Time Series Analysis with SAS® and R

Samuel T. Croker

General Ideas

R for Time series analysis when…

- SAS/ETS® is not available
- Time series techniques are needed that SAS/ETS does not handle (fractional differencing, wavelet decomposition, etc…)
- Exploratory data analysis

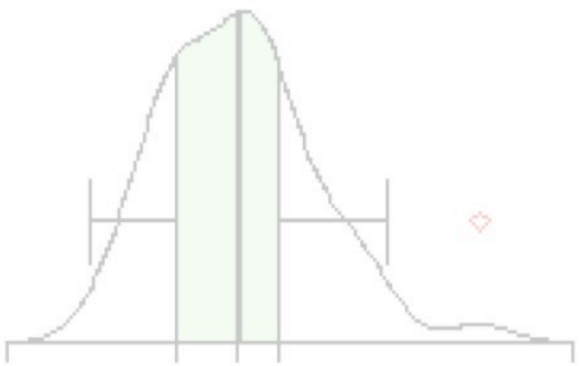
Probably not when…

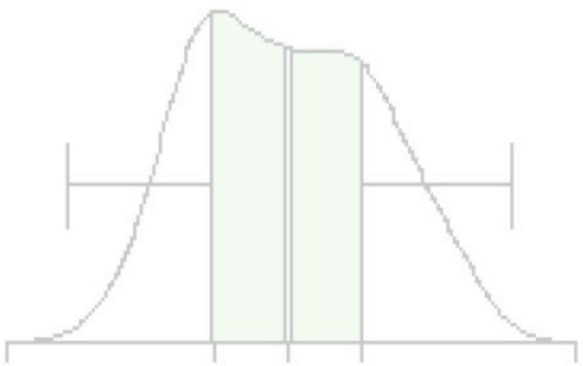
- Production environment
- Large quantities of data, many time series

What is R?

"R is a free software environment for statistical computing and graphics." – www.r-project.org

- Powerful mathematical, statistical and graphical problem solver
- Provides access to bleeding edge techniques
- Relatively easy to learn
- Data management and manipulation is not so easy
- Free!





General Techniques

- Use Base SAS for data preparation
- Use SAS to write and call R program
- Use called R program to perform the time series analysis
- Return the output and results to SAS
- Use SAS for final reporting

Data Preparation with SAS

- Extract, transform data using SAS
- Write out a text file of the prepared data
  - R can read delimited files fairly easily
  - Store data in databases (more later)

Extracting USGS Streamflow Data

```
filename congtrez url 'http://waterdata.usgs.gov/
   nwis/uv?
   cb_00065=on&format=rdb&period=31&site_no=02169810'
   ;
data congaree_trez;
   infile congtrez dlm='09'x;
   length agency $10 site $10 obsdatetime 8 stage 8;
   informat obsdatetime anydtdtm16.;
   format obsdatetime datetime28. obsdate mmddyy10.;
   input agency @;
   if agency ~= 'USGS' then delete;
   input site $ obsdatetime stage;
   obsdate=datepart(obsdatetime);
   obshour=hour(obsdatetime);
run;
```

