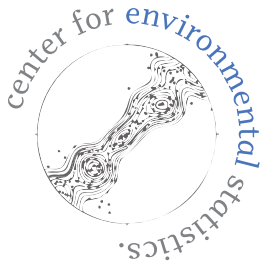

ELEVEN YEARS OF I-5 BETWEEN SR-14 AND SR-99

JAN DE LEEUW, IRINA KUKUYEVA, AND LINDA GHARIBANS

ABSTRACT. In this report we analyze hourly I-5 traffic counts over the period 1996-2006 at four locations. Thus for each location there are more than 95,000 time points. We analyze trend and seasonality, and at the same time illustrate various classes of time series techniques.

Date: June 3, 2008.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis.

CONTENTS

List of Tables

List of Figures

1. Introduction

1.1. **Loop counts.** Loop counts data are far from perfect. Individual loop counts can fail for various reasons and either give no data or data outside a reasonable range. In general, we expect $365 \times 24 = 8760$ observations in a year (8784 in a leap year). But we see far fewer.

Insert Table 1 about here

1.2. **Locations.**

2. Data

Insert Figure ?? about here

3. Imputation

The loop count data have 11 years of hourly counts at 4 locations. We use dummy regressors

effect	dfr	symbol
intercept	1	μ
Main Effects		
location	4-1=3	α_i
year	11-1=10	β_j
month	12-1=11	γ_k
day	7-1=6	δ_ℓ
hour	24-1=23	ϵ_m
Interactions		
year $\times$ location	10×3=30	π_{ij}
month $\times$ location	11×3=33	θ_{ik}
day $\times$ location	6×3=18	$\xi_{i\ell}$
hour $\times$ location	23×3=69	η_{im}

We fit a total of 59 parameters (54 dfr) by minimizing

$$\sigma(\mu,\alpha,\beta,\gamma,\delta,\epsilon) = \sum_{i=1}^{4}\sum_{j=1}^{11}\sum_{k=1}^{12}\sum_{\ell=1}^{7}\sum_{m=1}^{24} w_{ijk\ell m}(y_{ijk\ell m}-\mu-\alpha_i-\beta_j-\gamma_k-\delta_\ell-\epsilon_m)^2.$$

Here $z_{ijk\ell}$ are the loop counts (or, alternatively, logarithms of loop counts) and

$$w_{ijkl\ell m} = \begin{cases} 1 & \text{if } y_{ijk\ell m} \text{ is observed,} \\ 0 & \text{if } y_{ijk\ell m} \text{ is missing.} \end{cases}$$

Minimizations is done by alternating least squares, combining estimation and imputation. Thus we minimize

$$\sigma(Z,\mu,\alpha,\beta,\gamma,\delta,\epsilon) = \sum_{i=1}^{4}\sum_{j=1}^{11}\sum_{k=1}^{12}\sum_{\ell=1}^{7}\sum_{m=1}^{24} (z_{ijk\ell m}-\mu-\alpha_i-\beta_j-\gamma_k-\delta_\ell-\epsilon_m)^2.$$

over the six sets of parameters and over the elements of Z, requiring $z_{ijk\ell m} = y_{ijk\ell m}$ if $y_{ijk\ell m}$ is non-missing.

The algorithm starts with imputing the missing counts (for example, set them all to zero). Set the iteration counter ν equal to one. Then we compute

$$\begin{aligned}
\mu^{(\nu)} &= z^{(\nu)}_{\bullet\bullet\bullet\bullet\bullet}, \\
\alpha_i^{(\nu)} &= z^{(\nu)}_{i\bullet\bullet\bullet\bullet} - z^{(\nu)}_{\bullet\bullet\bullet\bullet\bullet}, \\
\beta_j^{(\nu)} &= z^{(\nu)}_{\bullet j\bullet\bullet\bullet} - z^{(\nu)}_{\bullet\bullet\bullet\bullet\bullet}, \\
\gamma_k^{(\nu)} &= z^{(\nu)}_{\bullet\bullet k\bullet\bullet} - z^{(\nu)}_{\bullet\bullet\bullet\bullet\bullet}, \\
\delta_\ell^{(\nu)} &= z^{(\nu)}_{\bullet\bullet\bullet\ell\bullet} - z^{(\nu)}_{\bullet\bullet\bullet\bullet\bullet}, \\
\epsilon_m^{(\nu)} &= z^{(\nu)}_{\bullet\bullet\bullet\bullet m} - z^{(\nu)}_{\bullet\bullet\bullet\bullet\bullet},
\end{aligned}$$

using bullets for averages. After this we update the $z_{ijk\ell m}$ for which the count is missing by

$$z^{(\nu+1)}_{ijk\ell m} = \mu^{(\nu)} + \alpha_i^{(\nu)} + \beta_j^{(\nu)} + \gamma_k^{(\nu)} + \delta_\ell^{(\nu)} + \epsilon_m^{(\nu)},$$

and we go back to update the parameters with these new values of Z. Continue until everything stabilizes.

