Econometric Analysis and Forecasting of Particulate Matter Concentration, Temperature, and Precipitation:

Comparative Forecasting Methods

KEITH ARANEO-YOWELL

UNIVERSITY OF UTAH

MASTER OF STATISTICS - ECONOMETRICS

FINAL PROJECT

Background

- Upward trend in global average temperature will have far reaching economic effects
- Winter Temperature inversion in the Salt Lake Valley (SLV) means over 35% of Utah residents regularly experience poor to dangerous air quality
- Associated with detrimental long term health effects
  - Asthma, heart and lung disease
- Has been shown to reduce worker productivity, while increasing missed school and work days, as well as ER visits, and hospital admissions
- It is yet unclear if climate change is expected to mitigate or exacerbate pollution events and therefore they are the focus of study

Forecasting

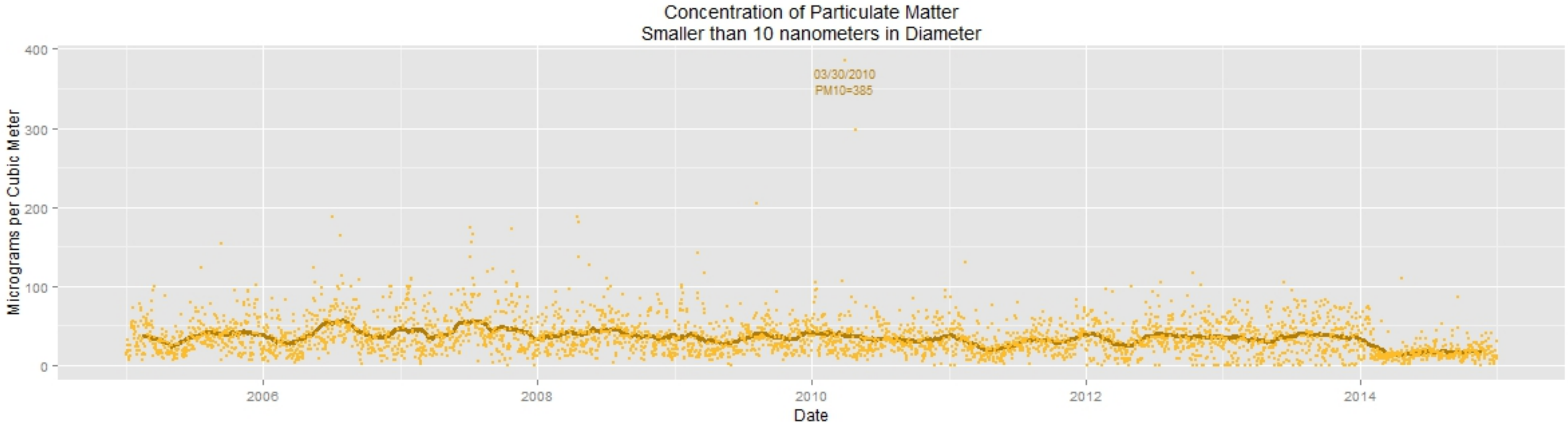
- Ultimately this project is a comparative study of macroeconomic forecasting methods and how they perform outside of their own wheelhouse
- The selection of our final model allows us to explore all our selected methods on multiple climate variables
- We test the macro time series forecasting methods of Autoregressive Integrated Moving Average (ARIMA), Vector Autoregression, and finally A Bayesian Vector Autoregression with the Minnesota Prior

Data

- Daily time series of climatological measures from January 1, 2008 to December 31,2016
- To keep the project manageable we limited the project to only three variables I thought may have strong interrelation with the main pollution variable of interest

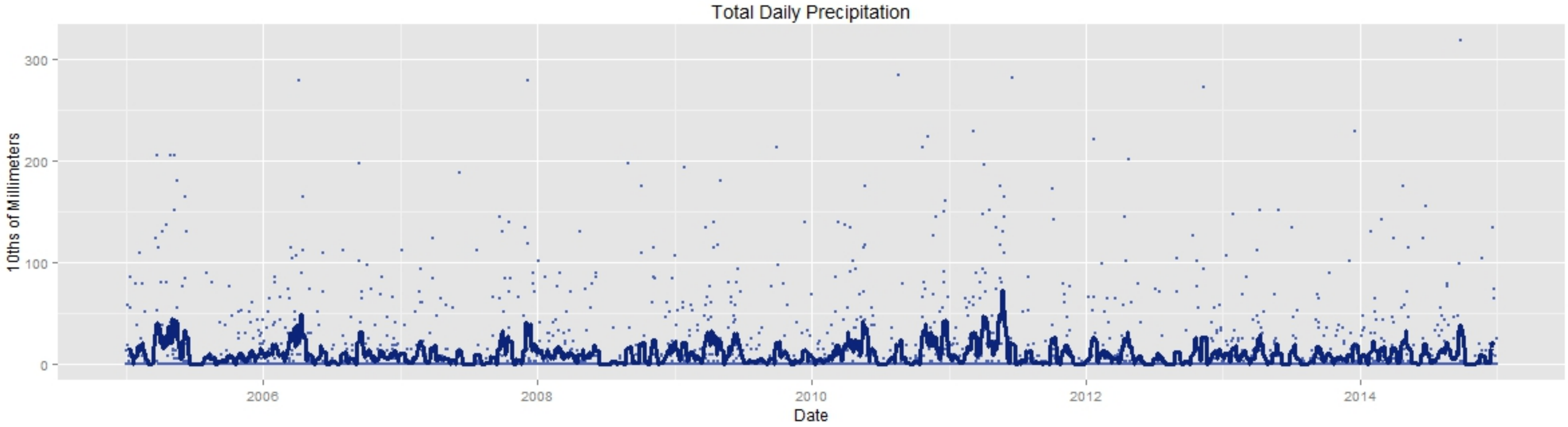
PM10

- Particulate Matter Less 10 nanometers in diameter
- Data were collected from the department of Environmental Quality’s air monitoring station at 1675 S. 660 East in Salt Lake City
- The measure is the concentration measured in the number of micrograms per cubic meter
  - At Standard Temperature and Pressure



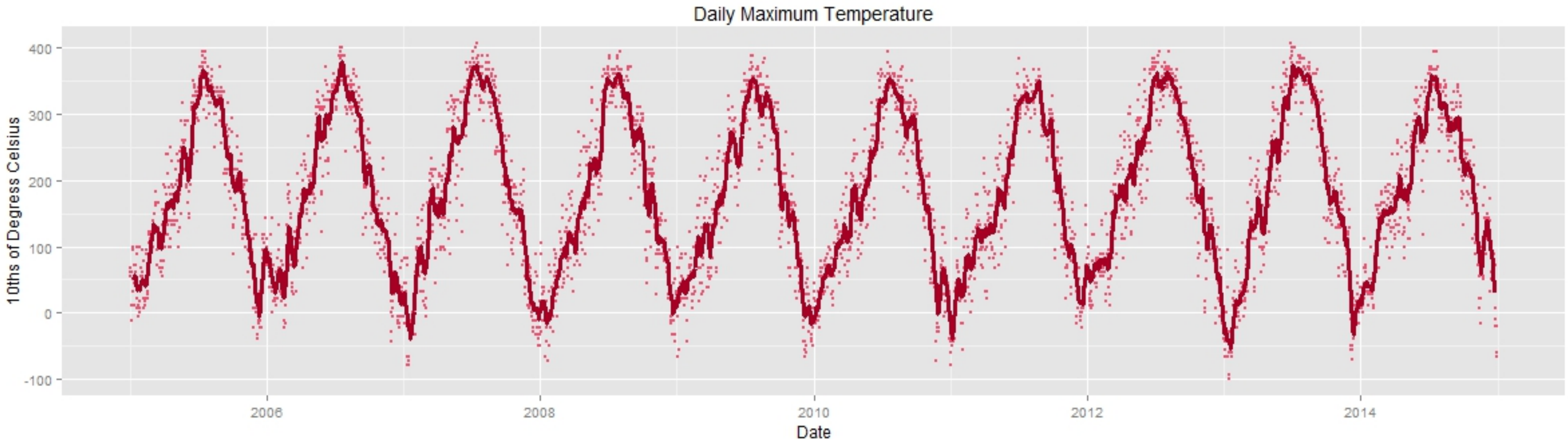
Precipitation

- Total Daily Precipitation in 10ths of millimeters
- Data were collected from the National Oceanic and Atmospheric Administration’s National Climate Data Center Online
- Data were measured at NOAA’s ground based weather station at the Salt Lake City International Airport



Temperature

- Total Daily Maximum Temperature in tenths of degrees celsius
- Data were collected from the National Oceanic and Atmospheric Administration's National Climate Data Center Online
- Data were measured at NOAA's ground based weather station at the Salt Lake City International Airport



Stationarity

- $y_t = \rho y_{t-1} + u_t$
- $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$
  - $\gamma = (\rho - 1)$
- $H_0\text{: } \gamma = 0$
  $H_a\text{: } \gamma < 0$
- OLS estimate for γ
- If $\gamma = 0$, the y_t process includes a unit root. When we estimate the value of γ and its associated standard error, and compare the resulting t-statistic against the critical τ value of the Dickey Fuller table to the right

Dickey-Fuller Unit Root Test for Daily Series					
	γ-hat	Standard Error	τ-Statistic	Critical Value	Rule
PM10	-0.1745	0.009345	-18.67308721	-1.95	REJECT
Precipitation	-0.7277	0.01593	-45.68110483	-1.95	REJECT
Temperature	-0.0166	0.00301	-5.514950166	-1.95	REJECT

ACFs and PACFs

- We can guess at the proper p, q, P, & Q parameters of the ARIMA specification by plotting a function of auto correlation and partial autocorrelation coefficients by by lag order
- The Autocorrelation coefficient indicates the degree to which all observations between the current time and that of k-periods before correlate
- The Partial Autocorrelation coefficient indicates the degree to which *only* the observation k-periods before influences the current observation

