T-61.6050 Special Course in Computer and Information Science V L: Nonlinear Dimensionality Reduction

Amaury Lendasse and Francesco Corona

Time Series Prediction and Chemoinformatics Group

Adaptive Informatics Research Centre

Helsinki University of Technology

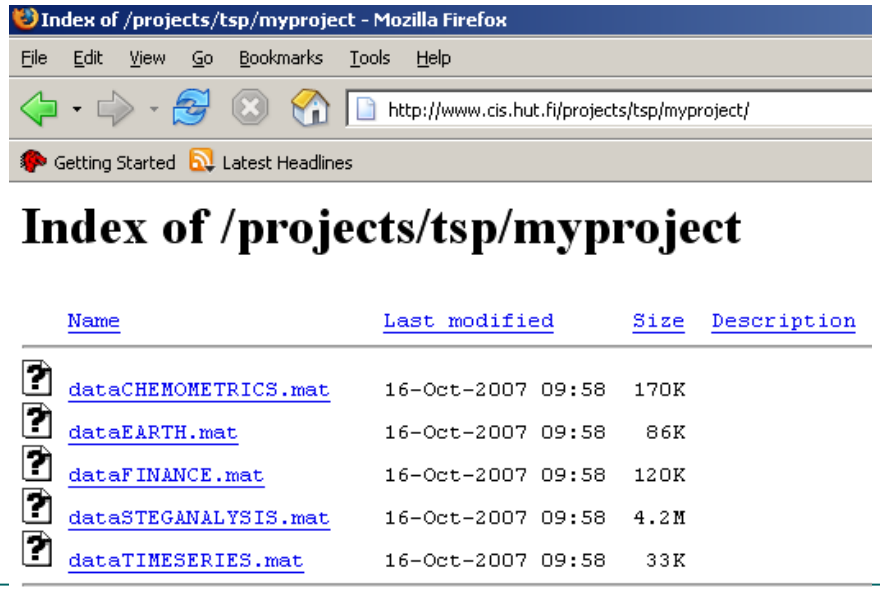
Project

- Deadline: Dec 21st at 3.45pm (by email to AL, FC and KN)
- Length: maximum 12 pages
- Latex Template:
  - http://www.estsp.org/files/ESTSP.cls
  - http://www.estsp.org/files/ESTSP.tex
- The report should be written like a scientific paper (Abstract, Introduction, Methods, Experiments, Conclusion, References, Appendix,…)
- NLDR Tools: http://www.dice.ucl.ac.be/~lee/software/nlp/main.html
- LSSVM: http://www.esat.kuleuven.ac.be/sista/lssvmlab/toolbox.html

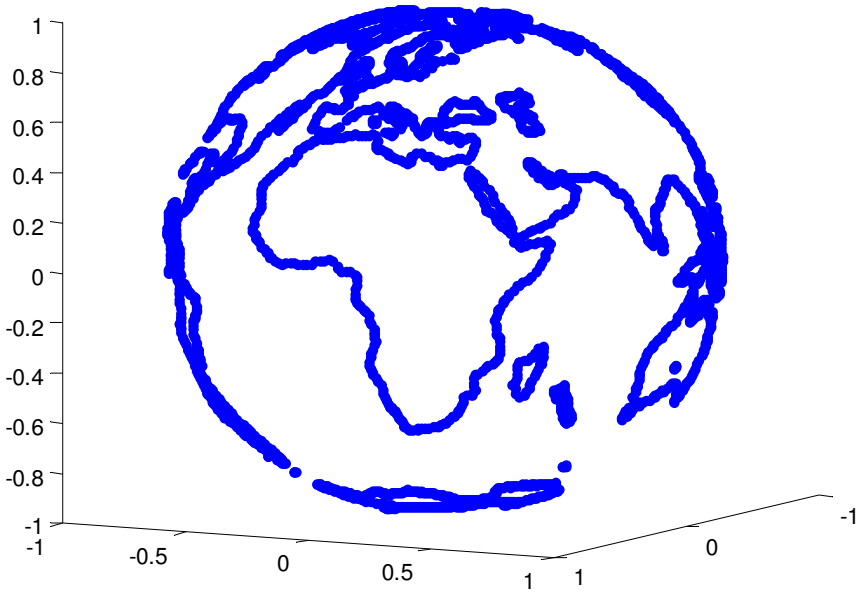
Project

- The datasets can be downloaded from:

