Generalized Significance in Scale Space: The GS3 Package

Daniel V. Samarov

Statistical Engineering Division
Information Technology Laboratory
National Institute of Standards and Technology

July 21, 2010

Table of Contents

Green House Gas (GHG) Emission Measurement

NIST developing technology & standards for remote sensing of GHG's

Green House Gas (GHG) Emission Measurement

NIST developing technology & standards for remote sensing of GHG's

DIAL for distributed sources

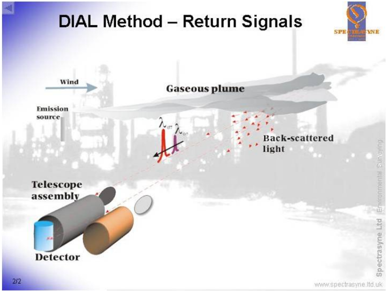
- **DI**fferential **A**bsorbtion **L**IDAR
- Range resolved, column integrated measurements

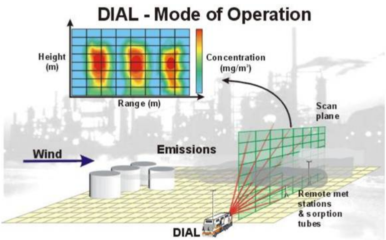
Green House Gas (GHG) Emission Measurement

NIST developing technology & standards for remote sensing of GHG's

DIAL for distributed sources

- **DI**fferential **A**bsorbtion **L**IDAR
- Range resolved, column integrated measurements





HSRL

DIAL technology not quite ready for primetime. Collaboration w/ NASA

HSRL

DIAL technology not quite ready for primetime. Collaboration w/ NASA

HSRL

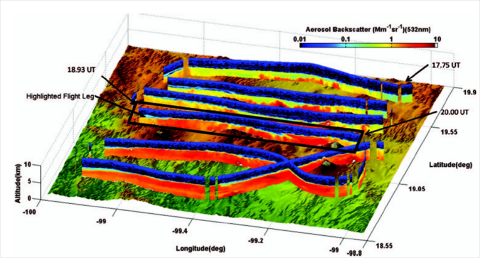
- **H**igh **S**pectral **R**esolution **L**IDAR
- Similar technology/data
- Validation of Calipso satellite measurements

HSRL

DIAL technology not quite ready for primetime. Collaboration w/ NASA

HSRL

- **H**igh **S**pectral **R**esolution **L**IDAR
- Similar technology/data
- Validation of Calipso satellite measurements



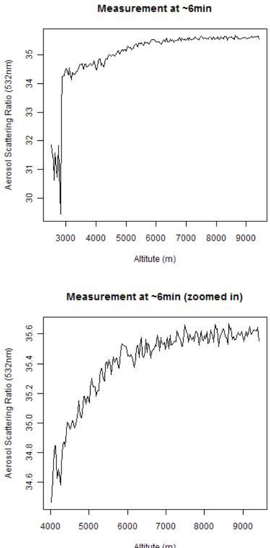
Hair et al. (2008)

Data graciously provided by NASA Langley Research Center

Challenges associated w/ HSRL & DIAL data

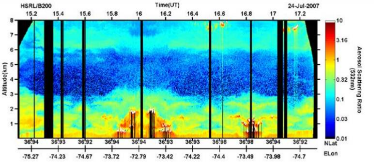
Challenges associated w/ HSRL & DIAL data

- Highly variable

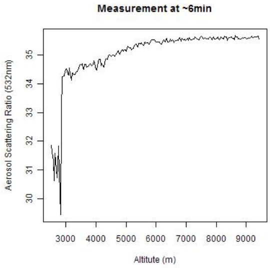


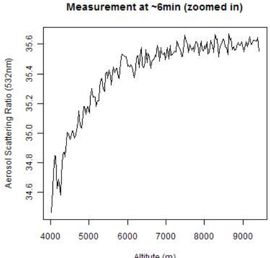
Challenges associated w/ HSRL & DIAL data