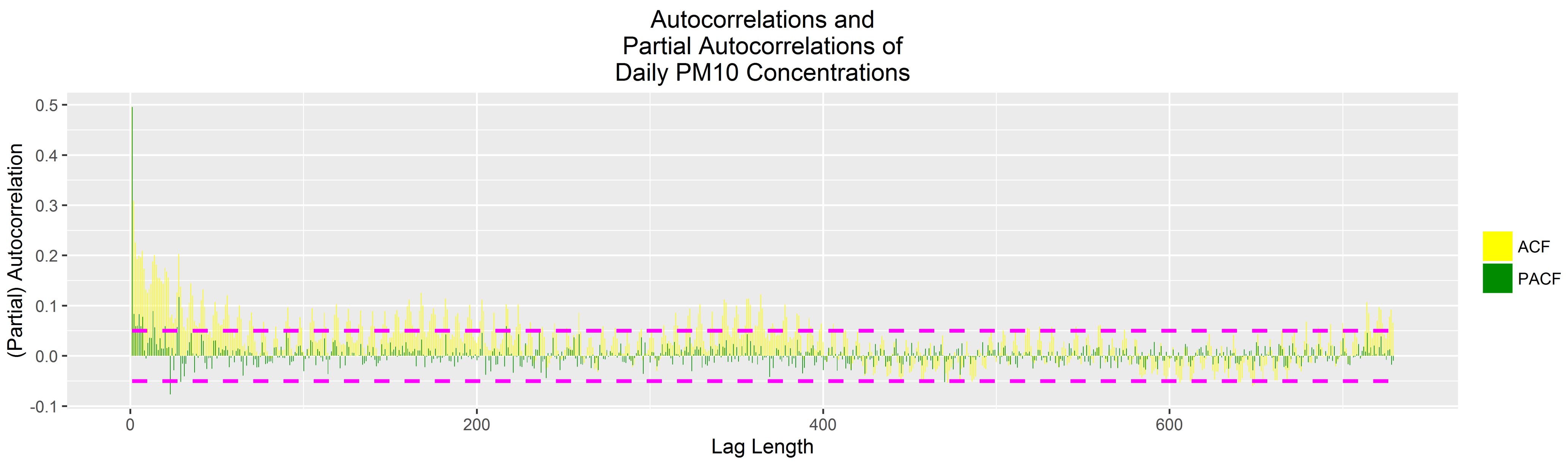
Autocorrelation

$$\rho_k = Corr(y_t, y_{t-k}) = \frac{cov(y_t, y_{t-k})}{var(y_t)}$$

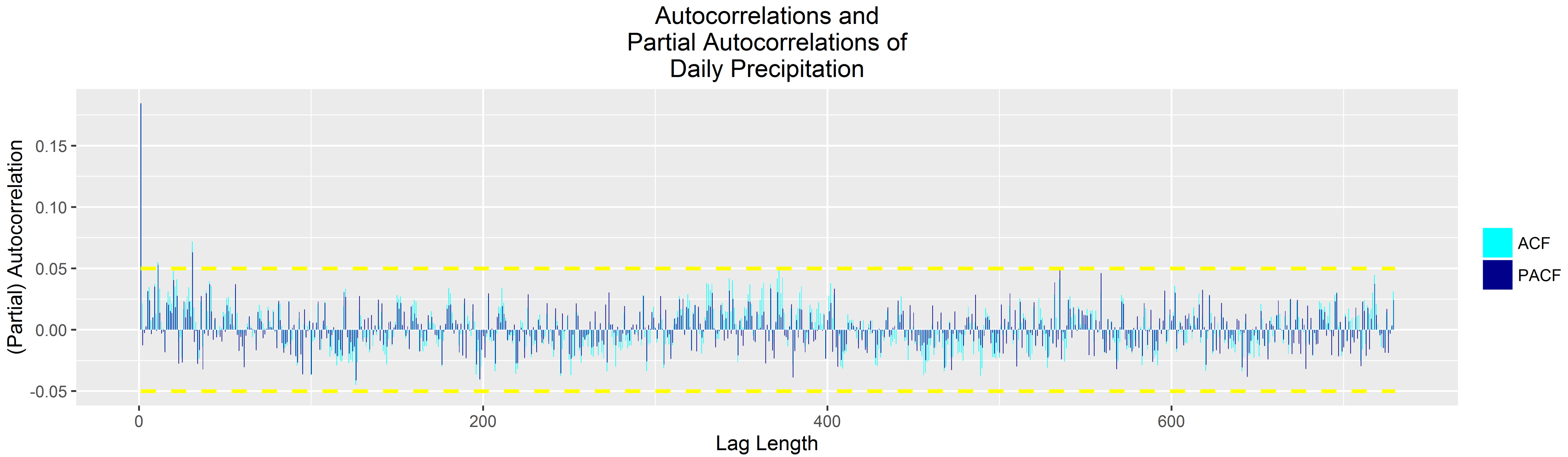
Partial Autocorrelation

$$\varphi_k = \frac{cov(y_t, y_{t-k} | y_{t-1}, y_{t-2}, \dots, y_{t-k+1})}{\sqrt{var(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-k+1}) var(y_{t-k} | y_{t-1}, y_{t-2}, \dots, y_{t-k+1})}}$$

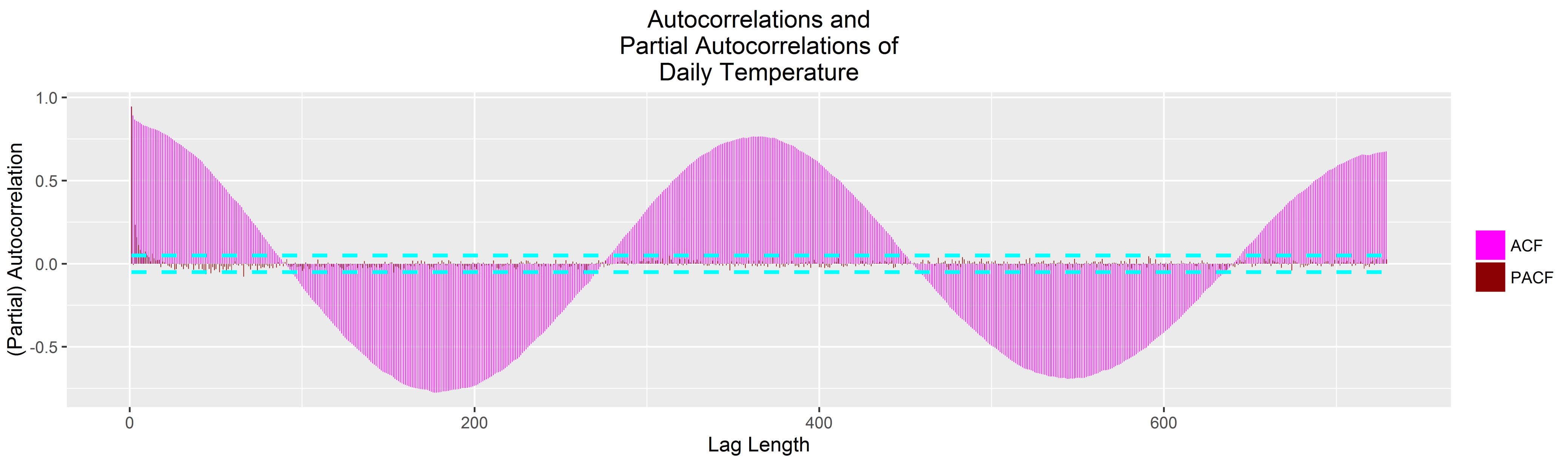
ACFs and PACFs of PM10 Series



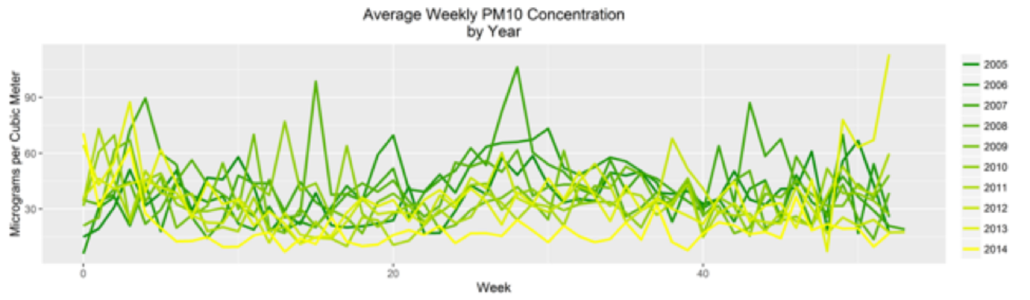
ACFs and PACFs of Precipitation Series

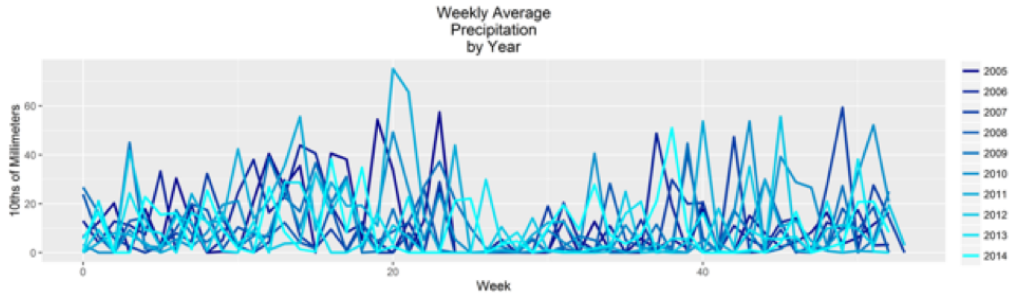


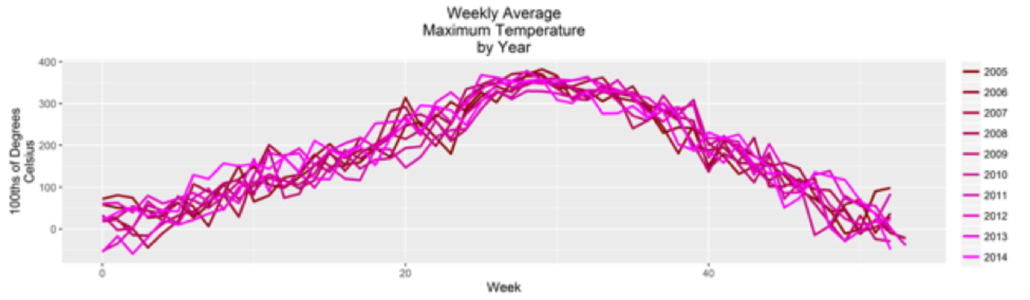
ACFs and PACFs of Temperature Series



Seasonality







Seasonal Arima Model

SEASONALITY

- Any Seasonality can be dealt with by differencing
- If we're taking a difference of any series we first define $y = \log(time\ series)$
- We then assess different multiplicative ARIMA(p,d,q)X(P,D,Q)$_s$ specifications by their Akaike, or Bayes Criteria
  - Very similar to standard ARIMA(p,d,q), where P,D, and Q are seasonal Autoregressive, Integrated, and Moving Average terms, and s is the number of time slices in a season

