Modeling results

Anirban Mitra

8/21/2020

Model building summary

The data corresponds to daily COVID incidents starting from May 5, 2020 until July 31, 2020. We fit Generalized Additive Models with Poisson family (log link) for each of the pollutants having their separate models. Using multiple pollutants in the same model is avoided because of multicollinearity and the possibility that they would overfit. Without loss of generality the model is fitted as follows:

$log(Y_t) = \alpha + t + log(Y_{t-1}) + Q_t + s(temp_t) + s(wind_t) + s(hum_t)$

Here Y_t denotes the number of confirmed COVID cases on day t. α is the intercept term. t is day t fitted linearly, where $t = 1, 2, ..., 87$. $log(Y_{t-1})$ is also used as feature to include the auto-regressive effect of incidents. Q_t represents the 8-day moving average of the pollutant under consideration, where the moving average is computed over observations ending on day t starting on day $t-7$. It is also fitted linearly and Q will represent one of AQI, PM2.5, PM10, NO2, NH3, SO2, CO, O3. $temp_t$, $wind_t$, hum_t represent the moving average of average temperature, average wind-speed and average humidity, calculated over the same period as Q. These three variables are fitted using a thin-plate regression spline. Moving average for the pollutant variables and the meteorological factors are computed because it has been found that the coronavirus has an inclubation period of about two to fourteen days.

The outputs after each of the following models represent the parameter estimates of the pollutant and the corresponding 95% Wald Confidence Intervals:

```
AQI_fit <- gam(daily_cases ~ t + caseprev + AQI + s(temp) + s(wind) + s(hum),
               data = mod_dat,family = poisson(link = "log"),
               drop.intercept = F, method = 'REML')
aqi_summ <- summary(AQI_fit)
est = round(aqi_summ$p.coeff[4],4)
lci = round(aqi_summ$p.coeff[4] -   1.96*aqi_summ$se[4],4)
uci = round(aqi_summ$p.coeff[4] + 1.96*aqi_summ$se[4],4)
paste("AQI:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "AQI: Parameter estimate =  0.0057 and CI : [ 6e-04 , 0.0107 ]"
```

```
PM2_fit <- gam(daily_cases ~ t + caseprev + PM2 + s(temp) + s(wind) + s(hum),
               data = mod_dat,family = poisson(link = "log"),
               drop.intercept = F, method = 'REML')
pm2_summ <- summary(PM2_fit)
est = round(pm2_summ$p.coeff[4],4)
lci = round(pm2_summ$p.coeff[4] - 1.96*pm2_summ$se[4],4)
uci = round(pm2_summ$p.coeff[4] + 1.96*pm2_summ$se[4],4)
paste("PM 2.5:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "PM 2.5: Parameter estimate =  0.0012 and CI : [ -0.0037 , 0.0062 ]"
```

```
PM10_fit <- gam(daily_cases ~ t + caseprev + PM10 + s(temp) + s(wind) + s(hum),
                data = mod_dat,family = poisson(link = "log"),
```

```
                drop.intercept = F, method = 'REML')
