Regression Retrieval Overview

Larry McMillin

Climate Research and Applications Division

National Environmental Satellite, Data, and Information Service

Washington, D.C.

Larry.McMillin@noaa.gov

Pick one - This is

- All you ever wanted to know about regression
- All you never wanted to know about regression

Overview

- What is regression?
- How correlated predictors affect the solution
- Synthetic regression or real regression?
- Regression with constraints
- Theory and applications
- Classification
- Normalized regression
- AMSU sample
- Recommendations

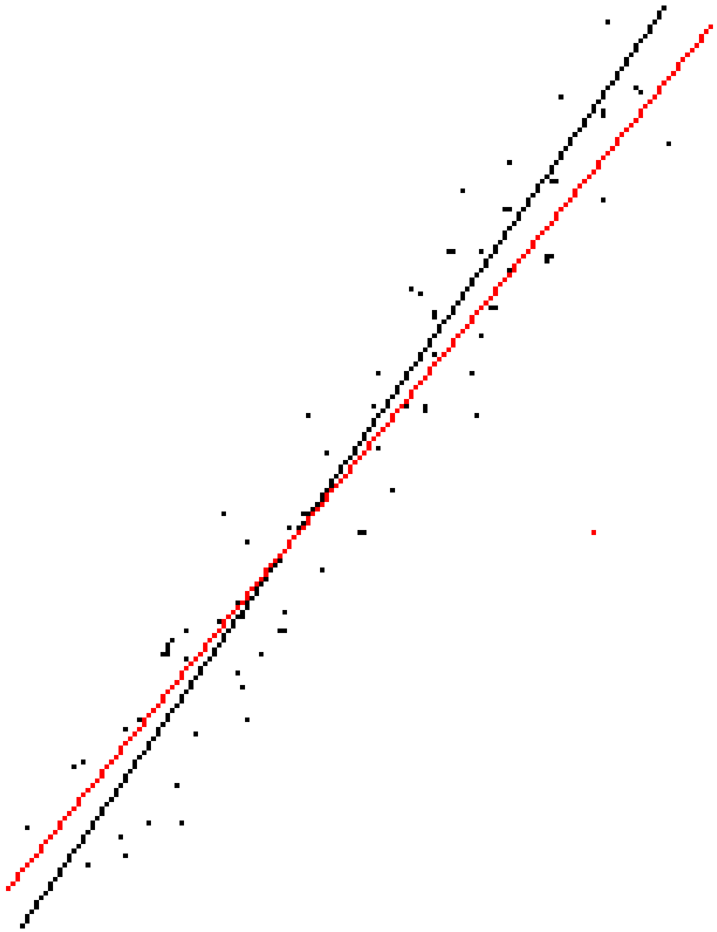
Regression - What are we trying to do?

- Obtain the estimate with the lowest RMS error.

- Or

- Obtain the true relationship

Which line do we want?



Considerations

- Single predictors
  - Easy
- Multiple uncorrelated predictors
  - Easy
- Multiple predictors with correlations
  - Assume two predictors are highly correlated and each has a noise
  - Difference is small with a larger noise than any of the two
    - And that is the problem
- Theoretical approach is hard for these cases
  - If there is an independent component, then you want the difference
  - If they are perfectly correlated, then you want to average to reduce noise

Considerations continued

- Observational approach using stepwise regression
- Stability depends on the ratio of predictors to the predictands
- Stepwise steps
  - 1. Find the predictor with the highest correlation with the predictand
  - 2. Generate the regression coefficient
  - 3. Make the predictands orthogonal to the selected predictor
  - 4. Make all remaining predictors orthogonal to the selected predictor
- Problem
  - When two predictors are highly correlated and one is removed, the calculation of correlation of the other one involves a division by essentially zero
  - The other predictor is selected next
  - The predictors end up with large coefficients of opposite sign

Considerations continued

- Consider two predictands with the same predictors
  - Stable case (temperature for example)
    - The correlation with the predictand is high
  - Unstable case (water vapor for example)
    - The correlation with the predictand is low
- Essentially when a selected predictor is removed from the predictand and the predictors
  - If the residual variance of the predictand decays at least as fast as the residual variance of the predictors, the solution remains stable

Considerations continued

- Desire
  - Damp the small eigenvectors but don't damp the regression coefficients
  - $C = YX^T(XX^T)^{-1}$
  - But when removing the variable want to use $(XX^T + \gamma I)^{-1}$
- Solutions
  - Decrease the contributions from the smaller eigenvectors
    - This damps the slope of the regression coefficients and forces the solution towards the mean value
  - Alternatives
    - Increase the constraint with each step of the stepwise regression
    - But no theory exists

Regression Retrievals

- $T = T_{guess} + C(R - R_{guess})$
- R is measured
- R_{guess}
  - Measured
    - Apples subtracted from apples (measured – measured)
  - Calculated
    - Apples subtracted from oranges (measured – calculated)
    - This leads to a need for bias adjustment (tuning)

Synthetic or Real

- Synthetic regression – use calculated radiances to generate regression coefficients
  - Errors
    - Can be controlled
    - Need to be realistic
  - Sample needs to be representative
  - Systematic errors result if measurements and calculations are not perfectly matched
- Real regression - uses matches with radiosondes
  - Compares measured to measured - no bias adjustment needed
  - Sample size issues - sample size can be hard to achieve
  - Sample consistency across scan spots - different samples for each angle
  - Additional errors - match errors, “truth” errors

Regression with constraints

- Why add constraints?
  - Problem is often singular or nearly so
- Possible regressions
  - Normal regression
  - Ridge regression
  - Shrinkage
  - Rotated regression
  - Orthogonal regression
  - Eigenvector regression
  - Stepwise regression
  - Stagewise regression
  - Search all combinations for a subset

Definitions

- Y = predictands
- X = predictors
- C = coefficients
- C_{normal} = normal coefficients
- C_0 = initial coefficients
- C_{ridge} = ridge coefficients
- $C_{shrinkage}$ = shrinkage coefficients
- $C_{rotated}$ = rotated coefficients
- $C_{orthogonal}$ = orthogonal coefficients
- $C_{eigenvector}$ = eigenvector coefficients