  http://www.cis.hut.fi/projects/tsp/myproject/



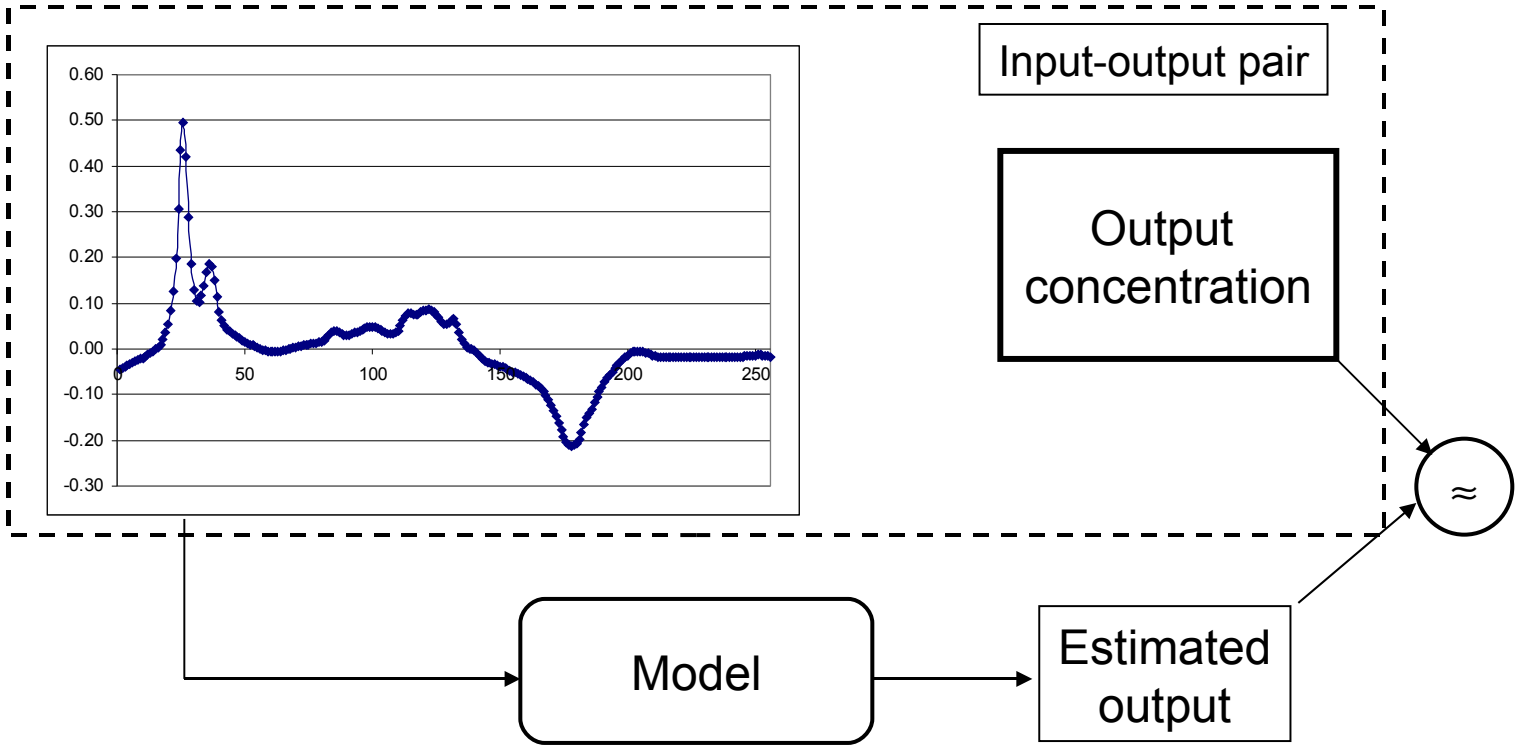
EARTH



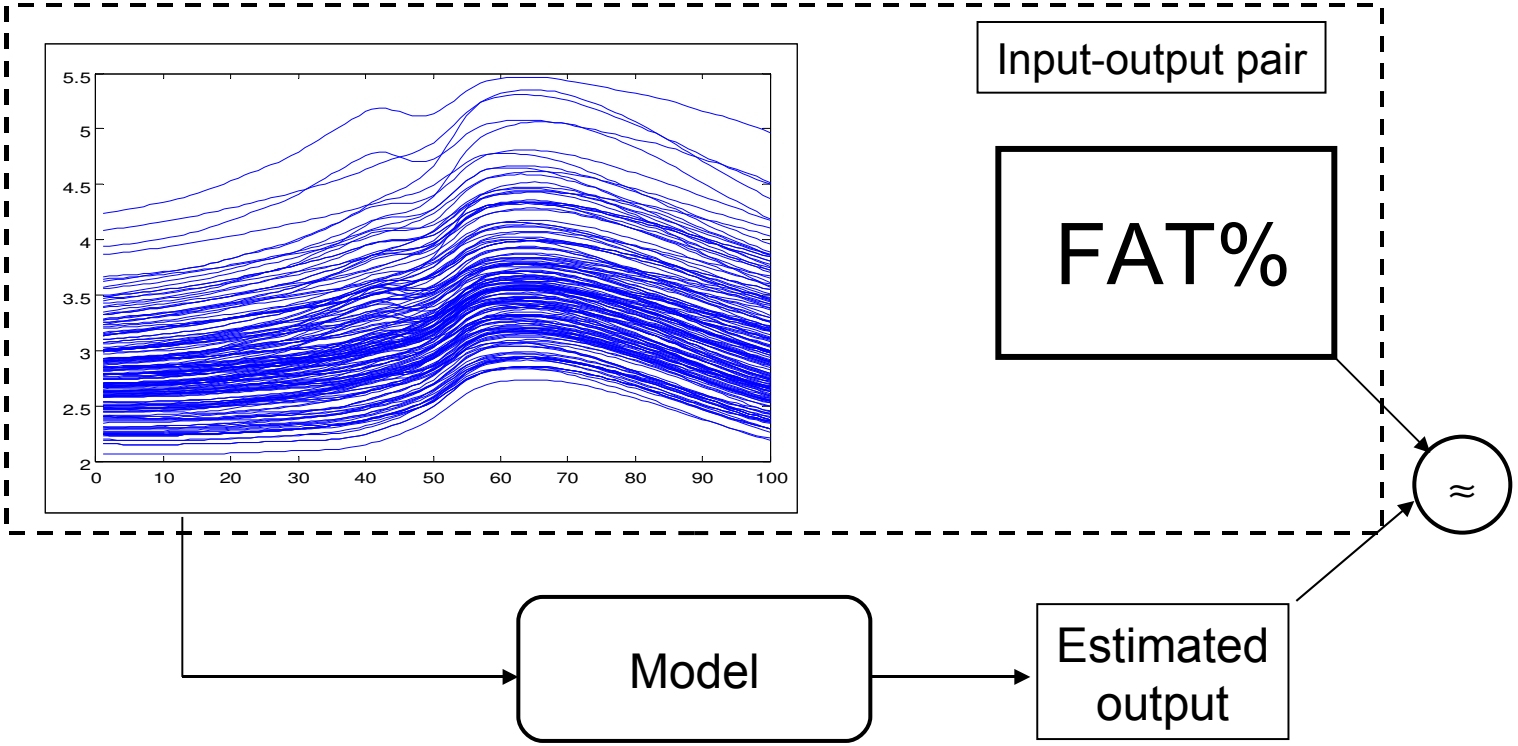
Intrinsic Dimension?

NLDR $\longrightarrow$ $z \in \Re^2$

CHEMOMETRICS



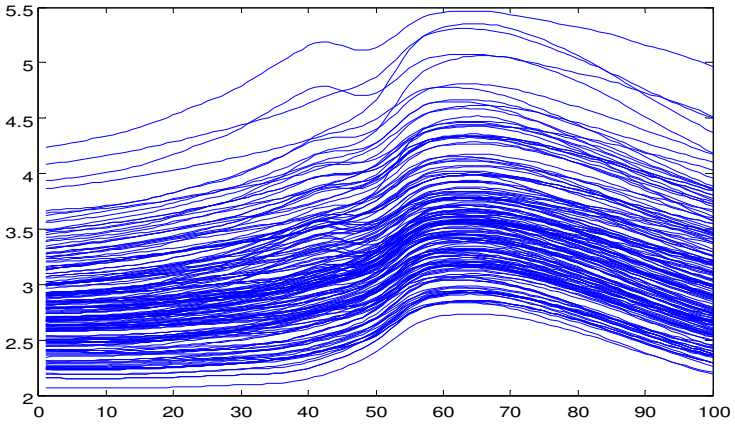
CHEMOMETRICS



TSPCi

CHEMOMETRICS: NLDR

Intrinsic Dimension?



NLDR → $z \in \Re^d$ → LSSVM → FAT%

With d = 1 to 5

Performance for Test Set?

TSPCi

FINANCE

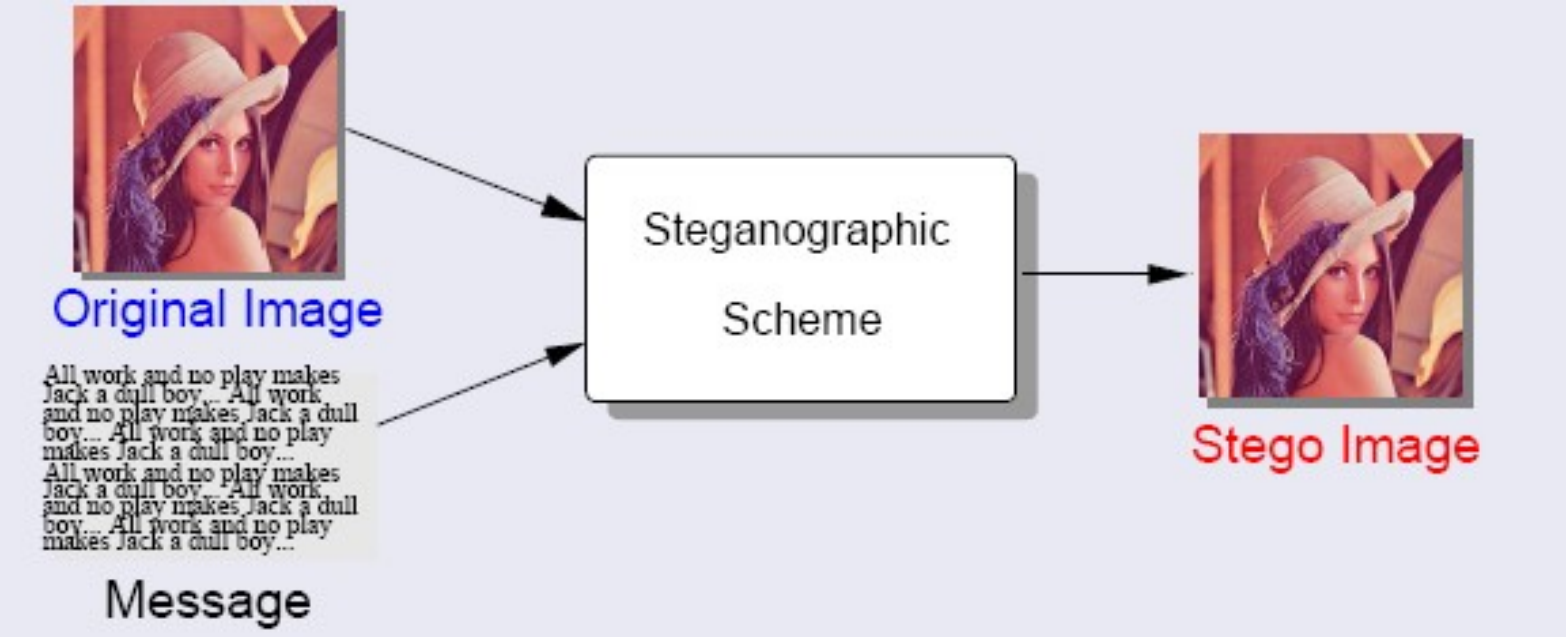
This data set contains 1053 samples which come from 117 companies during 9 years (1997 to 2005). Each samples has one output and 31 input variables.

