The Fitting of Multifunctions:
An approach to Nonparametric Multimodal Regression

Jochen Einbeck

National University of Ireland, Galway

joint work with Gerhard Tutz, University of Munich

COMPSTAT, Roma, 30th of August 2006

What is “Nonparametric Regression”?

Typical definition:

“Given observations from an explanatory variable X *and a response variable* Y*, construct a function, a “smoother”, which at point* x *estimates the average value of* Y *given that* $X = x$*.”*

(Holmström et.al, 2003)

Nonparametric regression

Given n data points $(X_i, Y_i), i = 1, \ldots, n$, one seeks a smooth function $m : \mathbb{R} \longrightarrow \mathbb{R}$ relating X and Y in a proper way, which can be generally expressed in the form

$$m(x) = \Phi(Y|X = x).$$

Possible choices of the operator $\Phi(\cdot)$ are obtained as solutions of the minimization problem

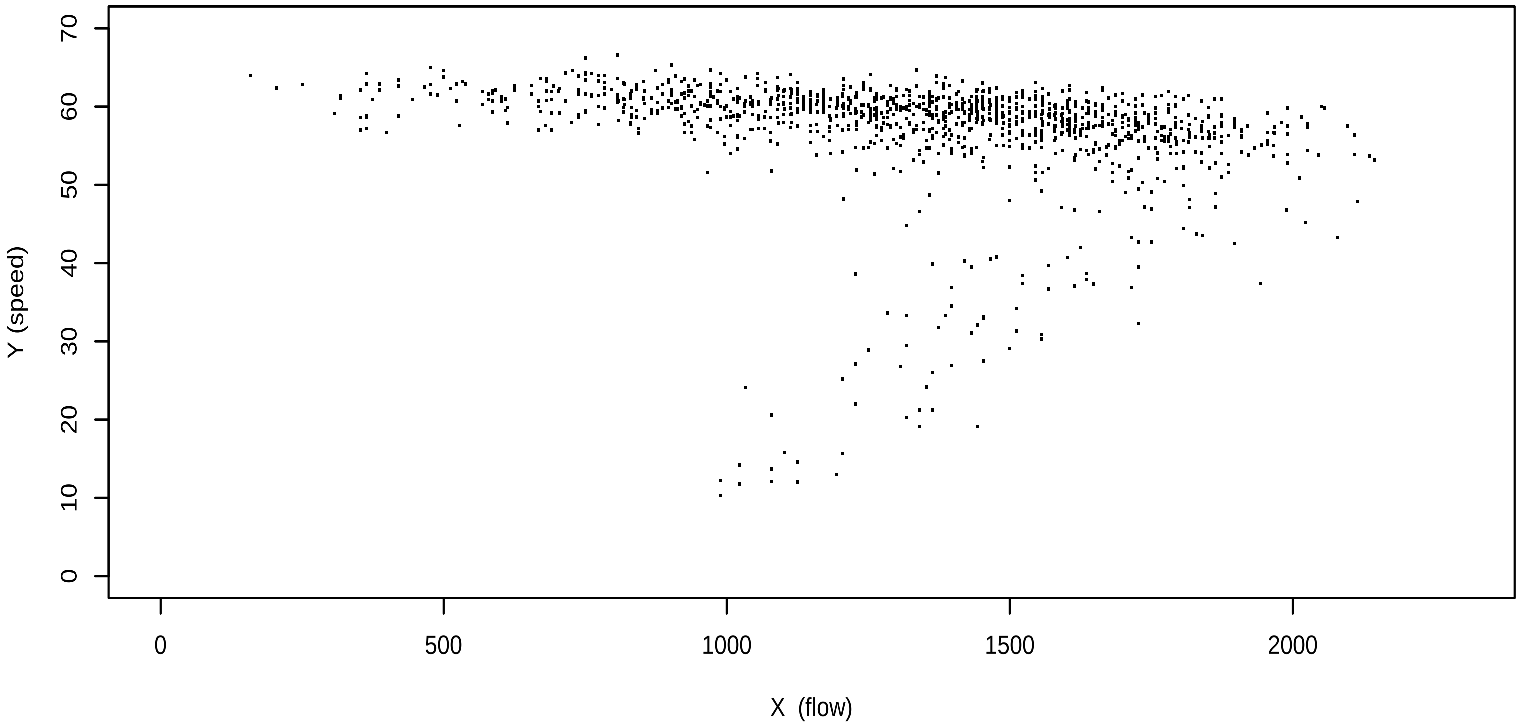
$$m_l(x) = \arg\min_a E(l(Y - a)|X = x),$$

yielding

$l(z)$	z^2	$\|z\|$	$-\delta(z)$
$\Phi(\cdot)$	$E(\cdot)$	$Median(\cdot)$	$Mode(\cdot)$