Definitions continued

γ = a constant

ε = errors in y

δ = errors in x

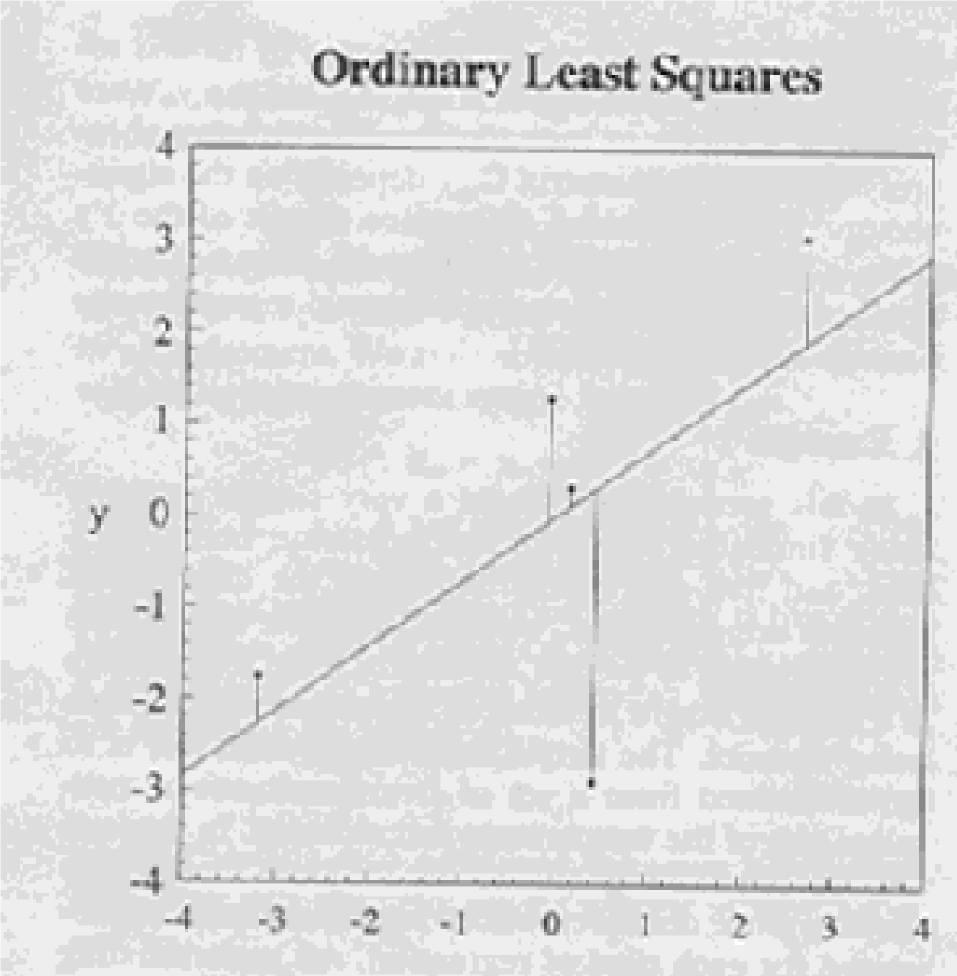
- X_t = true value when known

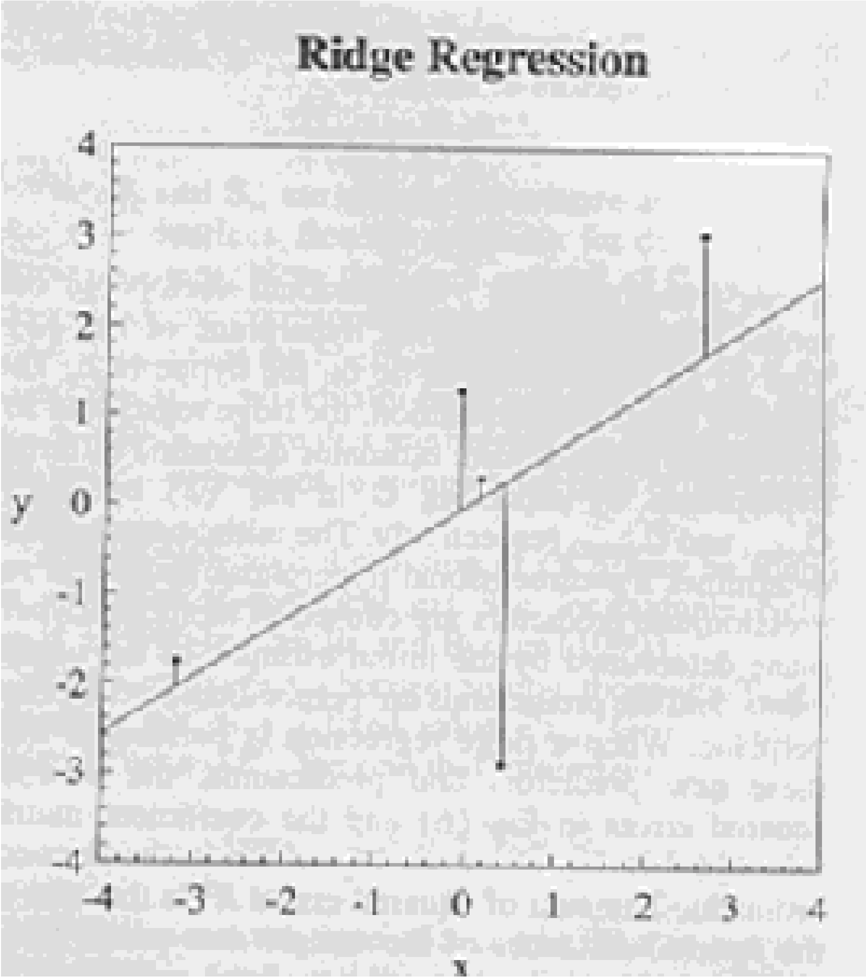
Equations

- $Y = C X$
- $C_{normal} = YX^T (XX^T)^{-1}$
- $C_{ridge} = YX^T (XX^T + \gamma I)^{-1}$
- $C_{shrinkage} = (YX^T + \gamma C_0)(XX^T + \gamma I)^{-1}$
- $C_{rotated} = (YY^T C_0^T + YX^T + \gamma C_0)(C_0^T YX^T + XX^T + \gamma I)^{-1}$
- $C_{orthogonal}$ = multiple rotated regression until solution converges

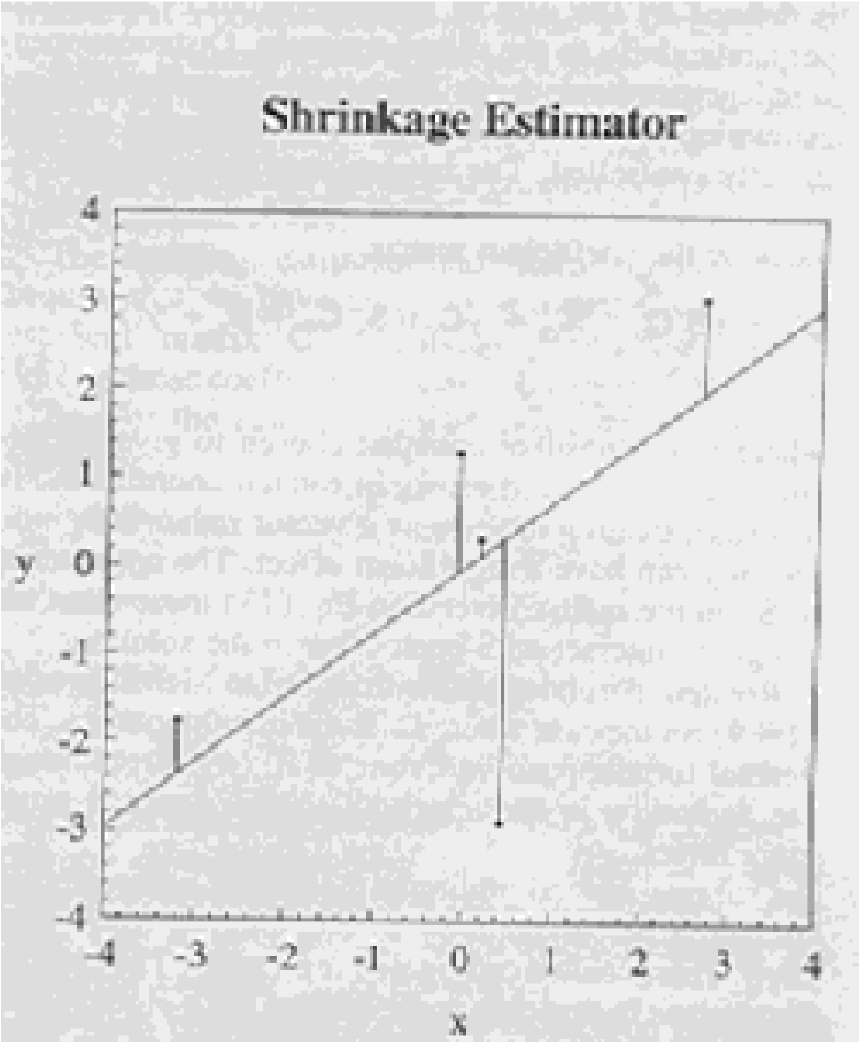
- Note many of these differ only in the directions used to calculate the components
  - The first 3 minimize differences along the y direction
  - Rotated minimizes differences perpendicular to the previous solution
  - Orthogonal minimizes differences perpendicular to the final solution