- Highly variable
- Subtle local structure



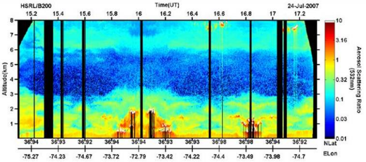
Hair et al. (2008)





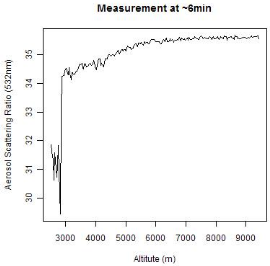
Challenges associated w/ HSRL & DIAL data

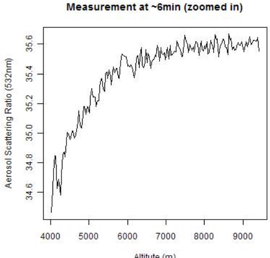
- Highly variable
- Subtle local structure



Hair et al. (2008)

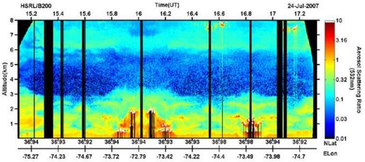
- Large ($\sim 300 \times 30,000$)





Challenges associated w/ HSRL & DIAL data

- Highly variable
- Subtle local structure

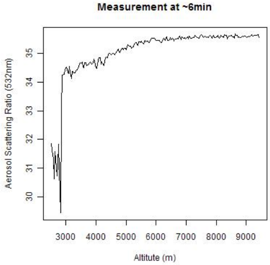


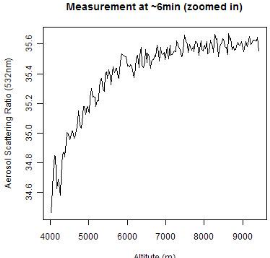
Hair et al. (2008)

- Large ($\sim 300 \times 30,000$)

Goals

- Estimate concentration (derivative)
- Calculate uncertainty





Local Polynomial Regression (LPR)

Natural choice for derivative estimation: LPR (Fan & Gijbels (1995))

Local Polynomial Regression (LPR)

Natural choice for derivative estimation: LPR (Fan & Gijbels (1995))

Pros

- Provides derivative estimate
- Locally adaptive
- Many other appealing properties

Local Polynomial Regression (LPR)

Natural choice for derivative estimation: LPR (Fan & Gijbels (1995))

Pros

- Provides derivative estimate
- Locally adaptive
- Many other appealing properties

Cons

- Challenge in 2($>$)-d: bandwidth choice (in particular *local*)
- Speed
- Exploratory analysis

Local Polynomial Regression (LPR)

Natural choice for derivative estimation: LPR (Fan & Gijbels (1995))

Pros

- Provides derivative estimate
- Locally adaptive
- Many other appealing properties

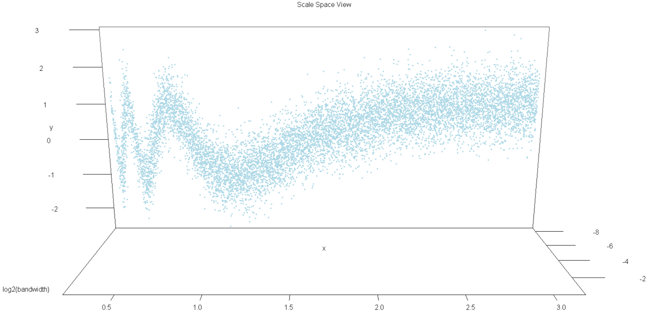
Cons

- Challenge in 2($>$)-d: bandwidth choice (in particular *local*)
- Speed
- Exploratory analysis

The GS3 package provides *a* solution

Scale Space

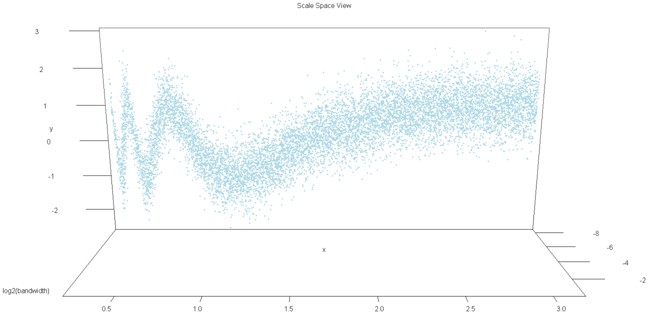
Scale space (Chaudhuri & Marron (2000)) a good starting point.
Consider the following example:



Scale Space

Scale space (Chaudhuri & Marron (2000)) a good starting point.
Consider the following example:

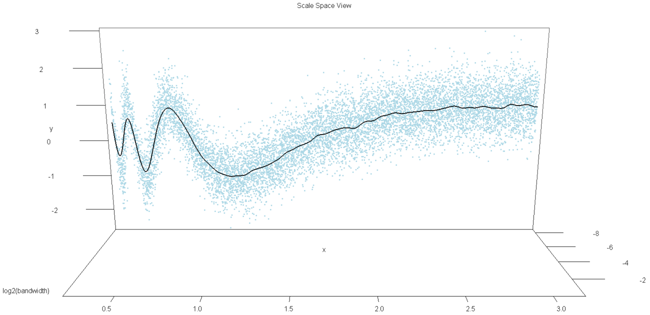
- Many instances where *a fit* desired



Scale Space

Scale space (Chaudhuri & Marron (2000)) a good starting point.
Consider the following example:

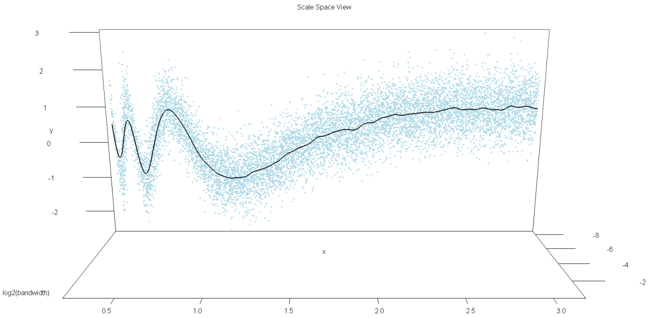
- Many instances where *a fit* desired



Scale Space

Scale space (Chaudhuri & Marron (2000)) a good starting point.
Consider the following example:

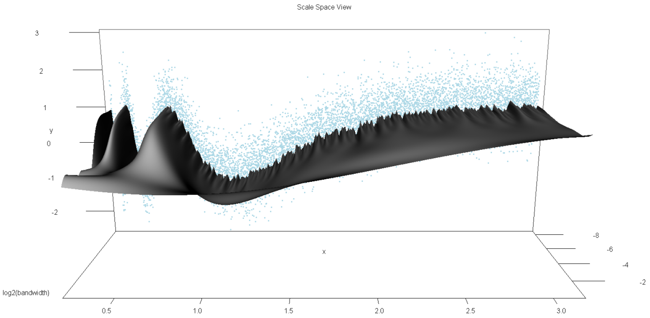
- Many instances where *a fit* desired
- However, good practice to look at multiple smooths



Scale Space

Scale space (Chaudhuri & Marron (2000)) a good starting point.
Consider the following example:

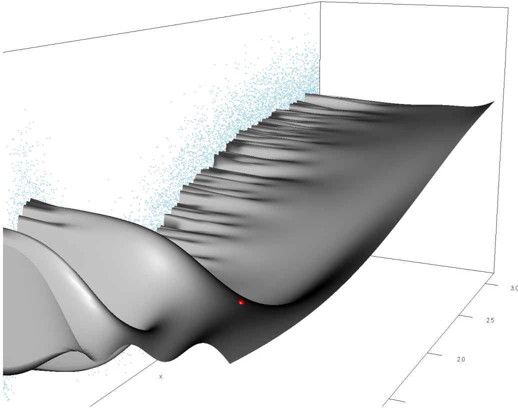
- Many instances where *a fit* desired
- However, good practice to look at multiple smooths
- Scale space studies a "family" of smooths



RODEO

RODEO (Wasserman & Lafferty (2008)) greedy algorithm for traversing "scale space surface"

Algorithm

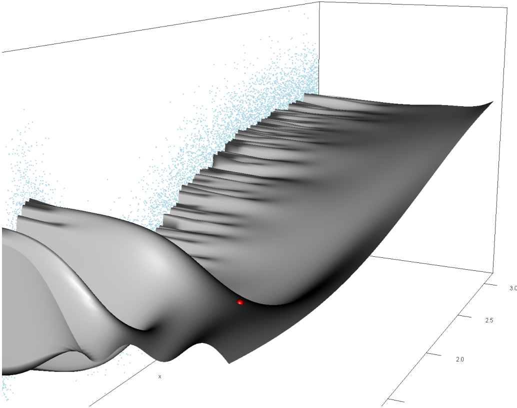


RODEO

RODEO (Wasserman & Lafferty (2008)) greedy algorithm for traversing "scale space surface"

Algorithm

- At a point, stay?

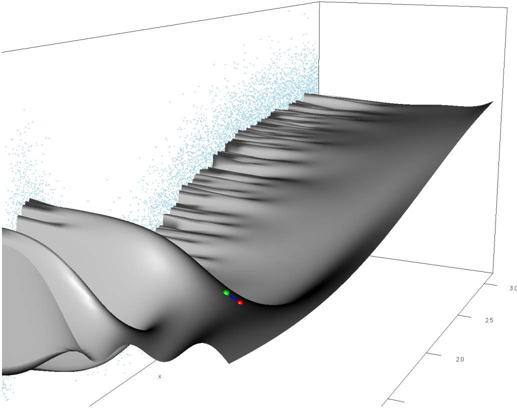


RODEO

RODEO (Wasserman & Lafferty (2008)) greedy algorithm for traversing "scale space surface"

Algorithm

- At a point, stay?
- or move?

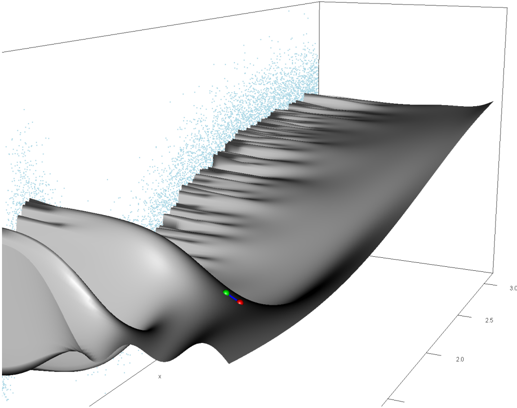


RODEO

RODEO (Wasserman & Lafferty (2008)) greedy algorithm for traversing "scale space surface"

Algorithm

- At a point, stay?
- or move?
- i.e. change from $\hat{m}_{h}(x)$ to $\hat{m}_{h}(x)$ significant?

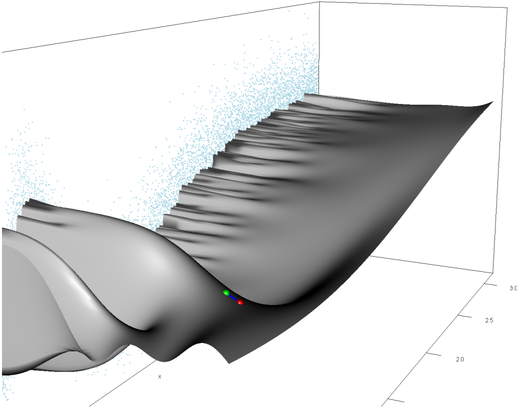


RODEO

RODEO (Wasserman & Lafferty (2008)) greedy algorithm for traversing "scale space surface"

Algorithm

- At a point, stay?
- or move?
- i.e. change from $\hat{m}_h(x)$ to $\hat{m}_h(x)$ significant?
- $Z = \frac{\partial \hat{m}_h(x)}{\partial h}$, test $|Z| > \sqrt{2\log(n)\mathrm{Var}(Z)}$



RODEO

RODEO (Wasserman & Lafferty (2008)) greedy algorithm for traversing "scale space surface"

Algorithm

- At a point, stay?
- or move?
- i.e. change from $\hat{m}_h(x)$ to $\hat{m}_h(x)$ significant?
- $Z = \frac{\partial \hat{m}_h(x)}{\partial h}$, test $|Z| > \sqrt{2\log(n)\mathrm{Var}(Z)}$
- NB: $\mathrm{Var}(Z) \sim \sigma^2$
  - σ^2 unknown population parameter

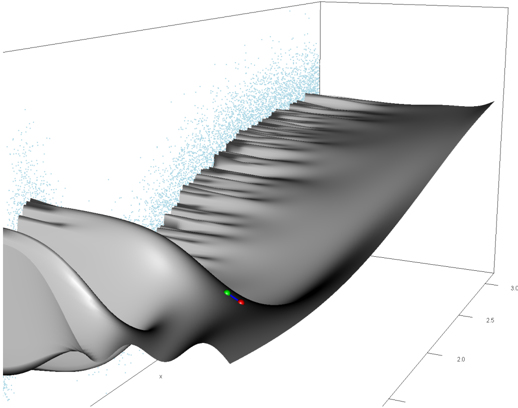
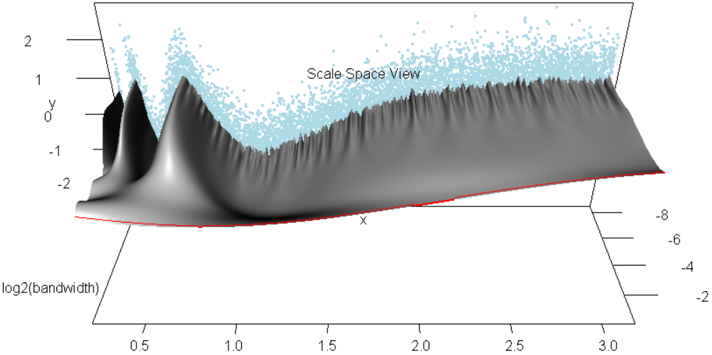


