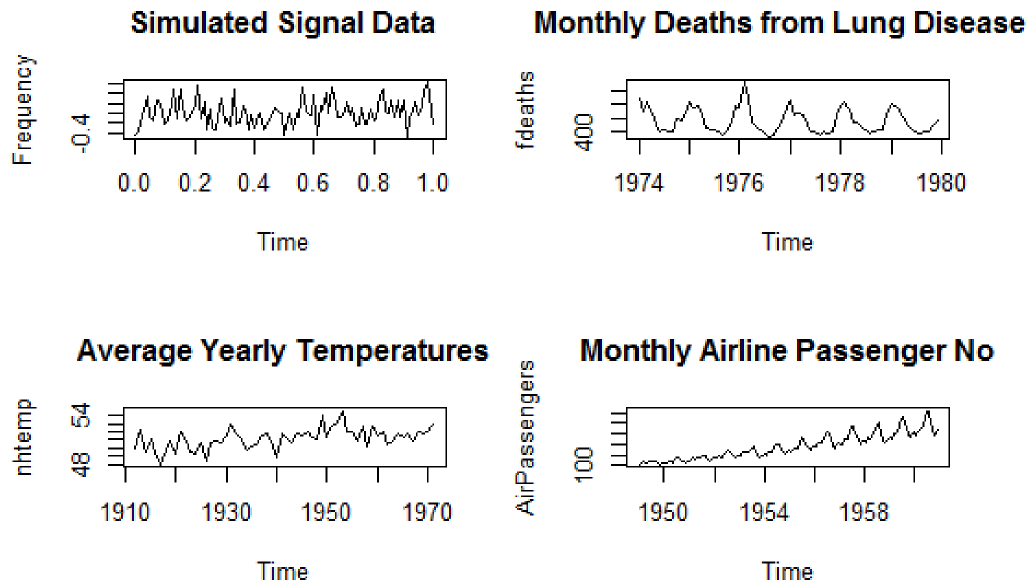


Figure 1: Examples of Time Series

1.9 Time Series Analysis and Forecasting

A time series is a sequence of values, measured at regular intervals over time. The motivation of time series analysis lies in the assumption that what happened in the past has an influence on what will happen in the future. Typically, time series are used for trends analysis and forecasting future values when there are good reasons to suspect the existence of cycles in the data; for instance, a time series analysis could be used to predict the number of passengers going through Canadian airports at various points in the future. An economist might be interested in forecasting the stock market, using time series analysis. Generally speaking, the forecast horizon is the length of the prediction period: predictions at shorter horizons tend to be more reliable and accurate than predictions at longer horizons.

Ideally, the reporting periods used in time series analysis should be identical (e.g. daily, monthly, quarterly or yearly), the measurements should be taken over discrete (exclusive), consecutive periods, and the concepts and the measurement approach should be consistent over time. Detection of periodicity should be done by graphical representation of the data (and the frequency of data collection) using logic (e.g., is there an expectation of hourly, weekly, monthly, quarterly, and/or x-year cycles).

Many traditional statistical methods assume that observations are independently and identically distributed, which is unlikely to happen in real life. At best, this assumption may be sufficiently accurate to allow for some predictive power; at worst, it can lead to wildly inaccurate insights and predictions. Figure 1 illustrates some examples of time series data. Various time series analysis methods and tests are found in applications and in the literature, including *Auto-Regressive* models (AR), *Smoothing and Filtering* models (such as *Moving Averages* (MA) and *Exponential Smoothing* (ES)), and various other *Detrending* models (ARMA, cubic splines, finite differences, etc.), *Seasonal Decomposition* models (such as the X11, X12, X13 and ARIMA models), and various linear and non-linear forecasting models (*Holt's Method*, *Winter's Method*, GARCH models, etc.). This list only skims the surface of available methods. A sample of these methods will be discussed here.

1.9.1 Automated Trend Extraction & Forecasting

Trend extraction is the detection of components within time series that can be combined in order to define the time series. Ideally, trend extraction should not be automated. There are ways to program methods (such as proc X12 in SAS or R) to automatically select the model (additive, multiplicative, etc.), to search for outliers and level shifts, etc., but the tests that are used are NOT perfect and visual examination is typically needed to confirm the procedure's decisions. That being said, it is possible to provide an algorithm that one would expect to de-trend and de-seasonalize a time series most of the time (the caveat being that unless one verifies the results on a given time series, one cannot be sure that the assumptions built into the algorithms apply to that time series). MATLAB has a time series module; some of the documentation gives suggestions as to how to automate trend extraction. But MATLAB is not a native time-series environment: SAS or R are preferred alternatives. Algorithms can be prepared once specific time series are exhibited.

Forecasting, is different from modeling the trend of the exist data. A good fitting model (or robust model, statistically speaking) is not necessarily good at forecasting. **Short-term forecasting:** There is no standard strategy or theory that can answer the question of how few observations can be used to build a model. It depends on the number of estimated parameters and also on how random the data is. With some methods, such as least squares estimation, it is feasible to create a model if the number of observation is greater than the number of estimated parameters. Some other approaches, such as LASSO, also allow analysts to build models with fewer observations than the parameters. **Long-term forecasting:** Due to the nature of real data, not generated from any models, ARIMA does not perform well on long-term forecasting, and neither do other approaches. With too much data, the amount of randomness and computational complexity increases, and thus forecasting is no longer effective and also time-consuming. To solve this issue, non-parametric models are more robust than parametric models, which take time-changing into consideration – that is, the earliest observation becomes the least important in the forecasting model.

Thus, automated trend extraction and forecasting is to some extent possible, but there is a high probability that the resulting automated system will make mistakes. To avoid misleading users of the automated system:

- The results generated by the automated system should be audited by people from time-to-time
- Metrics should be generated to convey the confidence that can be placed on the results produced by the automated system
- There must be a consideration of the cost that will be incurred if the automated system produces "wrong" results

1.9.2 Data Pre-Processing Strategies

Ideally, in time series analysis, we have "rectangular" data – each record having the same number of observation at each time step. In real situations, however, missing data is a very common and significant feature of the dataset. For that reason, the first attempt at solving the missing data problem is to impute these values (i.e. provide values for missing data), instead of removing them all (one exceptional case is when the Recompile data is sufficient and the missing data is limited – in this situation removing the missing cases could also be practical). Last observation carried forward, mean imputation, regression imputation and multiple imputation are all simple and effective approaches to replace the missing data. Analysts need to be cautious of this step since imputed values cannot add new knowledge to the dataset which was not originally present in the dataset. Furthermore, imputation can add bias to the data if not done properly. Outright removal (list deletion) should NOT be considered, as this can introduce bias from which the experiment simply cannot survive.

Multiple imputation extracts useful information from the available observations by completing the dataset (in potentially different ways) a number of times (typically 5-10). Analyses can then be run on each complete dataset, and the results are combined statistically: the estimates incorporate the uncertainty found in the dataset. This works well for up to 30-40 variables, but is poorly suited to datasets with endemic missing values. There exists software that can help with multiple imputation, notably Amelia. Certain time series methods also work with missing values (X12 in SAS or R, the Kalman filter); others do not. MATLAB's getdatasamples method allows the user to re-sample a time series at missing values (using various interpolation methods). When time series values are missing systematically or in large number, the problem of imputation may reduce to a problem of forecasting.

Missing imputation for data streams: the imputation approaches discussed above are not practical for streamed data, which is often incoming at a high frequency, for the following reasons: a) the statistical methods do not specify how much information analysts should know about the data environments; b) the similarities between different completed cases are very difficult to extract, and it is possible that one data stream is completely different from its neighbours; c) it is time consuming when analysts try to use all the available information to estimate the missing values, especially when the missing data is less relevant to the known information d) missing data may or may not appear at random, which requires analysts to obtain the property of the missing values first, in order to apply any statistical analysis.