Limits of ordinary nonparametric regression

Speed-Flow diagram:



(data from:

R package hdrcde)

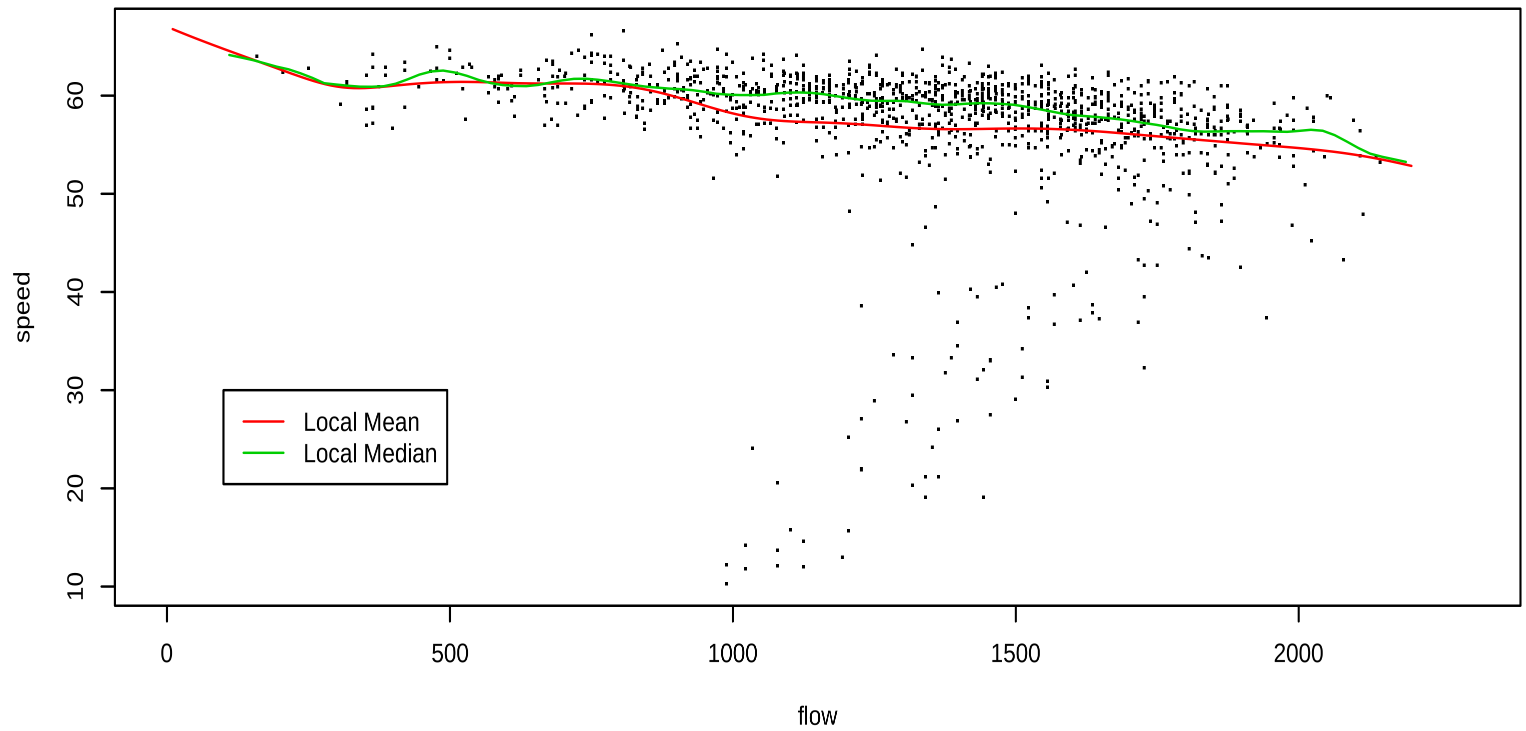
X: traffic flow in cars/hour,   Y: speed in Miles/hour

recorded on a Californian “freeway”.