- It will be useful to do this both on the counts and on their logs or square roots.
- It will also be useful to do some cross-validation experiments. We leave out, say, an additional 10% of the counts and see how well these are recovered by the algorithm.
- The parameter estimates themselves will also be useful to look at. The year effect, for example, measures growth.
- This is related to the way AADT is calculated.
- There may be some need for interactions. For instance, the day effect in the summer months may be different from that in the winter months.
- One logical next step is to do the analysis for each location separately. This amounts to using $4 \times 55 = 220$ parameters.
- And finally, this needs analysis of the residuals after fitting to see if there are systematic effects left, for example connected with time. We could use spectral analysis on the residuals, for example.
- Depending on what the residual analysis says, we can quite easily compute standard errors for the parameters.

APPENDIX A. CODE

```
impute<-function(design,counts,eps=1e-6,itmax=1000,verbose
    =FALSE) {
design<-apply(design,2,as.character); m<-ncol(design)
ii<-is.na(counts); impcnt<-ifelse (ii,0,counts)
itel<-1; oldssq<-Inf
repeat {
        z<-array(0,dim(design)); mn<-mean(impcnt)
        for (k in 1:m) {
                aux<-tapply(impcnt,design[,k],mean)
                z[,k]<-aux[design[,k]]-mn
                }
        pred<-mn+rowSums(z)
        impcnt<-ifelse(ii,mn+rowSums(z),counts)
        newssq<-sum((impcnt-pred)^2)
        if (verbose)
                cat("Iteration:␣",formatC(itel,digits=3,
                    width=3),
                        "Previous␣Loss:␣",formatC(oldssq,
                            digits=4,width=20,format="f"),
                        "Current␣Loss:␣",formatC(newssq,
                            digits=4,width=20,format="f"),
                        "\n")
        if (((oldssq-newssq) < eps) || (itel == itmax))
            break()
                oldssq<-newssq; itel<-itel+1
        }
return(list(z=z,mn=mn,pred=pred,ssq=newssq))
}
```