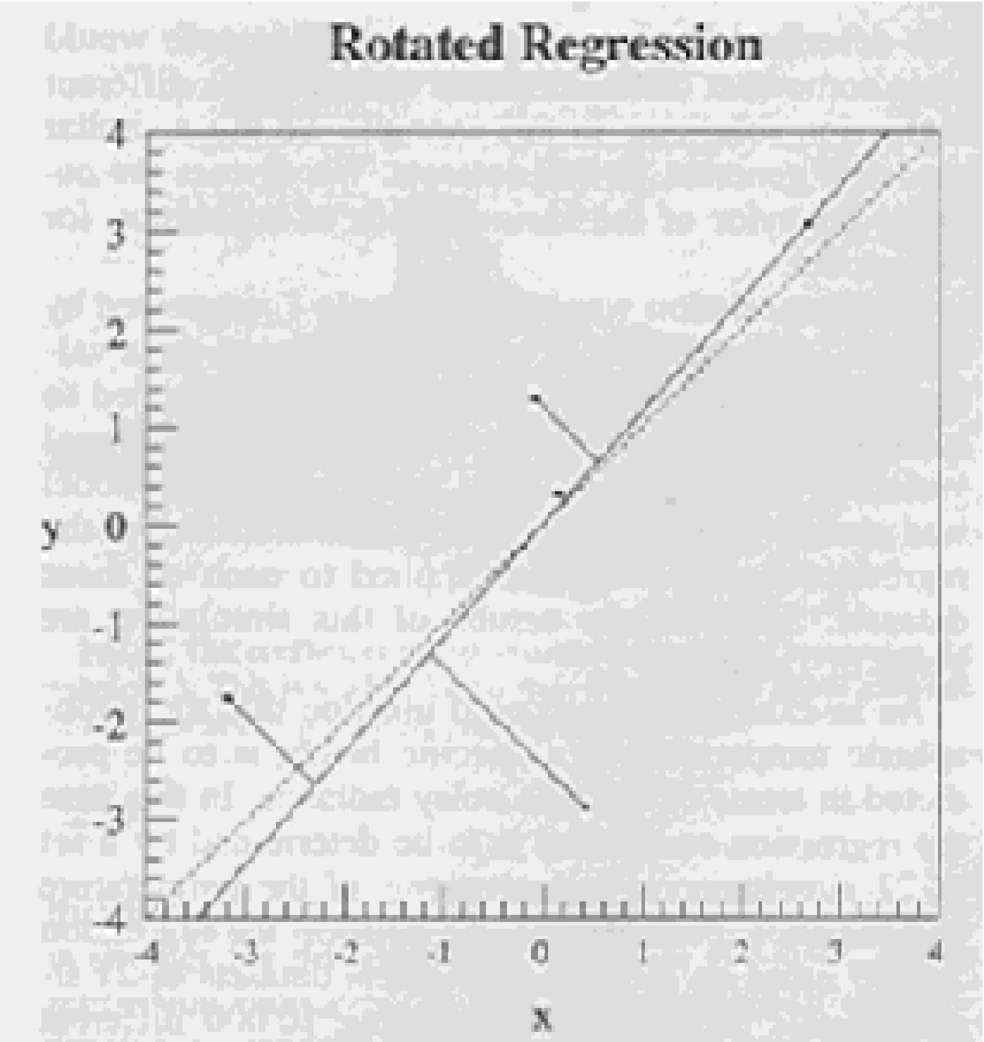
Regression examples



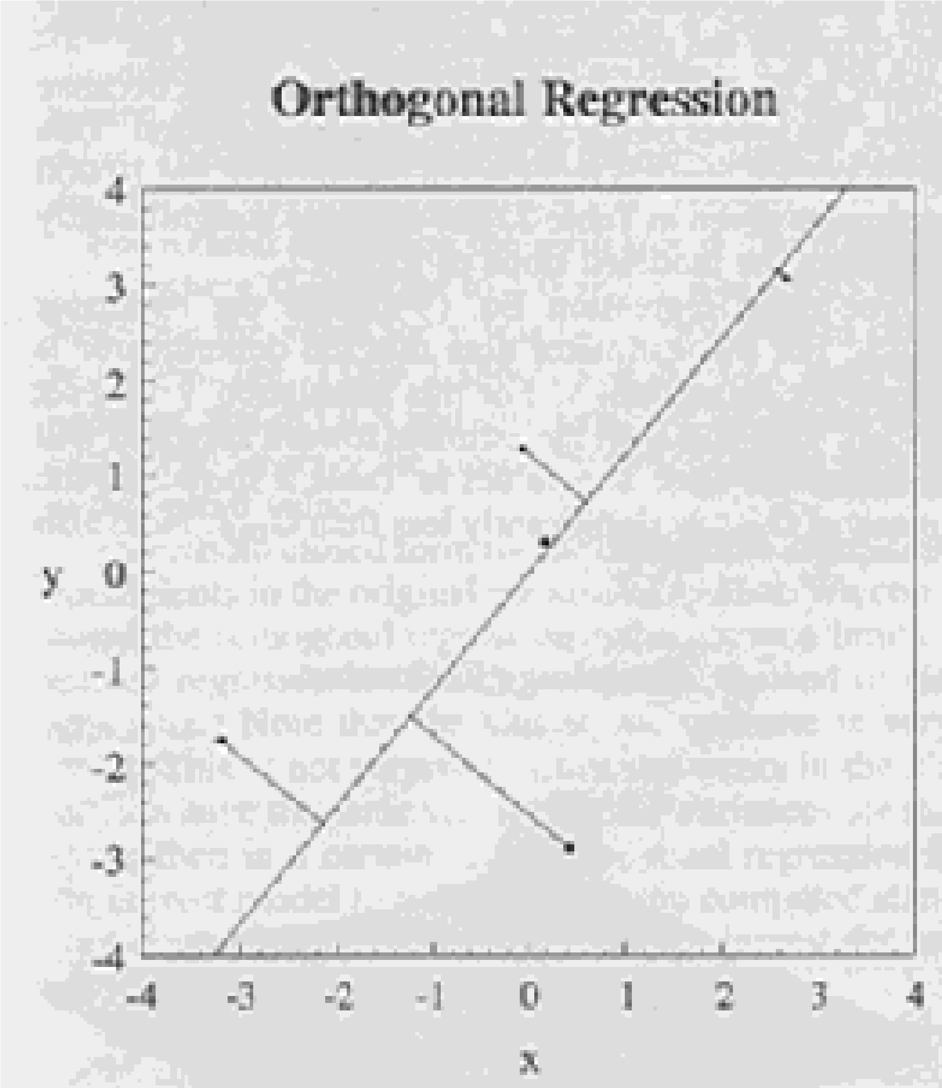


Regression examples





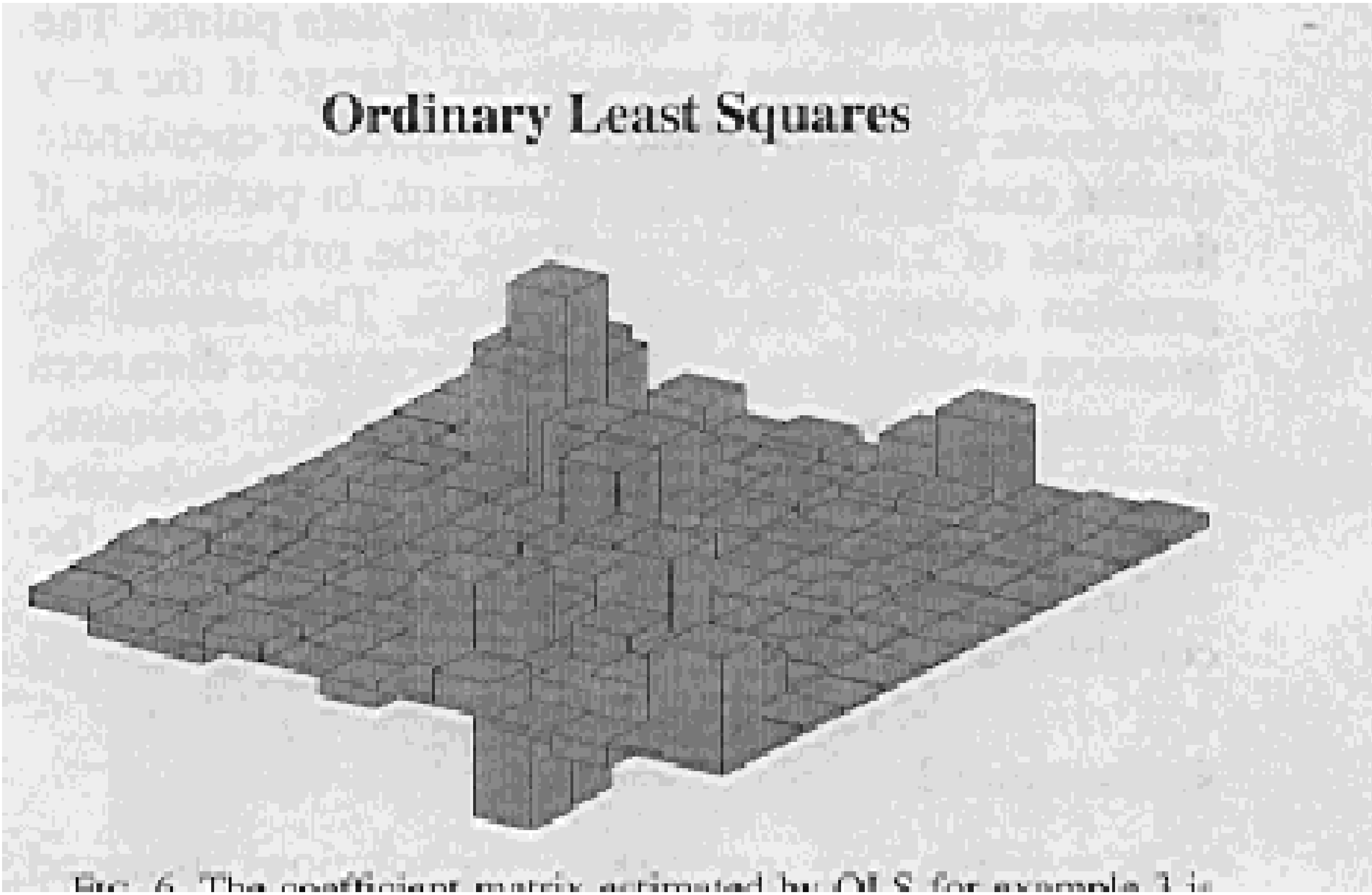
Regression Examples



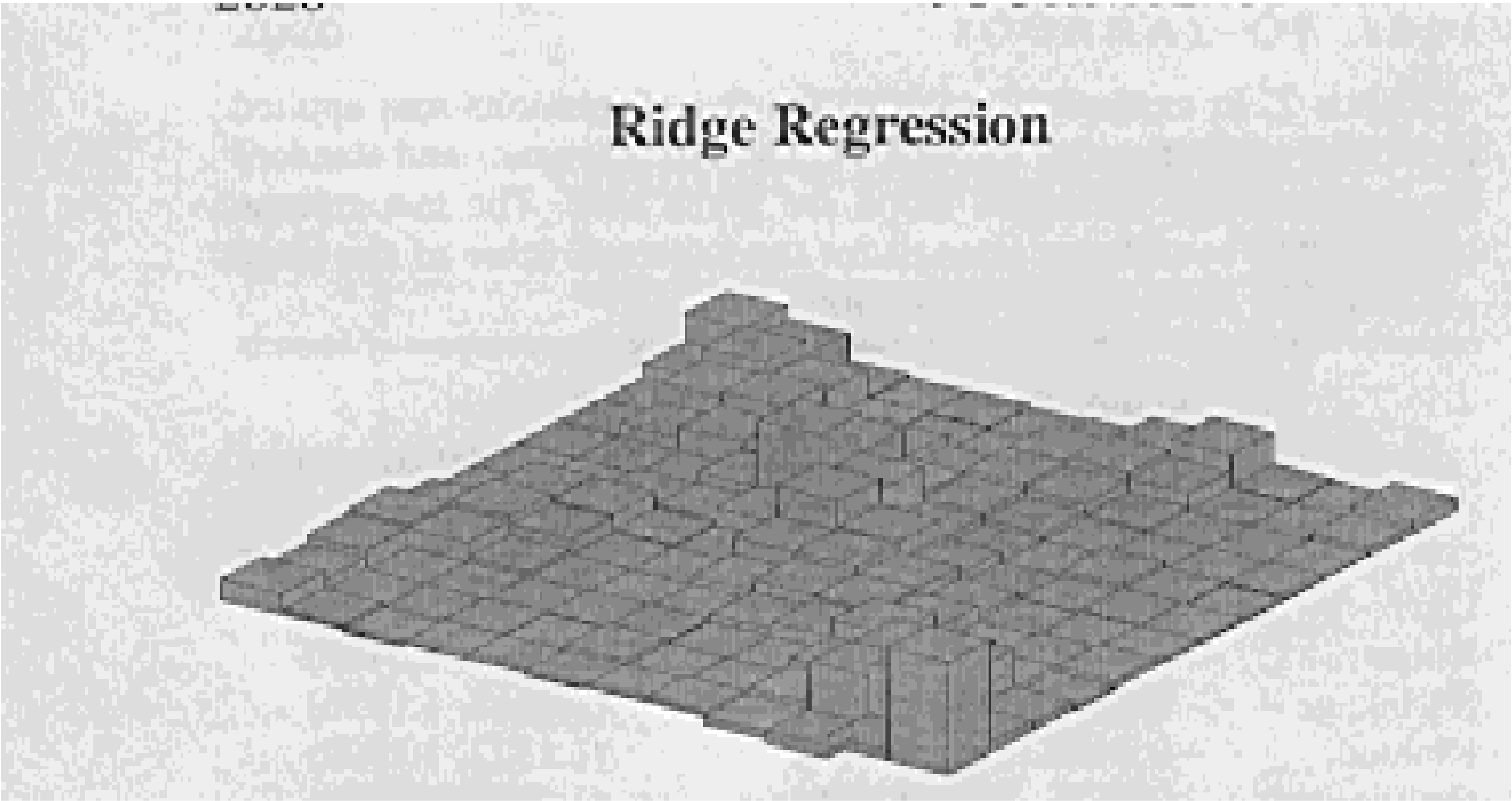
Constraint summary

- True relationship Y = 1.2 X   Guess Y = 1.0 X   ss = 17.79
- Ordinary Least Squares
  - Y = 0.71 X   ss = 13.64
- Ridge - gamma = 2
  - Y = 0.64 X   ss = 13.73
- Shrinkage - gamma = 2
  - Y = 0.74 X   ss = 13.65
- Rotated - gamma = 4 (equivalent to gamma = 2)
  - Y = 1.15 X   ss = 16.94   ss = 7.35 in rotated space
- Orthogonal - gamma = 4
  - Y = 1.22 X   ss = 18.14   ss = 7.29 in orthogonal space

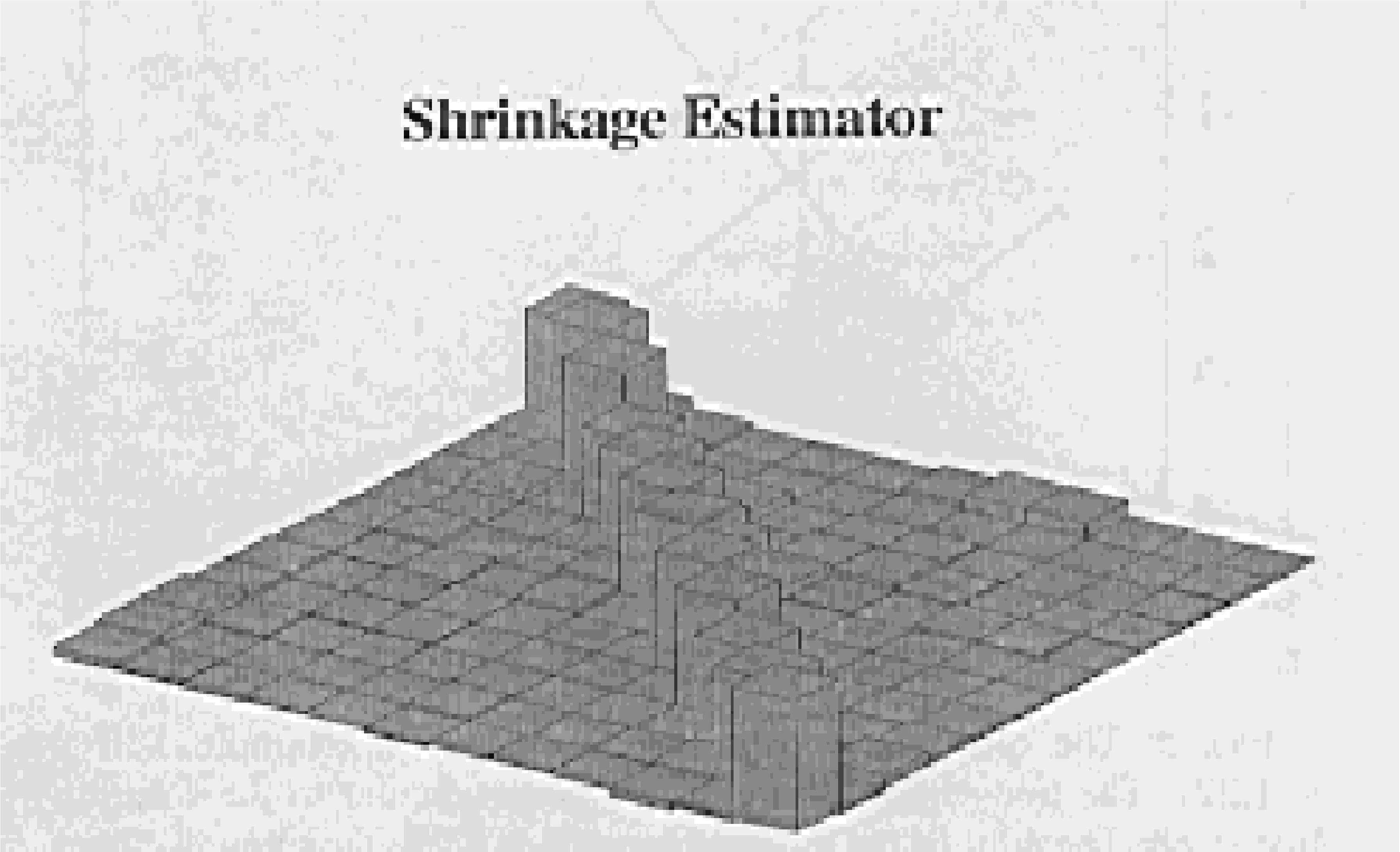
Regression Examples



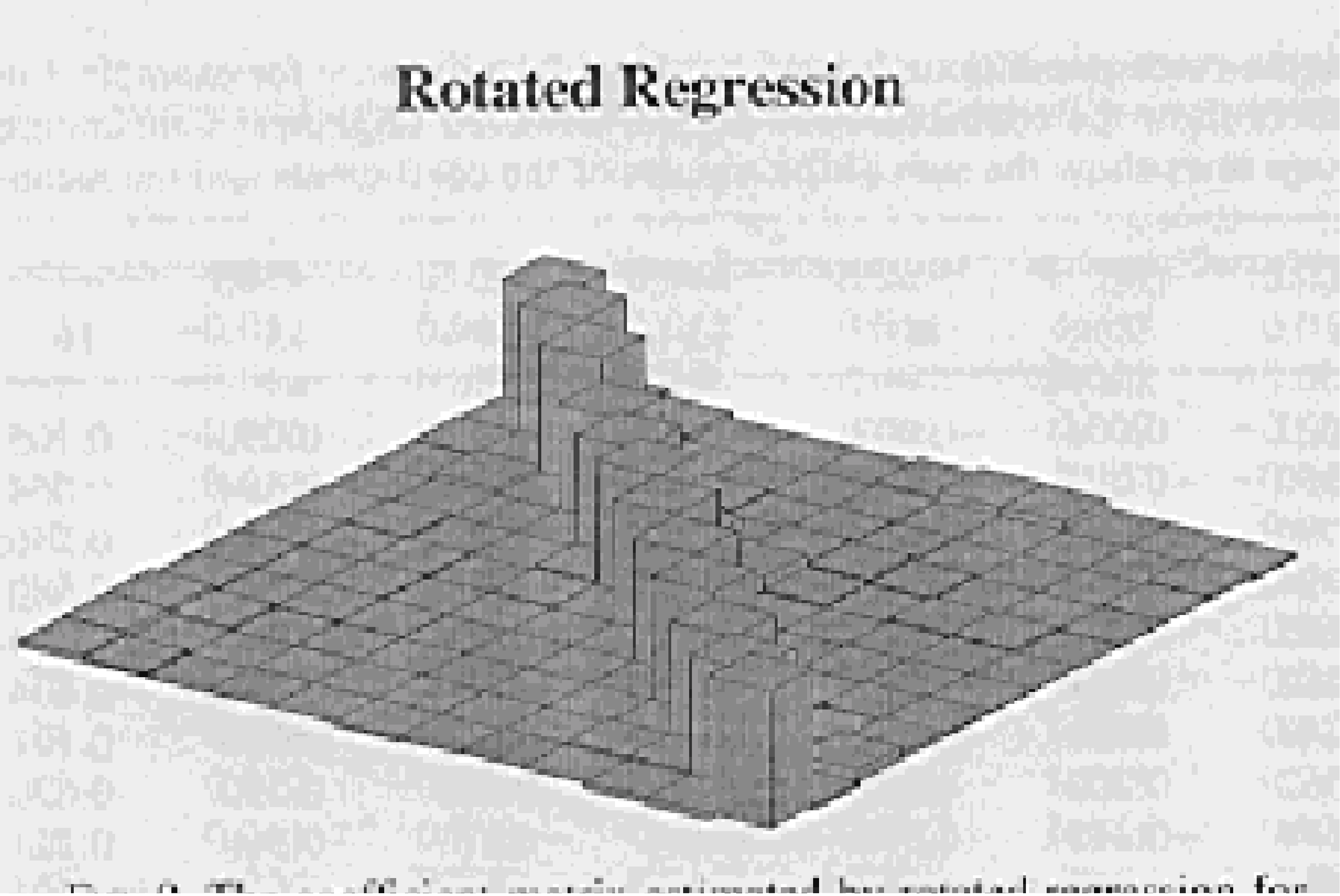
Regression Examples



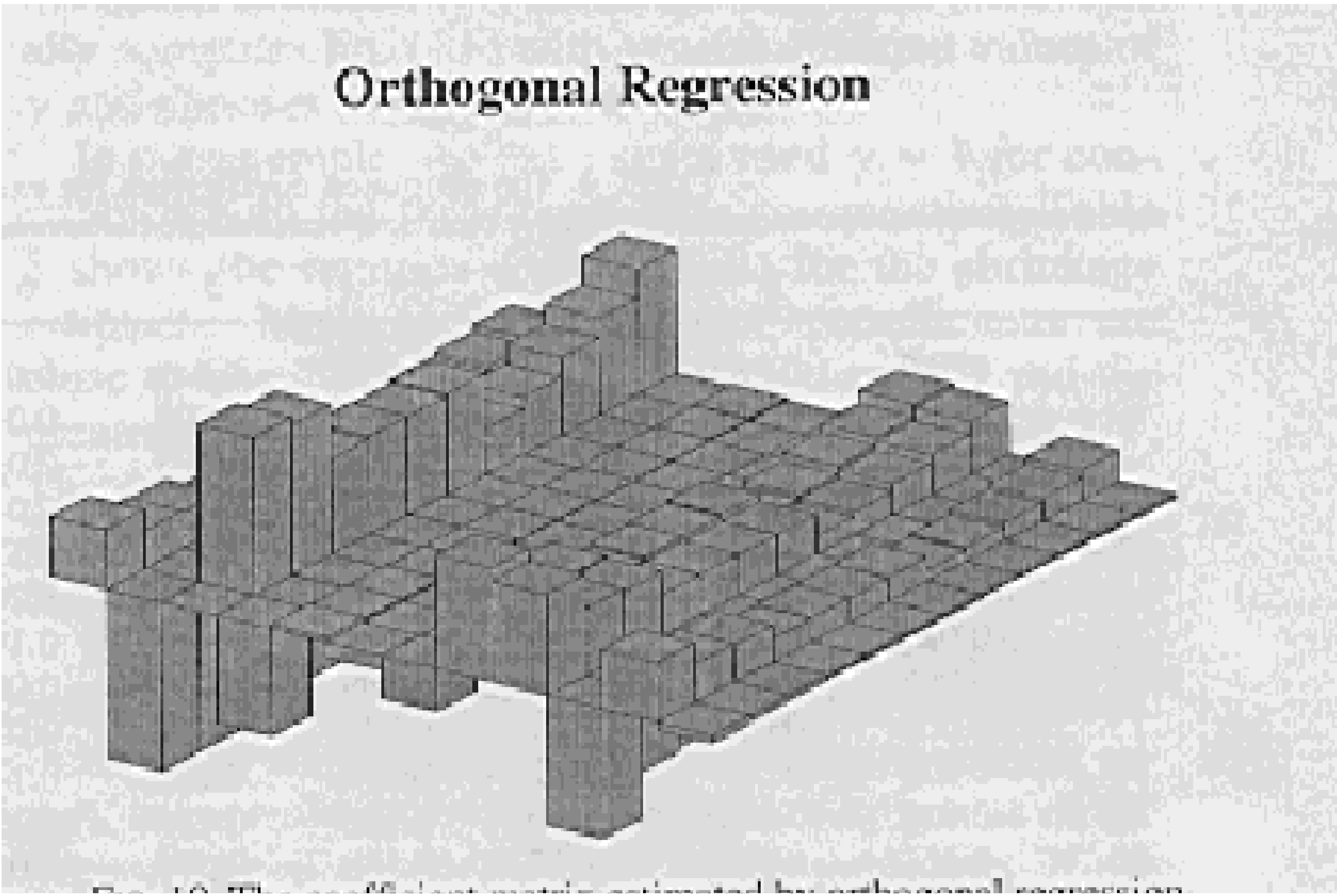
Regression Examples



Regression Examples



Regression Examples



Popular myth - Or the devil is in the details

- Two regression can be replaced by a single one
- Y = C X
- X = D Z
- Y = E Z
- Then Y = C D Z and E = C D
- True for normal regression but false for any constrained regression
- In particular, if X is a predicted value of Y from Z using an initial set of coefficients and C is obtained using a constrained regression, then the constrain is in a direction determined by D. If this is iterated, it becomes rotated regression.

Regression with Classification

- Pro
  - Starts with a good guess
- Con
  - Decreases the signal to noise ratio
  - Can get a series of means values
  - With noise, the adjacent groups have jumps at the boundaries