Nan Jiang and Le Gruenwald proposed an estimation technique, using associating rule mining on streaming data based on closed frequent itemsets (CARM), and validated their approach by comparing the performance between their method to other techniques, including Window Size (AWS), the Simple Linear Regression (SLR), the Curve Regression (CE), and Multiple Regression (MR). The results generated from CARM were more accurate.

1.9.3 Parameter Estimation

In the case of parametric algorithms, an assumption is made in advance about the structure of the model that can best be applied to the data (i.e. which model is most appropriate). The chosen model will then have particular parameters, the number and nature of which are pre-defined based on the model choice. The values of these parameters are then estimated according to a best fit strategy. In this case, if an incorrect assumption is made about which model is best suited to the data, then the model may fail to fit the data well, regardless of how the parameter values are varied. Non-parametric models make few structural assumptions and thus have fewer constraints with respect to the number and nature of the model parameters themselves- in this case, the selection of these, as well as the values assigned to the parameters, are driven by the nature of the data itself.

We are going to discuss parameters setting for some selective non-parametric algorithms in this section, for the reasons that i) these models are flexible in the sense that they can change over time; ii) they could be the initial step to finding an underlying parametric model, where parametric algorithms are those for which parameters can be estimated, which can in turn be used to make predictions (linear regression vs. smoothing methods). The intention of this section is not to provide a comprehensive review on non-parametric methods, but to give ideas on the concept of the subject. More importantly, the questions of how to estimate parameters, how to utilize data to forecast, etc., depend on which model being selected.

Non-parametric methods: threshold autoregression (TAR), ARCH & GARCH, etc. No assumption made on the model shape but sometimes requires smoothness conditions. The complexity depends on the data, which is also a trade-off of models being very flexible.

Estimating parameters in this context is to select a good kernel estimator and as well as a suitable smoothing parameter (referred to as the bandwidth). Significant efforts have been put in to studying this field and thus it is not feasible to give an overview on each strategy here.

Kernel functions include uniform, triangle, bisquare, Epanecknikov, Gaussian, etc., where in most practical cases, Epanecknikov kernel generates a better result and sometimes, analysts prefer using higher ordered kernel, such as bisquare.

Bandwidth selection is the essential task in kernel smoothing. Its goal is to balance between bias and variance, that is, to minimize the mean square error, given by $\text{MSE}(x) = \text{Bias}^2(x) + \text{Var}(x)$, where x is the observation in general. Apparently, increasing bandwidths results in a low variance and a high bias, while small bandwidth will increase the variance. An ideal bandwidth is the balance between bias and variance.

1.9.4 Time Series Analysis Strategies and Algorithms

The basic principle of component decomposition is central to time series analysis. A discussion of relevant concepts and strategies relating to this technique is presented first, followed by an extensive review of important considerations and issues which need to be kept in mind during time series analysis. Finally, two specific time series analysis techniques are considered: TAR and GARCH.

The Components of Time Series Time series analysis is helpful in detecting different patterns in time series and then further categorise the patterns or behaviours which can be explained by the time series. To do this, the time series is decomposed into component parts. Displaying the components of a time series is also helpful in understanding the data. Each of the components represents a category of patterns.

Generally speaking, there are three common components of time series, which are trend, seasonal, irregular. Here we will briefly discuss other potential components as well, but for the sake of simplicity, only the three main components will be illustrated in the modeling section.

- **Trend** component describes the overall "changing direction" of the data, either increase or decrease or flat, which is a long-term effect and not necessarily linear. For example, in Figure 1, the bottom right graph shows the trend going up, which means the overall monthly airline passenger number increased from 1948 to 1960.

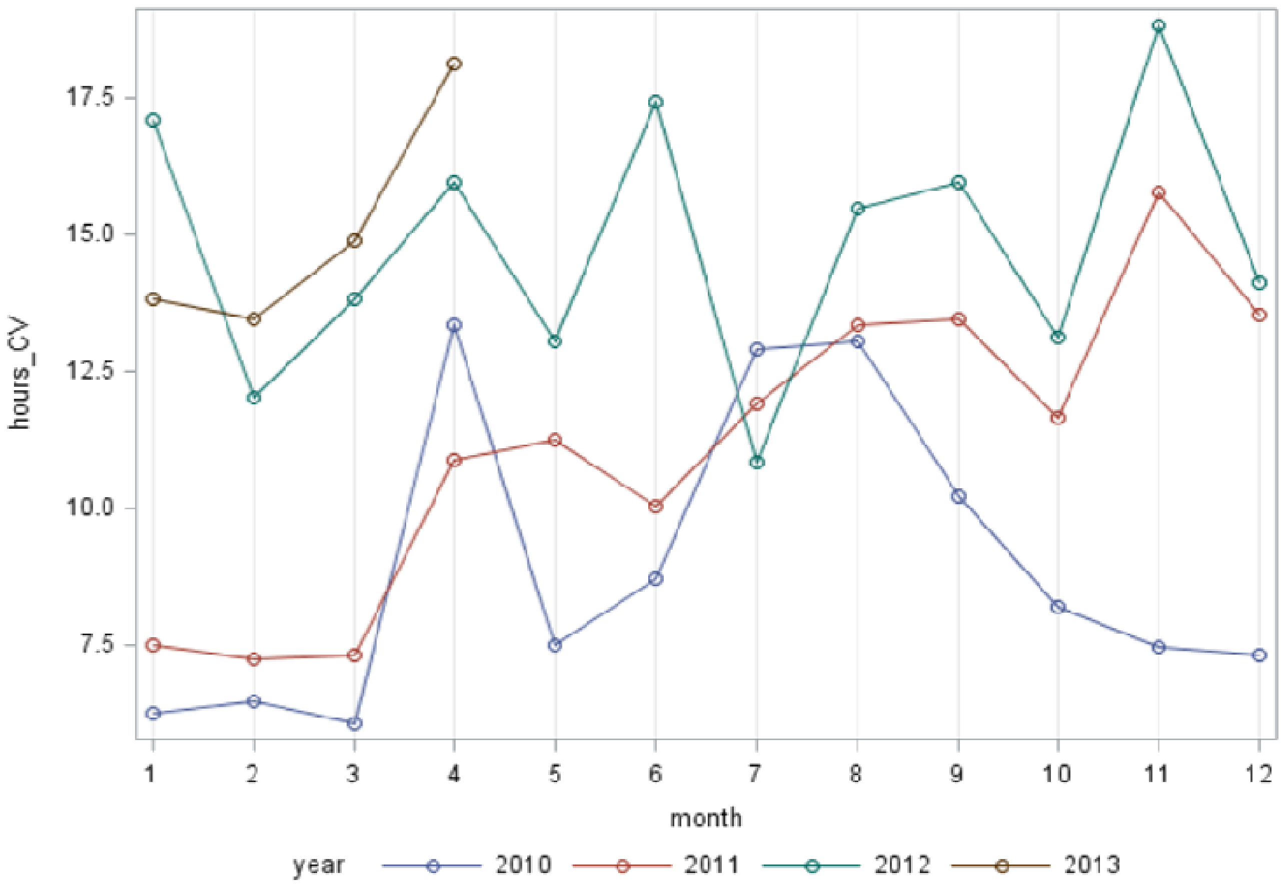
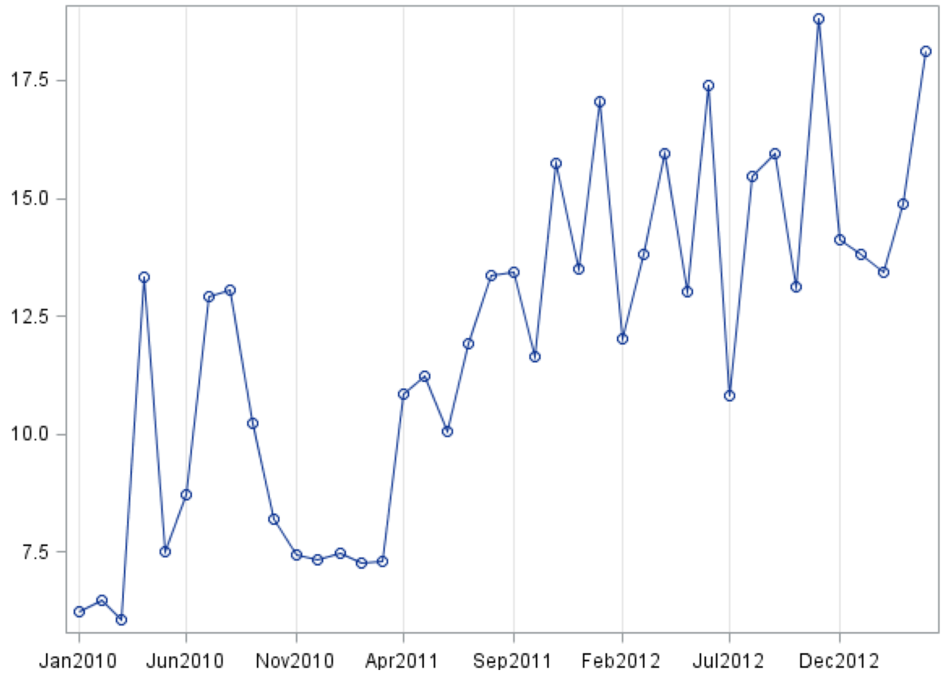


