Plug-in Bandwidth Selectors for Bivariate Kernel Density Estimation

Tarn Duong
Supervisor: Dr Martin Hazelton

May 2002

Outline

1. Motivation

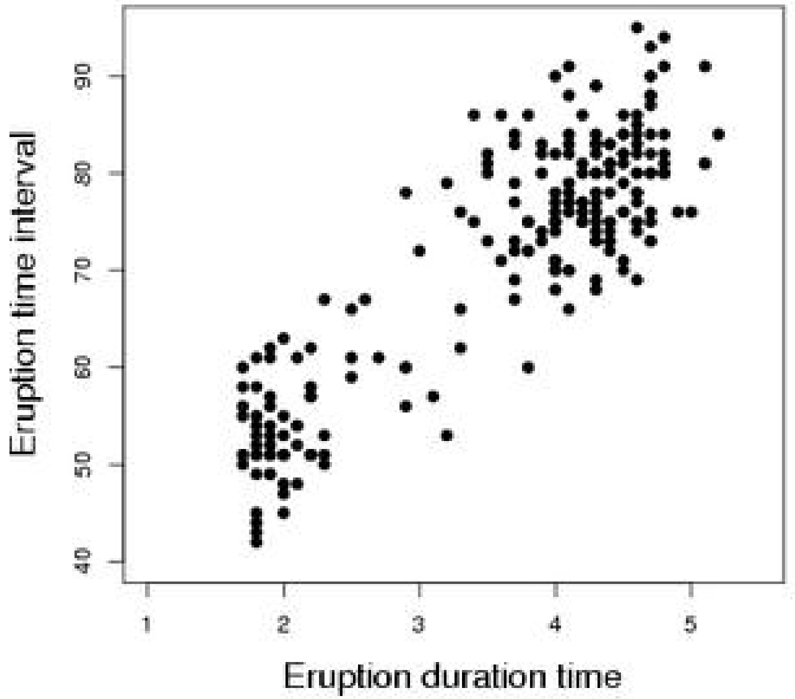
2. Basics of KDE

3. Plug-in bandwidth selectors
   - Existing and proposed methods

4. Simulation and real data results

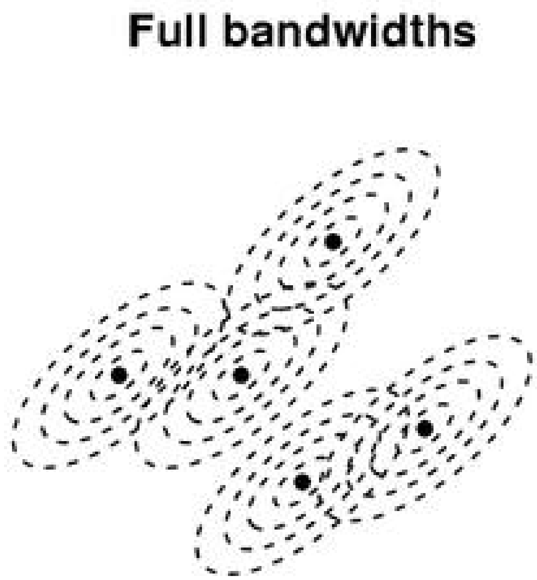
5. Summary

Motivation

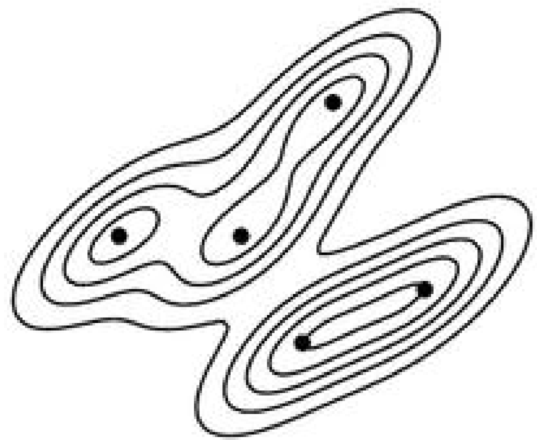


Eruption time interval
40
50
60
70
80
90
1
2
3
4
5
Eruption duration time

Forming KDE (1)



Forming KDE (2)



Equation for KDE

Kernel density estimate is

$$\hat{f}(\boldsymbol{x};\mathbf{H}) = n^{-1}\sum_{i=1}^{n} K_{\mathbf{H}}(\boldsymbol{x}-\boldsymbol{X}_i)$$

where

- $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is sample of n data points
- $\mathbf{H}$ is **bandwidth** matrix
- $K_{\mathbf{H}}(\boldsymbol{x}-\boldsymbol{X}_i)$ is normal pdf with mean $\boldsymbol{X}_i$ and variance $\mathbf{H}$

Advantages of KDE

- does not require parametric form i.e. is non-parametric
- always guaranteed to be a proper pdf

MISE

Mean Integrated Squared Error (MISE) is

$$\begin{aligned}\mathrm{MISE}\,\hat{f}(\cdot;\mathbf{H}) &= \int_{\mathbb{R}^2} \mathrm{MSE}\ \hat{f}(\boldsymbol{x};\mathbf{H})\ d\boldsymbol{x} \\ &= \int_{\mathbb{R}^2} \mathbb{E}[\hat{f}(\boldsymbol{x};\mathbf{H}) - f(\boldsymbol{x})]^2\ d\boldsymbol{x}\end{aligned}$$

where

- $\hat{f}(\boldsymbol{x};\mathbf{H})$ is kernel density estimate
- $f(\boldsymbol{x})$ is target density

AMISE

Asymptotic Mean Integrated Squared Error (AMISE) is

$$\text{AMISE}\,\hat{f}(\cdot;\mathbf{H}) = n^{-1}(4\pi)^{-1}|\mathbf{H}|^{-1/2} + \tfrac{1}{4}(\text{vech}^T\,\mathbf{H})\boldsymbol{\Psi}(\text{vech}\,\mathbf{H})$$

where

- $\text{vech}\,\mathbf{H}$ is vector of lower triangular half of $\mathbf{H}$
  i.e. if $\mathbf{H} = \begin{bmatrix} h_1^2 & h_{12} \\ h_{12} & h_2^2 \end{bmatrix}$ then $\text{vech}\,\mathbf{H} = [h_1^2\ h_{12}\ h_2^2]$

- $\boldsymbol{\Psi}$ is matrix of functionals that depend on f

Plug-in selector

1. choose diagonal or full bandwidth matrix

2. estimate $\mathbf{\Psi}$

3. PI is AMISE with $\mathbf{\Psi}$ replaced with $\hat{\mathbf{\Psi}}$

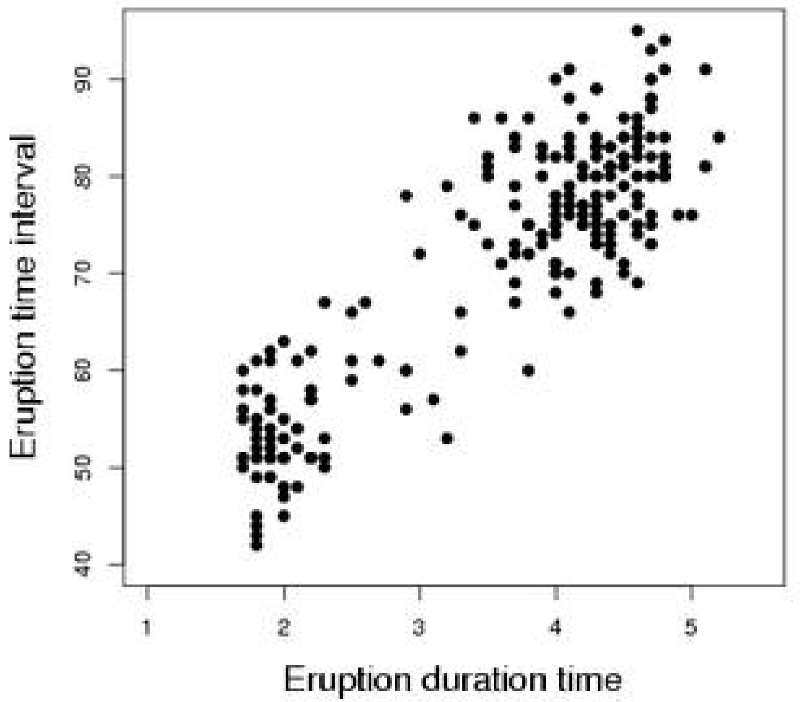
$$\mathrm{PI}(\mathbf{H}) = n^{-1}(4\pi)^{-1}|\mathbf{H}|^{-1/2} + \tfrac{1}{4}(\mathrm{vech}^T\,\mathbf{H})\hat{\mathbf{\Psi}}(\mathrm{vech}\,\mathbf{H})$$