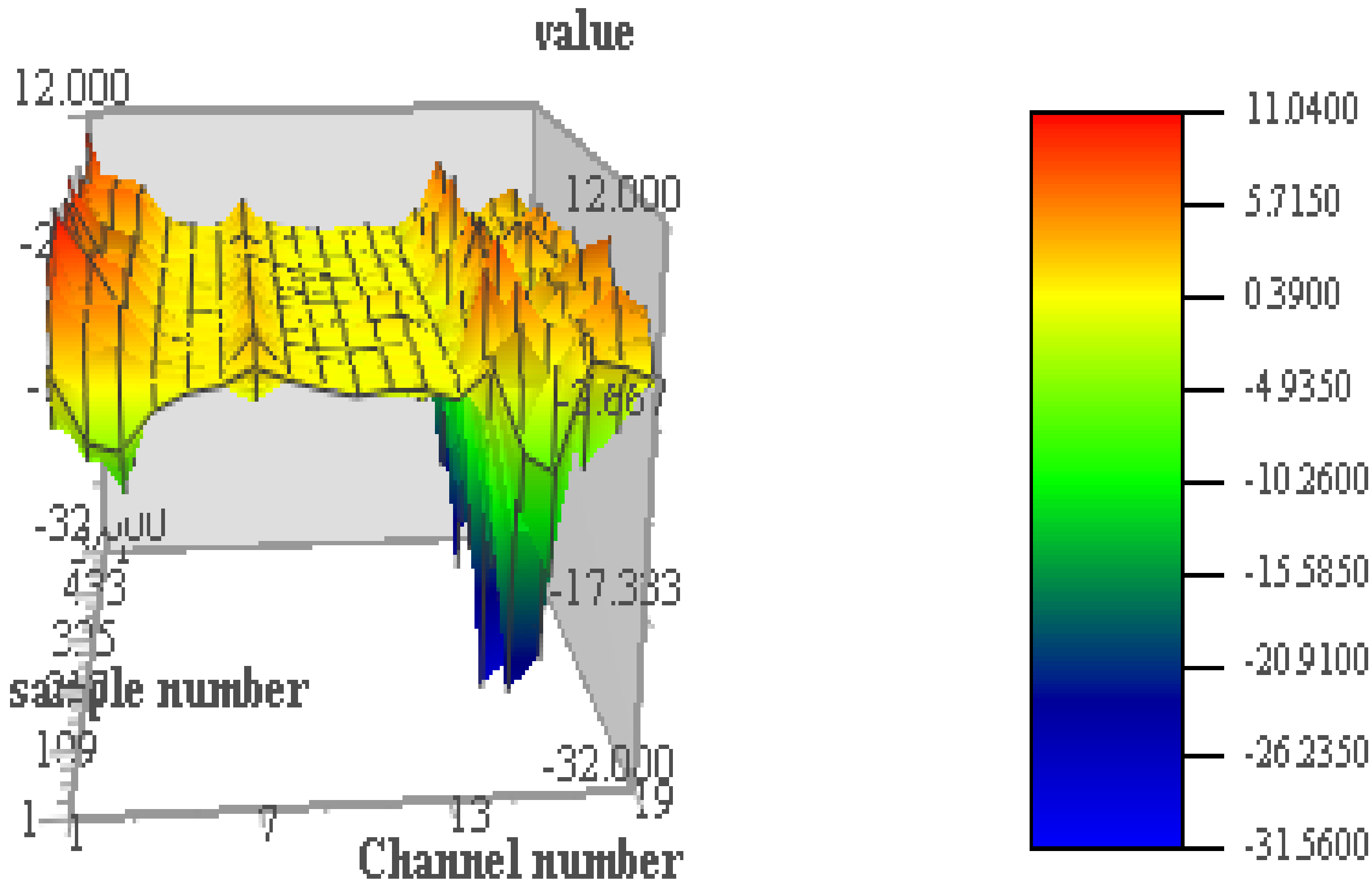
normalized regression

- Subtract the mean from both X an Y
- Divide by the standard deviation
- Theoretically makes no difference
- But numerical precision is not theory
- Good for variables with large dynamic range
- Recent experience with eigenvectors suggests dividing radiances by the noise

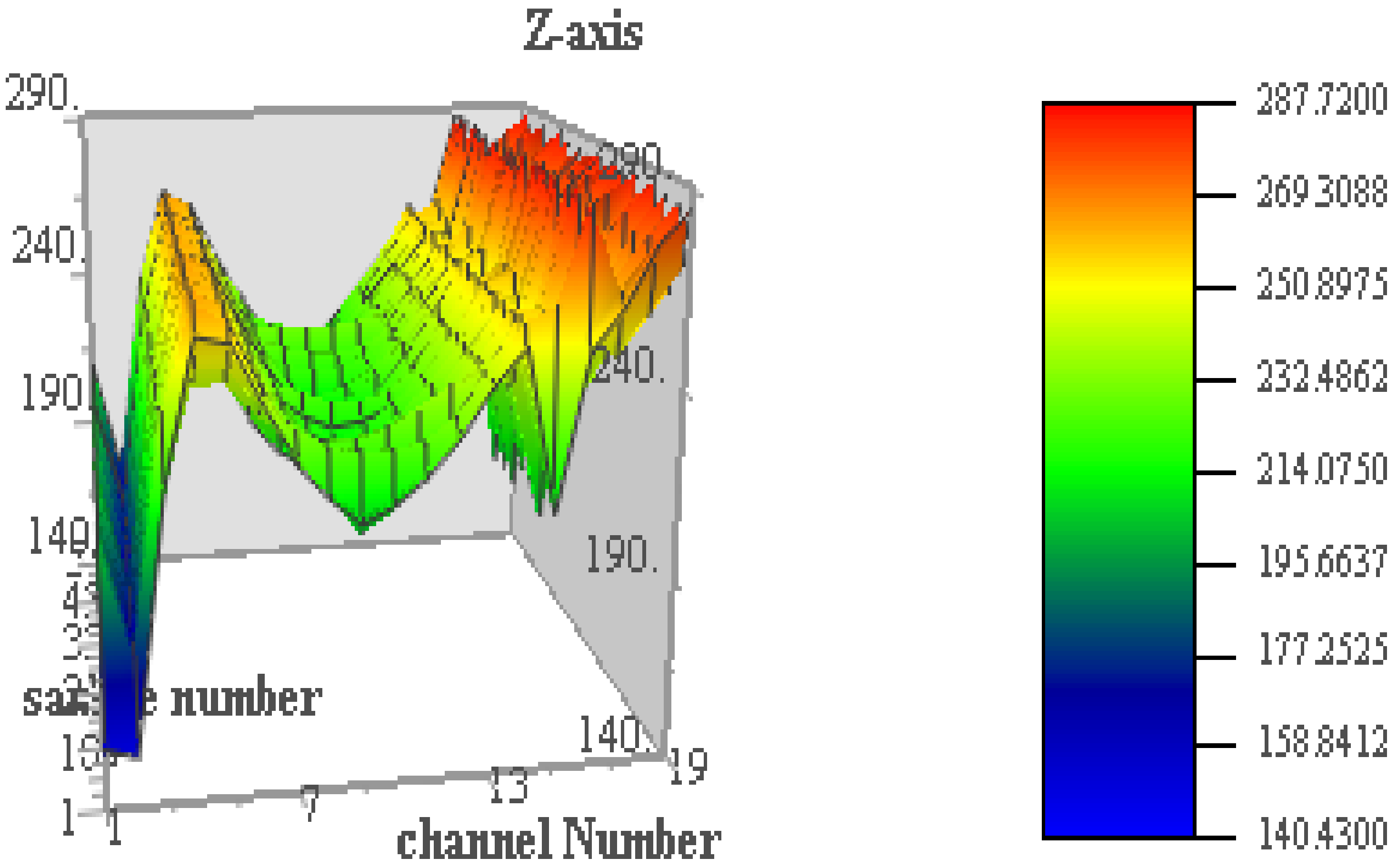
Example - Tuning AMSU on AQUA

- Predictors are the channel values
- Predictands are the observed minus calculated differences