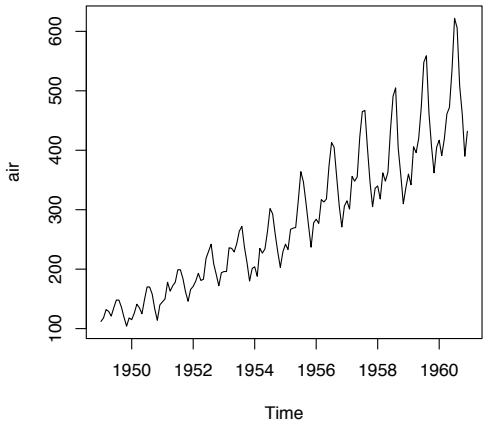
Continue Data Manipulation

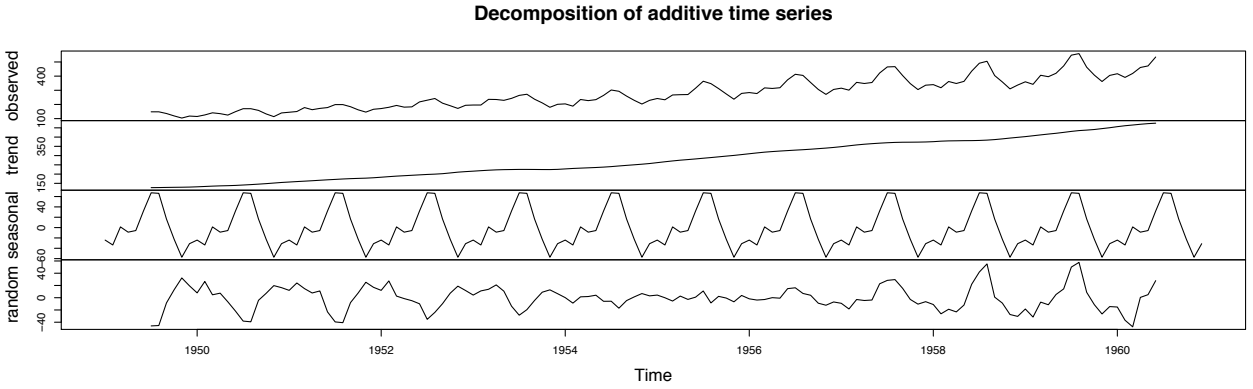
```
proc means data=congaree_trez nway noprint;
   class agency site obsdate obshour;
   output out=congaree_trez_hourly_avg
   mean(stage)=hourly_mean_stage;
run;

%let rsourcepath=c:\r_out;

data _null_;
   set congaree_trez_hourly_avg;
   /*write out only the hourly data for simplicity - you can make
   this as detailed as you want*/
   file "&rsourcepath.\congaree_trez_hourly.dat";
   put hourly_mean_stage;
run;
```

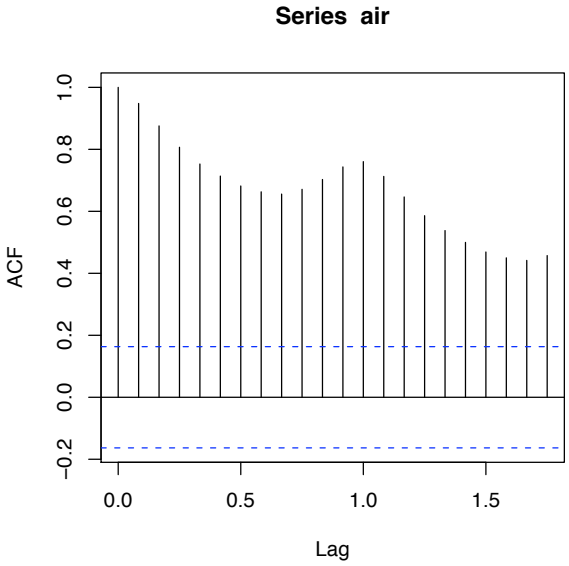
The R Program

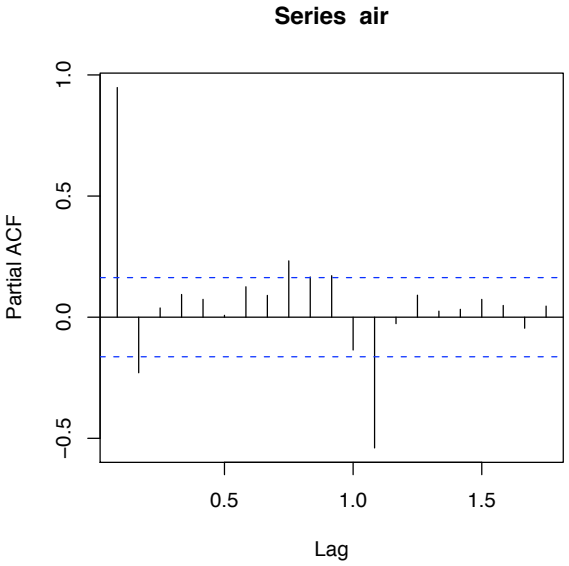
```
library(forecast)
cong_trez <- read.table('c:\\r_out\\congaree_trez_hourly.dat')
cong_trez.ts <- ts(cong_trez)
cong_trez.fit <- arima(cong_trez.ts,order=c(1,1,0))
png(filename="c:\\r_out\\plot.png ")
tsdiag(cong_trez.fit,6)
dev.off()
cong_trez.fcst<-forecast.Arima(cong_trez.fit)
png(filename="c:\\r_out\\fcst.png ")
plot.forecast(cong_trez.fcst)
dev.off()
png(filename="c:\\r_out\\acf.png ")
acf(cong_trez.ts)
dev.off()
png(filename="c:\\r_out\\pacf.png ")
pacf(cong_trez.ts)
dev.off()
png(filename="c:\\r_out\\spect.png ")
spectrum(cong_trez.ts)
dev.off()
```