4. optimal bandwidth is $\hat{\mathbf{H}}_{\mathrm{PI}} = \underset{\mathbf{H}}{\mathrm{argmin}}\ \mathrm{PI}(\mathbf{H})$

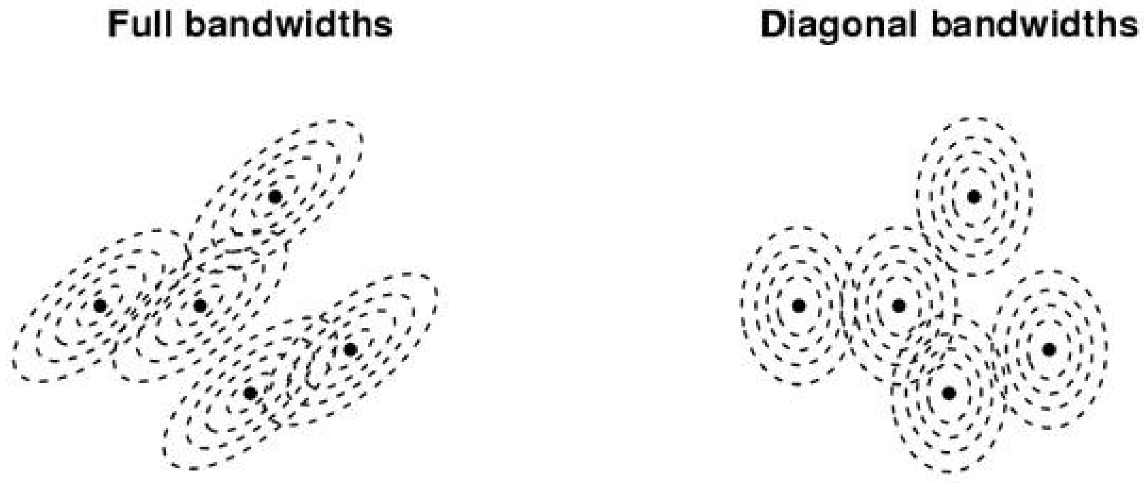
Existing and proposed methods

- **Existing**
  1. use diagonal bandwidth matrix
  2. estimate $\boldsymbol{\Psi}$ element-wise
- **Proposed**
  1. use full bandwidth matrix
  2. estimate $\boldsymbol{\Psi}$ matrix-wise

‘Old Faithful’ geyser data



Diagonal and full bandwidth matrices



Explicit expression for $\boldsymbol{\Psi}$

- partial derivative: $f^{(r_1,r_2)}(\boldsymbol{x}) = \dfrac{\partial^{(r_1+r_2)}}{\partial x_1^{r_1}\,\partial x_2^{r_2}} f(\boldsymbol{x})$

- ψ functional: $\psi_{r_1,r_2} = \displaystyle\int_{\mathbb{R}^2} f^{(r_1,r_2)}(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x}$

- matrix: $\boldsymbol{\Psi} = \begin{bmatrix} \psi_{40} & 2\psi_{31} & \psi_{22} \\ 2\psi_{31} & 4\psi_{22} & 2\psi_{13} \\ \psi_{22} & 2\psi_{13} & \psi_{04} \end{bmatrix}$

Estimating ψ_{r_1,r_2}

- $\psi_{r_1,r_2} = \mathbb{E}\, f^{(r_1,r_2)}(\boldsymbol{X})$

- $\hat{\psi}_{r_1,r_2} = n^{-1} \sum_{i=1}^{n} \hat{f}^{(r_1,r_2)}(\boldsymbol{X}_i; g)$
  where g is a (scalar) **pilot** bandwidth

- choose g via minimising $\mathrm{MSE}\ \psi_{r_1,r_2}$

Estimating ψ_{r_1,r_2} - existing method (1)

Asymptotic MSE of $\hat{\psi}_{r_1,r_2}$ is

$$\mathrm{AMSE}\,\hat{\psi}_{r_1,r_2}(g) = (4\pi)^{-1} n^{-2} g^{-10} + \left[n^{-1} g^{-6} K^{(r_1,r_2)}(\mathbf{0}) + \frac{1}{2} g^2 (\psi_{r_1+2,r_2} + \psi_{r_1,r_2+2}) \right]^2$$

Estimating ψ_{r_1,r_2} - existing method (2)

If r_1, r_2 are both even then

$$g_{r_1,r_2} = \left[\frac{-2K^{(r_1,r_2)}(\mathbf{0})}{(\psi_{r_1+2,r_2} + \psi_{r_1,r_2+2})n}\right]^{1/8}$$

If r_1, r_2 are both odd then

$$g_{r_1,r_2} = \left[\frac{5\psi_{0,0}}{2\pi(\psi_{r_1+2,r_2} + \psi_{r_1,r_2+2})^2 n^2}\right]^{1/14}$$

Estimating Ψ - existing method

$$\hat{\mathbf{\Psi}} = \begin{bmatrix} \hat{\psi}_{40}(g_{40}) & 2\hat{\psi}_{31}(g_{31}) & \hat{\psi}_{22}(g_{22}) \\ 2\hat{\psi}_{31}(g_{31}) & 4\hat{\psi}_{22}(g_{22}) & 2\hat{\psi}_{13}(g_{13}) \\ \hat{\psi}_{22}(g_{22}) & 2\hat{\psi}_{13}(g_{13}) & \hat{\psi}_{04}(g_{04}) \end{bmatrix}$$

$\hat{\mathbf{\Psi}}$ can be

- not $+$ve definite so no finite min for $\mathrm{PI}(\mathbf{H})$
- nearly singular so numerical instability in $\mathrm{PI}(\mathbf{H})$

Estimating ψ_{r_1,r_2} - proposed method (1)

Sum of AMSE is

$$\begin{aligned}\mathrm{SAMSE}\,(g) &= \sum_{r_1=0}^{4} \mathrm{AMSE}\,\hat{\psi}_{r_1,r_2}(g) \quad \text{where } r_2 = 4 - r_1 \\ &= n^{-2} g^{-12} A_2 + n^{-1} g^{-4} A_3 + \frac{1}{4} g^4 A_4\end{aligned}$$

where $A_2, A_4 > 0$ and $A_3 < 0$.

Estimating ψ_{r_1,r_2} - proposed method (2)

$$g = \left[\frac{6A_2}{(-A_3 + \sqrt{A_3^2 + 3A_1A_2})n} \right]^{1/8}$$

Estimating Ψ - proposed method

$$\hat{\mathbf{\Psi}} = \begin{bmatrix} \hat{\psi}_{40}(g) & 2\hat{\psi}_{31}(g) & \hat{\psi}_{22}(g) \\ 2\hat{\psi}_{31}(g) & 4\hat{\psi}_{22}(g) & 2\hat{\psi}_{13}(g) \\ \hat{\psi}_{22}(g) & 2\hat{\psi}_{13}(g) & \hat{\psi}_{04}(g) \end{bmatrix}$$

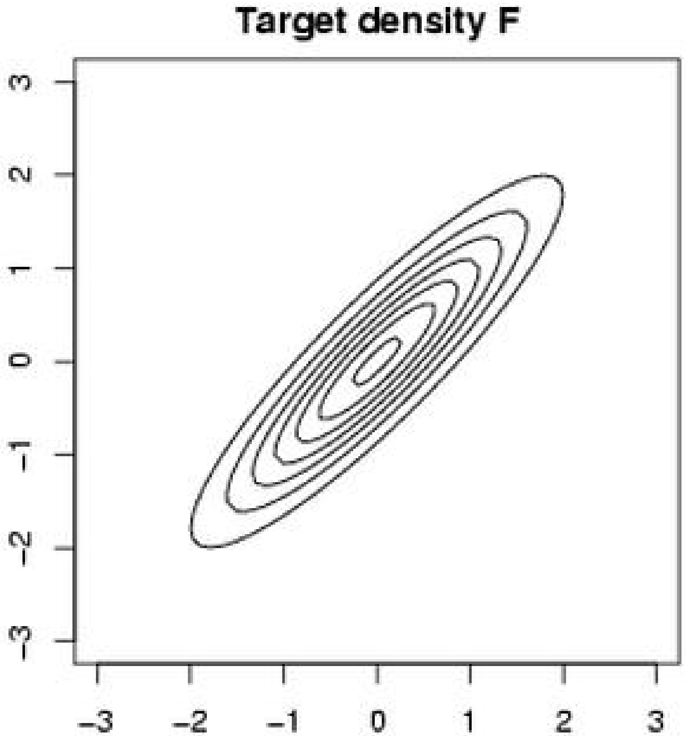
- $\hat{\mathbf{\Psi}}$ is $+$ve definite.
- lose a little efficiency compared to exisiting method

Simulation results (1)

- normal mixture densities
  - wide range of features
  - closed form for $\mathrm{ISE}\hat{f}(\cdot;\mathbf{H}) = \int_{\mathbb{R}^2} [\hat{f}(\boldsymbol{x};\mathbf{H}) - f(\boldsymbol{x})]^2 d\boldsymbol{x}$
- use $\log(\mathrm{ISE})$ to compare performance between
  - diagonal selector with different pilots
  - full selector with different pilots
  - full selector with single pilot
- 400 simulations, each of size $n = 1000$

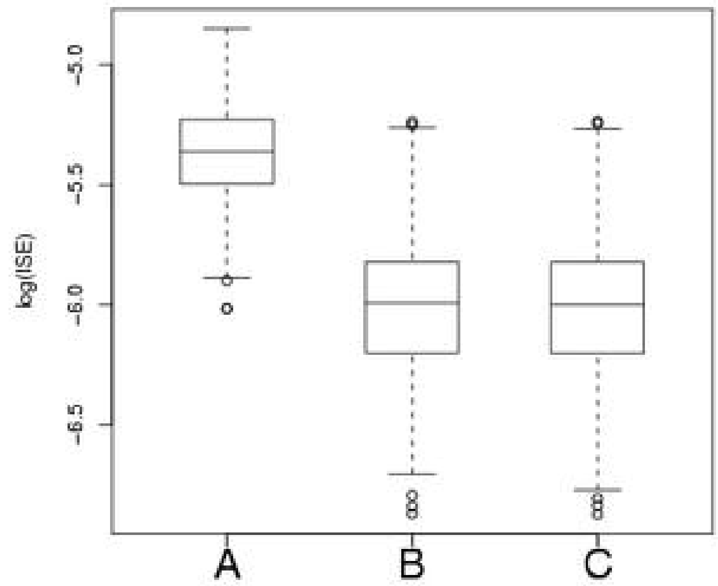
Simulation results (2)

