SUPPORTING MATERIAL

Combining continuous near-road monitoring and inverse modeling to isolate the effect of highway expansion on a school in Las Vegas

Ram Vedantham [1*], Gary Norris [1], Steven G. Brown [2], Paul Roberts [2]

[1] *United States Environmental Protection Agency, Office of Research and Development*
[2] *Sonoma Technology, Inc., 1455 N. McDowell Blvd, Suite D, Petaluma, CA, 94954-1169, USA*

CONTENT

*Corresponding Author: Tel: (919) 541 5039, vedantham.ram@epa.gov

Non-Parametric Regression

The basic idea in Non-Parametric Regression (NPR) is to construct an estimate of the underlying density function using data collected during an intensive. Using this function and its accompanying mathematical framework, means and variances can be computed. This approach is called the non-parametric density regression method. The term non-parametric refers to modeling of the underlying density using representations that are different from the parameterized distributions such as the Gaussian distribution.

Non-Parametric Wind Regression

The Non-parametric Wind Regression (NWR) utilizes the NPR techniques to derive an estimate of the underlying density function and its related quantities (Henry et al., 2009). The kernel density function is estimated to be:

$$\hat{f}(\theta,u) = \frac{1}{N\sigma h}\sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u - U_i)}{h}\right) \quad (1)$$

where $K_1(\theta,\sigma)$ and $K_2(u,h)$ are the kernel functions of direction θ and speed u respectively with smoothing parameters σ and h. Also, U_i and W_i are the observed wind speed and direction associated with C_i the i^{th} observation ($(1 \leq i \leq N)$ in a time period starting at time t_i. The average concentration of a pollutant for a particular wind speed and direction pair (θ, u) is calculated as a weighted average of the concentration data and is estimated by the following expression:

$$E(C \mid \theta,u) = \frac{\sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u - U_i)}{h}\right) C_i}{\sum_{i=1}^{N} K_1\left(\frac{(\theta - W_i)}{\sigma}\right) K_2\left(\frac{(u - U_i)}{h}\right)} \quad (2)$$

Using these definitions, the fraction of the weighted average associated with winds from the sector defined by the intervals *U* and *Θ*, defined as the sector apportionment is given by:

$$S(\Theta,U) = \int_{u_1}^{u_2}\int_{\theta_1}^{\theta_2} \hat{f}(\theta,u)\, E(C \mid \theta,u)\, d\theta\, du \quad (3)$$

Using the definitions in (1) and (2), we obtain:

$$S(\Theta,U) = \frac{\Delta u \Delta\theta}{N\sigma h}\sum_{\theta_k=\theta_1}^{\theta_2}\sum_{u_j=u_1}^{u_2}\sum_{i=1}^{N} K_1\left(\frac{(\theta_k - W_i)}{\sigma}\right) K_2\left(\frac{(u_j - U_i)}{h}\right) C_i$$

Sector apportionment can be used as a measure of the contribution of a specific local source or source type to the air shed. Uncertainty associated with the sector apportionment is given by:

$$\mathrm{var}(S(\Theta,U)) = \sum_{\theta_k,\theta_m \in \Theta,\, u_k,u_n \in U} \sum s(\theta_k,u_j)s(\theta_m,u_n)\exp\left(-\frac{1}{2}\left(\frac{\theta_k-\theta_m}{\sigma}\right)^2\right) \tag{4}$$

where,

$$\begin{aligned} s^2(\theta,u) &= \mathrm{var}(\hat{f}(\theta,u)E(C\,|\,\theta,u)) \\ &\approx \mathrm{var}(\hat{f}(\theta,u))E(C\,|\,\theta,u)^2 + \mathrm{var}(E(C\,|\,\theta,u))\,\hat{f}(\theta,u)^2. \end{aligned} \tag{5}$$

The same kernel functions are used for wind direction and wind speed. They are the Gaussian kernel for wind direction given by:

$$K_1(x) = (2\pi)^{-1/2}\exp(-0.5x^2), \quad -\infty < x < \infty$$

and the Epanechnikov kernel for wind speed given by

$$K_2(x) = \begin{cases} 0.75 * (1-x^2), & -1 < x < 1 \\ 0, & \mathit{otherwise.} \end{cases}$$

