Predicting aboveground biomass from field data and hyperspectral image information

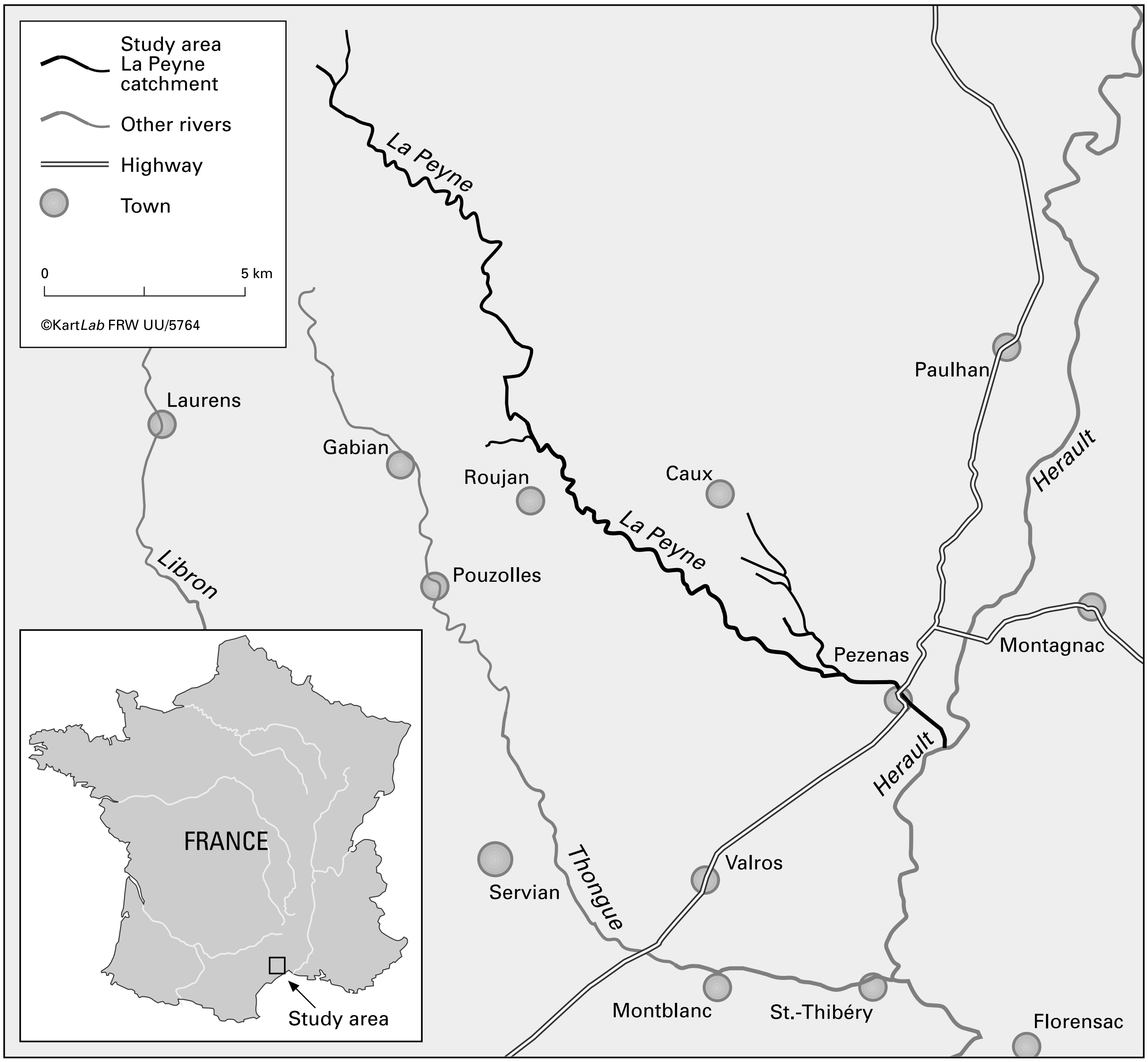
Edzer J. Pebesma,
Steven M. de Jong, Raymond Sluiter
Dept. of Physical Geography
Utrecht University

TIES 2002

Overview

- Study area and data
- Previous work
- Methods
- Results
- Conclusions/discussion

Study area


Study area La Peyne catchment
Other rivers
Highway
Town
0
5 km
©KartLab FRW UU/5764
La Peyne
La Peyne
Laurens
Gabian
Roujan
Caux
Pouzolles
Libron
Paulhan
Herault
Pezenas
Montagnac
Herault
Thongue
Valros
Servian
Montblanc
St.-Thibéry
Florensac
FRANCE
Study area

Digital Airborne Imaging Spectrometer

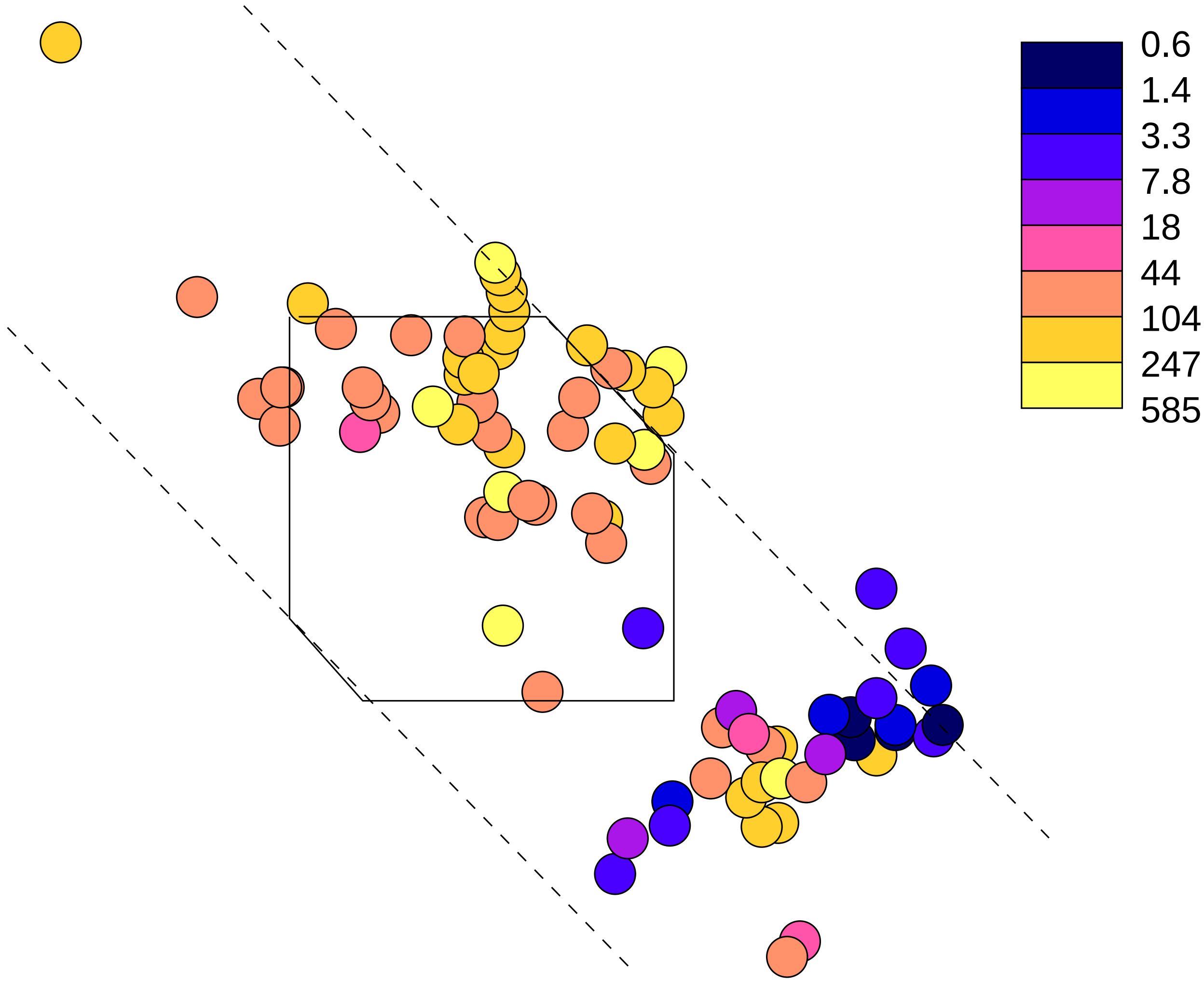
Wavelength range	# Bands	Band width	Detector
0.4-1 μm	32	12-35 nm	Si
1.5-1.8 μm	8	36-56 nm	InSb
2-2.5 μm	32*	20-40 nm	InSb
3-5 μm	(1)	2.0 μm	InSb
8.7-12.3 μm	(6)	0.6-1 μm	MCT

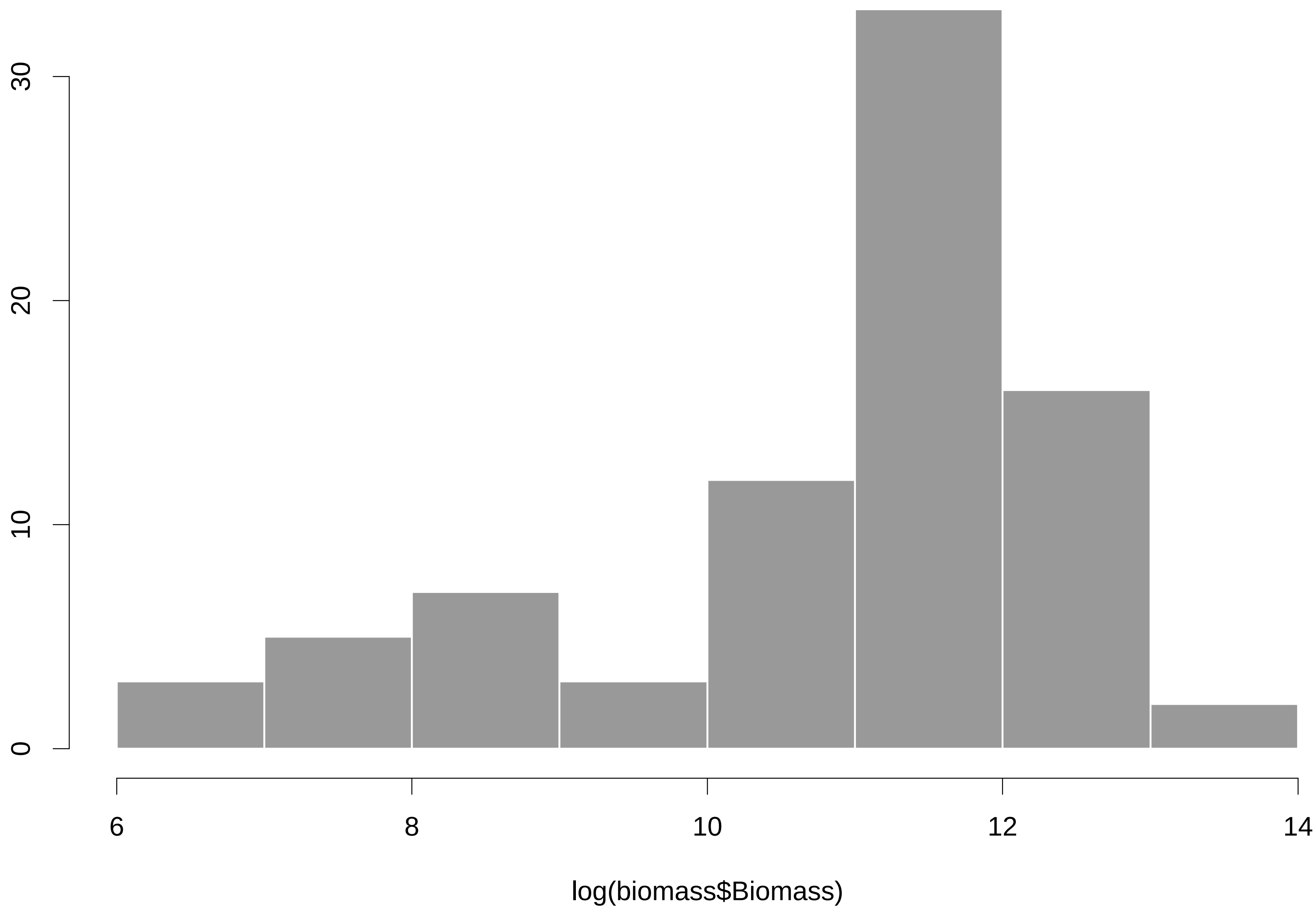
Spectral band characteristics of the DAIS7915 (Strobl et al., 1996)

- pixels: 6 $\times$ 6 m, 71 bands useful

Biomass data

- For 30 $\times$ 30 m plots, estimated from:
- number of trees, stem diameters, tree height
- aggregated over trees, shrubs, plants
- data collection random stratified: forest, maquis, garrigue
- 81 observations within flight strip: $y = f(X) + e$


0.6
1.4
3.3
7.8
18
44
104
247
585



30
20
10
0
6
8
10
12
14
log(biomass$Biomass)

Previous work: mapping biomass

- stepwise forward selected X_i: bands at 636, 773, 895, 985, 1004, 1563, 2199 and 2343 nm.

- correlations over .99: bands 773, 985, and 2343 were dropped manually

- $y(s) = \beta_0 + \sum_{i=1}^{5} \beta_i X_i(s) + e(s)$

- $e(s)$ second order stationary $\Rightarrow$ kriging/BLUP

- regression coefficients: -29, -44, 73, -74, 76

- problem: multiple collinearity
  - stepwise variable selection is unstable
  - variable selection: can/should it be avoided?

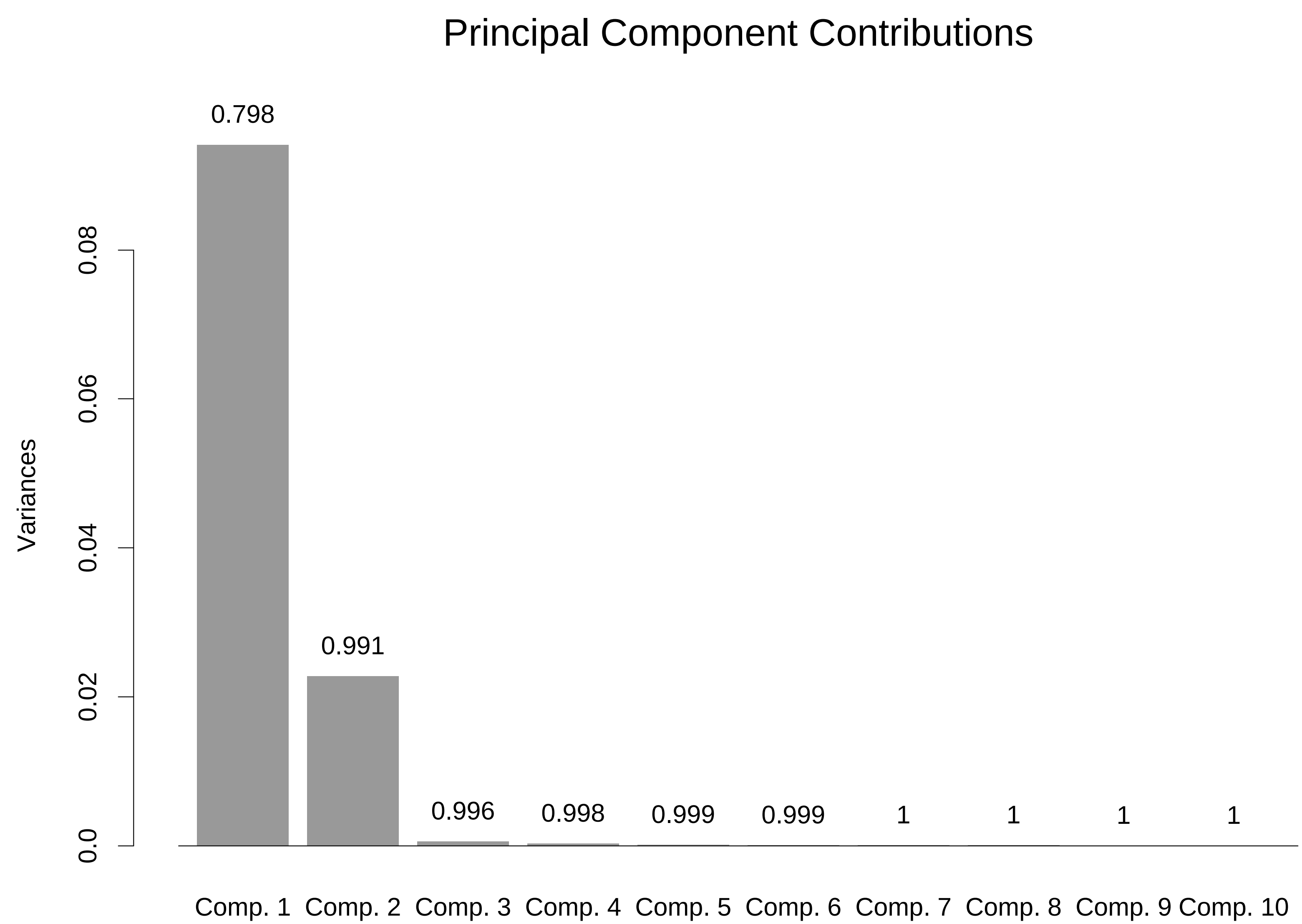
Methods

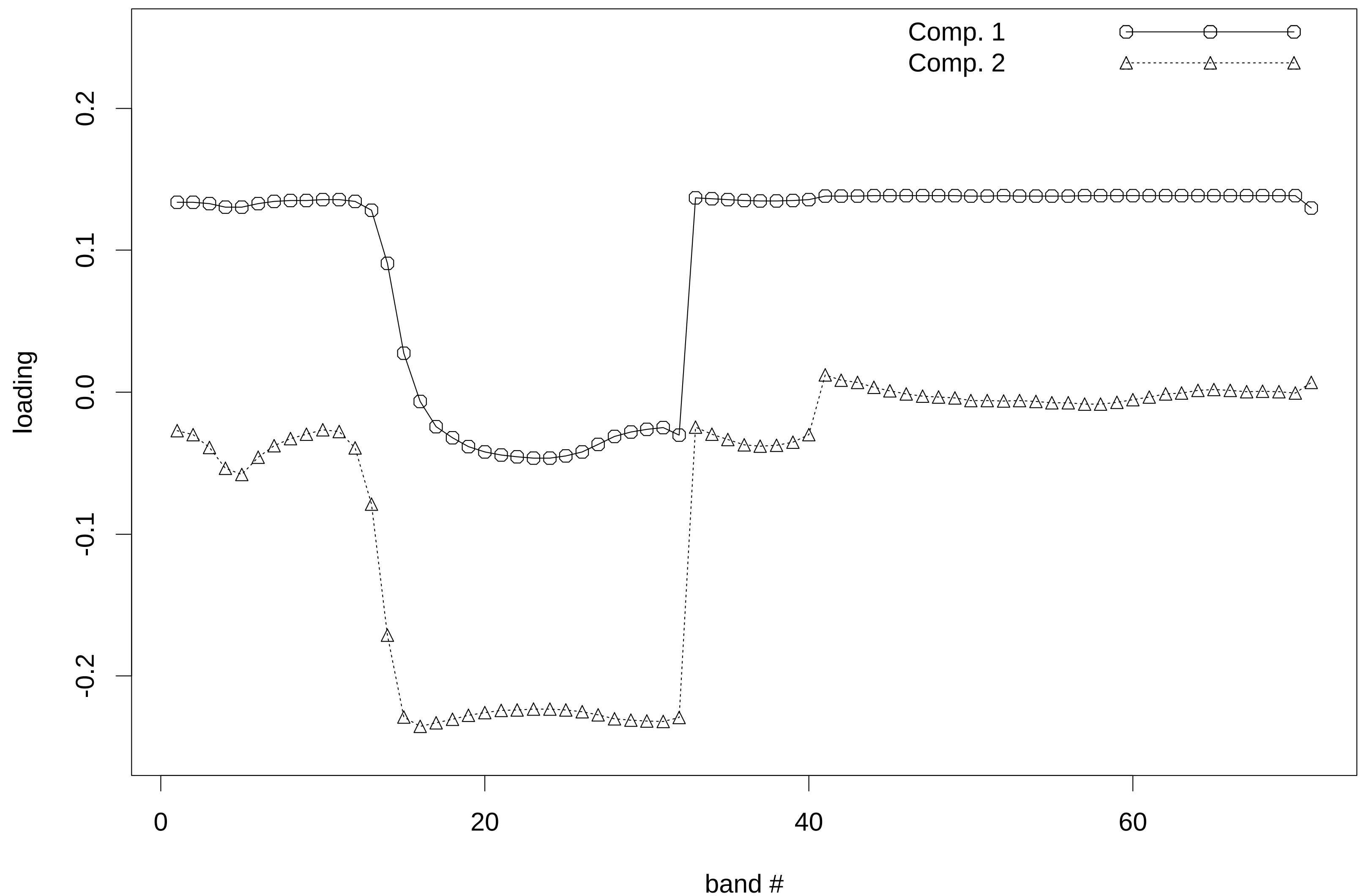
- Principal component regression (PCA)
- Partial least squares (PLS)
- Ridge regression (RR)
- Neural network (NN)

Principal component regression

Idea:

- find linear combinations $f_j = \sum_i \alpha_i X_i$ that
  - explain maximum variability of X
  - are orthogonal (uncorrelated)
- use a small number of f_j as regressors


Principal Component Contributions
Variances
0.0
0.02
0.04
0.06
0.08
0.798
0.991
0.996
0.998
0.999
0.999
1
1
1
1
Comp. 1
Comp. 2
Comp. 3
Comp. 4
Comp. 5
Comp. 6
Comp. 7
Comp. 8
Comp. 9
Comp. 10


Comp. 1
Comp. 2
loading
0.2
0.1
0.0
-0.1
-0.2
0
20
40
60
band #

Partial least squares

Idea:

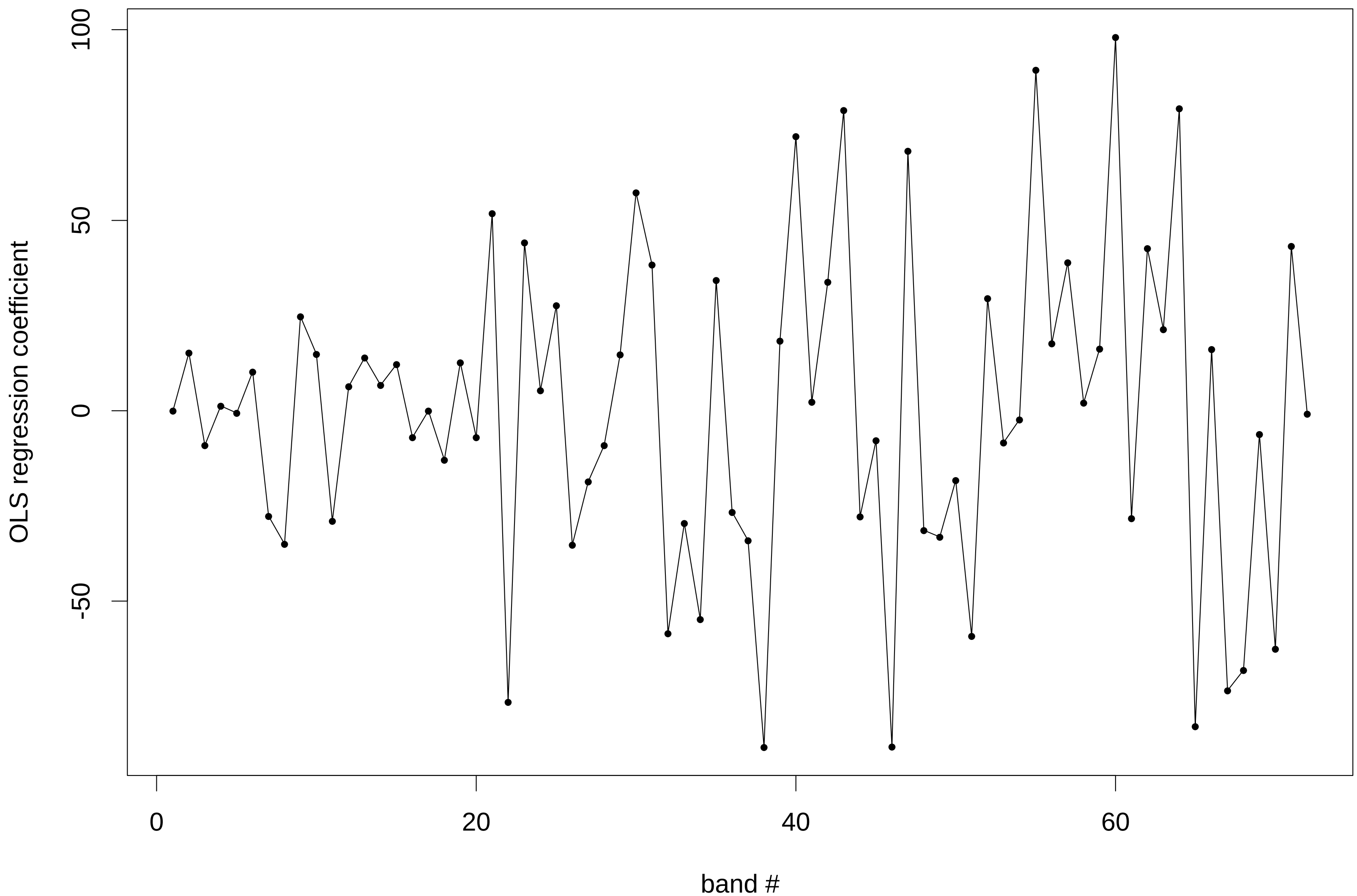
- find linear combinations $g_j = \sum_i \alpha_i X_i$ that
  - maximally correlate with y *and* explain maximum variability of X
  - are orthogonal (uncorrelated)
- use a small number of g_j as regressors

Ridge regression

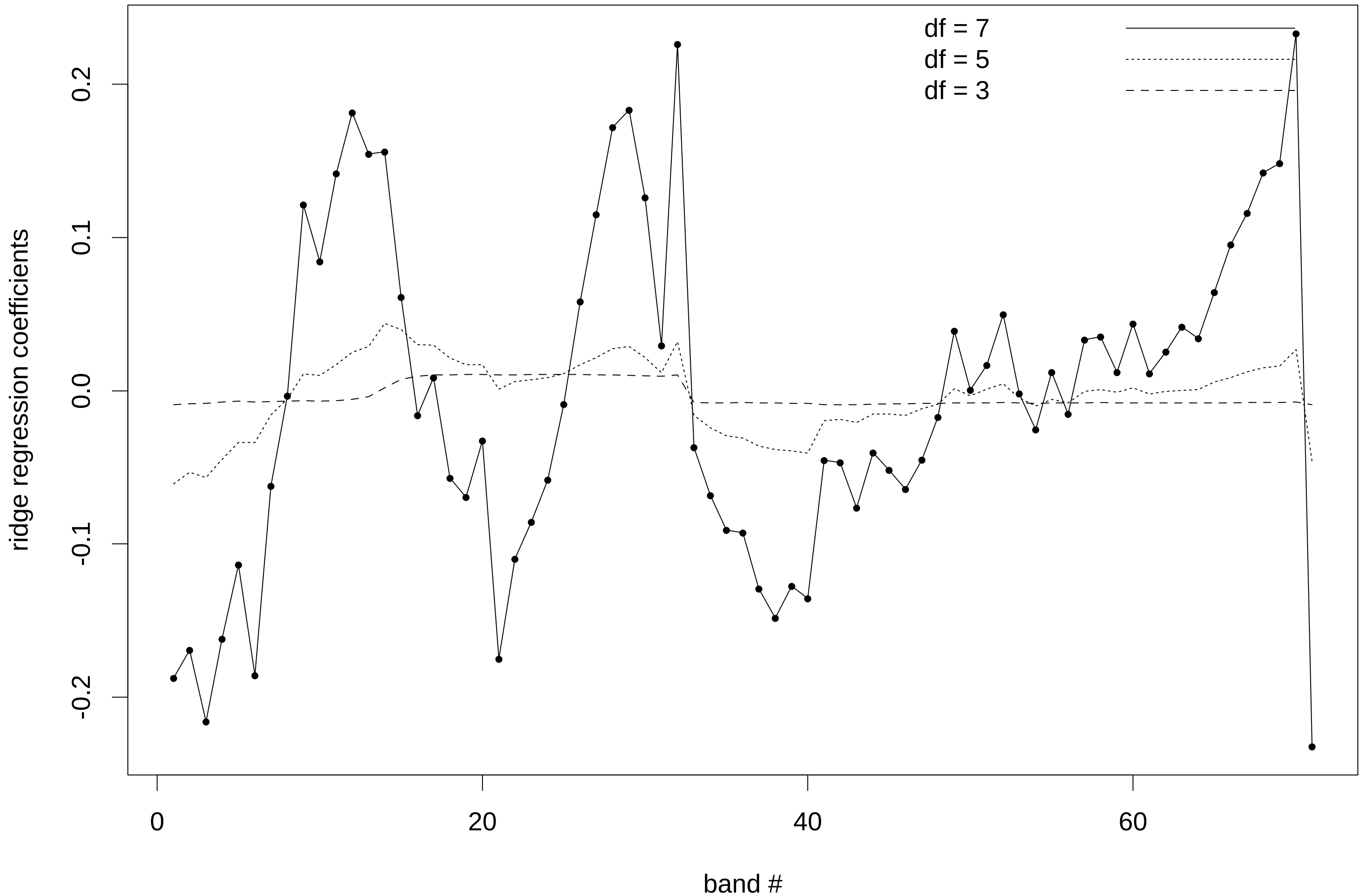
- OLS: $\hat{\beta} = (X'X)^{-1}X'y$
- ridge: $\hat{\beta} = (X'X + \lambda I)^{-1}X'y$

$$\text{RSS}(\lambda) = (y - X\beta)'(y - X\beta) + \lambda\beta'\beta$$

- large λ shrinks coefficients to zero
- *effective* degrees of freedom $\text{df}(\lambda) = f(X, \lambda)$



OLS regression coefficient
100
50
0
-50
0
20
40
60
band #



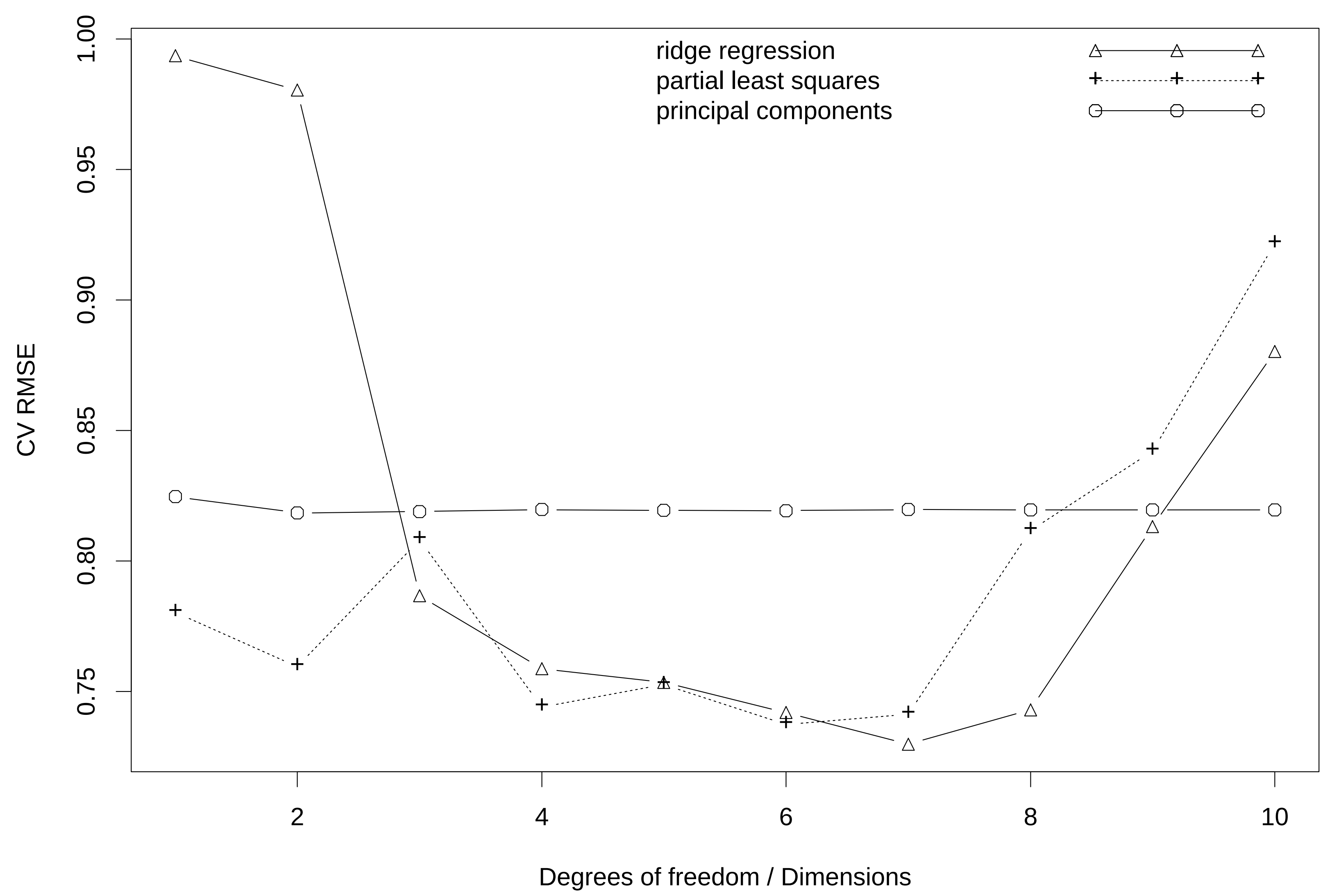
df = 7
df = 5
df = 3
ridge regression coefficients
-0.2
-0.1
0.0
0.1
0.2
0
20
40
60
band #

Neural network

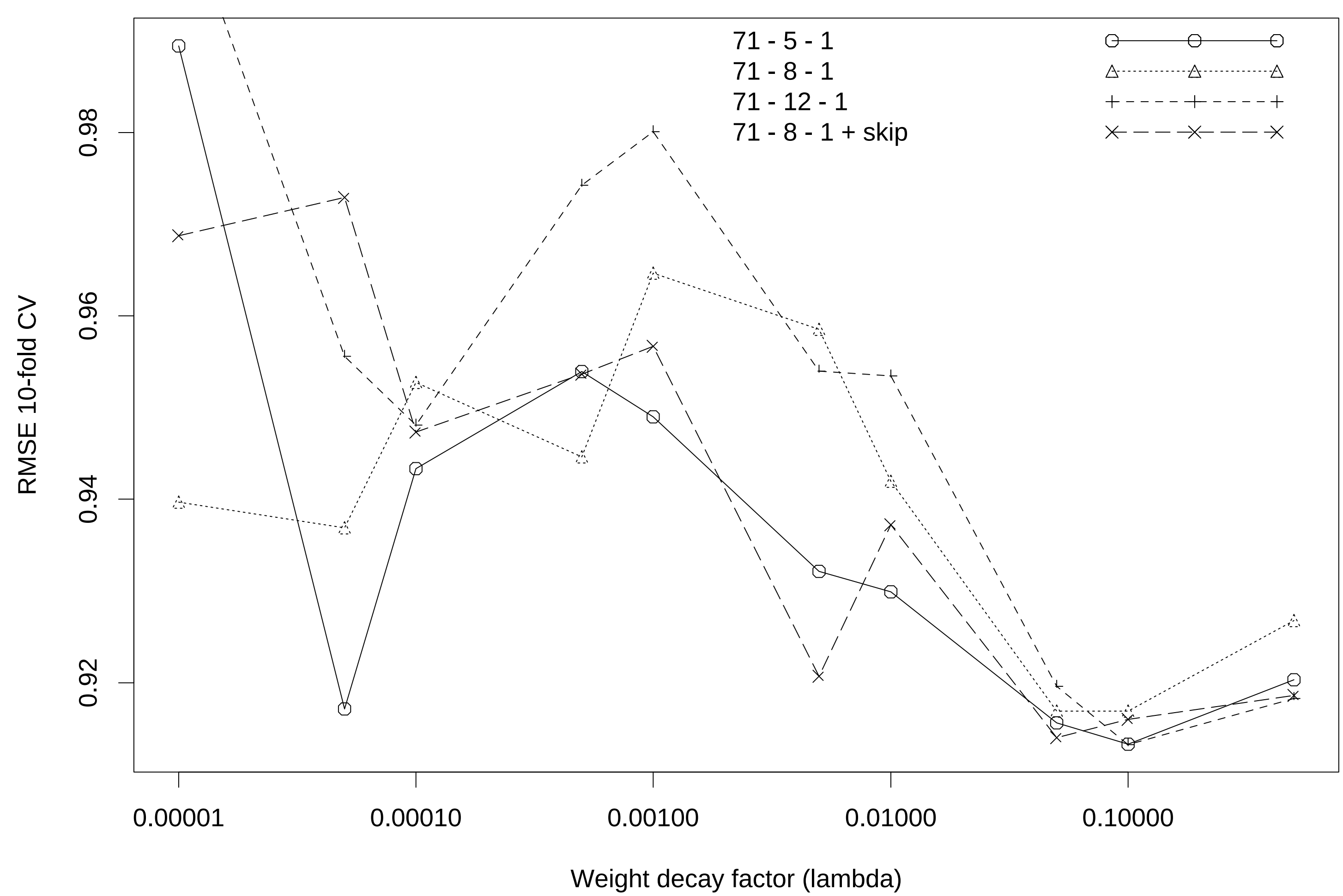
- picks up nonlinear dependencies of log(biomass) to reflectances
- single hidden layer, feed-forward network (Ripley's `nnet`)
- 71 inputs (X_j), 1 output (y), 5, 8 or 12 hidden layers
- combines several sigmoid curves of linear combinations of the X_j
- find optimal weight decay factor to avoid overfitting by CV
- (optional: skip layer connection)