Figure 2: CV by year

- **Seasonal** component reveals the seasonal effect on a series of data, such as that passengers in the airport will increase during summer vacation season. Similarly, in Figure 1 top right graph, the peaks occurs at the beginning of each year. Since the data was collected in London, this implies that winter time had a higher rate of deaths from lung disease than summer.
- **Irregular** is a short-term effect, which can vary considerably from period to period, and includes measurement errors, unseasonal change, etc. Mathematically, we use the residual of the time series after removing trend, seasonal, and cyclical effects, to identify the irregular effect.
- **Cyclical** component usually lasts at least two years. The exact length of the current cycle cannot be predicted. For example, the global financial crisis in 2008 lasted about 5 years. The difference between seasonal and cyclical is that the former displays the change in a fixed time period.
- **Other** components may include calender effect (trading day, leap year, etc.), government policies, strike actions, exceptional events, inclement weather, etc.

Let us illustrate the process of decomposition with an arbitrary time series recording the monthly number of hours for a variable called CV, whose values are shown in the Figure 2. The continuous plot, Figure 3 shows that the size of the peaks and troughs does not seem to change with changing trends: the additive model is thus selected. The SAS procedure X12 agrees with that assessment,



Estimation Summary	
For Variable hours_CV	
Number of Observations	40
Number of Residuals	27
Number of Parameters Estimated	3
Variance Estimate	5.6E-02
Standard Error Estimate	2.4E-01
Standard Error of Variance	1.5E-02
Log likelihood	0.4658
Transformation Adjustment	-69.5685
Adjusted Log likelihood	-69.1027
AIC	144.2053
AICC (F-corrected-AIC)	145.2488
Hannan Quinn	145.3613
BIC	148.0928

Results of Automatic Transformation Selection	
For Variable hours_CV	
AICC (with aicdiff=-2.00) prefers	No transformation
Adjustment will be	Additive

Figure 3: Continuous CV; estimation summary.

and further suggests no data transformation. The diagnostic plots are shown in Figure 4: the 2010 CV series is prior-adjusted from the beginning until OCT2010 after the detection of a level shift. The SI (Seasonal-Irregular) chart shows that there are more than one irregular component which exhibits volatility. The adjusted series is shown below in Figure 5 (the trend and irregular components are shown separately for readability).

Traditionally, decomposition follows one of three models: multiplicative, additive, and pseudo-additive.

- **Multiplicative** – This modeling approach assumes that a) the magnitude of the seasonal spikes/troughs increases when the trend increases (and vice versa); b) the trend T_t has the same dimensions as the original series O_t, and the seasonal component S_t and the irregular component I_t are dimensionless and centered around 1; c) the seasonal fluctuation $\sum_{j=1}^{n} S_{t+j} = 0$, where $n = 365$ for daily series, $n = 12$ for monthly series,$n = 4$ for quarterly series,etc.,and d) the original series O_t does not contain zero values. Mathematically, the model is expressed as:

  $$O_t = T_t \times S_t \times I_t$$

  All components share the same units. After seasonality adjustment, the seasonality adjusted series is

  $$SA_t = \frac{O_t}{S_t} = T_t \times I_t$$

  To transfer multiplicative to additive model, we could take a logarithm transformation, such as:

  $$\log O_t = \log T_t + \log S_t + \log I_t$$

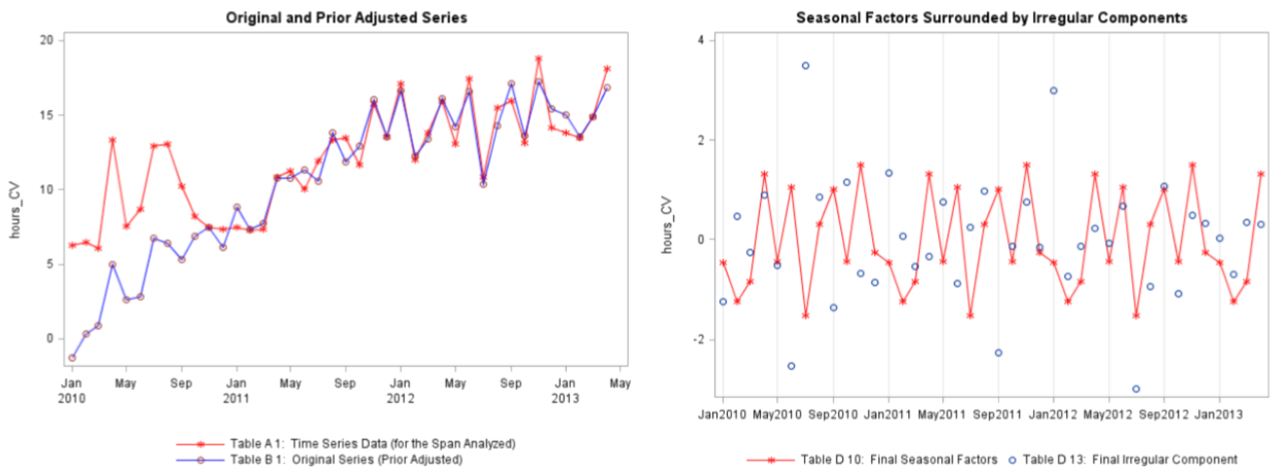


Figure 4: Diagnostic plots. Note here that the analysis of a time series starts with estimation of the effects of festivals and trading days. These precalculated estimates are then used for prior adjustment of the series. The prior adjusted original series is subsequently analyzed using the seasonal adjustment.