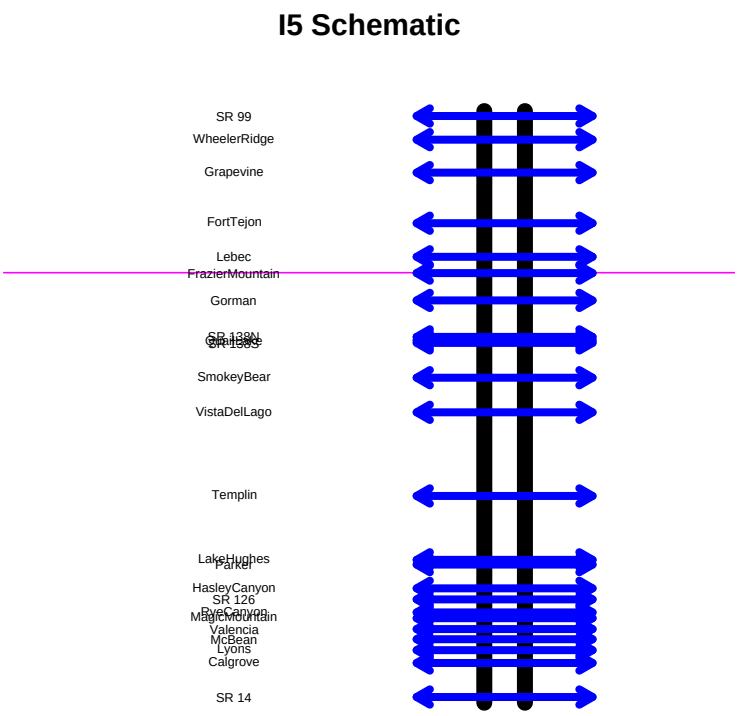

FIGURE 1. I-5 Schematic

FIGURE 2. Northbound Time Series

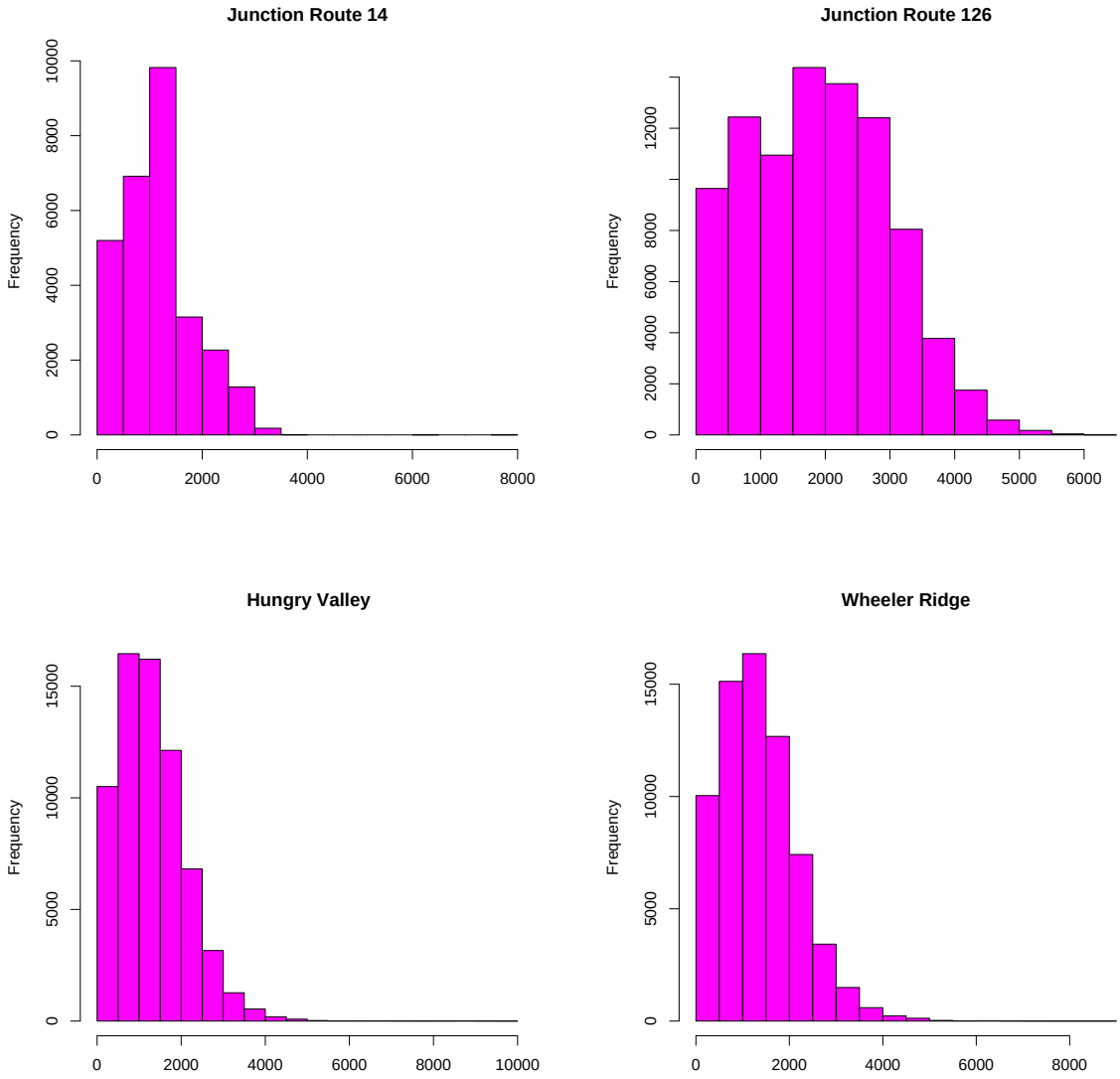

FIGURE 3. Frequency Distribution Northbound Traffic

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
junction route 14	0	0	2208	3695	0	24	4162	8640	3215	6123	757
junction route 126	6000	8257	7142	8733	7867	8315	7843	8204	8160	8696	8712
hungry valley	6720	3797	3748	4322	7272	8005	7311	6862	6187	5388	7821
wheeler ridge	0	168	5412	7739	7998	7838	8230	8088	8599	5818	7667

TABLE 1. Number of Observations, Northbound Traffic

CENTER FOR ENVIRONMENTAL STATISTICS, DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu