1. **STANDARDIZATION**:

The model parameters are calculated for different data standardizations a) b) in c). With the LSE method the following model with the residual part is considered:

y = a1 x1 + … + r

With the “PCA” method the model is calculated by getting the data in matrix:

X = [A\_LOW T\_H2O C\_ACID I\_EFF]).

Next the covariance matrix F = X'X/(N-1) is calculated and svd is done: ([P,D] = svd(F)).

You have to choose the proper eigen vector P1 = P(:,i), to get the model equation in form of

P1' \* (x-V) = 0, where V is the center of data.

In order to compare the model parameters with the LSE method the implicit model must be converted into explicit form:

Example:

[p1 p2] \*([x1 y]' – [v1 v2]' ) = 0

p1 x1 + p2 y – p1 v1 –p2 v2 = 0

y = -p1/p2 x1 + p1/p2 v1 + p2/p2 v2

To compare the obtained model parameters for all three cases the ones obtained with the normalized data should be transformed back into the unnormalized space:

Example:

y\_n = (y-m\_y)/n\_y x\_n = (x –m\_x) / n\_x

y\_n = a\_n \* x\_n + r\_n

(y-m\_y)/n\_y = a\_n \* (x –m\_x) / n\_x + r\_n

…

y = (a\_n \* n\_y / n\_x) \* x - ( (a\_n \* n\_y / n\_x) \*m\_x - r\_n \* n\_y – m\_y )

Now you can fairly compare the model parameters by showing them in table (and comment the results).

Check which of the two methods is less sensitive to data standardization.

For each model you can calculate the error: e = y – ksi\*theta. What is the mean error for each model, what is the standard deviation of the error. Based on that you can judge on Consistency and bias of the model. Together with the normalized square error the quality of the model can be estimated. Good model is not biased and is consistent. Nonbiased model has mean error close to zero and consistent model has low error variance.

NRMSE:

Sqrt(Sum((yest-y)^2) /n)/std(y)

1. **Colinearity**:

A new measurement is introduced which is collinear with one of the existing measurements.

The data should be standardized (mean and std). With LSE method no residual part is needed since the data is centered.

With the svd the eigen values and vectors are obtained. Next the PCA procedure is done with transformation of the original data to targets. You can remove only one eigen vector:

T = X\*Ps.

Parameters are calculated THETA = (T'T)^-1 T' Y

And transformed back to original space THETA = Ps THETA.

Run the identification multiple times and observe the changes in parameters:

a)PCR and LSE parameters are different from the model error we can’t judge which ones are best.

b)Check the parameters for the two collinear variables. What can you observe?

c)What is the sum of the two parameters?

d)You can make a table of eigen vectors (transformation matrix Ps). Note that the collinear variables have approximately the same effect on each of the main components of the new space.

Example of table

|  |  |  |
| --- | --- | --- |
| Variable | P\_1 | P\_2 |
| X1 | P11 | P21 |
| X2 | P12 | P22 |
| X3 | P13 | P23 |
| … | .. |  |
|  |  |  |

P11 is the element of the first eigen vector. For better representation we can check the percentage of variable X1 in each component ( abs(P\_1)./sum(abs(P\_1)) ).