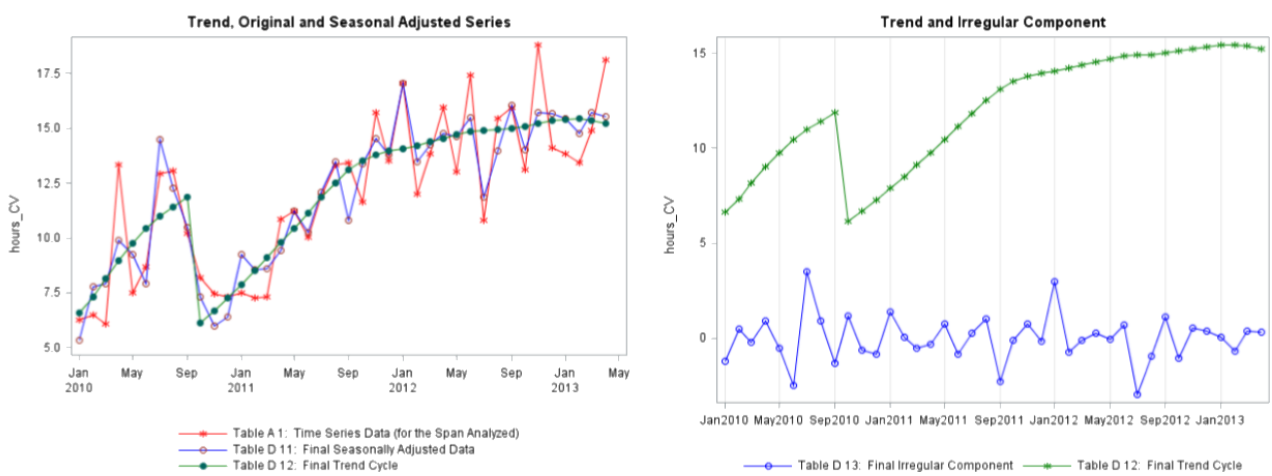


Figure 5: Adjusted plots

- **Additive** – This modeling approach assumes that: a) the seasonal component S_t and the irregular component I_t are independent of the trend behaviour T_t; b) the seasonal component S_t remains stable from year to year; and c) the seasonal fluctuation $\sum_{j=1}^{n} S_{t+j} = 0$, where $n = 365$ for daily series, $n = 12$ for monthly series, $n = 4$ for quarterly series, etc. Mathematically, the model is expressed as:

  $$O_t = T_t + S_t + I_t$$

  All components share the same dimensions and units. After seasonality adjustment,the seasonality adjusted series is:

  $$SA_t = O_t - S_t = T_t + I_t$$

- **Pseudo-additive** – This modeling approach assumes that some of the values of the original series O_t are 0 (or very close to 0) and that a) the seasonal component S_t and the irregular component I_t are both dependent on the trend level T_t, but independent of each other, and b) the trend T_t has the same dimensions as the original series O_t, and the seasonal component S_t and the irregular component I_t are dimensionless and centered around 1. Mathematically, the model is expressed as:

$$O_t = T_t + T_t \times (S_t - 1) + T_t \times (I_t - 1) = T_t \times (S_t + I_t - 1)$$

All components share the same units. After seasonality adjustment, the seasonality adjusted series is:

$$SA_t = O_t - T_t \times (S_t - 1) - T_t \times (D_t - 1) = T_t \times I_t$$

- **How to choose the model?** – The choice of a model is driven by data behaviour and choice of assumptions. The analyst needs to plot the time series graph and test a range of models, selecting the one which stabilized the seasonal component.
- **Suggested methods of estimating if a time series is multiplicative, additive or pseudo-additive?** – The simplest way to determine whether to use multiplicative or additive decomposition, is by graphing the time series. If the size of the seasonal variation increases/decreases over time, multiplicative decomposition should be used. On the other hand, if the seasonal variation seems to be constant over time, additive model should be used. A pseudo-additive model should be used when the data exhibits the characteristics of the multiplicative series, but parameter values are close to zero.
- **how to automate the selection of multiplicative/additive/pseudo-additive model?** – Complete automation of the model selection will be very difficult (as there is no prior knowledge about the shape of the data). Under the assumption that the data behaves nicely, one may be able to test whether the magnitude of seasonal variability remains constant over time. Suppose that there is data from January 2010 to December 2015: if the magnitude remains constant, the ratio of seasonal component over time should be one. Comparing data points from January 2010, 2011, ..., 2015, the ratios of January 2011/January 2010, January 2012/January 2010, and so on should all be one. Otherwise, Multiplicative decomposition should be used. If the multiplicative method turns out to be ineffective (since estimating a component that is virtually 0 in multiplication is unstable), a pseudo-additive model may be chosen.

Notes, Challenges and Pitfalls

- Time series are said to be (weakly) **stationary** when time series values are independent of time (i.e. the mean and variance are stable over time). Autocorrelation function (ACF) and Partially ACF (PACF) are commonly used to determine whether a time series is stationary. The explanation is very straightforward, as shown in Figure 6. The lower ACF is the desired plot, implying the time series is stable. When the empirical data's ACF does not match the model's ACF, the model being considered is unlikely to be adequate. For example, if a model (such as SARIMA) is fitted, then the residuals should exhibit the ACF of a white noise series (i.e., the residuals are uncorrelated.) If the ACF/PACF show some spikes in the resulting residuals, it is an indication that the fitted model is missing some components (hence inadequate).
- In order to identify correlations and root causes *via* analysis of the relationships between the variables, one must assume that the time series values are stationary. Statistical tools which assume data independence are invalidated if the data is not actually independent of time,

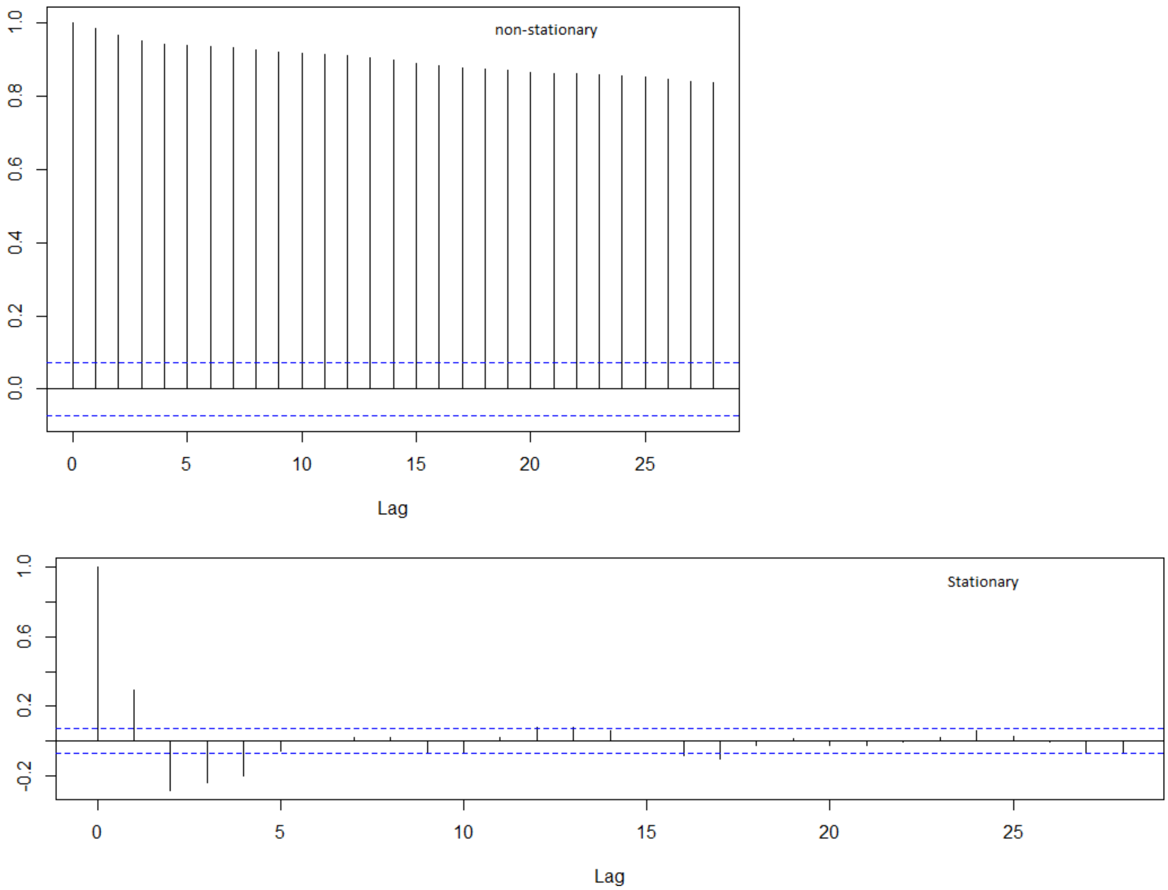


Figure 6: ACF plots for a model which is not matching the empirical data.

that is if it is **serially dependent**. Classical statistical methods are built on the assumption of independence of data points (y1, y2, ...). If they are dependent of each other (in the form of time or others), then the fundamental assumption fails, and hence these statistical methods must be adjusted for time dependence.