Attempt: Ordinary nonparametric regression

One assumes an asymmetric relationship between X and Y:

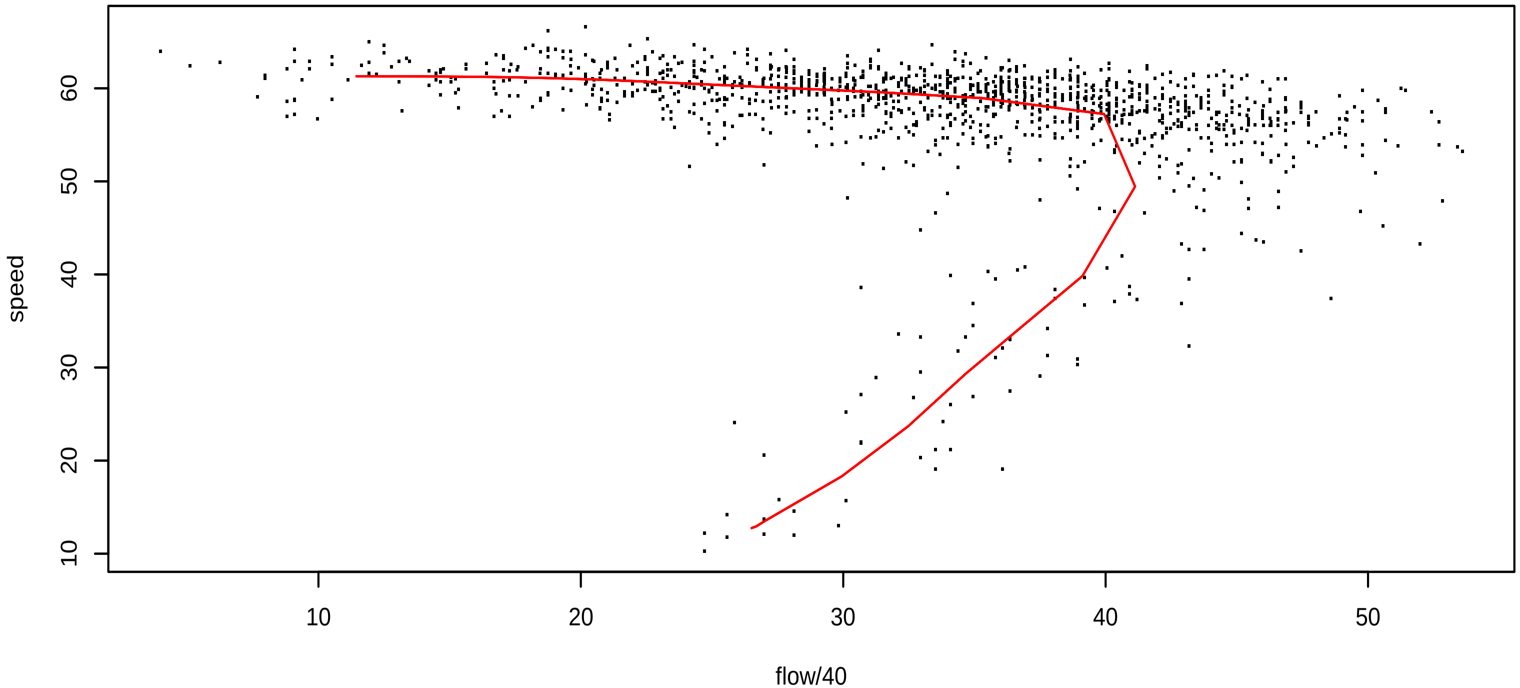
$$Y = m(X) + \epsilon\,; \text{ with a function } m : \mathbb{R} \longrightarrow \mathbb{R}$$



$\longrightarrow$ Obviously some information is discarded!

Possible solution: Principal curves

$$\begin{pmatrix} X \\ Y \end{pmatrix} = f(t) + \epsilon\,;\ f : \mathbb{R} \longrightarrow \mathbb{R}^2 \qquad \text{(X and Y symmetric!)}$$



- Principal curves are smooth functions passing through the middle of the data cloud.
- They uncover nicely the underlying branched structure.
- However, they are not suitable for prediction of Y from a given $X = x$.