ARIMA(P,D,Q)X(P,D,Q)$_S$

- $$\Delta^2 y_t = y_t - y_{t-1}$$
  $$\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2}$$
  - This, recall is the first difference for a daily series

$$\Delta_{s=365}^{D=2} y_t = \Delta_{356}(y_t - y_{t-365})$$

- $= y_t - y_{t-365} - (y_{t-365} - y_{t-730})$
  $$= y_t - 2y_{t-365} + y_{t-730}$$
  - This is the *seasonal* difference for a daily series
- $y_{sfd} = (1 - L)(1 - L^{365})y_t$
  - This is the First Seasonal Difference multiplication of the two in lag notation

PM10 ARIMA Specification

$$ARIMA(1,1,2)\times(0,1,0)_{365}$$

$$\Delta\Delta_{365}\mathrm{P}M10_t = \alpha_1 PM10_{t-1} + \gamma_1\varepsilon_{t-1} + \gamma_2\varepsilon_{t-2} + \varepsilon_t$$

$$\Delta\Delta_{365}\mathrm{P}M10_t = \underset{(15.6021)}{.2902} PM10_{t-1} - \underset{(-17.89)}{1.95}\varepsilon_{t-1} + \underset{(133.8)}{.9519}\varepsilon_{t-2} + \varepsilon_t,$$

$$AIC = 29348.37$$

Precipitation ARIMA Specification

$$ARIMA(1,0,0)\times(0,1,1)_{365}$$

$$\Delta_{365}Precip_t = \alpha_1 Precip_{t-1} + \mathrm{B}_1\varepsilon_{t-365} + \varepsilon_t$$

$$\Delta_{365}Precip_t = \underset{(9.91)}{.1764}Precip_{t-1} - \underset{(-221.57)}{.9954}\varepsilon_{t-365}, \qquad AIC = 31578.46$$

Temperature ARIMA Specification

$$ARIMA(2,1,1)\times(1,1,1)_{365}$$

$$\Delta\Delta_{365}Temp_t = \alpha_1 Temp_{t-1} + \alpha_2 Temp_{t-2} + \beta_1\varepsilon_{t-1} + \mathrm{A}_1 Temp_{t-365} + \mathrm{B}_1\varepsilon_{t-365}$$

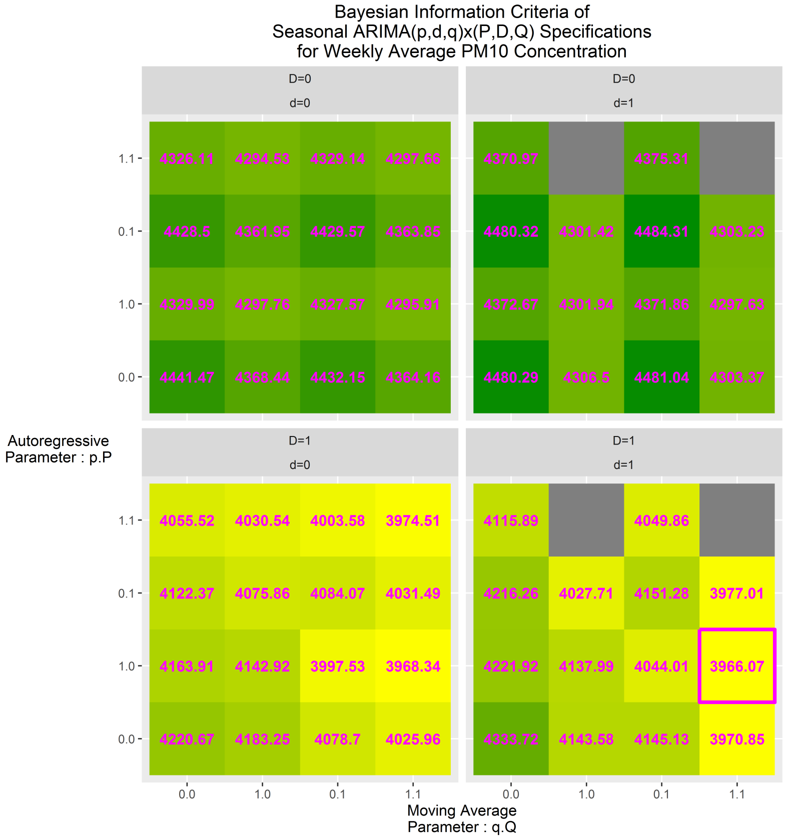
$$\Delta\Delta_{365}Temp_t = \frac{.2954}{(.012)} Temp_{t-1} - \frac{.1102}{(-4.61)} Temp_{t-2} - \frac{.9445}{(-66.04)} \varepsilon_{t-1} + \frac{.04672}{(5.697)} Temp_{t-365} - \frac{.9445}{(66.04)} \varepsilon_{t-365}$$

$$AIC = 32970.33$$

Weekly Data

- Due to software constraints, seasonal models with daily data are infeasible, as such, half way through the data were aggregated to weekly averages and the models re-estimated
- We wrote an algorithm to select the ARIMA specification with the best post estimation residual fit based on the Schwartz-Bayesian Criterion
- $BIC = T\ln(SSR) + n\ln(T)$
  - Where n is the number of parameters estimated (p+q+P+Q) and T is the number of usable observations

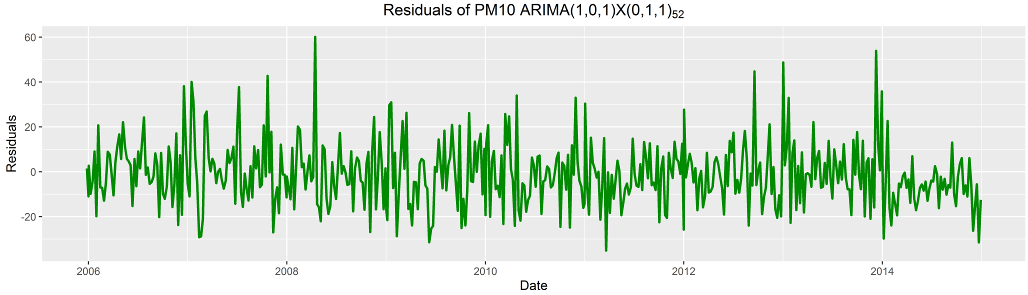
Best PM10 Weekly Specification



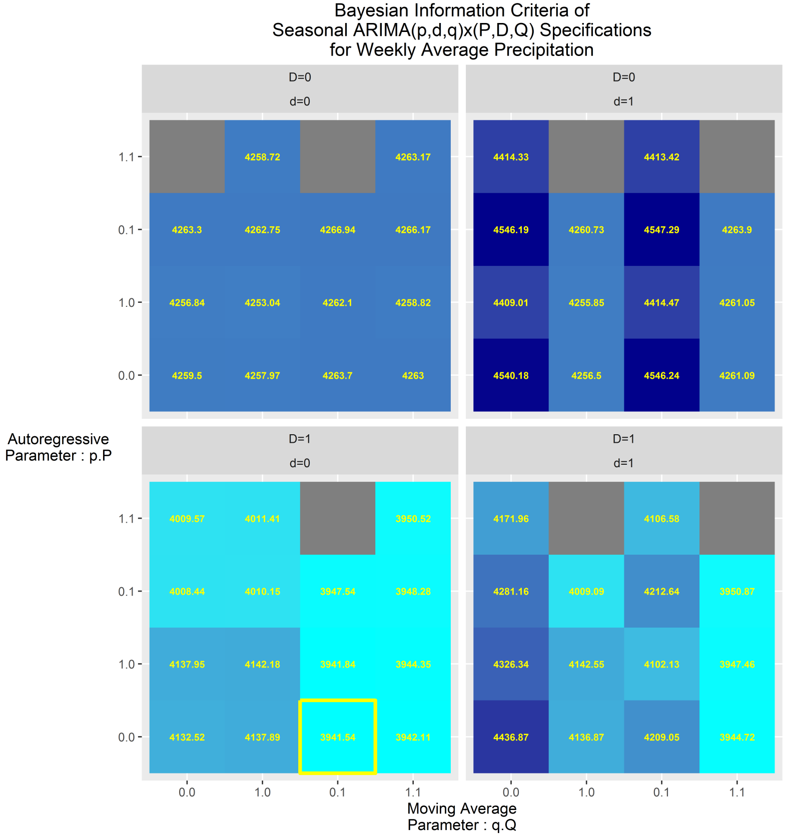
$$\boldsymbol{ARIMA(1,1,1)\times(0,1,1)_{52}}$$

$$\Delta\Delta_{52}PM10_t = a_1 PM10_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + B_1\, \varepsilon_{t-52}$$

$$\Delta\Delta_{52}PM10_t = \frac{0.2919}{(5.84)} PM10_{t-1} + \varepsilon_t - \frac{0.938}{(-37.37)}\varepsilon_{t-1} - \frac{0.7268}{(-16.11)}\varepsilon_{t-52}$$