$X \in \Re^{31}$ Intrinsic Dimension?

$$\xrightarrow{\text{NLDR}} \quad Z \in \Re^{2}$$

Do you see some clusters?
What about the target?
Plot the target with respect to *z*

STEGANALYSIS



STEGANALYSIS

TSPCi

STEGANALYSIS

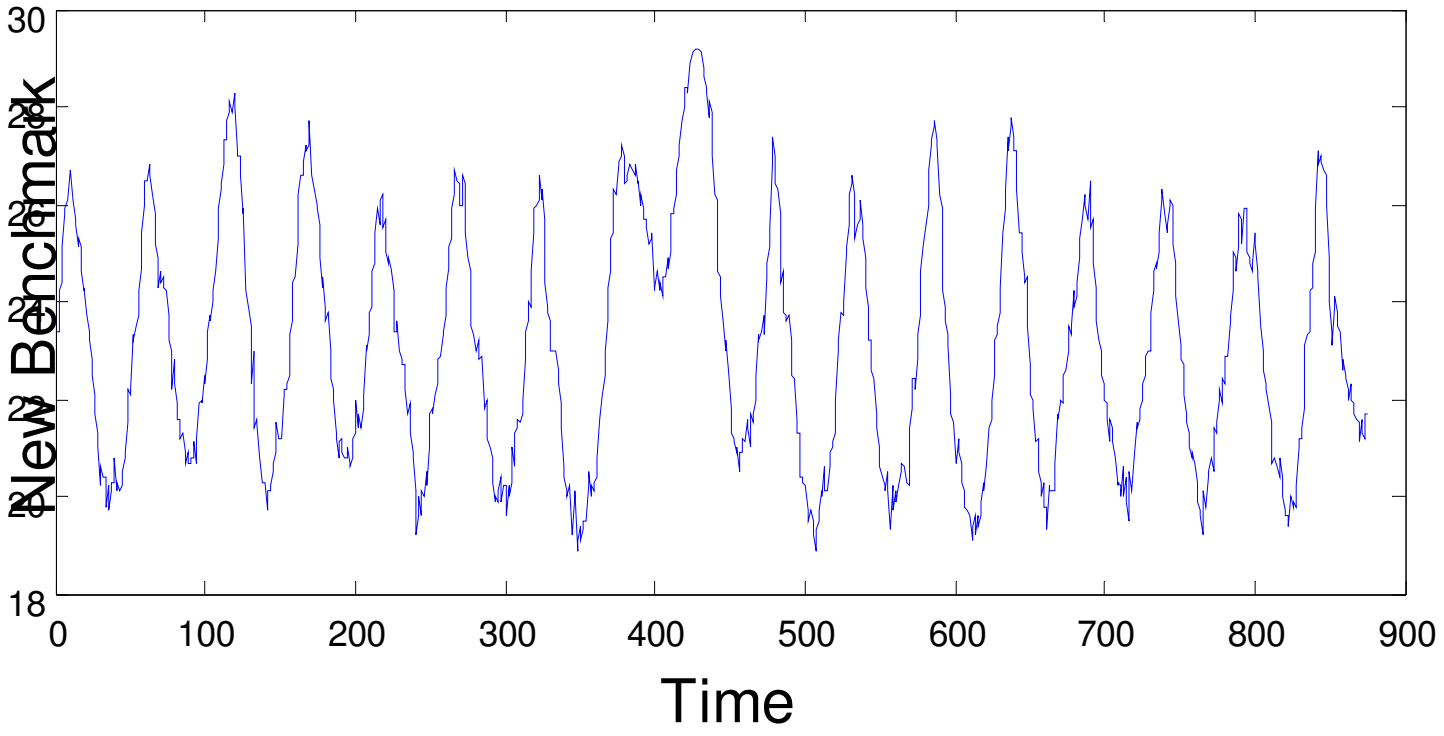
$X \in \Re^{192}$ Intrinsic Dimension?