New approach: Multi-valued nonparametric regression

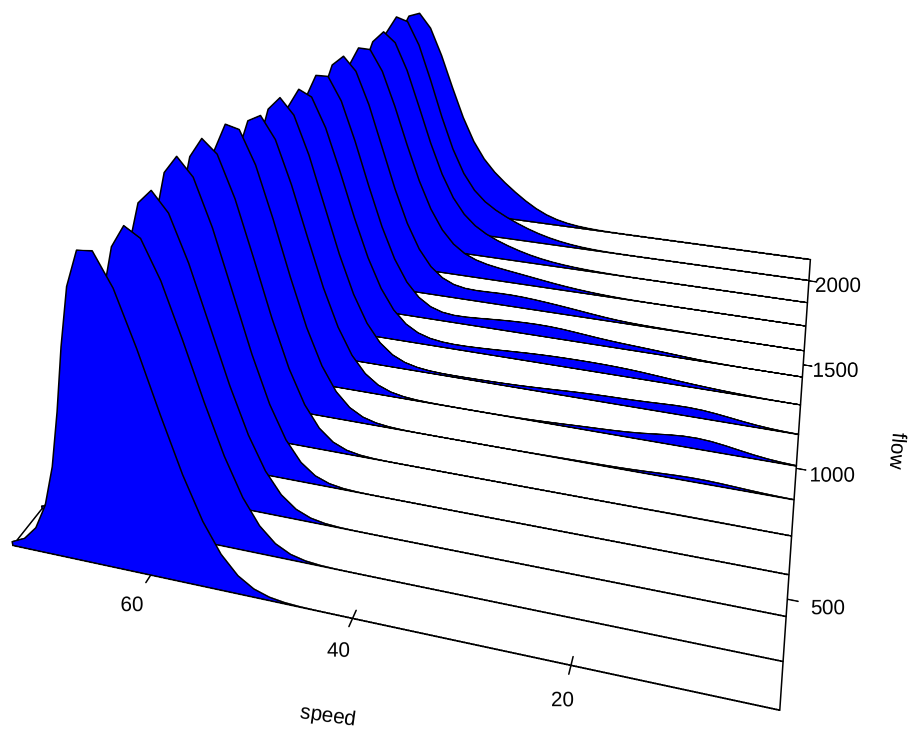
As for ordinary regression, we assume an asymmetric relationship

$$Y = M(X) + \epsilon,$$

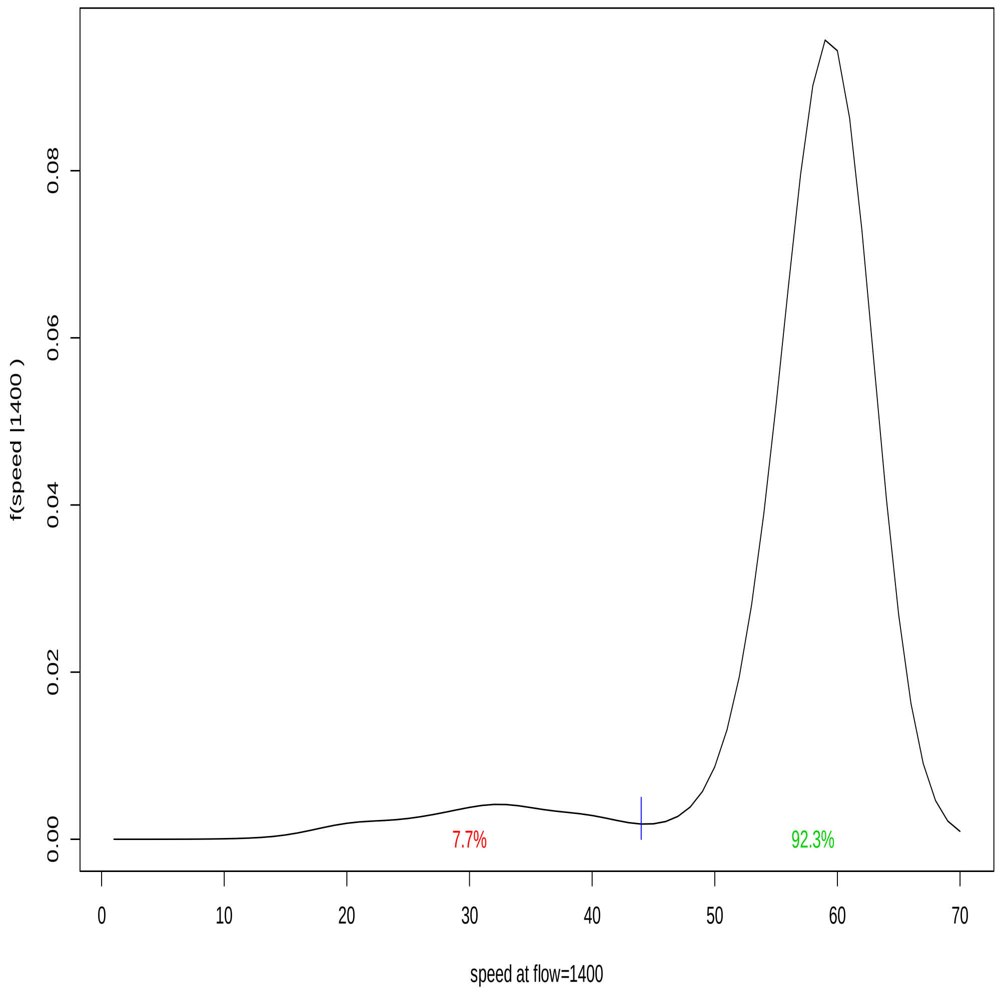
but now employing a multifunction (multi-valued 'function') $M : \mathbb{R} \longrightarrow \mathbb{R}$.