pm10_summ <- summary(PM10_fit)
est = round(pm10_summ$p.coeff[4],4)
lci = round(pm10_summ$p.coeff[4] - 1.96*pm10_summ$se[4],4)
uci = round(pm10_summ$p.coeff[4] + 1.96*pm10_summ$se[4],4)
paste("PM 10:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "PM 10: Parameter estimate =  0.0046 and CI : [ -2e-04 , 0.0094 ]"
```

```
NO2_fit <- gam(daily_cases ~ t + caseprev + NO2+ s(temp) + s(wind) + s(hum),
               data = mod_dat,family = poisson(link = "log"),
               drop.intercept = F, method = 'REML')
no2_summ <- summary(NO2_fit)
est = round(no2_summ$p.coeff[4],4)
lci = round(no2_summ$p.coeff[4] - 1.96*no2_summ$se[4],4)
uci = round(no2_summ$p.coeff[4] + 1.96*no2_summ$se[4],4)
paste("NO2:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "NO2: Parameter estimate =  0.034 and CI : [ 0.0137 , 0.0544 ]"
```

```
NH3_fit <- gam(daily_cases ~ t + caseprev + NH3 + s(temp) + s(wind) + s(hum),
               data = mod_dat,family = poisson(link = "log"),
               drop.intercept = F, method = 'REML')
nh3_summ <- summary(NH3_fit)
est = round(nh3_summ$p.coeff[4],4)
lci = round(nh3_summ$p.coeff[4] - 1.96*nh3_summ$se[4],4)
uci = round(nh3_summ$p.coeff[4] + 1.96*nh3_summ$se[4],4)
paste("NH3:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "NH3: Parameter estimate =  0.573 and CI : [ 0.4538 , 0.6922 ]"
```

```
SO2_fit <- gam(daily_cases ~ t + caseprev + SO2 + s(temp) + s(wind) + s(hum),
               data = mod_dat,family = poisson(link = "log"),
               drop.intercept = F, method = 'REML')
so2_summ <- summary(SO2_fit)
est = round(so2_summ$p.coeff[4],4)
lci = round(so2_summ$p.coeff[4] - 1.96*so2_summ$se[4],4)
uci = round(so2_summ$p.coeff[4] + 1.96*so2_summ$se[4],4)
paste("SO2:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "SO2: Parameter estimate =  -0.0031 and CI : [ -0.0123 , 0.0061 ]"
```

```
CO_fit <- gam(daily_cases ~ t + caseprev + CO + s(temp) + s(wind) + s(hum),
              data = mod_dat,family = poisson(link = "log"),
              drop.intercept = F, method = 'REML')
co_summ <- summary(CO_fit)
est = round(co_summ$p.coeff[4],4)
lci = round(co_summ$p.coeff[4] - 1.96*co_summ$se[4],4)
uci = round(co_summ$p.coeff[4] + 1.96*co_summ$se[4],4)
paste("CO:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "CO: Parameter estimate =  -0.0547 and CI : [ -0.0683 , -0.0411 ]"
```

```
O3_fit <- gam(daily_cases ~ t + caseprev + O3 + s(temp) + s(wind) + s(hum),
              data = mod_dat,family = poisson(link = "log"),
              drop.intercept = F, method = 'REML')
o3_summ <- summary(O3_fit)
```

```r
est = round(o3_summ$p.coeff[4],4)
lci = round(o3_summ$p.coeff[4] - 1.96*o3_summ$se[4],4)
uci = round(o3_summ$p.coeff[4] + 1.96*o3_summ$se[4],4)
paste("O3:","Parameter estimate = ",est, "and CI : [",lci,",",uci,"]")
```

```
## [1] "O3: Parameter estimate =  0.0091 and CI : [ 0.0035 , 0.0148 ]"
```