Highly correlated normal density - contour plot



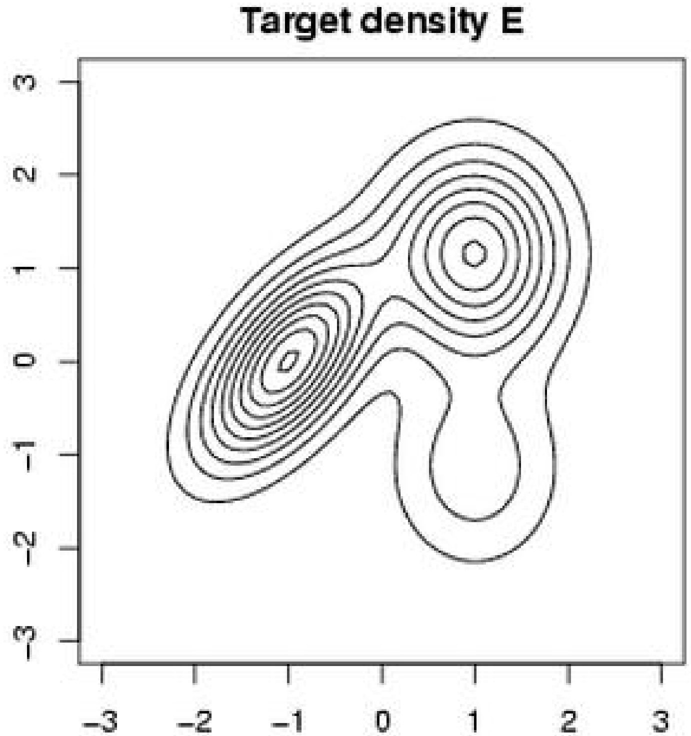
Simulation results (3)

Highly correlated normal density - boxplot



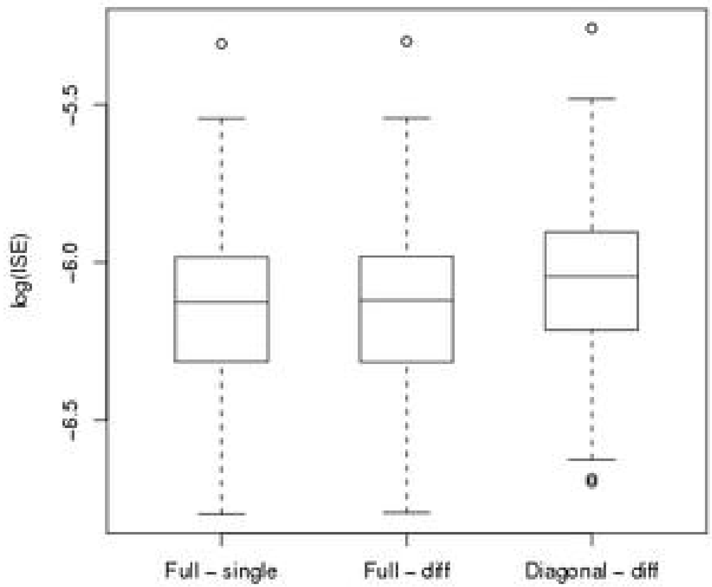
Simulation results (4)