- The data cloud is thought of to consist of several smooth branches (regimes).
- For every $X = x$, more than one predicted value is possible.
- The estimator for M at x is a random set $\hat{M}(x) = \{\hat{m}_1(x), \ldots, \hat{m}_k(x)\}$, with k being the number branches.

Basic idea: Consider the conditional densities.

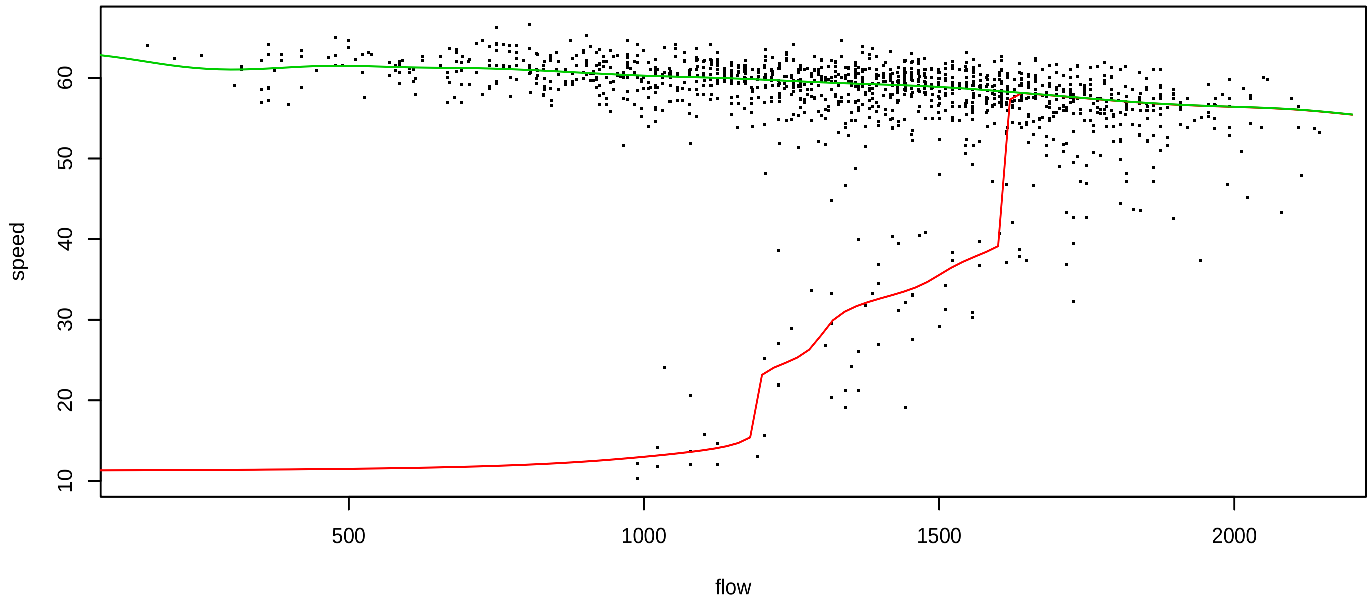


For instance, conditional density at a flow $= 1400$.