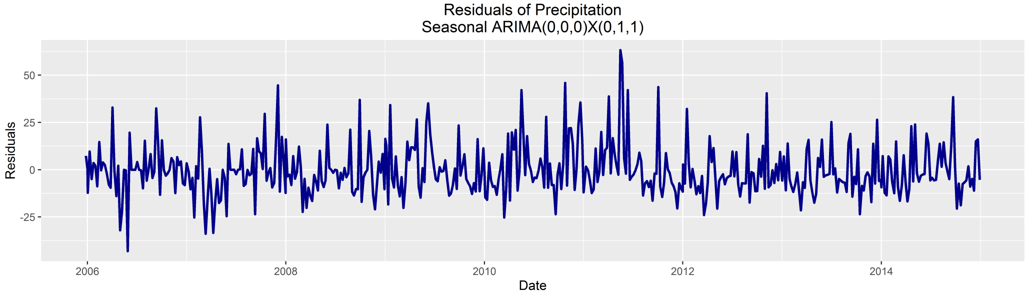
Best Precipitation Weekly Specification



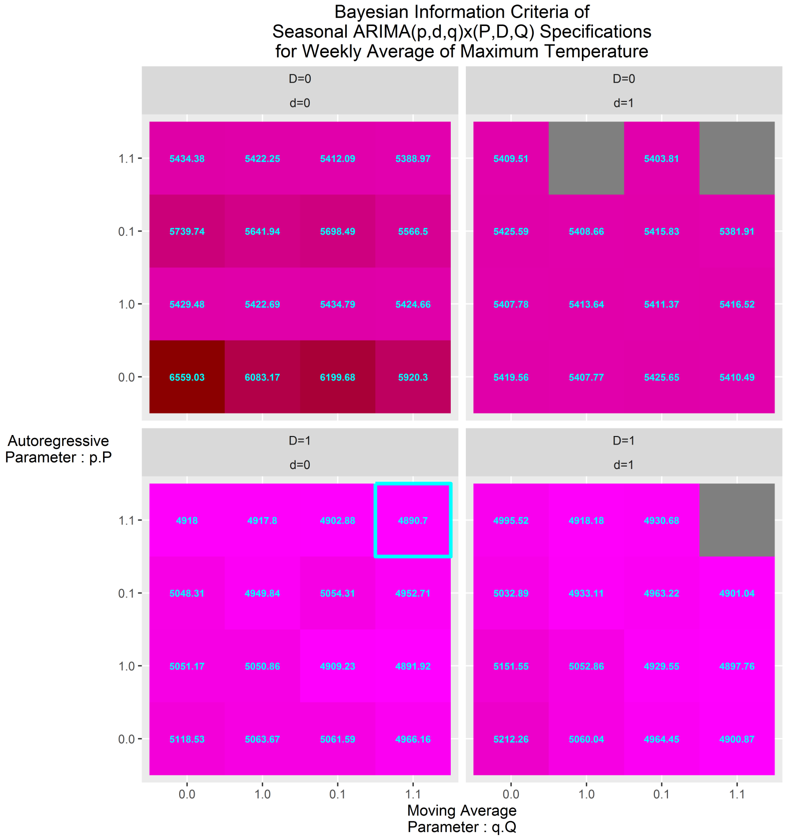
$$ARIMA(0,0,0)\times(0,1,1)_{52}$$

$$\Delta_{52}\text{Precip}_t = \varepsilon_t + B_1\ \varepsilon_{t-52}\Delta_{52}\text{Precip}_t = \varepsilon_t + B_1\ \varepsilon_{t-52}$$

$$\Delta_{52}\text{Precip}_t = \varepsilon_t - \frac{0.8823}{(-10.08)}\varepsilon_{t-52}$$



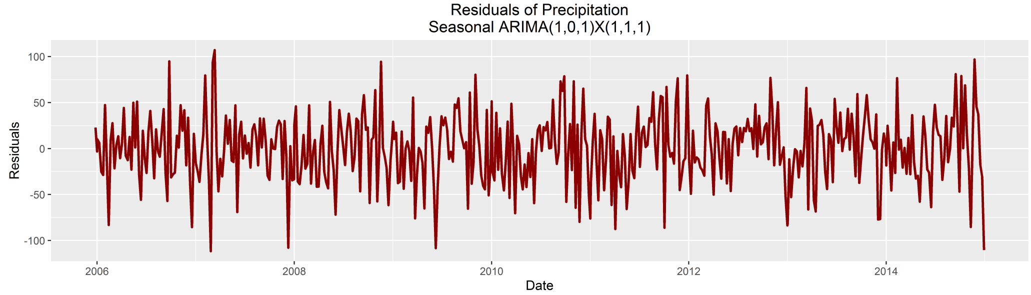
Best Temperature Weekly Specification



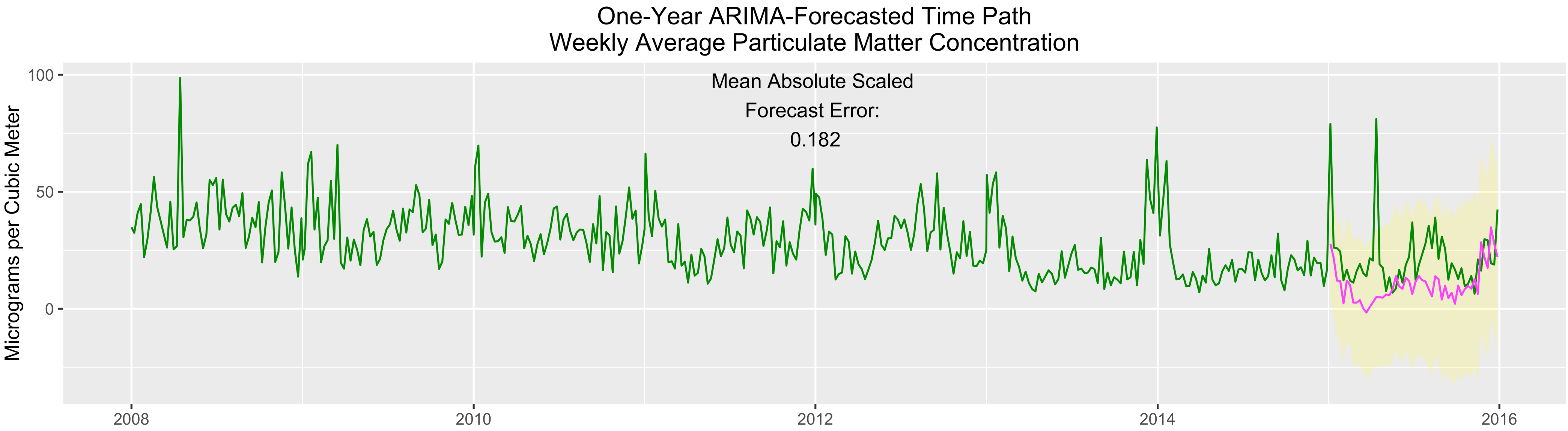
$$\boldsymbol{ARIMA(1,0,1)\times(1,1,1)_{52}}$$

$$\Delta_{52}\mathrm{Temp}_t = a_1 Temp_{t-1} + A_1 Temp_{t-52} + \varepsilon_t + \beta_1\varepsilon_{t-1} + B_1\,\varepsilon_{t-52}$$

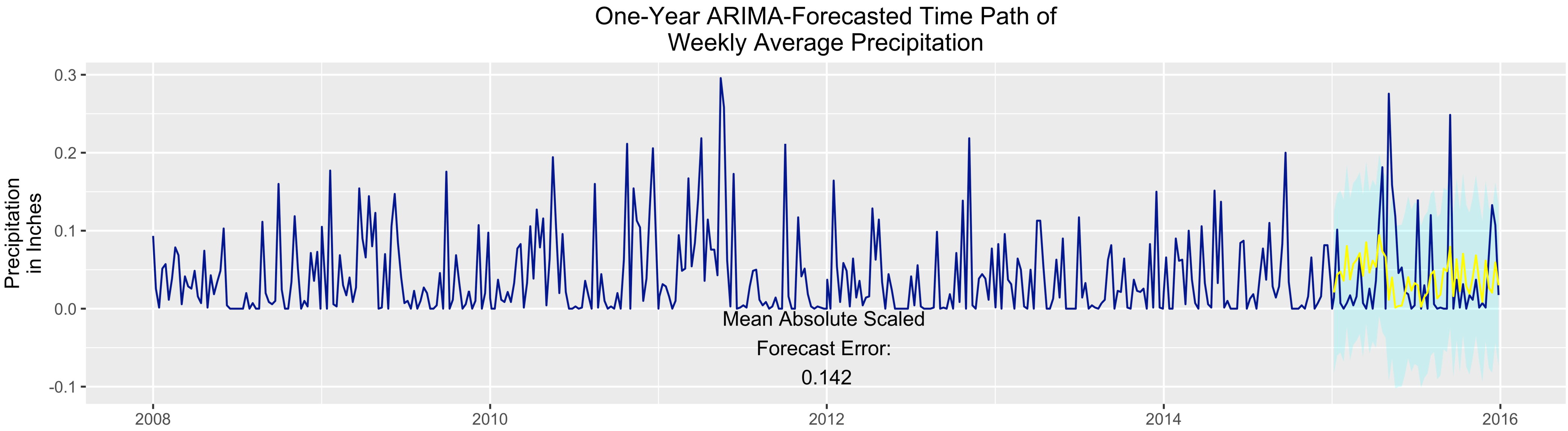
$$\Delta_{52}\mathrm{Temp}_t = \frac{0.8577}{(21.55)}Temp_{t-1} - \frac{0.2106}{(-2.79)}Temp_{t-52} + \varepsilon_t - \frac{0.3897}{(-5.04)}\varepsilon_{t-1}\frac{0.5631}{(-6.67)}\varepsilon_{t-52}$$



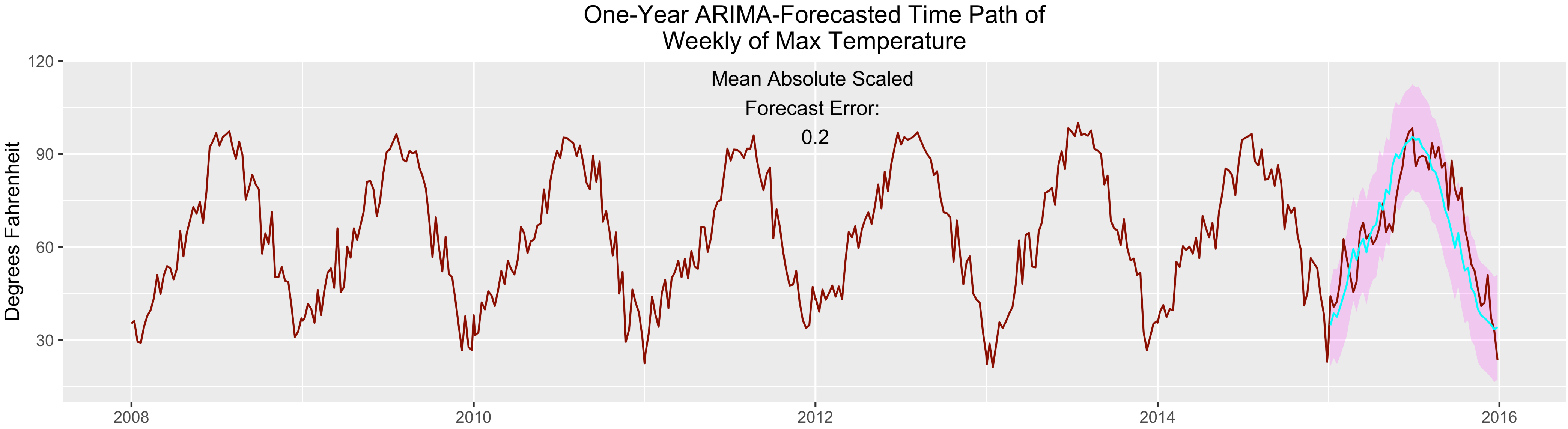
Weekly Best ARIMA Forecast PM10



Weekly ARIMA Forecast Precipitation



Weekly ARIMA Forecast Temperature



Vector Autoregression

- VAR is a multivariate generalization of the univariate AR

$$y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + e_t$$

$$E[e_t] = 0$$

$$\mathrm{Var}[e_t] = \Sigma = E[e_t{}' e_t] = \begin{matrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_k^2 \end{matrix}$$

$$E[e_t{}' e_s] = 0\,;\forall\, s \neq t$$

in scalar form :

$$y_{i,t} = c_i + a_{i,1}^1 y_{1,t-1} + \cdots + a_{i,k}^1 y_{k,t-1} + \cdots + a_{i,1}^p y_{1,t-p} + \cdots + a_{i,k}^p y_{k,t-p} + e_{i,t}$$