Summaries for all the models are below:

```r
aqi_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + AQI + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.6434211  0.1220813  21.653  < 2e-16 ***
## t           0.0358185  0.0015748  22.745  < 2e-16 ***
## caseprev    0.0003947  0.0001163   3.394 0.000689 ***
## AQI         0.0056744  0.0025644   2.213 0.026910 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq  p-value
## s(temp) 5.362  6.560  66.69 6.56e-12 ***
## s(wind) 7.894  8.619 164.51  < 2e-16 ***
## s(hum)  8.332  8.856 135.22  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.976   Deviance explained =   98%
## -REML = 550.29  Scale est. = 1         n = 87
```

```r
pm2_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + PM2 + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.8357375  0.0918808  30.863  < 2e-16 ***
## t           0.0360885  0.0015692  22.997  < 2e-16 ***
## caseprev    0.0003693  0.0001155   3.197  0.00139 **
## PM2         0.0012443  0.0025042   0.497  0.61928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq  p-value
## s(temp) 5.199  6.386  75.39 9.66e-14 ***
## s(wind) 7.992  8.675 155.70  < 2e-16 ***
## s(hum)  8.274  8.832 130.99  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.975   Deviance explained = 97.9%
## -REML = 552.56  Scale est. = 1         n = 87
```

```
pm10_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + PM10 + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.6990391  0.1130057  23.884  < 2e-16 ***
## t           0.0356844  0.0015855  22.506  < 2e-16 ***
## caseprev    0.0003941  0.0001165   3.382  0.00072 ***
## PM10        0.0046070  0.0024695   1.866  0.06210 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq  p-value
## s(temp) 5.358  6.557  69.02 2.31e-12 ***
## s(wind) 7.914  8.630 163.05  < 2e-16 ***
## s(hum)  8.323  8.852 134.62  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.976   Deviance explained =   98%
## -REML = 551.02  Scale est. = 1         n = 87
```

```
no2_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + NO2 + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.7471481  0.0796346  34.497  < 2e-16 ***
## t           0.0338665  0.0017046  19.867  < 2e-16 ***
## caseprev    0.0004142  0.0001166   3.552 0.000382 ***
## NO2         0.0340064  0.0103821   3.275 0.001055 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq p-value
## s(temp) 5.685  6.854  73.02 4.9e-13 ***
## s(wind) 7.855  8.596 110.77 < 2e-16 ***
## s(hum)  8.087  8.751  97.18 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.976   Deviance explained =   98%
## -REML = 546.23  Scale est. = 1         n = 87
```

```
nh3_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + NH3 + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.390e+00  8.648e-02  27.637   <2e-16 ***
## t            3.268e-02  1.598e-03  20.452   <2e-16 ***
## caseprev    -8.838e-05  1.250e-04  -0.707     0.48
## NH3          5.730e-01  6.083e-02   9.419   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq  p-value
## s(temp) 5.542  6.715  62.18 6.24e-11 ***
## s(wind) 7.606  8.437 112.92  < 2e-16 ***
## s(hum)  8.060  8.737  96.14  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.983   Deviance explained = 98.4%
## -REML = 505.24  Scale est. = 1         n = 87
```

```
so2_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + SO2 + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.8624446  0.0706733  40.502   <2e-16 ***
## t            0.0366030  0.0017602  20.795   <2e-16 ***
## caseprev     0.0003258  0.0001333   2.444   0.0145 *
```

```
## SO2         -0.0030897  0.0046836  -0.660   0.5095
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq  p-value
## s(temp) 5.155  6.324  76.91 4.26e-14 ***
## s(wind) 8.038  8.698 146.01  < 2e-16 ***
## s(hum)  8.276  8.830 129.12  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.975   Deviance explained = 97.9%
## -REML = 551.84  Scale est. = 1         n = 87
```

```
co_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + CO + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.117e+00  7.726e-02  40.341  < 2e-16 ***
## t            4.534e-02  1.961e-03  23.116  < 2e-16 ***
## caseprev    -1.053e-05  1.256e-04  -0.084    0.933
## CO          -5.468e-02  6.926e-03  -7.895  2.9e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq p-value
## s(temp) 6.912  7.962  120.0  <2e-16 ***
## s(wind) 7.925  8.640  112.0  <2e-16 ***
## s(hum)  7.704  8.550  112.3  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.981   Deviance explained = 98.3%
## -REML = 522.86  Scale est. = 1         n = 87
```

```
o3_summ
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## daily_cases ~ t + caseprev + O3 + s(temp) + s(wind) + s(hum)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.4960469  0.1368294  18.242  < 2e-16 ***
```

```
## t           0.0389828  0.0018219  21.397  < 2e-16 ***
## caseprev    0.0002596  0.0001207   2.150  0.03152 *
## O3          0.0091361  0.0028982   3.152  0.00162 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df Chi.sq  p-value
## s(temp) 5.133  6.307  75.11 8.58e-14 ***
## s(wind) 8.000  8.678 154.68  < 2e-16 ***
## s(hum)  8.313  8.849 126.35  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.977   Deviance explained =   98%
## -REML = 547.57  Scale est. = 1         n = 87
```

The conclusions are as follows: AQI, NO2, NH3, CO, O3 are significant at 5% level. PM 2.5 and SO2 are not significant. PM 10 is not significant at 5% level but significant at 10% level.

Interpretation of parameters in terms of daily COVID cases:

- Increase in AQI by 1 unit increases COVID incident by 0.5%.
- Increase in NO2 by 1 unit increases COVID incident by 3%.
- Increase in NH3 by 1 unit increases COVID incident by 77%.
- Increase in CO by 1 unit decreases COVID incident by 5%.
- Increase in O3 by 1 unit increases COVID incident by 0.9%.

The above are calculated as follows. Let's say for any pollutant, without loss of generality, $\hat{\beta}$ be its estimated regression coefficient. Then for unit change in the pollutant controlling for other variables in its model, the incident rate changes by $(\exp(\hat{\beta}) - 1)\%$.