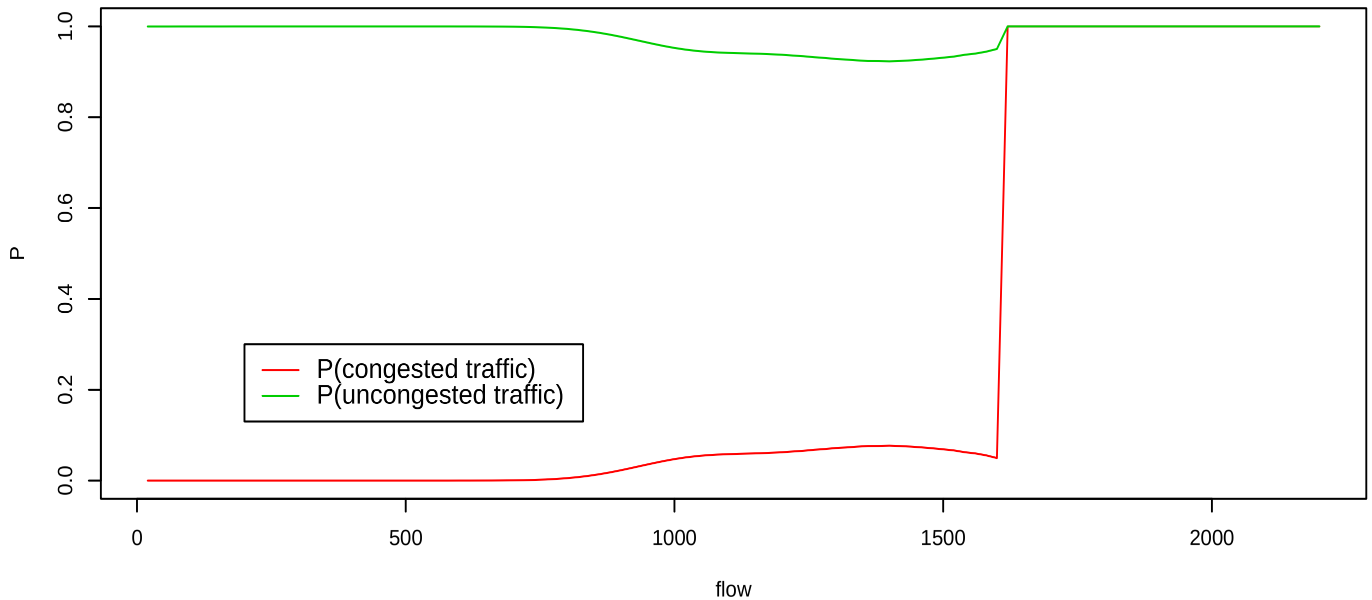
- For estimation of $M(x)$, compute the modes of the estimated conditional densities $\hat{f}(y|x)$.
- The area between a mode and the neighboring 'antimode' serves as estimated probability, that, given x, a value on the corresponding branch is attained.

Estimated multifunction



The estimated branches correspond to the uncongested and congested regimes.

Relevance assessment



The relevance of the branches varies smoothly with x, as long as as the branches can be separated.

Estimation of conditional modes

We are interested in all local maxima of the estimated conditional densities

$$\hat{f}(y|x) = \frac{\hat{f}(x,y)}{\hat{f}(x)} = \frac{\sum_{i=1}^{n} K_1\left(\frac{x-X_i}{h_1}\right) K_2\left(\frac{y-Y_i}{h_2}\right)}{h_2 \sum_{i=1}^{n} K_1\left(\frac{x-X_i}{h_1}\right)}$$

with kernels K_1, K_2 and bandwidths h_1, h_2. We assume, that a profile $k(\cdot)$ for kernel

K_2 exists such that

$$K_2(\cdot) = c_k k((\cdot)^2),$$

holds. One calculates

$$\frac{\partial \hat{f}(y|x)}{\partial y} = \frac{2c_k}{h_2^3} \sum_{i=1}^{n} K_1\left(\frac{x-X_i}{h_1}\right) k'\left(\left(\frac{y-Y_i}{h_2}\right)^2\right)(y-Y_i) \overset{!}{=} 0$$