- When **non-stationary** serial dependence is suspected or expected to exist, **time series decomposition** is required to extract the data's **trend**, **cyclical** and **seasonal components**, **calendar effects**, and **structural breaks**, in addition to providing accurate forecasts. The time series seasonal adjustment enables the identification of **turning points** and provides consistent comparisons of indicators across time periods.
- Traditionally, time series decomposition is built on one of three models: **multiplicative**, **additive**, or **pseudo-additive**. The choice of a model is driven by data behaviour and choice of assumptions.
- Time series decompositions, and hence any activity depending on them (such as forecasting, for instance), **ultimately rely on the quality of the underlying data**. In particular, there are a number of well-known data quality issues which affect the results of the analysis:

  1. the method of data collection may lead to unusual effects, especially if collection is made on a non-calendar basis or if there is a lag between activity and measurement;
  2. any change to the method or timing of data collection could lead to the false identification of trend or seasonal breaks;

3. some series are sensitive to events such as extreme weather, strikes, wars, etc., which could cause breaks or outliers of large magnitude;
4. at least 5 years' worth of data are required to insure stability on future updates, and
5. at least 10 years' worth of data are required to insure that the adjustment of the first year is unlikely to be revised.

- Forecasts tend to be **wrong**: aggregated forecasts (using more than one model and aggregating the results, instead of using a single model) are usually more accurate. Emphasis should not be placed on a **single estimate** (such as the mean): they should also include the **standard deviation** or an **accuracy range**.
- Identification of trend in time series is subjective because what appears to be a trend over a short time period may prove to simply be a **small fluctuation** which could form part of a cycle over the long-term horizon of the series.
- Regression models of various complexity levels can be fitted (against time and/or auxiliary variables) to identify possible trends. At long horizons, polynomial response functions explode: if such models must be used, **it is recommended to use linear or quadratic response functions**, as slope and concavity might be the best we can hope to detect in light of the previous remark.
- In combination with **appropriate data transformations** (e.g. logarithm, square root, inverse, Box-Cox, etc.), low order regression models can achieve good results.
- **power transformation** is the simple and effective way to stabilize the variances, including square root, cube root, logarithm, etc.
- **Fourier transforms** can help identify potential trend and cycles (as well as their respective frequencies), so can a variety of statistical tests (like the **Mann-Kendall** test, for instance).
- De-trending methods come with strengths and weaknesses:

1. **Finite Differences** are iterated differences between subsequent time series observations, which can remove polynomial trends. They are useful if the exact shape of trend cannot be estimated, but too high an order may introduce variance inflation. They ignore the potential effect of any variable over the trend, save for the passage of time.
2. **Curve Fitting** includes regressions against time as well as more complicated models involving auxiliary variables. Prior knowledge of the situation under consideration can be used to provide an acceptable model which naïve analysis of the data might not be able to suggest, but a simple regression model may be unrealistic.
3. **Filtering and Smoothing** consists of various weighted averages of the time series data which are used to compute a filtered series. Their advantages and disadvantages are discussed below. Note that filtering and smoothing methods preclude the possibility of finding an explicit functional form for the trend.

Moving average of order N is the arithmetic average of the most recent N observations. It tends to provide **stable forecasts**, and bad data (e.g. irregular points, bad stretches) is **eventually removed** from the prediction process; however, it requires saving a potentially large number of past data points, it can **lag** behind the actual trend, and it ignores complex relationships in data.

Weighted moving average of order N attaches importance to certain observations in the form of weights (recent observations could have more influence than older observations, for example); such weighted average may **reduce the lag** behind the trend, but there's no natural and universal way to select the weights.

Exponential Smoothing is a weighted moving average with declining weights for past data which carries the entire past history of the series, without the need to save past data points. They tend to produce **stable forecasts**, but may **increase the lag**.

Holt's Method is a double exponential smoothing which allows for quick multi-step forecasts in the presence of linear trend.

Winter's Method is a triple exponential smoothing with can take seasonal factors into consideration. At least two complete data cycles are required to provide initial estimates, and a third cycle is required for fine-tuning.

4. **Cubic Splines** fit the data to a sequence of separate cubic polynomials for every sequence of three points in the series (the first and second derivatives are continuous at each point, ensuring a well-behaved curve). A parameter has to be specified which depends on the relative importance given to "smoothness" and "goodness-of-fit" of the spline. These methods are prone to overfitting the data.

Non-Linear Models Linear models are not always appropriate to describe the data. In this section, we introduce two (of many) potential non-linear models: TAR (Threshold Autoregressive) and GARCH. Non-parametric models allow data to have different states or regimes, and models to be dynamic in these different regimes.

Threshold autoregression is the extended application of the autoregressive model (AR), defined as

$$y_t = \beta_0^{(1)} + \beta_1^{(1)} y_{t-1} + \epsilon_t \quad \text{if } y_{t-1} \leqslant \gamma;$$
$$y_t = \beta_0^{(2)} + \beta_1^{(2)} y_{t-1} + \epsilon_t \quad \text{if } y_{t-1} > \gamma.$$

where $\beta^{(1)}$ and $\beta^{(2)}$ are less than 1; ϵ_t is the error term; $\gamma \in \mathbb{R}$. There is one model for each side of the threshold. Without two different models, this is simply an AR model. The above equation is a simple example of the TAR model and in general, analysts could set different thresholds. When the threshold $\gamma = y_{t-d}$, with a delayed parameter $d > 0$, the TAR model is called self-exciting TAR or SETAR.

An autoregressive conditionally heteroscedastic (ARCH) model is used to model a changing variance per time period, and in most cases, it is used when increased variation occurs in short periods. Applications are common in economic and financial domains. The GARCH model is the generalized version of the ARCH model; GARCH(1,1), for instance, looks like:

$$\sigma^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

The initial step of finding the "correct" model is to plot the time series and, depending on the properties of the plot, to set up up parameters and candidate models. For instance, to identify which ARCH model would be ideal, one could look at the autocorrelation function plots of y_t and y_t^2 – an ARCH(1,1) model would be appropriate if y_t seems to be a white noise and y_t^2 follows an AR(1) model. Other tests exist (see Shumway and Stoffer's *Time Series Analysis and its Applications* for more examples).

1.9.5 Results Evaluation

In the previous sections, we have discussed how to select parameters and generate models. The next task is to evaluate the models: to apply the candidates to the data and compare their results. Evaluation measures the "difference" between estimated and true values (the residual). Usually, analysts separate the data into training and testing sets, retaining a small portion for the purpose of validation. The optimal model parameters include the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The two parameters measure not only the goodness-of-fit of a model, but also its simplicity. Generally speaking, the lower AIC/BIC a model possesses, the better the model would be, but the value in isolation is of little use. Furthermore, the performance of forecasting on the training data can also be evaluated by model forecasting accuracy matrices (such as a confusion matrix). The criteria are defined as follows:

$$\text{AIC}(p) = n\ln(\sigma_e^2/n) + 2p$$

$$\text{BIC}(p) = n\ln(\sigma_e^2/n) + p + p\ln(n)$$

where p is the number of the parameters in the model, n is the number of observations, and σ_e^2 is the sum of squared residuals using the fitted time series given by the model .

Other approaches also exist. Let y_t denote the actual value of the time series, $\hat{y}_t$ be its forecast, and $e_t = y_t - \hat{y}_t$ be the error.