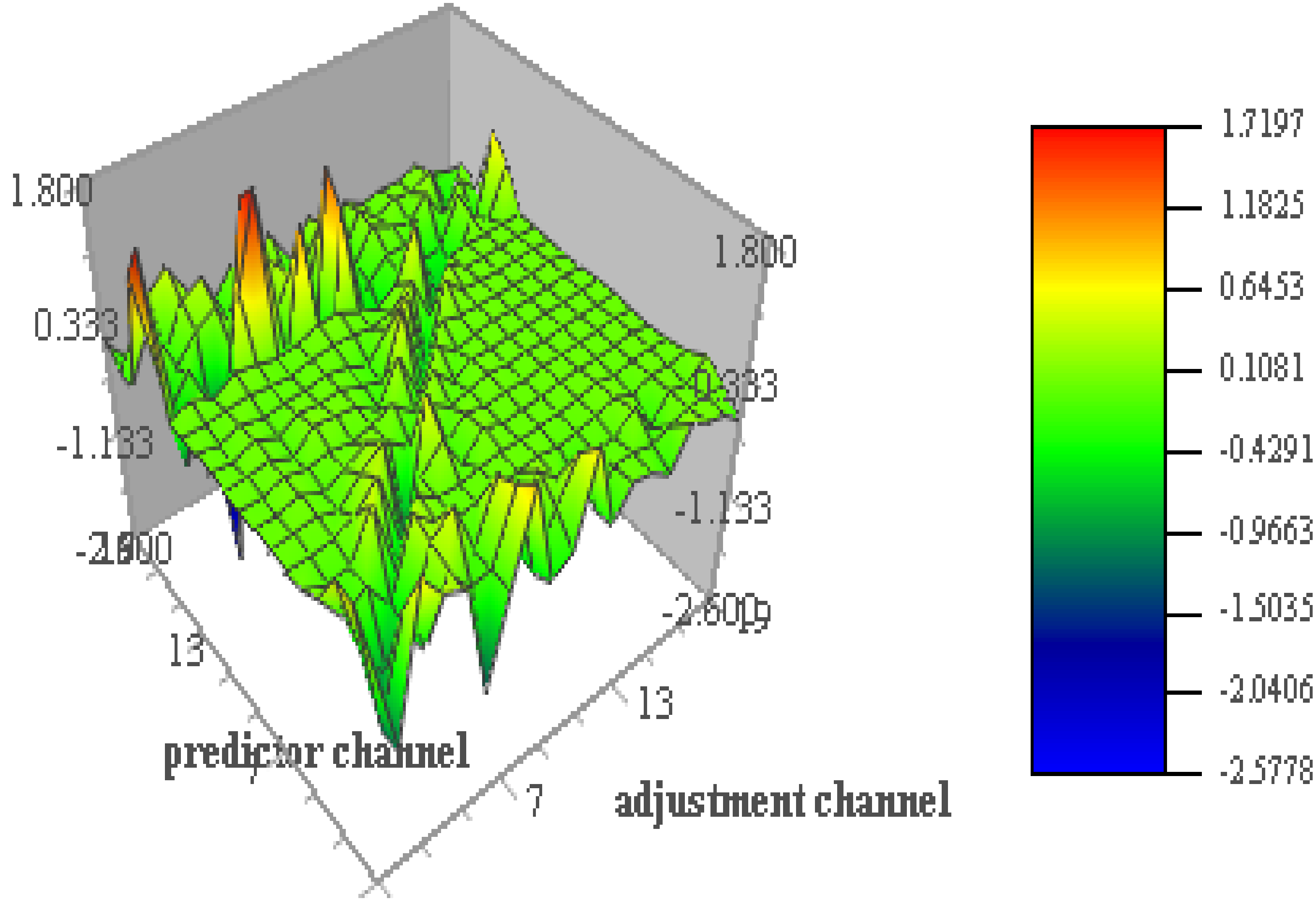
Measured minus calculated



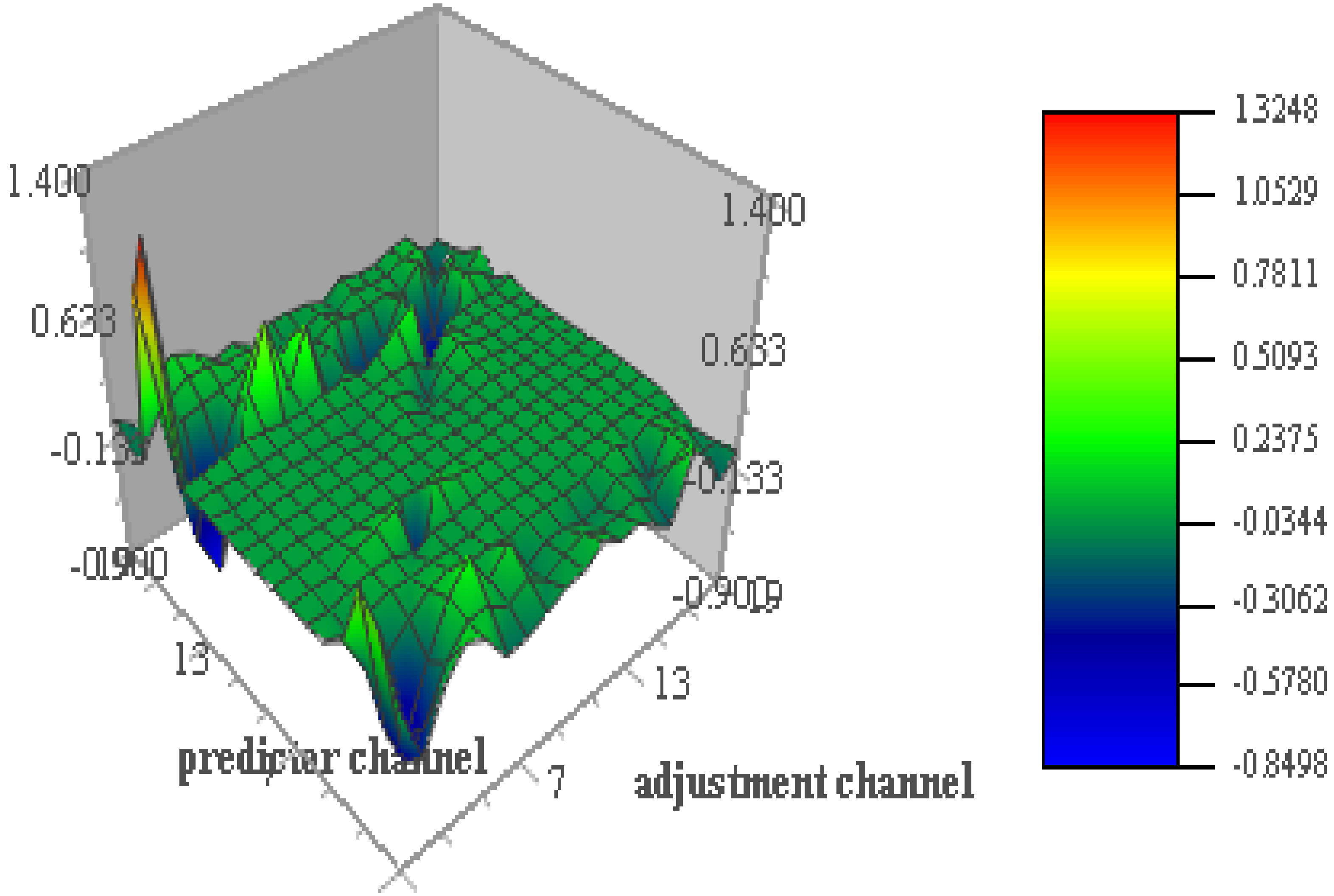
The predictors



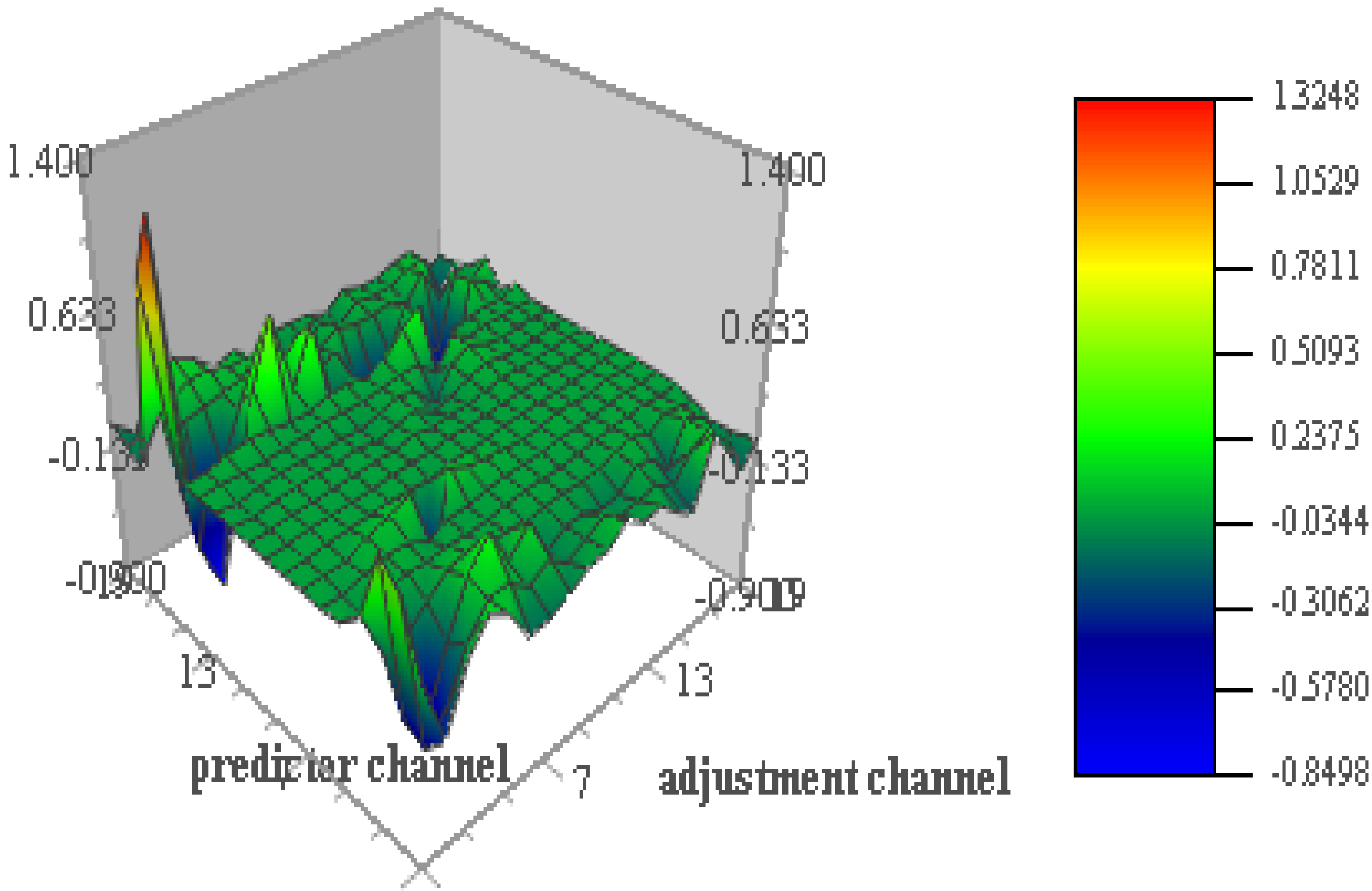
Ordinary Least Squares



Ridge Regression

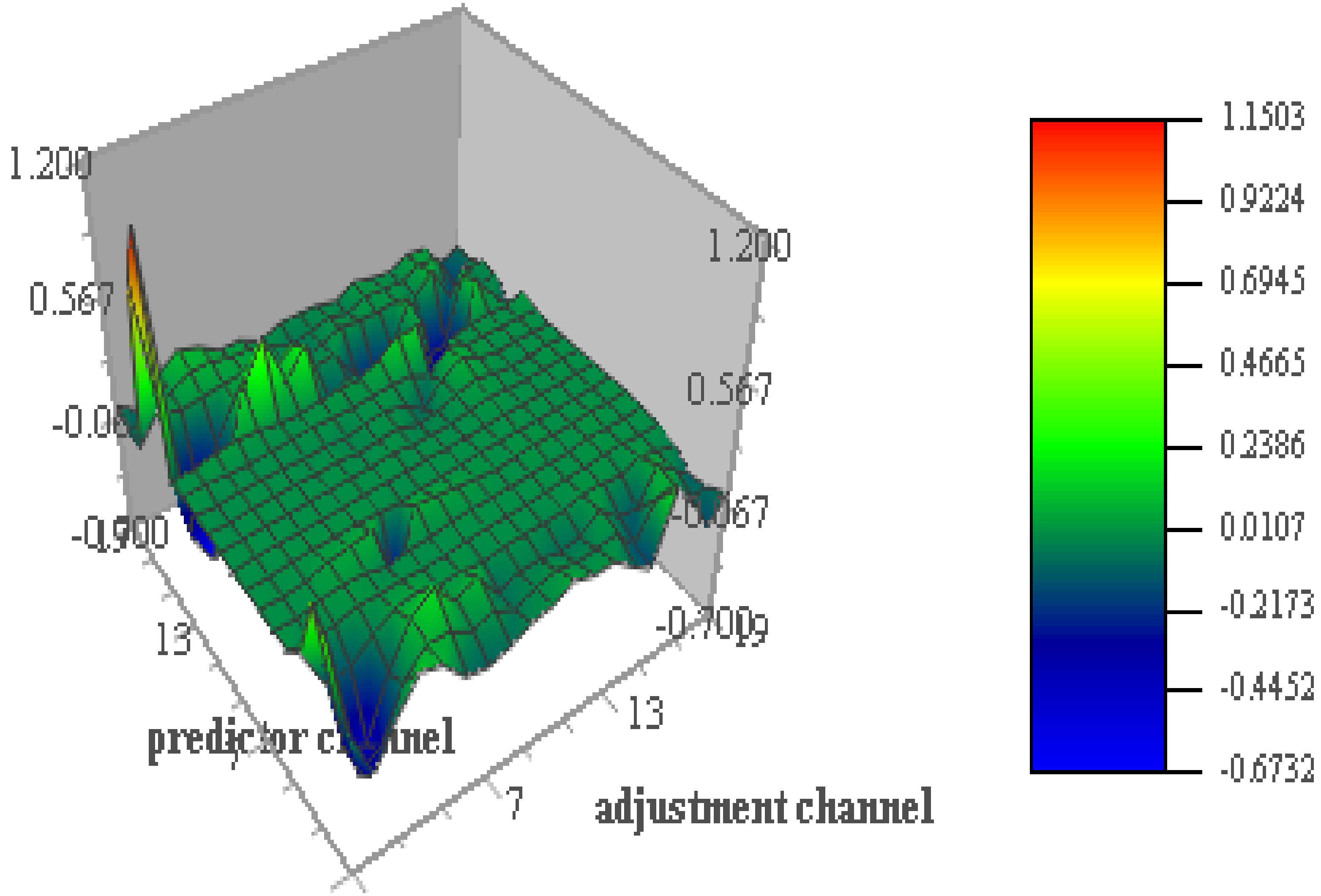


Shrinkage

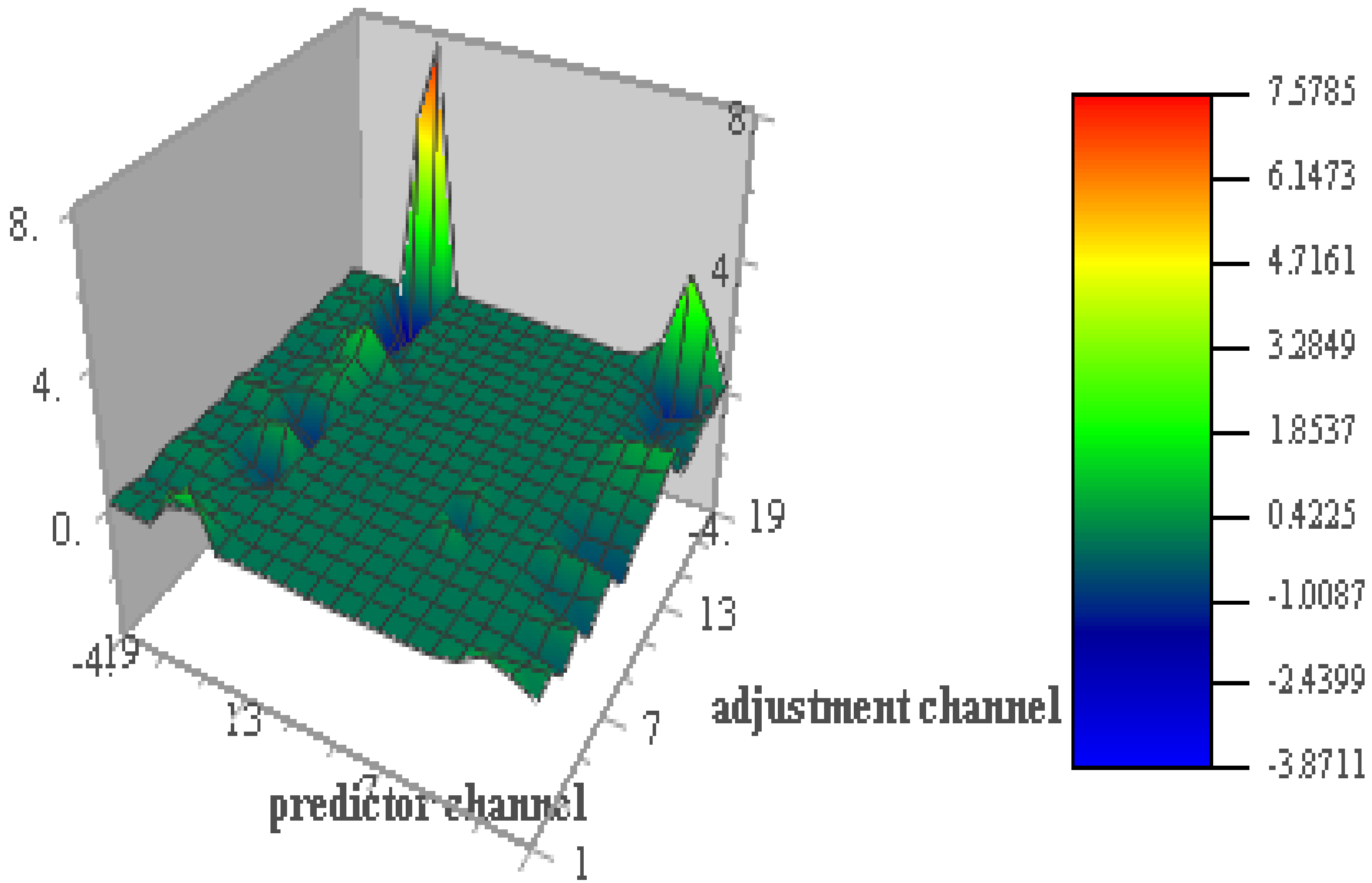


1.400
0.633
-0.133
1.400
0.633
-0.133
13
13
7
predictor channel
adjustment channel
1.3248
1.0529
0.7811
0.5093
0.2375
-0.0344
-0.3062
-0.5780
-0.8498

Rotated Regression



Orthogonal Regression



Results Summarized

- Maximum means maximum absolute value
- Ordinary least squares - max coefficient = -2.5778
- Ridge regression - max coefficient = 1.3248
- Shrinkage - max coefficient = 1.3248
  - Shrinkage to 0 as the guess coefficient is the same as ridge regression
- Rotated regression - max coefficient = 1.1503
  - Rotated to the ordinary least squares solution
- Orthogonal Regression - 7.5785

Recommendations

- Think
- Know what you are doing