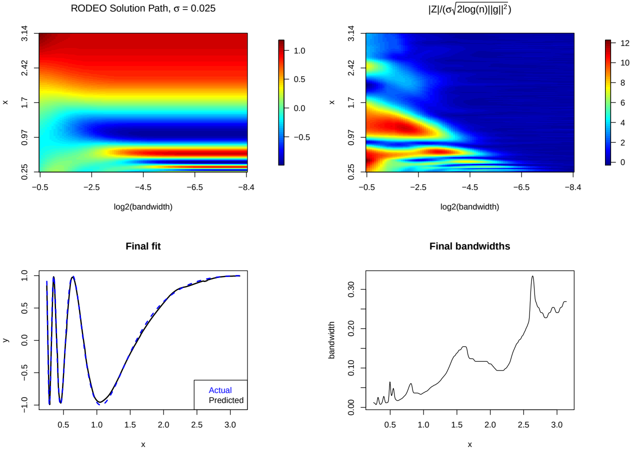
Illustration of RODEO in Scale Space

For $\sigma = 0.025$



Results

- Produces sensible smooth & bandwidths

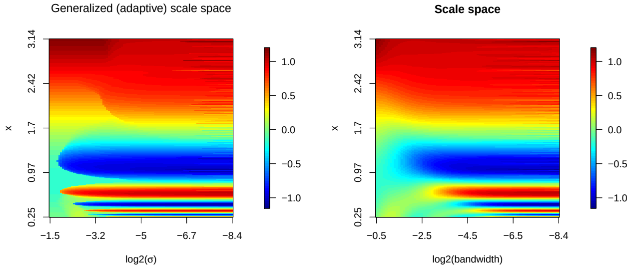


Back to σ

- Looking for *a* fit, could estimate σ

Back to σ

- Looking for *a* fit, could estimate σ
- However, can also intrepet σ in similar way to h

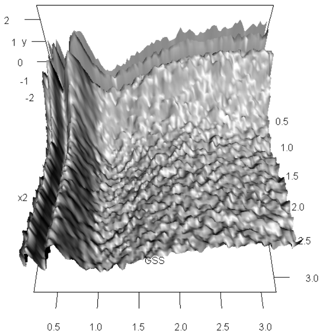


Higher dimensions

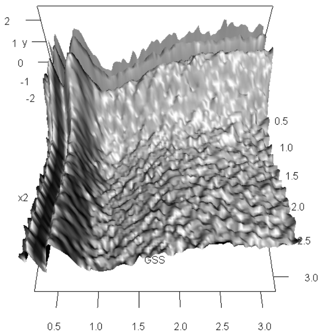
$d > 1$

- Straightforward, calculate Z_j, $j = 1, \ldots, d$ & $\mathrm{Var}(Z_j)$ $(\sim \sigma)$
- Same greedy approach used along each dimension

Generalized Scale Space



Scale Space

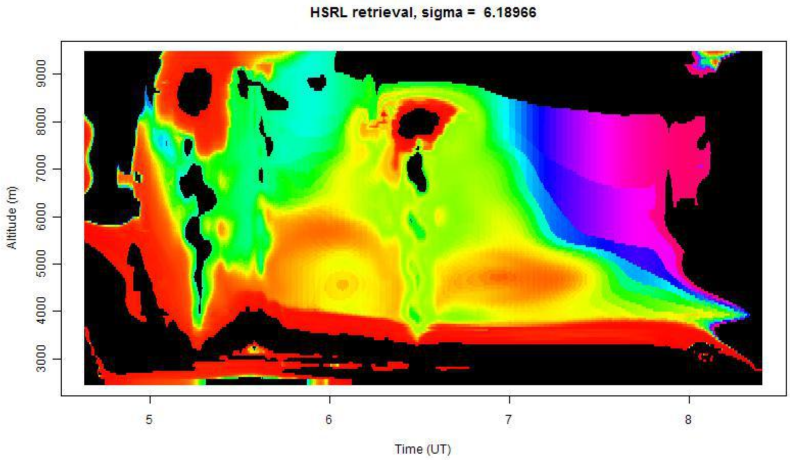


Computation

Speed of Bandwidth Selection

- Use linear binning ideas (Fan & Marron (1994), Wand (1995))
- M_i, $i = 1, \ldots, d$ grid size & $T = \#$ candidate bandwidths
- Selection takes $O(dTM_1 \log M_1 \ldots M_d \log M_d)$
- Compared to smooth w/ global bandwidth, $O(M_1 \log M_1 \ldots M_d \log M_d)$

Illustration of Generalized Scale Space on HSRL data



Conclusions & Future Work

Conclusions

- Provided R functions for 1, 2 & 3D binned LLR
- Implementation/development of method for fast local bandwidth estimation
- Generalization of scale space

Conclusions & Future Work

Conclusions

- Provided R functions for 1, 2 & 3D binned LLR
- Implementation/development of method for fast local bandwidth estimation
- Generalization of scale space

Future Work (currently in progress)

- Local variance function estimation
- Spatial dependence