- **The Mean Forecast Error (MFE)** is defined as MFE $= \frac{1}{n}\sum_{t=1}^{n} e_t$ where n is the number of observations. Clearly, a desired MFE will be close to zero. Even though it is easy to compute, it has many disadvantages: an MFE near 0 does not necessarily lead to good forecasts due to positive and negative errors potentially cancelling out; the MFE is easily affected by extreme values, and it is unduly influenced by data transformations. Similarly, the **mean absolute error** (MAE) and the **mean absolute deviation** (MAD) measure the average absolute difference between forecasts and the real values. Unlike MFE, MAE takes into account the magnitude of overall error.
- **The Mean Absolute Percentage Error (MAPE)** is given by, MAPE $= \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{|y_t|}\times 100$, which measures the percentage of average absolute error. MAPE is independent of the scale; but it is unduly affected by data transformation and extreme cases, and opposite errors cancel out.
- **The Mean Squared Error (MSE)** $= \frac{1}{n}\sum_{t=1}^{n} e_t^2$ measures the average squared deviation of the forecast. Mathematically, a squared root version of MAPE is given by

  $$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}.$$

  Unlike MAPE, y_t being zero or very small does not affect the MSE much, but MSE measurement is not scale-free.
- **The Normalized Mean Squared Error (NMSE)** is defined as

  $$\textbf{NMSE} = \frac{MSE}{\sigma^2} = \frac{1}{\sigma^2 n}\sum_{t=1}^{n} e_t^2,$$

which normalizes the MSE. It has similar properties to the MSE, but is more effective. The smaller the NMSE of a model is, the better the forecast will be.

- **Theil's U-statistic** is given by

$$U = \frac{\sqrt{\sum_{t=1}^{n} e_t^2/n}}{\sqrt{\sum_{t=1}^{n} \hat{y}_t^2/n \sqrt{\sum_{t=1}^{n} y_t^2/n}}}.$$

  It compares the actual values with the results of forecasting with historical data. Due to the squared value, large errors have more weights. Its interpretations are that if $U < 1$, then the forecast is better than mere guessing; if $U = 1$, there is no difference; if $U > 1$, then the forecast is worse than guessing, and if $u = 0$ the fit is near perfect.

We have presented some selected metrics to evaluate the forecasting results and each of them has its own set of properties. In practice, a model's validity is gauged by looking at more than one metric, in order to mimize the odds of making a mistake.

1.9.6 Case Study: Multi-Modal Supply Chains

The project's description is provided in the Project Summary, in the following pages.

References

[1] A.Milhoj, Practical Time Series Analysis Using SAS, SAS Institute.
[2] Time Series Analysis Branch, Guide to Seasonal Adjustment with X-12-ARIMA, Office for National Statistics, UK.
[3] W.S.Cleveland, S.J.Devlin [1982], Calendar Effects in Monthly Time Series: Modeling and Adjustment, J.Am.Stat.Assoc., Vol. 77, no. 279, pp. 520-528.
[4] Estimating and Removing the Effects of Chinese New Year and Ramadan to Improve the Seasonal Adjustment Process, Australian Economic Indicators, November 2005.
[5] P.Boily, Y.Huang [2013], Analysis of Fluidity Indicators and Seasonality Adjustment for Containers Transit Times in a Multi-Modal Supply Chain Network, Centre for Quantitative Analysis and Decision Support.
[6] R.H.Shumway, D.S.Stoffer [2010], Time Series Analysis and its Applications, Springer.
[7] R.Lindeke, Forecasting Models, Lecture Notes.
[8] N.Silver [2012], The Signal and the Noise: Why So Many Predictions Fail - and Others Don't, Penguin Publishing.
[9] D.Meko, Notes on Detrending, University of Arizona (online).
[10] Wikipedia's entry on the Kalman Filter
[11] D.S.C.Fung, Methods for the Estimation of Missing Values in Time Series
[12] O.Anava, E.Hazan, A.Zeevi, Online Time Series Prediction with Missing Data
[13] MATLAB's fillts method (http://www.mathworks.com/help/finance/fillts.html)
[14] J.Honaker, G.King, What to do about Missing Values in Time Series Cross-Section Data
[15] J.Honaker, G.King, M.Blackwell, Amelia II: a Program for Missing Data
[16] B.Pekar, Automating Time Series Analysis - A Case-based Reasoning and Web Services

[17] R.J.Hyndman, Y.Khandakar, Automatic Time Series Forecasting: The forecast Package in R
[18] J.D.Ashley, Why Seasonal Adjustment, Catherine Hood Consulting (online).
[19] C.C.Hood, Seasonal Adjustment and Time Series FAQ, Catherine Hood Consulting (online).
[20] B.C.Monsell, A Painless Introduction to Seasonal Adjustment, U.S. Census Bureau.
[21] H.A.Latane [1942], Seasonal Factors Determined by Difference from Average of Adjacent Months, J.Am.Stat.Assoc, Dec.
[22] D.F.Findley, C.C.Hood, X-12-ARIMA and its Application to Some Italian Indicator Series, U.S. Bureau of the Census.
[23] D.F.Findley, B.C.Monsell, Bell, Otto and Chen [1998], New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program, U.S. Bureau of the Census.
[24] An Introductory Course on Time Series Analysis, Australian Bureau of Statistics.
[25] Seasonal Adjustment of Economic Time Series, Singapore Department of Statistics.
[26] T.Jackson, M.Leonard, Seasonal Adjustment Using The X12 Procedure, SAS Institute.
[27] A.J.Calise, J.Earley, Detecting Structural Change Using SAS/ETS Procedures.
[28] Ratnadip Adhikari, R. K. Agrawal An Introductory Study on Time Series Modeling and Forecasting
[29] Siegfried Heiler, A Survey on Nonparametric Time Series Analysis
[30] Rui Ding, Qiang Wang, Yingnong Dang, Qiang Fu, Haidong Zhang, Dongmei Zhang, YADING: Fast Clustering of Large-Scale Time Series Data
[31] Paul L. Anderson and Mark M. Meerschaert, Parameter Estimation for Periodcally Stationary Time Series
[32] Shiming Yang, Konstantinos Kalpakis, Colin F. Mackenzie, Lynn G. Stansbury, and Deborah M. Stein, Online recovery of missing values in vital signs data streams using low-rank matrix completion
[33] Nan Jiang and Le Gruenwald, Estimating Missing Data in Data Streams
[34] Thomas Url, Rob J Hyndman, Alexander Dokumentov, Long-term forecasts of age-speciïňĄc participation rates with functional data models
[35] **https://onlinecourses.science.psu.edu/stat510/node/78**
[36] Rob J Hyndman 's blog
[37] S.Lenser and M.Veloso, *Non-Parametric Time Series Classification*, Carnegie-Mellon University, 2007
[38] H.F.F.Mahmoud, *Parametric versus Semi/nonparametric Regression Models*, Virginia Polytechnic Institute and State University Department of Statistics, LISA short course series- July 23, 2014

Project Summary – Transport Canada

Analysis of Fluidity Indicators and Seasonality Adjustments for Containers Transit Times in a Multi-Modal Supply Chain Networks

by Patrick Boily, Yue Huang
Centre for Quantitative Analysis and Decision Support, Carleton University

CLIENT ORGANIZATION

Now, more than ever, Canadians need a safe and secure transportation system. Transport Canada (TC) is a government agency that is responsible for transportation systems, policies and programs. It promotes safe, secure, efficient and environmentally-responsible transportation within Canada and reports to Parliament and Canadians through the Minister of Transport.

As a result of ensuring a safe and secure transportation system, TC's work protects people from accidents and exposure to dangerous goods, protects the environment from pollution, and contributes to a healthy population and economy. TC is also responsible for the safety and security of activities including: aircraft services, rail, road and marine safety, and transportation of dangerous goods.

PROJECT INTENT, SCOPE, AND OBJECTIVES

Supply chains play a crucial role in the transportation of goods from one part of the world to another. As the saying goes, "a given chain is only as strong as its weakest link" – in a multi-modal context, comparing the various transportation segments is far from an obvious endeavour.