NLDR → $Z \in \Re^{d}$

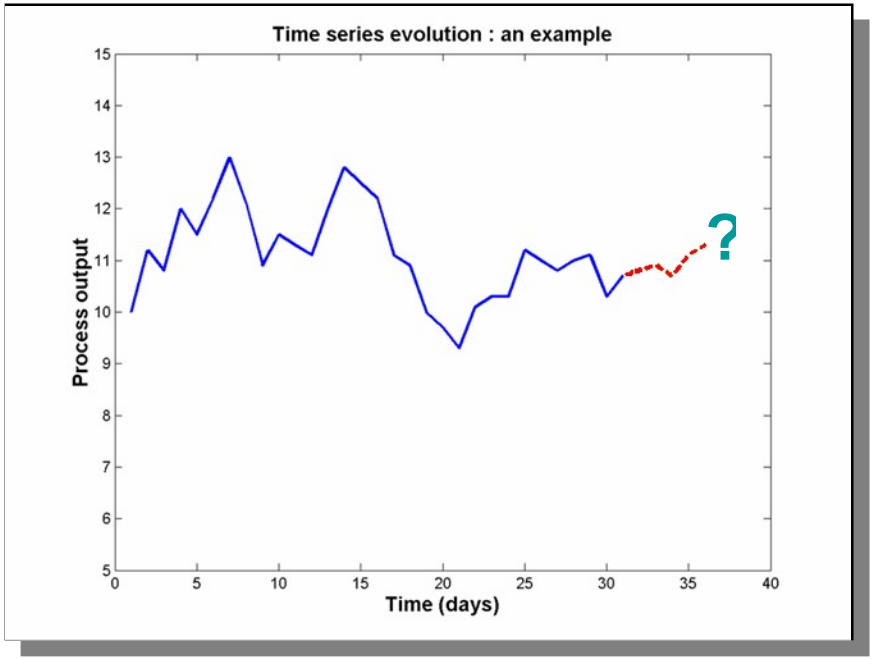
With d = 1 to 3

Do you see some clusters?
What about the Class?
The Class is the first column, remove it before NLDR

Time Series



Time Series



$$\hat{y}_{t+1} = \mathrm{f}(\, y_t, y_{t-1}, \ldots, y_{t-n}, \theta)$$

Time Series

$$y_1, y_2, \ldots, y_n$$

$$\downarrow$$

$$z_1, z_2, \ldots, z_k$$

With k << n

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, \ldots, y_{t-n}, \theta)$$

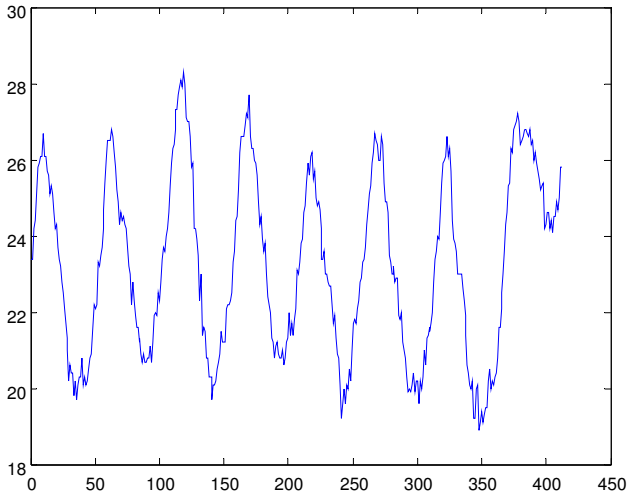
$$\downarrow$$

$$\hat{y}_{t+1} = g(z_1, z_2, \ldots, z_k)$$

Time Series

$x \in \Re^{52}$

Intrinsic Dimension?



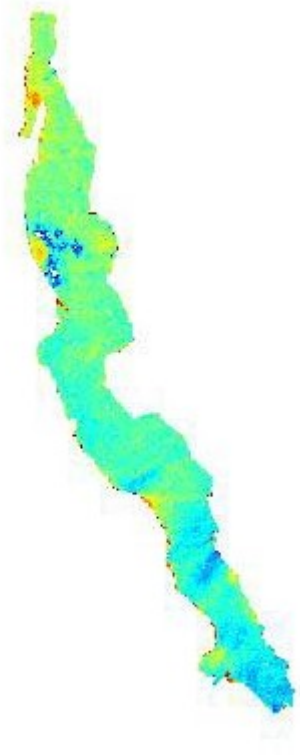
NLDR → $z \in \Re^{d}$ → LSSVM → Next Value

With d = 1 to 5

Performance for Test Set?

Climate

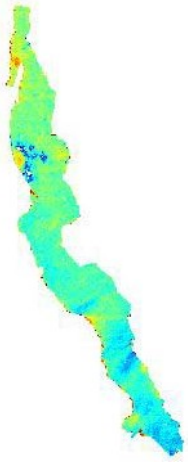

Bujumbura
BURUNDI
Gitega
Ruyigi
Rutana
Bururi
Kigoma
Lake Tanganyika
Kalemie
TANZ
MAHALI MOUNTAINS
MITUMBA MOUNTAINS
EGION
1470 m
F THE CONGO
Lake Rukwa
Sumbawanga
Lake Mweru
Mbala



Climate

$x \in \Re^{?}$

Intrinsic Dimension?



NLDR

$z \in \Re^{d}$

With d = 1 to 10

Reconstruction error?