and obtains

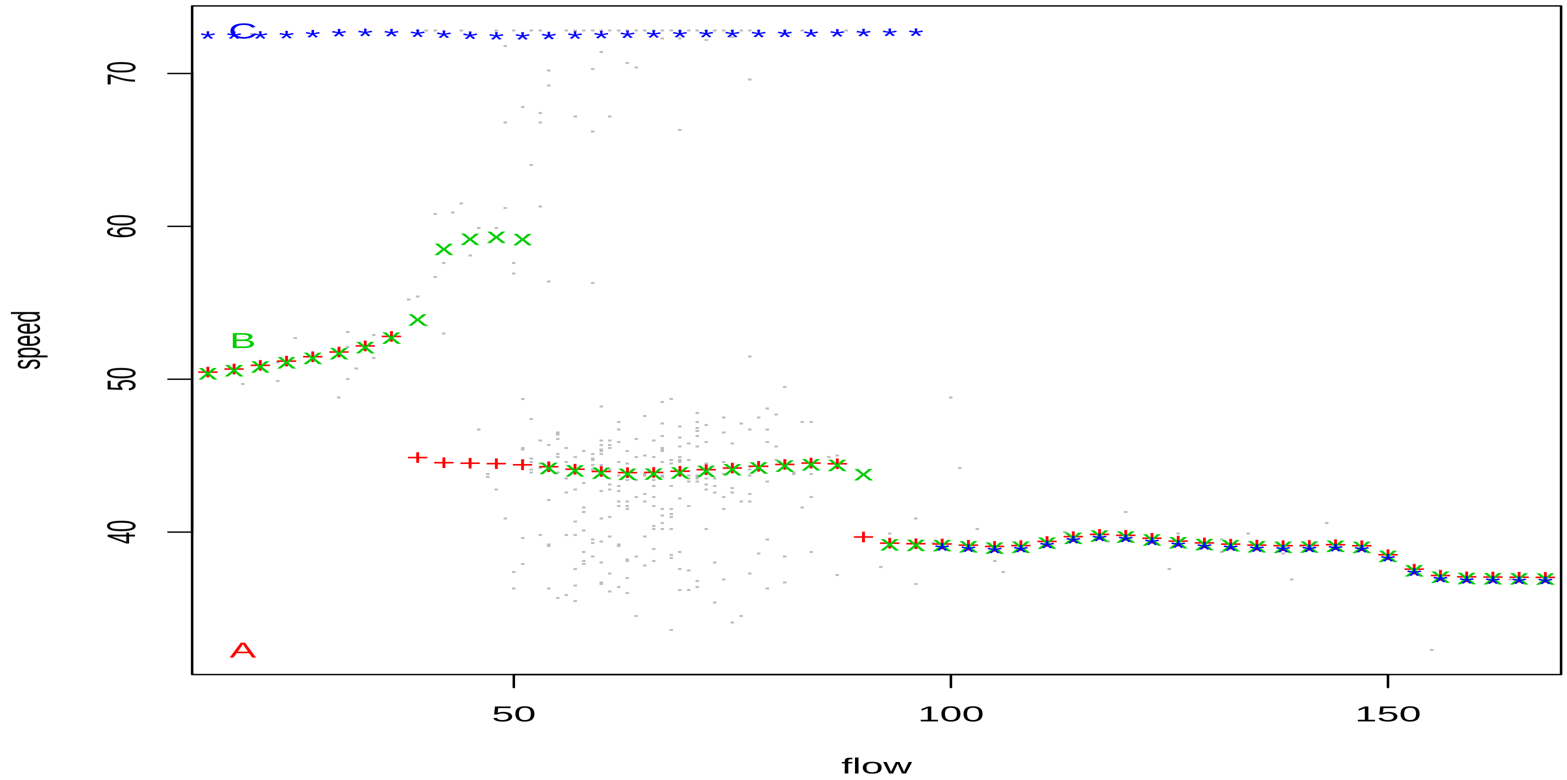
$$y = \frac{\sum_{i=1}^{n} K_1\left(\frac{x-X_i}{h_1}\right) G\left(\frac{y-Y_i}{h_2}\right) Y_i}{\sum_{i=1}^{n} K_1\left(\frac{x-X_i}{h_1}\right) G\left(\frac{y-Y_i}{h_2}\right)}. \tag{1}$$

with $G(\cdot) = -k'((\cdot)^2)$.

Remarks:

- Applying (1) iteratively, the sequence (y_j) corresponds to a series of 'local centers of mass' and converges to a conditional mode of $Y|X = x$ ("mean shift algorithm", Comaniciu & Meer, 2002).
- The right side of (1) corresponds to the "Sigma-Filter" used in digital image smoothing. Hence, the sigma filter is a one-step approximation to the conditional mode.