In matrix notation:

$$Y = BZ + U$$

$Y = BZ + U$

$$Y = \left[y_p, y_{p+1}, \dots, y_T\right] = \begin{pmatrix} y_{1,p} & y_{1,p+1} & \cdots & y_{1,T} \\ y_{2,p} & y_{2,p+1} & \cdots & y_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,p} & y_{k,p+1} & \cdots & y_{k,T} \end{pmatrix}$$

$$B = [c\ A_1 A_2 \cdots A_p] = \begin{pmatrix} c_1 & a_{1,1}^1 & a_{2,1}^1 & \cdots & a_{1,k}^1 & \cdots & a_{1,1}^p & a_{1,2}^p & \cdots & a_{1,k}^p \\ c_2 & a_{2,1}^1 & a_{2,2}^1 & \cdots & a_{2,k}^1 & \cdots & a_{2,1}^p & a_{2,2}^p & \cdots & a_{2,k}^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ c_k & a_{k,1}^1 & a_{k,2}^1 & \cdots & a_{k,k}^1 & \cdots & a_{k,1}^p & a_{k,2}^p & \cdots & a_{k,k}^p \end{pmatrix}$$

$Y = BZ + U$ Ctd.

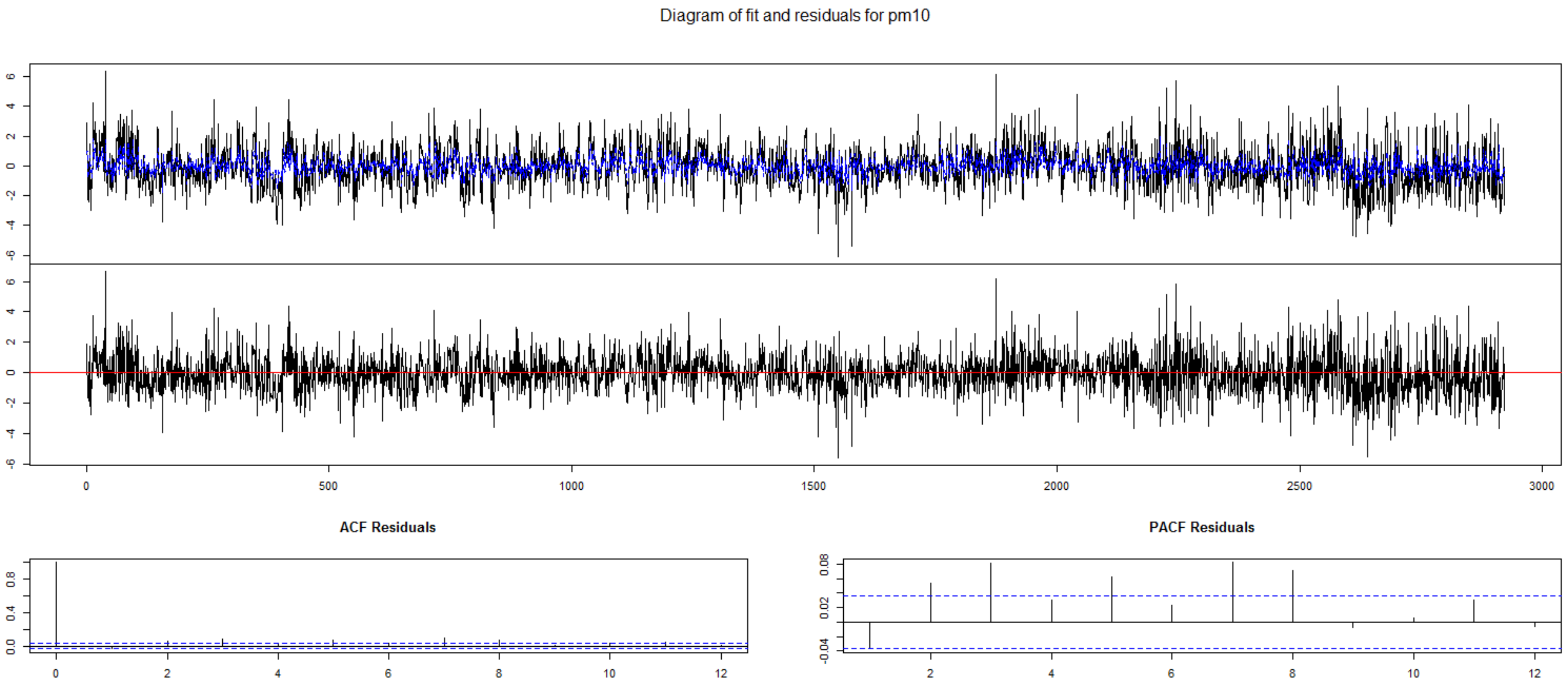
$$Z = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ y_{p-1} & y_p & \cdots & y_{T-1} \\ y_{p-2} & y_{p-1} & \cdots & y_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_0 & y_1 & \cdots & y_{T-p} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ y_{1,p-1} & y_{1,p} & \cdots & y_{1,T-1} \\ y_{2,p-1} & y_{2,p} & \cdots & y_{2,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,p-1} & y_{k,p} & \cdots & y_{k,T-1} \\ y_{1,p-2} & y_{1,p-1} & \cdots & y_{1,T-1} \\ y_{2,p-2} & y_{2,p-1} & \cdots & y_{2,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,p-2} & y_{k,p-1} & \cdots & y_{k,T-2} \\ y_{1,p-3} & y_{1,p-2} & \cdots & y_{2,T-3} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ y_{1,0} & y_{1,0} & y_{1,0} & y_{1,T-p} \\ y_{2,0} & y_{2,0} & y_{2,0} & y_{2,T-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,0} & y_{k,1} & \cdots & y_{k,T-p} \end{pmatrix}, U = [\varepsilon_p, \epsilon_{p+1}, \cdots, \varepsilon_T] = \begin{pmatrix} \epsilon_{1,p} & \epsilon_{1,p+1} & \cdots & \varepsilon_{1,T} \\ \varepsilon_{2,p} & \epsilon_{2,p+1} & \cdots & \varepsilon_{2,T} \\ \cdots & \cdots & \ddots & \vdots \\ \varepsilon_{k,p} & \varepsilon_{k,p+1} & \cdots & \varepsilon_{k,T} \end{pmatrix}$$

VAR(1)	PM10		
	Variable Name	Estimate	T-Statistic
	PM10 (1 Lag)	0.2839	**15.994**
	Precipitation (1-Lag)	-0.1485	**-11.828**
	Temperature (1 Lag)	0.0012	0.035
	R-Squared:	**0.149**	
	Adj. R-Squared	**0.148**	
	Precipitation		
	Variable Name	Estimate	T-Statistic
	PM10 (1 Lag)	0.0525	2.039
	Precipitation (1-Lag)	0.27	**14.845**
	Temperature (1 Lag)	0.282	**5.524**
	R-Squared:	**0.07674**	
	Adj. R-Squared	**0.0579**	
	Temperature		
	Variable Name	Estimate	T-Statistic
	PM10 (1 Lag)	-0.0128	-2.1
	Precipitation (1-Lag)	-0.0582	**-13.46**
	Temperature (1 Lag)	0.07399	0.035
	R-Squared:	**0.5876**	
	Adj. R-Squared	**0.5872**	
	Characteristic Roots of Lag Polynomial		
	0.7065	0.31	0.31

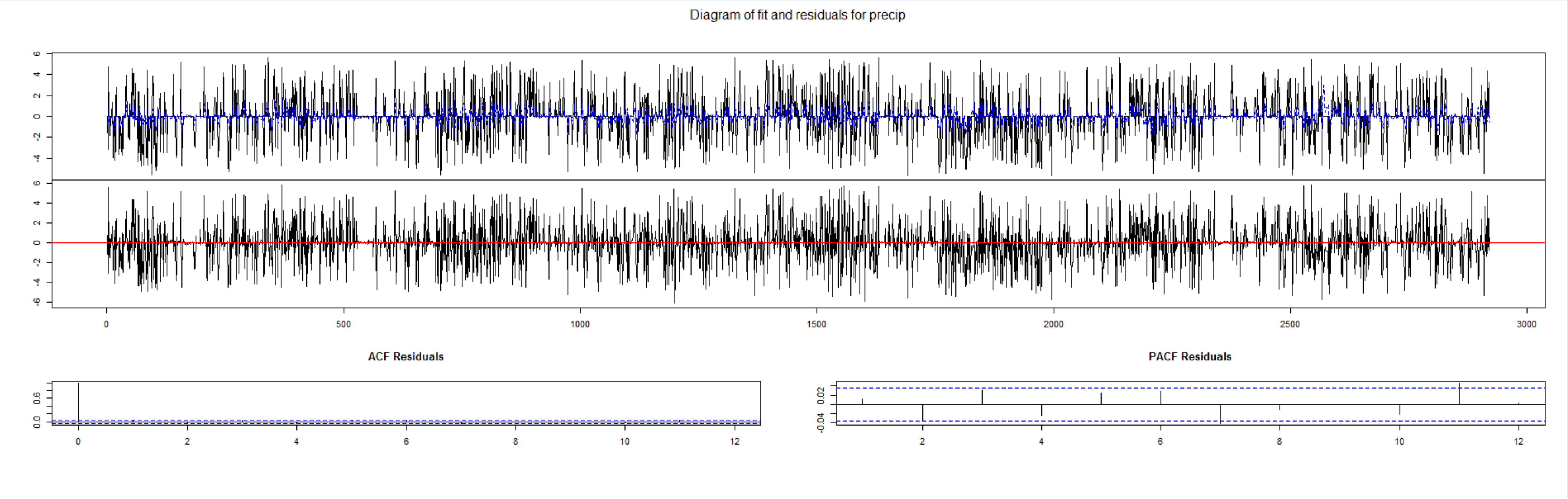
Covariance Matrix of Residuals:			
Variable	**PM10**	**Precipitation**	**Temperature**
PM10	1.8597	-0.4938	0.1064
Precipitation	-0.4938	3.9309	-0.1709
Temperture	0.1064	-0.1064	0.2213