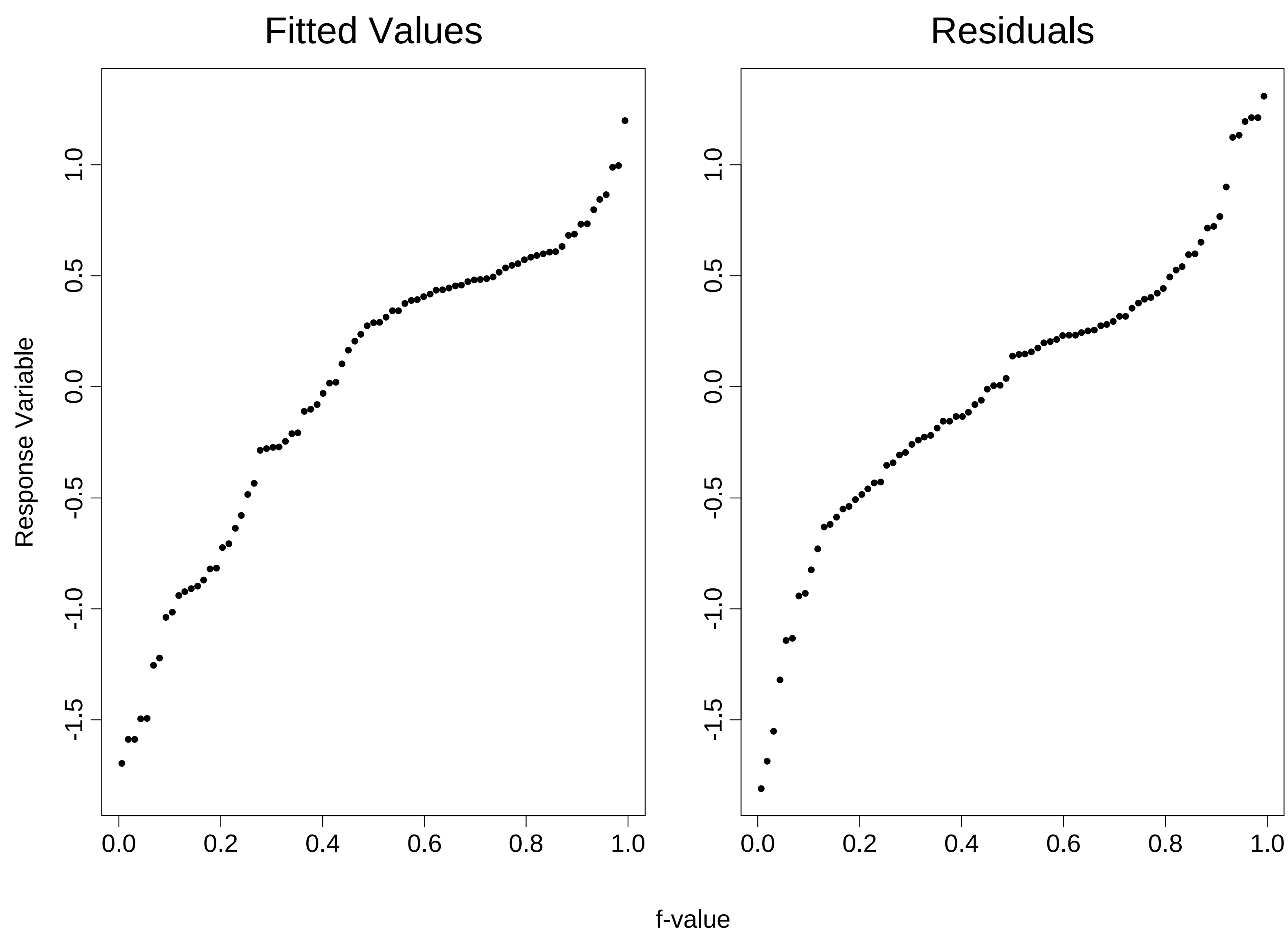
Results

- y and X_j scaled to zero mean, unit std.dev.
- 10-fold cross validation: detects overfitting
- $$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - p_i)^2}$$