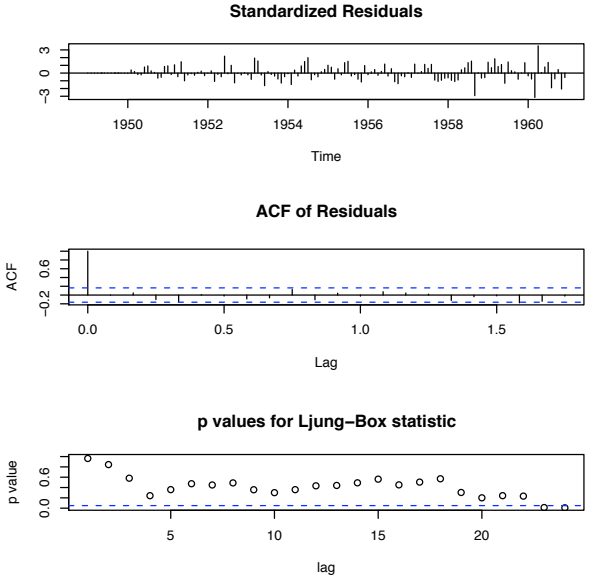
R Time Series Analysis











Building the R Program

```
data _null_;
    file "&rsourcepath.\tsdiag.r";
    fcst=tranwrd("&rsourcepath\fcst.png",'\','\\');
    diag=tranwrd("&rsourcepath\plot.png",'\','\\');
    spect=tranwrd("&rsourcepath\spect.png",'\','\\');
    acf=tranwrd("&rsourcepath\acf.png",'\','\\');
    pacf=tranwrd("&rsourcepath\pacf.png",'\','\\')
    put "library(forecast)";
    put "cong_trez <- read.table('c:\\r_out\\congaree_trez_hourly.dat');
    put 'cong_trez.ts <- ts(cong_trez)';
    put 'cong_trez.fit <- arima(cong_trez.ts,order=c(1,1,0))';
```

<<continued>>

Building the R Program

```
 /* redirect graphs to a png file */
    put 'png(filename="' diag '")';
    put 'tsdiag(cong_trez.fit,6)';
    put 'dev.off()';
    put 'cong_trez.fcst<-forecast.Arima(cong_trez.fit)';
    put 'png(filename="' fcst '")';
    put 'plot.forecast(cong_trez.fcst)';
    put 'dev.off()';
    put 'png(filename="' acf '")';
    put 'acf(cong_trez.ts)';
    put 'dev.off()';
    put 'png(filename="' pacf '")';
    put 'pacf(cong_trez.ts)';
    put 'dev.off()';
    put 'png(filename="' spect '")';
    put 'spectrum(cong_trez.ts)';
    put 'dev.off()';
run;
```

Getting the Data Into R

```
library(forecast)
cong_trez <- read.table('c:\\r_out\\congaree_trez_hourly.dat')
cong_trez.ts <- ts(cong_trez)
```

Executing the R Program

```
options xwait xsync;

x "'C:\Program Files\R\R-2.7.1\bin\r.exe'
     --no-save --no-restore <""&rsourcepath\tsdiag.r"">
     ""&rsourcepath\tsdiag.out""";
```

Including the R Log in the SAS Log

```
data _null_;
   infile "&rsourcepath\tsdiag.out";
   file log;
   input;
   put 'R LOG: ' _infile_;
run;
```

```
R LOG:
R LOG: R version 2.7.1 (2008-06-23)
R LOG: Copyright (C) 2008 The R Foundation for Statistical Computing
R LOG: ISBN 3-900051-07-0
R LOG:
R LOG: R is free software and comes with ABSOLUTELY NO WARRANTY.
R LOG: You are welcome to redistribute it under certain conditions.
R LOG: Type 'license()' or 'licence()' for distribution details.
R LOG:
R LOG:   Natural language support but running in an English locale
R LOG:
R LOG: R is a collaborative project with many contributors.
R LOG: Type 'contributors()' for more information and
R LOG: 'citation()' on how to cite R or R packages in publications.
R LOG:
R LOG: Type 'demo()' for some demos, 'help()' for on-line help, or
R LOG: 'help.start()' for an HTML browser interface to help.
R LOG: Type 'q()' to quit R.
R LOG:
R LOG: > library(forecast)
R LOG: This is forecast 1.13
R LOG: > cong_trez <- read.table('c:\\r_out\\congaree_trez_hourly.dat')
R LOG: > dev.off()
NOTE: 23 records were read from the infile "c:\r_out\tsdiag.out".
      The minimum record length was 0.
      The maximum record length was 64.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

Getting the R Results in the SAS Output

```
ods html;
data _null_;
    file print;
    put "<IMG SRC='" "&rsourcepath\plot.png" "' border='0'>";
    put "<IMG SRC='" "&rsourcepath\acf.png" "' border='0'>";
    put "<IMG SRC='" "&rsourcepath\pacf.png" "' border='0'>";
    put "<IMG SRC='" "&rsourcepath\spect.png" "' border='0'>";
    put "<IMG SRC='" "&rsourcepath\fcst.png" "' border='0'>";
run;
ods html close;
```

Other Ideas

- R RODBC package provides ODBC access to databases via SQL queries
- read.csv is a read.table wrapper for CSV files

Questions

Thanks!

Full code and data in proceedings and on
www.sascommunity.org

Feel free to contact me with questions

Sam Croker
– scoyote@scoyote.net