Trimodal density - contour plot

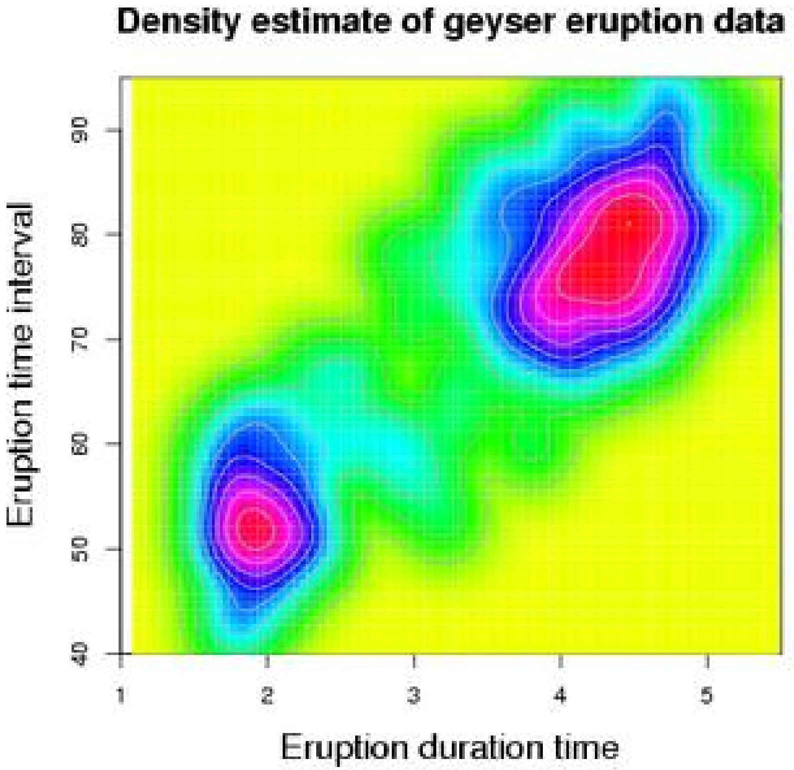


Simulation results (5)

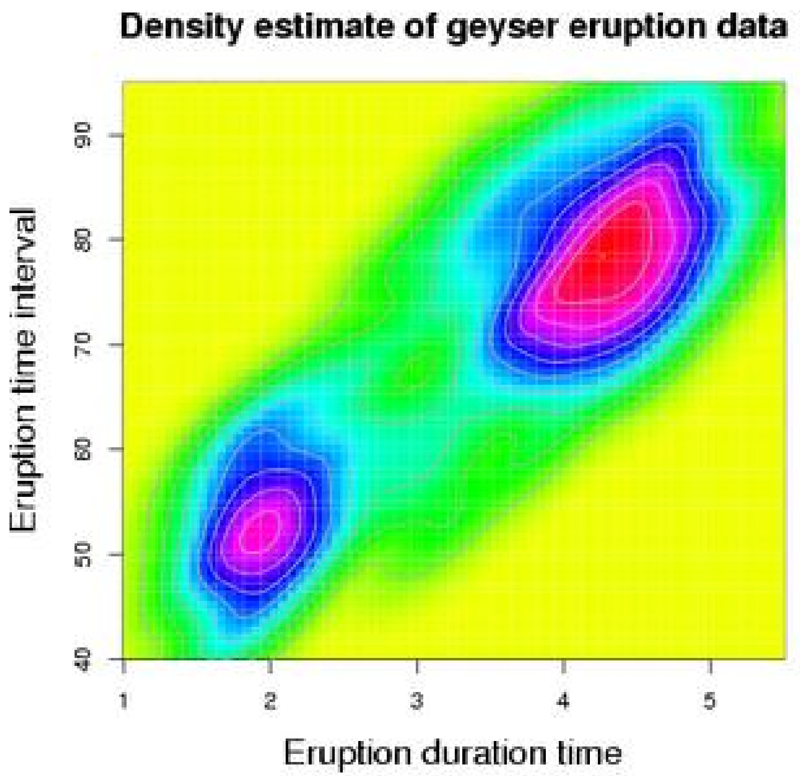
Trimodal density - boxplot



Real data KDE - existing method



Real data KDE - proposed method



Summary

Existing method	Proposed method
Diagonal $\mathbf{H}$	Full $\mathbf{H}$
Different pilots	Single pilot
$\hat{\boldsymbol{\Psi}} \not> 0$	$\hat{\boldsymbol{\Psi}} > 0$
Poor estimation of data not aligned to axes	Good estimation of data not aligned to axes