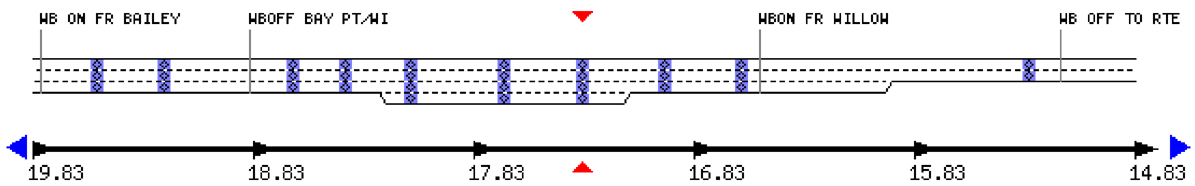
Example with 3 branches: Atypical speed-flow curves



recorded on a line segment for low-speed vehicles in Contra Costa County, California.

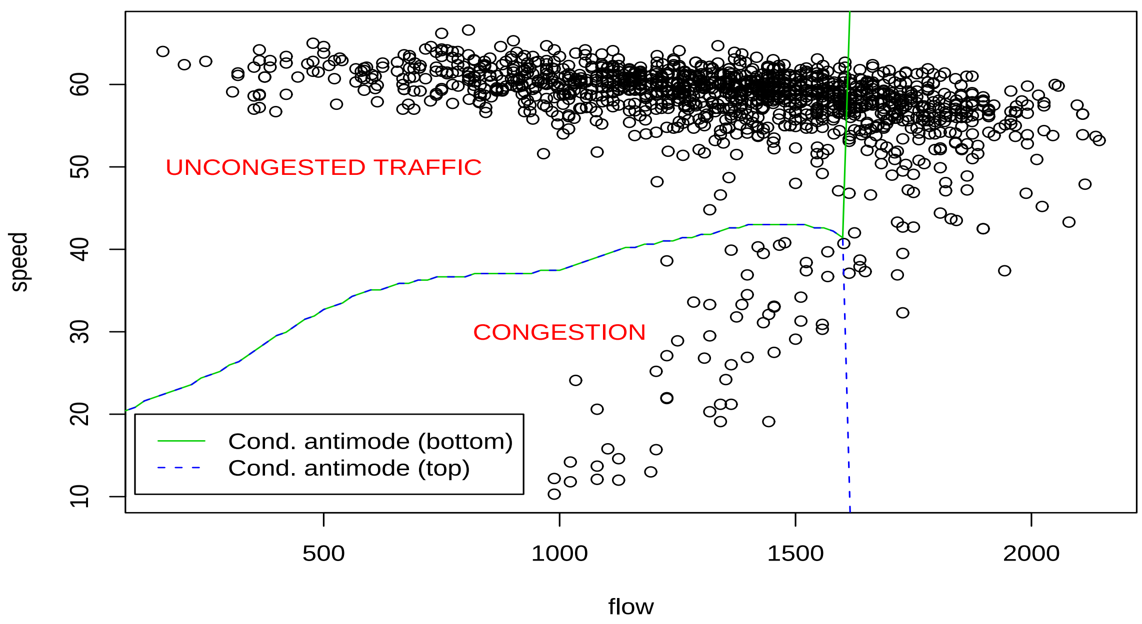
(Data from: ‘PemS’)

Highway Layout:



Outlook: Antiregression and Classification

If one plots the antimodes, which are obtained as a by-product of the computation of the relevances, one obtains an antiprediction or antiregression curve.



This curve serves as a separator between the branches, and thus as a tool to classify observations to the uncongested or congested regime.

Summary:

- The conditional mode is more useful for the analysis of multimodal data than the conditional mean or median.
- Maxima of the conditional density can be calculated fast and easily via a conditional mean shift procedure.
- The resulting curves are smooth, but edge-preserving.
- Sometimes problems for very large slopes.
- Function *modalreg* integrated into R package **hdrcde** (maintained by R. Hyndman), available on CRAN.
- To do: Automatic detection of the number of branches.

Literature

Bashtannyk & Hyndman (2001): Bandwidth selection for kernel conditional density estimation. *Comp. Stat. Data Analysis*, **36**, *279–298*.

Carreira-Perpiñan & Williams (2003): On the number of modes of a Gaussian mixture. *Lecture Notes in Computer Science*, **2695**, *625–640*.

Comaniciu & Meer (2002): Mean shift: A robust approach towards feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, **24**, *603-619*.

Berlinet, Gannoun & Matzner-Løber (1998): Normalité asymptotique d'estimateurs convergents du mode conditionnel. *Canadian Journal of Statistics*, **26**, *365–380*.

Einbeck & Tutz (2006): Modelling beyond regression functions: An application of multimodal regression to speed-flow data. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **55**, 461–475.

Silverman & Wood (1987): The nonparametric estimation of branching curves. *JASA*, **82**, *551-558.*