Sustained Wind Incidence Method

The Sustained Wind Incidence Method (SWIM) is a modified version of NWR. Changes were made to NWR to make the methodology more robust in analyzing the high time resolution data. The modifications and their justifications will be discussed as they are introduced. The kernel density estimate used by SWIM has the same functional representation as NWR but with one crucial change. It is now given by

$$\hat{f}(\theta,u) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i}K_1\left(\frac{(\theta-W_i)}{\hat{\sigma}_i}\right)\ \frac{1}{\hat{h}_i}K_2\left(\frac{(u-U_i)}{\hat{h}_i}\right). \tag{1a}$$

This definition differs from (1) in using the wind direction standard deviation $\hat{\sigma}_i$ associated with the wind event instead a fixed smoothing parameter σ and similarly for h. This change is both necessary and advantageous as will be shown and is motivated by one of the basic assumptions in kernel density estimate theory as $N \to \infty$, the estimated density will converge to the actual density function. The mathematical statement that summarizes this requirement is expressed as $\hat{\sigma}_i \to 0$ and $N\hat{\sigma}_i \to \infty$ as $N \to \infty$. Similar assumptions are also made for h_i. An offshoot of this assumption is the ability to approximate certain means and variances. We will show that without these assumptions, the expressions for means and variances of the various quantities that have been derived for NWR are not bounded as $N \to \infty$ and their approximations are not accurate.

For the sake of brevity, the following derivations will be presented with $\hat{f} = \hat{f}(\theta)$, that is, as a function of a single variable. These results can be extended to multiple variables to obtain equivalent results. Accordingly,

$$\hat{f}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i}K\left(\frac{(\theta - W_i)}{\hat{\sigma}_i}\right) \tag{1b}$$

Then,

$$E\,\hat{f}(\theta) = \int \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i}K\left(\frac{(\theta - W_i)}{\hat{\sigma}_i}\right)\, f(\theta)\, d\theta$$

where $f(\theta)$ is the unknown density. By a change of variables and using the Taylor series expansion, we obtain:

$$E\,\hat{f}(\theta) = f(\theta) + \frac{1}{2}\sum_{i=1}^{N}\mu_i^2 f''(\theta) + \sum_{j=1}^{\infty}\sum_{i=1}^{N} o(\hat{\sigma}_i^{\,2j}) \tag{6}$$

where we have used the standard assumptions for the kernels, $\int K(z)\,dz = 1$,

$\int z^{2j+1}K(z)\,dz = 0,\ j = 0,1,....$ and $\int z^2 K(z)\,dz = \mu^2 < \infty$. By defining $\hat{\sigma}_i = \dfrac{\sigma_i}{N^{\alpha}}$ with $0 < \alpha < 1$, we not only satisfy the assumptions made earlier about σ_i, but also approximate,

$$E\,\hat{f}(\theta) \approx f(\theta) + \frac{1}{2}\sum_{i=1}^{N}\mu_i^{\,2}\, f''(\theta) \tag{7}$$

for large enough values of *N*. This implies that $\mathrm{var}(\hat{f}(\theta)) \approx \sum_{i=1}^{N}\dfrac{1}{N\hat{\sigma}_i}\mu_i^2 f(\theta)$. The use of an arbitrary fixed smoothing parameter in (6), instead of the scaled variable $\hat{\sigma}_i$, the expression in (7) does not allow (7) to approximate (6) since the higher order terms not included in (7) are not bounded as $N \to \infty$.

Apart from the mathematical need for such as definition for σ, there is an advantageous interpretational consequence if σ_i were to be viewed as a weight associated with quality of the ith observation. Larger values of σ_i imply large variation in the wind direction over the period in which the values were averaged which implies a comparably higher inconsistency. By using a kernel such the Gaussian kernel, the contribution from such observations will be proportionally down-weighted.

Another instance of use of fixed smoothing parameter is found in the definition of the uncertainty associated with the sector apportionment in equation (4). Although, the use of the smoothing parameter is not mathematically flawed, it allows the user to manipulate an important result. The uncertainty estimates are directly proportional to the choice of the smoothing parameter. Thus, in a effort to let the data speak for itself, the smoothing parameter is replaced with the average of the wind direction standard deviation. Thus,

$$\operatorname{var}(S(\Theta,U)) = \sum_{\theta_k,\theta_m\in\Theta,\, u_k,u_n\in U} \sum s(\theta_k,u_j)s(\theta_m,u_n)\exp\left(-\frac{1}{2}\left(\frac{\theta_k-\theta_m}{\overline{\sigma}}\right)^2\right) \tag{4a}$$

where $\overline{\sigma}$ is the overall average of the wind direction standard deviation.