Correlation Matrix of Residuals			
Variable	**PM10**	**Precipitation**	**Temperature**
PM10	1	-0.1826	0.1659
Precipitation	-0.1826	1	-0.1832
Temperture	0.1659	-0.1832	1

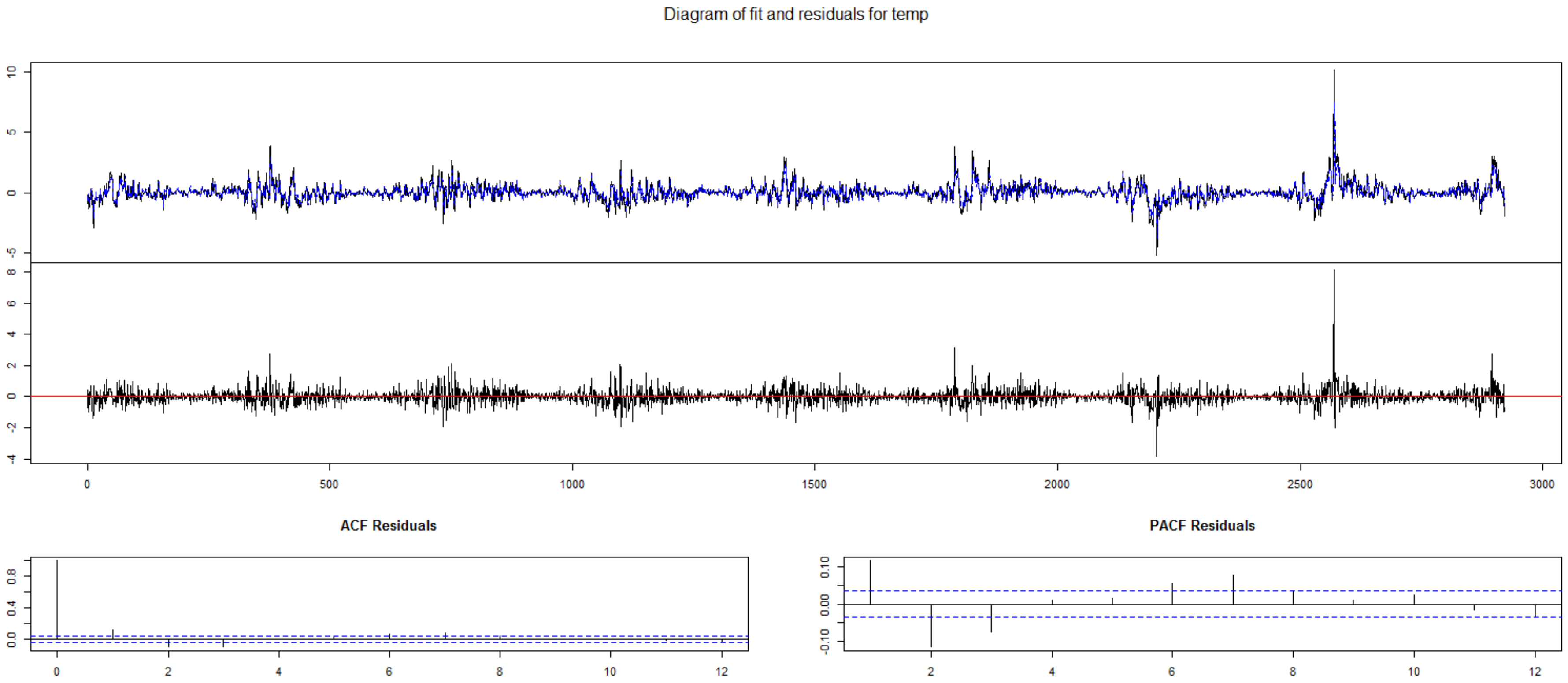
Residual Plots PM10



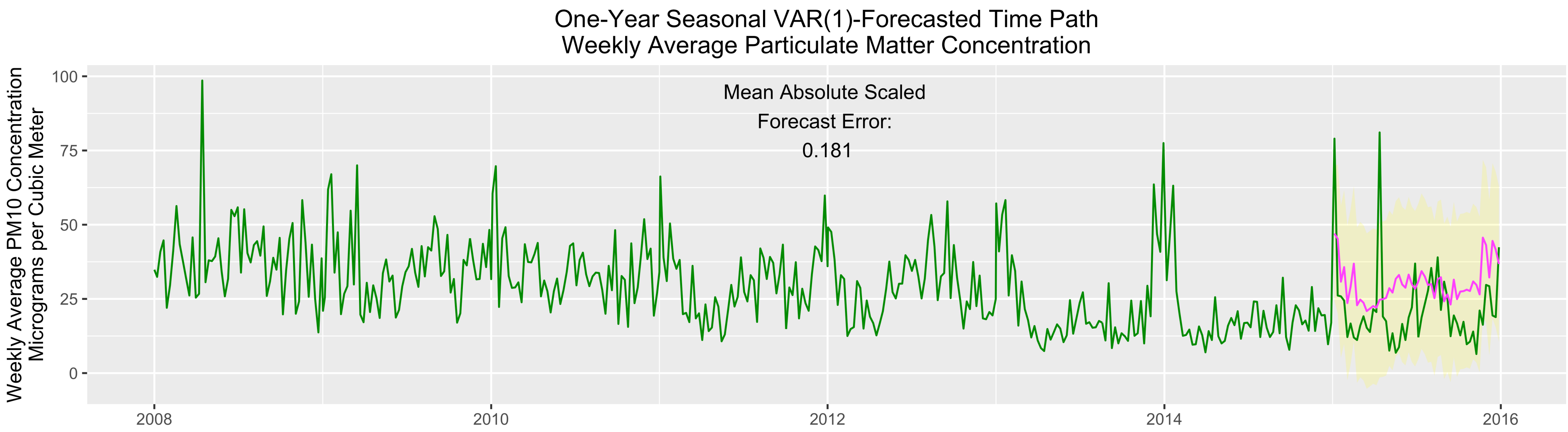
Residual Plots Precipitation



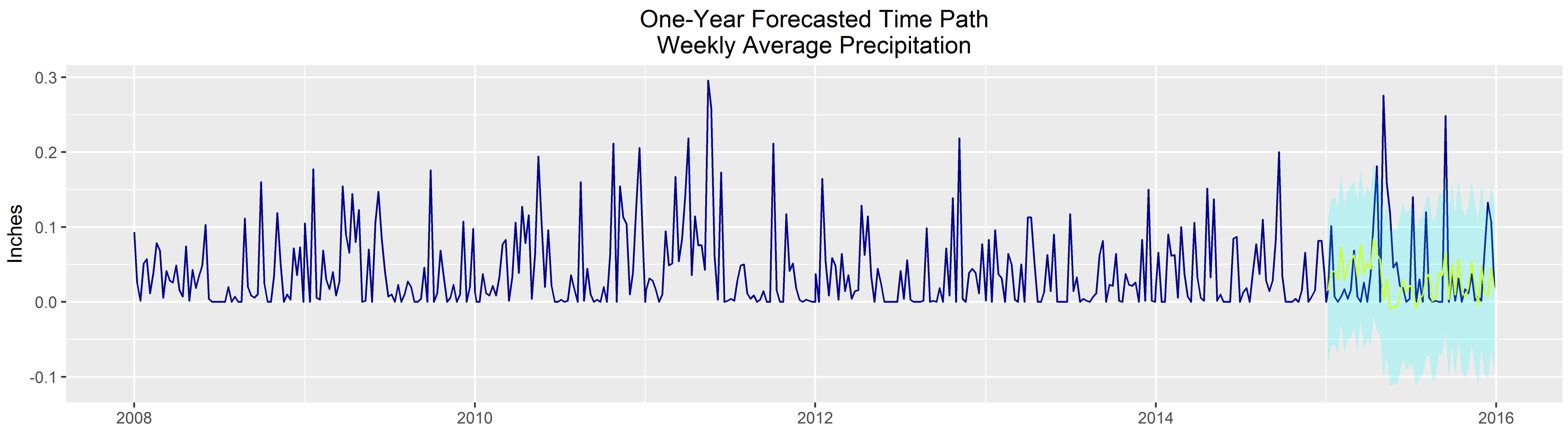
Residual Plots Temperature



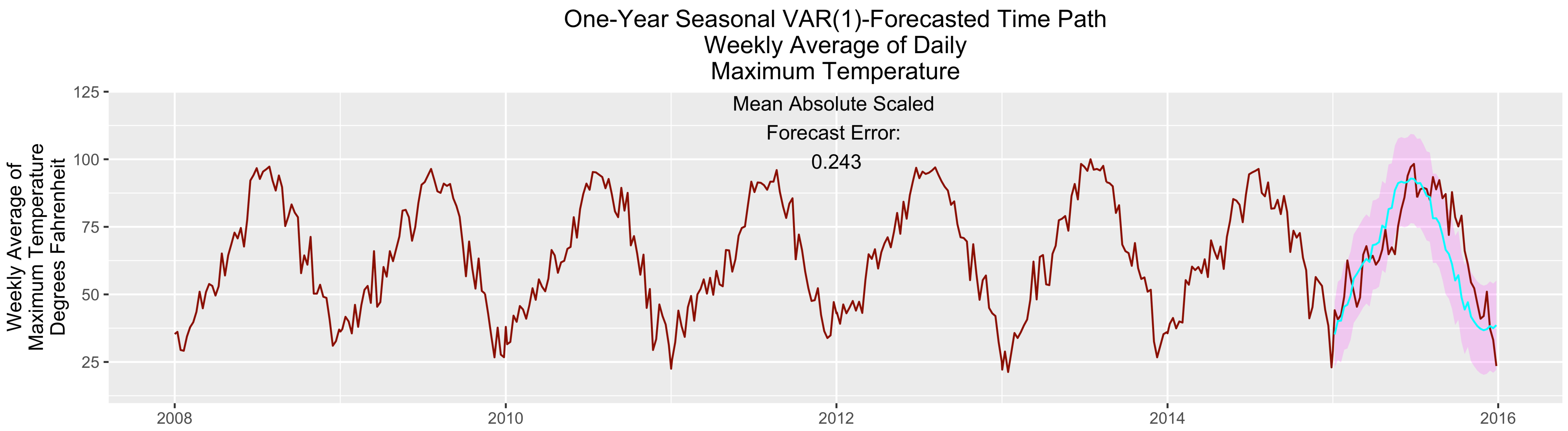
SVAR Forecasts PM10



SVAR Forecasts Precipitation



SVAR Forecasts Temperature

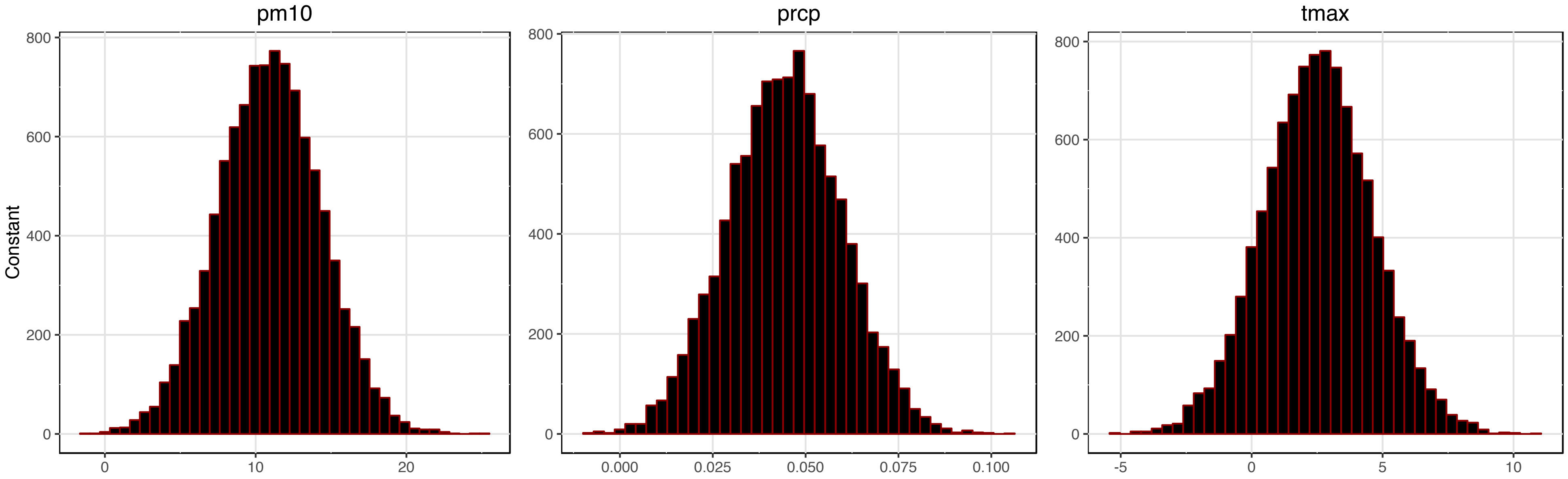


Bayesian Vector Autoregression Climate Forecasting (simplified)

- Rather than estimating a rigid parameter value like the VAR method Bayesian methods concern itself more with the shape of the distribution of the parameter values
- Start by generating a huge (10,000) sample of plausible values from a prior distribution for slope parameters of interest
- For each of these slope parameters it calculates the probability of all the data given each particular set of *randomly generated* slope estimates

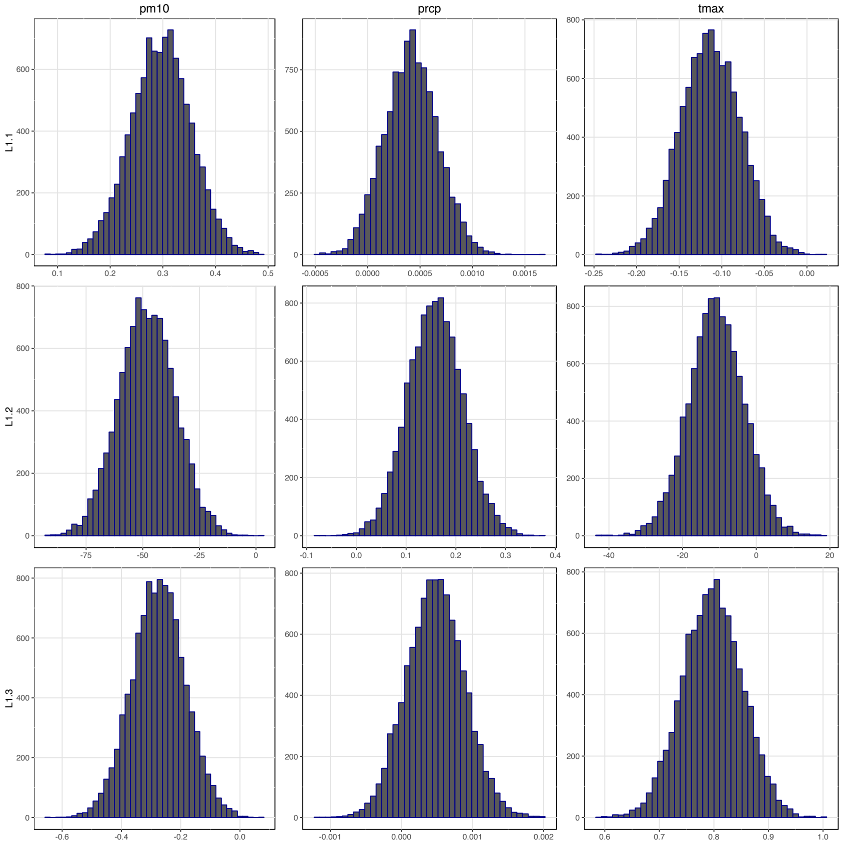
The ’BMR’ package in R allows us to calculate these empirical posterior distributions for our models with the Minnesota Prior

$$pm10_t = \phi_{pm10} + \beta_{1,pm10} pm10_{t-1} + \beta_{1,precip} precip_{t-1} + \beta_{1,temp} temp_{t-1} + e_{pm10,t}$$
$$precip_t = \phi_{precip} + \beta_{2,pm10} pm10_{t-1} + \beta_{2,precp} precip_{t-1} + \beta_{2,temp} temp_{t-1} + e_{precip,t}$$
$$temp_t = \phi_{temp} + \beta_{3,pm10} pm10_{t-1} + \beta_{3,precip} precip_{t-1} + \beta_{3,temp} temp_{t-1} + e_{temp,t}$$



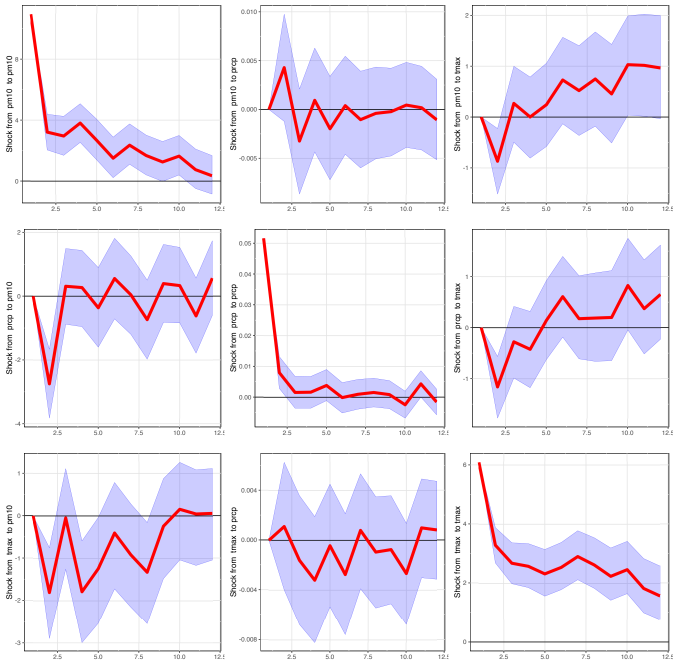
ϕ_{PM10}	$\phi_{Precipitation}$	$\phi_{Temperature}$
$N(10.979,11.747)$	$N(.0435,2.41e^{-04})$	$N(2.41,4.28)$

$$pm10_t = \phi_{pm10} + \beta_{1,pm10}pm10_{t-1} + \beta_{1,precip}precip_{t-1} + \beta_{1,temp}temp_{t-1} + e_{pm10,t}$$
$$precip_t = \phi_{precip} + \beta_{2,pm10}pm10_{t-1} + \beta_{2,precp}precip_{t-1} + \beta_{2,temp}temp_{t-1} + e_{precip,t}$$
$$temp_t = \phi_{temp} + \beta_{3,pm10}pm10_{t-1} + \beta_{3,precip}precip_{t-1} + \beta_{3,temp}temp_{t-1} + e_{temp,t}$$

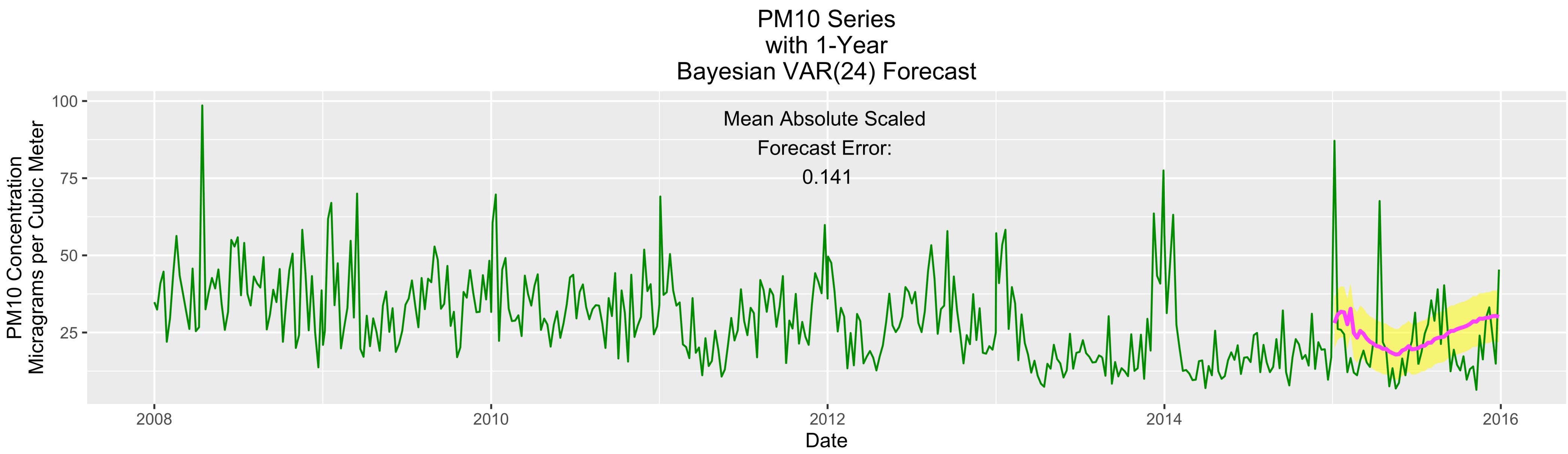


Posterior Distribution of β Coefficients

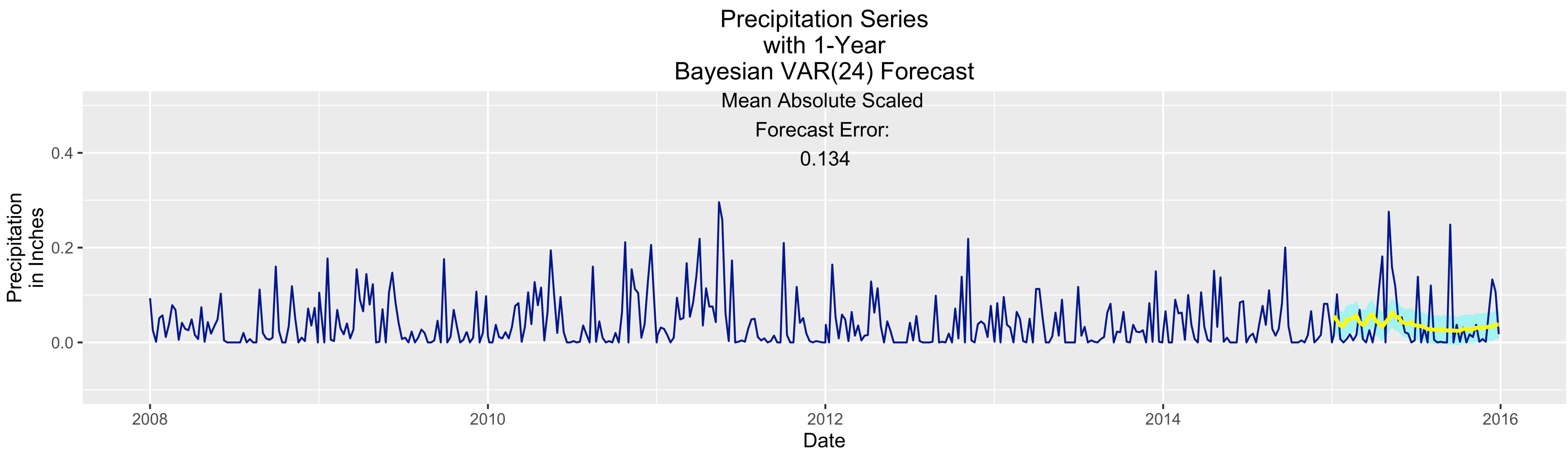
$Effect\ of\ \downarrow on \rightarrow$	$PM10_t$	$Precipitation_t$	$Temperature_t$
$PM10_{t-1}$	$N(0.305,.003)$	N(3.24e-4,6.35e-4)	$N(-.108,.0011)$
$Precipitation_{t-1}$	$N(-46.508,157.05)$	$N(0.154,3.10e-03)$	N(-10.55,56.18)
$Temperature_{t-1}$	$N(-0.27,8.19e-03)$	N(5.17e-04,1.64e-07)	N(.797,3.02e-03)



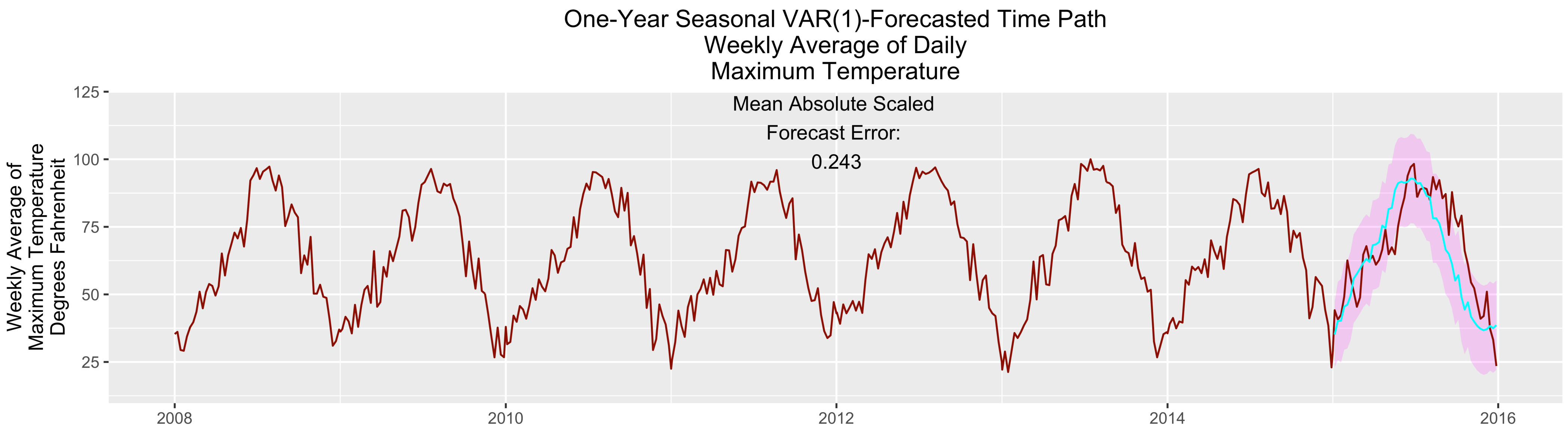
BVAR Forecast PM10



BVAR Forecast Precipitation



BVAR Forecast PM10



Forecasts Summary

M.A.S.E.	Seasonal ARIMA	Seasonal VAR(1)	Bayesian VAR(1)	Bayesian VAR(4)	Bayesian VAR(24)
PM10	0.182	0.181	.204	.189	0.14
Precipitation	0.142	0.142	.143	.146	0.134
Maximum Temperature	0.199	0.243	.562	.569	0.208

Conclusions

- In the end we’ve provided an example where Bayesian forecasting outperforms traditional methods
- Through Bayesian forecasting and impulse response prediction, we were able to provide evidence that absent any other factor, rising temperatures will reduce the number of winter time temperature inversions
- Comparative methods Demonstrates how these methods fit more with modelling macroeconomy
- With better selection of data beforehand, Macroeconometric methods could be informative in a broader range of applications

The End

References

1. **The World Bank.** *Turn Down the Heat: Climate Extremes, Regional Impacts, and the case for Resilience.* Washington, D.C. : International Bank for Reconstruction and Development, 2013. 978-1-4648-0056-6.
2. **the Utah Environmental Epidemiology Program.** *Air Pollution and Public Health in Utah.* Environmental Public Health Tracking, Bureau of Epidemiology. Salt Lake City : Utah Department of Health, 2015.
3. **Utah Department of Environmental Quality.** Emissions Sources of Winter PM2.5. *deq.utah.gov.* [Online] 2016. http://www.deq.utah.gov/FactSheets/fspages/sources.htm.
4. **Adar, S.D.** *Ambient Coarse Particulate Matter and Human Health: A Systematic Review and Meta-Anal ysis.* Current Environmental Health Reports. s.l. : Springer, 2014. pp. 258-274.
5. **American Heart Association.** *AHA Scientific Statement: Particulate Matter Air Pollution and Cardiovascular Disease.* American Heart Association.
6. *The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City.* **National Bureau of Economic Research.** s.l. : Elsevier, 2015, Journal of Public Economics, Vol. 122(C), pp. 68-79.
7. **Utah Asthma Task Force.** *Salt Lake Valley Asthma Report.* Utah Asthma Task Force, Utah Department of Health. Salt Lake City : State of Utah, 2009.
8. **Enders, Walter.** *Applied Econometric Time Series.* 3rd Edition. Hoboken : Wiley, 2010.
9. **Pfaff, Bernhard and Tanus, Kronberg.** VAR, SVAR, and SVEC Models: Implementation Within R Package 'vars'. [R Package Vignette]. 2008.
10. *Forecasting With Bayesian Vector Autoregression.* **Litterman, Robert B.** 1985, American Statistical Association.
11. **Todd, Richard M.** *Improving Economic Forecasting With Bayesian Vector Autoregression.* Research Department, Federal Reserve Bank of Minneapolis. Minneapolis : Economist, 1984. Quarterly Review.
12. **O'Hara, Keith.** Bayesian Macroeconomics in R. [R Package Vignette]. July 20, 2015.
13. **Lopes, Hedibert Freitas.** Lecture 4: Bayesian VAR. *hedibert.org/.* [Online] 12 11, 2013. [Cited: 7 10, 2016.] http://hedibert.org/wp-content/uploads/2013/12/BVAR-part11.pdf.
14. *Macroeconomics and Reality.* **Sims, Christopher A.** 48, 1980, Econometrica, pp. 1-48.
15. **Cornwall, Warren.** Efforts to link Climate Change to Sever Weather Gain Ground. *Science.* 2016, Vol. 351, 6279, pp. 1249-1250.
16. **Whiteman, C. David, et al.** Relationship Between Particulate Air Pollution and Meteorological Variables in Utah's Salt Lake Valley. *Atmospheric Environment.* 2014, 94, pp. 742-753.
17. **Wooldridge, Jeffery M.** *Introductory Econometrics.* 4th Edition. Lansing : Cengage Learning, 2009.
18. —. *Econometric Analysis of Cross Section and Panel Data.* Cambridge : Massachusetts Institute of Technology, 2010.
19. *Generalizing the Bayesian Vector Autoregression Approach for Regional Interindustry Employment Forecasting.* **Patridge, Mark D. and Rickman, Dan S.** 1, s.l. : American Statistical Association, 1998, Journal of Business & Economic Statistics, Vol. 16, pp. 62-72.
20. **Canova, Fabio.** *Methods for Applied Macroeconomic Research: Chapter 10.* Hoboken : Princeton University Press, 2007. 9781400841028