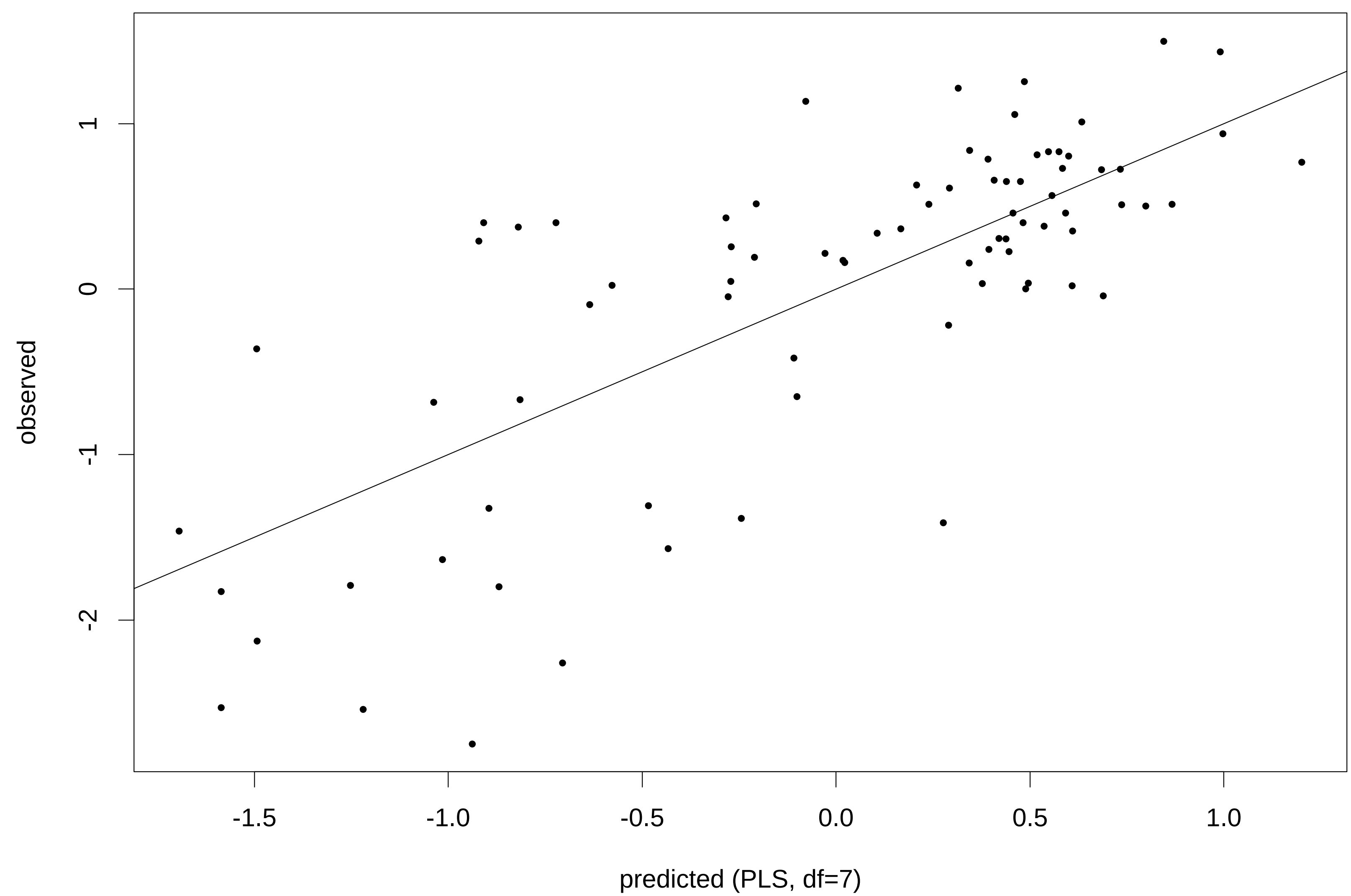


CV RMSE
1.00
0.95
0.90
0.85
0.80
0.75
ridge regression
partial least squares
principal components
2
4
6
8
10
Degrees of freedom / Dimensions


RMSE 10-fold CV
0.92
0.94
0.96
0.98
71 - 5 - 1
71 - 8 - 1
71 - 12 - 1
71 - 8 - 1 + skip
0.00001
0.00010
0.00100
0.01000
0.10000
Weight decay factor (lambda)



Fitted Values
Residuals
Response Variable
f-value
-1.5
-1.0
-0.5
0.0
0.5
1.0
0.0
0.2
0.4
0.6
0.8
1.0

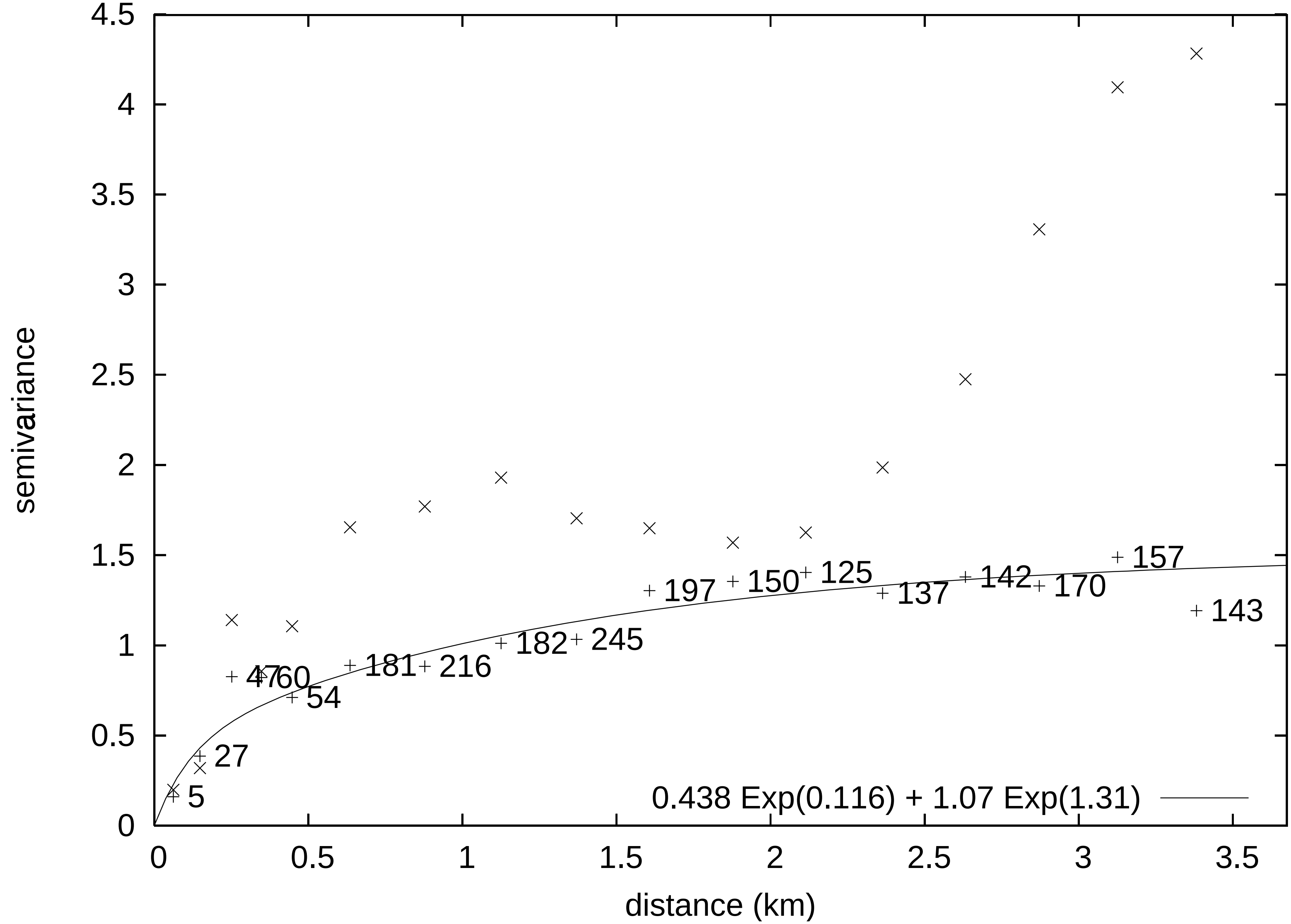

observed
1
0
-1
-2
-1.5
-1.0
-0.5
0.0
0.5
1.0
predicted (PLS, df=7)

Conclusion

- PLS and ridge regression both solve the problem
- although the reflectance data seem to lie in a plane, ridge and PLS perform best in a 7-dimensionsional subspace

Discussion

- Rather poor overall performance:
  - do other physical factors dominate image variability?
  - do images still need more correction?
  - poor quality of data (measurement errors)?
  - spatial matching field to image data?



semivariance
distance (km)
0.438 Exp(0.116) + 1.07 Exp(1.31)
5
27
47
60
54
181
216
182
245
197
150
125
137
142
170
157
143

Extensions

- collect more/other field data
- extend X:
  - $$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}$$
  - soil and geology
- extend method space (e.g., use k-NN or SVM's)