TC is looking to produce an index to track container transit times in multi-modal chain networks. This index should depict the reliability and the variability of transit times but in such a way as to be able to compare performance between differing time periods. The seasonal variability of performance is relevant to supply chain monitoring and the ability to quantify and account for the severity of its impact on the data is thus of great interest.

The ultimate goal of this project was to compare quarterly and/or monthly performance data, irrespective of the transit season, in order to determine how well the network is performing, as it applies to the *Shanghai → Port Metro Vancouver/Prince Rupert → Toronto* corridors, and to produce a scoring methodology which could then be applied to other corridors.

METHODOLOGY

In order to complete the assignment, CQADS used the following methodological steps:

1. *Review transportation literature*, in order to develop a scoring methodology to determine which is most relevant. The scoring methodology was applied to several proposed indicators that were developed for the *Shanghai – Port Metro Vancouver – Toronto* corridor.
2. *Review and explore available transit time data* to identify seasonality components, leading to adjusted data elements and to the elimination the variability component attributable to such trends.
3. *Testing various pre-existing reliability/variability indicators* against collected container transit time data, which identified promising leads.
4. *Development of the conceptual time-series model,* which established the logic and interaction between the proposed model indicators, and identified the essential data elements providing the best fit to the available data.
5. *Implementation of conceptual model* on a SAS platform, which allowed for the recognition that indicators which best reflected the performance of the chain in a given link were not necessarily the best choice for other links, and led to a new iteration of the prototype model.
6. *Final validation of the prototype model* using collected container transit time data, adjusted to reflect all the underlying trends that had been discovered.

7. *Documentation of the final model*: a technical report providing a detailed description of the model, as well as a number of useful scoring examples, was written and delivered to TC stakeholders. Quality assurance was insured by getting the report summarized and reviewed by a third party, external to the project.
8. *Knowledge transfer* was achieved through regular phone meetings and email exchanges detailing the project progress, and by getting the report reviewed and summarized by external parties.

PROJECT SUMMARY

The supply chain under investigation has Shanghai as the point of origin of shipments, with Toronto as the final destination; the containers enter the country either through Port Vancouver or Prince Rupert. Containers leave their point of origin by boat, arrive and dwell in either of the two ports before reaching their final destination by rail. The situation is illustrated in Figure 1 below.

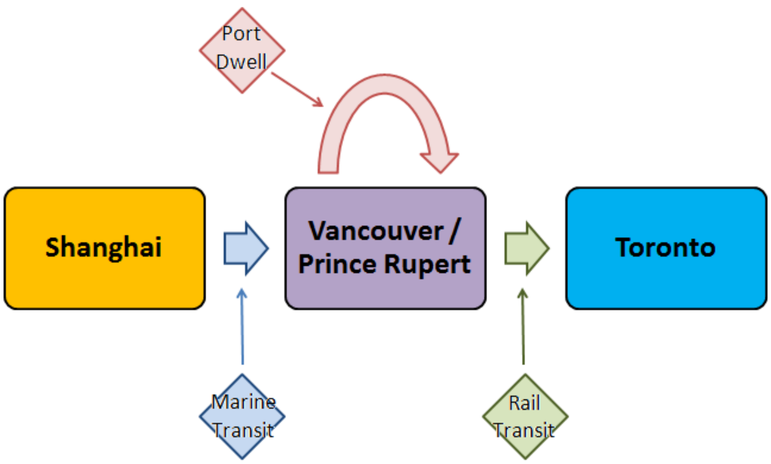


Figure 1 – The Shanghai → Vancouver/Prince Rupert → Toronto supply chain**.**

For each of the three segments (Marine Transit, Port Dwell, or Rail Transit), the data consists of the monthly empirical distribution of transit times from January 2010 to March 2013 (for Port) or April 2013 (for Marine and Rail), built from sub-samples (assumed to be randomly selected and fully representative) of all containers entering the appropriate segment.

Each segment's performance was measured using Fluidity Indicators, which are computed using various statistics of the transit/dwelling time distributions for each of the supply chain' segments. The main indicators under consideration were:

- the *Reliability Indicator* (RI) is the ratio of the 95th percentile to the 5th percentile of transit/dwelling times. A high RI indicates high volatility, whereas a low RI (≈ 1) indicates a reliable corridor;
- the *Buffer Index* (BI) is the ratio of the positive difference between the 95th percentile and the mean, to the mean. A small BI (≈ 0) indicates that the mean and the 95th percentile transit times are roughly the same, and so that there is only slight variability in the upper (longer) transit/dwelling times; a large BI indicates that the variability of the longer transit/dwelling times is high, and that outliers might be found in that domain;
- the *Coefficient of Variation* (CV) is the ratio of the standard deviation of transit/dwelling times to the mean transit time.

The time series of monthly indicators (which are derived from the monthly transit/dwelling time distributions in each segment) were then decomposed into their Trend, Seasonal (Seasonality, Trading-day, Moving-holiday), and Irregular components (see Figure 2, next page, for an example).

The Trend and Seasonal components provided the expected behaviour of the indicator time series; the Irregular components arose as a consequence of supply chain volatility.

Broadly-speaking, the decomposition involved three main steps:

1. the selection and application of a **seasonal decomposition model** (either additive or multiplicative), through *graphical inspection* (multiplicative if the size of seasonal peaks and troughs changes as the trend changes, additive otherwise) and/or *AICC comparison* (using the SAS procedure `X12`);

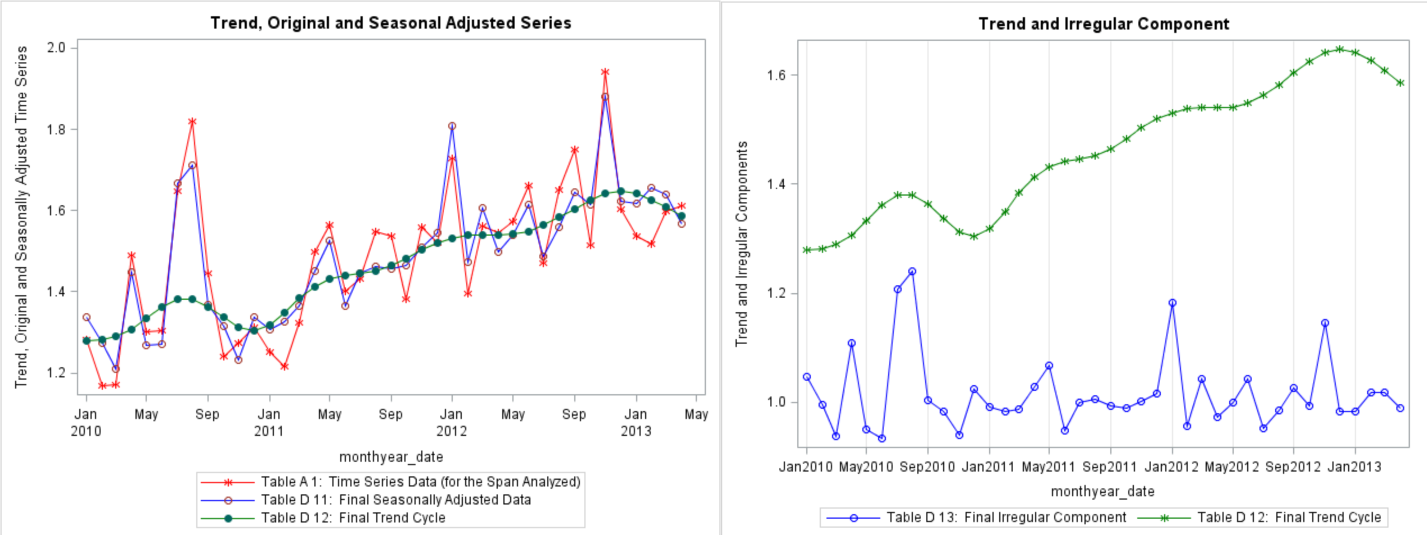


Figure 2 – Indicator time series for Marine RI (Shanghai → Vancouver) in red**.** The Trend component is shown in green, the Irregular one in blue.

2. the identification of, and adjustment (as required) for **calendar effects** such as **trading-day effects** (the effect of the monthly number of weekend days) or **moving-holiday effects** (due to Easter, for instance, in the Western world, or the Chinese New Year in the Pacific Rim), using *spectral plots*, *AICC tests* and *graphical inspection of diagnostic plots;*
3. the identification of, and adjustment (as required) for **trend level shifts** (abrupt but sustained changes in the underling level of the time series which usually have an identifiable cause, such as an increase in shipments due to an extra terminal having opened) and **outliers** (extreme values which fall outside the general trend pattern which can be caused either by an extreme random effect or an identifiable reason such as a short strike or a bad weather event), using *month-to-month percentage changes* and *residual patterns*.

Time series decompositions, and hence any activity depending on them (such as forecasting, for instance), ultimately rely on the quality of the underlying data. In particular, there are a number of well-known data quality issues which affect the results of the analysis:

- the method of data collection may lead to unusual effects, especially if collection is made on a non-calendar basis or if there is a lag between activity and measurement;
- any change to the method or timing of data collection could lead to the false identification of trend or seasonal breaks;
- some series are sensitive to events such as extreme weather, strikes, wars, etc., which could cause breaks or outliers of large magnitude;
- at least 5 years' worth of data are required to insure stability on future updates, and
- at least 10 years' worth of data are required to insure that the adjustment of the first year is unlikely to be revised.

The last two of these did apply to the indicator time series, but were not deemed problematic in the long term as the data collection program from which the data was obtained was slated to continue indefinitely.

A larger issue, and one that it is more difficult to ignore, is that transit/dwelling times are only available for a sample of all containers going through the supply chain, and it is not at all obvious that this sample is randomly selected by the relevant authorities (and so may fail to be representative). Even if it was randomly selected, there is no guarantee that the sampling has remained (or will remain) the same over time.

Consequently, the analysis results were only ever as good as the quality of the data that went into the model. Since the aim of the project was solely to provide a methodology for time series decomposition – rather than to highlight particular irregular values on specific indicator time series – the issue is of lesser consequence at this stage. But this will have to be resolved one way or another by TC and its clients.

There were no overarching results that apply to all indicator time series, on each segment (save for the lack of effect of the Chinese New Year, surprisingly enough): there were series with an Easter effect for a given indicator but not for another; series with a trading-day effect in a segment but not in another; series with outliers and series without, series with a trend level shift and series without.

Another issue is that the reliability of a supply chain is a function of the total transit time from its origin to its destination. The importance of obtaining end-to-end data (i.e., of following a container from one end to another) has been recognized recently, and this data will be used in the analyses when enough of it becomes available. For the time being, however, the segmented data must somehow be joined together, one after the other, in order to provide approximate end-to-end data.

Once the seasonal adjustments are made on the segmented data (i.e. the Marine transit, Port dwelling, and Rail transit time data), we construct an **aggregate indicator** (or index) using these seasonally adjusted segmented indicators as a (necessarily poor) substitute for a direct end-to-end indicator, the underlying argument being that the total fluctuation for the supply chain should be the sum of the fluctuations from the segments since the supply chain as a whole is made up of the individual transportation modes.

Hence, the aggregate index is conceptualized as a weighted average of the indices of the component transportation modes:

$$I_t^A = \frac{\sum_j I_{j,t} \times W_{j,t}}{\sum_j W_{j,t}},$$

where

- I_t^A is the aggregate index for the entire supply chain at time t;
- $I_{j,t}$ is the component index for the specific transportation mode (or segment) j at time t (one of RI, BI or CV, say), and
- $W_{j,t}$ is the weight assigned to mode j at time t (the weights must be internally consistent from mode to mode in order for the weighted average to have meaning).

For a given supply chain, the average transit time in each mode is considered a good candidate for the weights $W_{j,t}$ since it functions as a reflection of the importance of the specific transportation mode to the entire supply chain, and since the average transit time of the entire supply chain is the sum of the average transit time of the individual component transportation modes. Had financial data been available, the **value-added** (the product of quantity and price) would also have been a good choice.

In the absence of financial data, however, we may suppose that the cost of a container spending a certain amount of time in a given mode is proportional to the amount of time spent in that mode (with the understanding that the proportionality constant may differ from mode to mode); as such, it makes sense to use $W_{j,t} = q_{j,t} \times T_{j,t}$ where

- $q_{j,t}$ is the quantity of containers through mode j at time t, and
- $T_{j,t}$ is the average amount of time spent in mode j at time t.

With these assumptions, the aggregate index is eventually defined as

$$I_t^A = \frac{\sum_j I_{j,t} \times q_{j,t} \times T_{j,t}}{\sum_j q_{j,t} \times T_{j,t}},$$

although it is important to note that there is no easy way to validate this formula without end-to-end data.

After seasonality adjustments, it also became possible to compare the performance of (and hence to attempt to differentiate) the various indicators (RI, BI, CV) on a segment-by-segment basis:

- *Marine Transit* – all indicators show increasing trends in the Shanghai → Vancouver channel, and they all identify an outlier for MAY2011 in the Shanghai → Prince Rupert. In both channels, RI had a less volatile seasonal component, while BI had a less abnormally irregular component. It was thus not possible to cleanly rank RI and BI, but the adjusted data suggested that CV would be a poor selection as the indicator of choice.
- *Port Dwelling* – BI was seen to be the less volatile of all indicators, in both Vancouver and Prince Rupert.
- *Rail Transit* – RI was shown to be less volatile in the Prince Rupert → Toronto channel, but not enough data was available to come to a conclusion in the Vancouver → Toronto channel.

The importance of eventually collecting end-to-end data was made clear, as no clear-cut consensus for all segments emerged, apart from the unsuitability of CV as an indicator.

A number of supply chain scoring metrics (scaled scores, comparison scores) were also provided (pitting the adjusted expected data against the actual data in different ways), but before 5-years' worth of data is available, it is a somewhat artificial endeavour to select the optimal one.

ISSUES

The short timeline allocation (see Project Logistics, below) was a consequence of a now-discontinued internship program, in which promising graduate students were hired as interns by CQADS to work on small projects and paid at a discounted rate in exchange for course credit. Consequently, the project dollar value (see Project Logistics, below) was about a fourth as large as it would have been under normal conditions.

Under these same normal conditions, the Centre would have negotiated a 3-month period over which to complete the project, rather than the agreed-upon 4 weeks. The total level of effort would not have changed.

Furthermore, some unforeseen data quality issues emerged (with respect to the Port Dwelling time in Vancouver) and as a result, the deadline was pushed back, upon mutual agreement.

RESULTS AND RELEVANCE

The suggested scoring methodology provided TC's Economic Analysis and Research (EAR) group with a basis for implementing seasonality identification, and compensation methods. It is also known that the final report was circulated by EAR to a select group of academic and industry contacts.

It eventually made its way into the hands of Prof. Ata Khan, of the Faculty of Engineering and Design at Carleton University, who was impressed with the work and as a result enlisted CQADS's assistance for a study on the transportation of dangerous goods on behalf of the Nuclear Waste Management Organization.

PROJECT LOGISITCS

Timeline 10-May-13 to 12-Jun-13

(the original deadline of 07-Jun-13 was pushed back upon mutual agreement from both parties, given unexpected data issues)

Resources/Personnel Patrick Boily, Ph.D.
Managing Consultant, CQADS
Project Lead / SME / Senior Analyst

Yue Huang
Consultant, CQADS
Analyst (Time Series Analysis and Forecasting)

Total Effort Level

250 hours (estimate)	**Boily**	**Huang**
Total:	125	125

Dollar Value $4,424.73 (+ HST)

TC Project Authority Alexander Gregory
Economic Analyst

Economic Analysis and Research
4900 Young Street
North York, Ontario M2N 6A5
Canada

416 973-2444
alexander.gregory@tc.gc.ca