An Example of `plspolychaos` Use: ishigami20000

J.P. Gauchi and A. Bouvier

MaIAGE, INRA, Université Paris-Saclay,
78350 Jouy-en-Josas,
France

April 27, 2016

Abstract

The **plspolychaos** R package computes sensitivity indexes from polynomial chaos expansions and regression PLS, for computer models with correlated continuous inputs. The functionalities and method are explained in the in-line manual of the package.

This paper illustrates the way of using the package on an example dataset, named `ishi20000`, which has 20000 rows and 3 correlated inputs. The dataset is stored in the file `ishigami20000.Rda` in the directory `extdata` of the package. We analyze the full polynomial of degree 6, which corresponds to 83 monomials, then the polynomial reducted to 50 monomials by selection of the most significant ones.

Contents

1 Read Data

```
> library("plspolychaos")
> load(system.file("extdata", "ishigami20000.Rda", package="plspolychaos"))
> X <- ishi20000[, -ncol(ishi20000)] #inputs
> Y <- ishi20000[,  ncol(ishi20000)] #response
```

2 Building Legendre Polynomial

```
> degree <- 6 # polynomial degree
> pce <- polyLeg(X, Y, degree)
> print(pce)
```

```
Total number of monomials: 83
Number of inputs:  3
Polynomial degree:  6
Number of rows: 20000
```

3 Computations

```
> nc <- 30 # number of components
> ret <- calcPLSPCE(pce, nc=nc)
> print(ret)
```

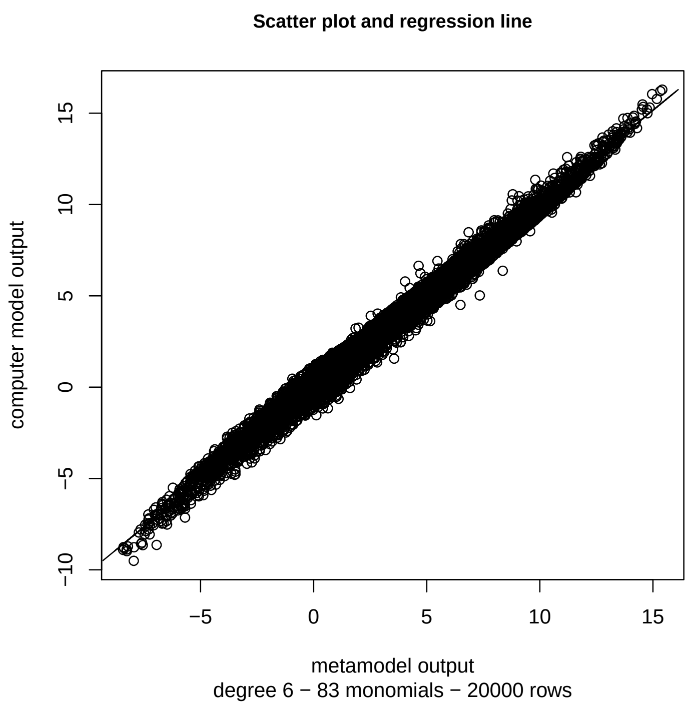
```
Explanation level of the response (R2, its percentage and cumulated percentage)
        R2     %R2 %R2cumulated
c1  0.3044 30.8609      30.8609
c2  0.4181 42.3873      73.2482
c3  0.0914  9.2668      82.5150
c4  0.1084 10.9885      93.5035
c5  0.0282  2.8629      96.3665
c6  0.0142  1.4361      97.8026
c7  0.0100  1.0103      98.8128
c8  0.0024  0.2414      99.0542
c9  0.0015  0.1487      99.2029
c10 0.0023  0.2370      99.4399
c11 0.0012  0.1236      99.5634
c12 0.0010  0.1021      99.6655
c13 0.0005  0.0549      99.7204
c14 0.0006  0.0583      99.7787
c15 0.0002  0.0192      99.7978
c16 0.0005  0.0534      99.8512
c17 0.0003  0.0286      99.8798
c18 0.0003  0.0269      99.9067
c19 0.0002  0.0200      99.9267
c20 0.0002  0.0175      99.9442
c21 0.0002  0.0178      99.9620
c22 0.0001  0.0151      99.9771
c23 0.0001  0.0054      99.9825
c24 0.0000  0.0026      99.9852
c25 0.0000  0.0038      99.9889
c26 0.0000  0.0033      99.9923
c27 0.0000  0.0021      99.9944
c28 0.0000  0.0022      99.9966
c29 0.0000  0.0018      99.9984
c30 0.0000  0.0016     100.0000
```

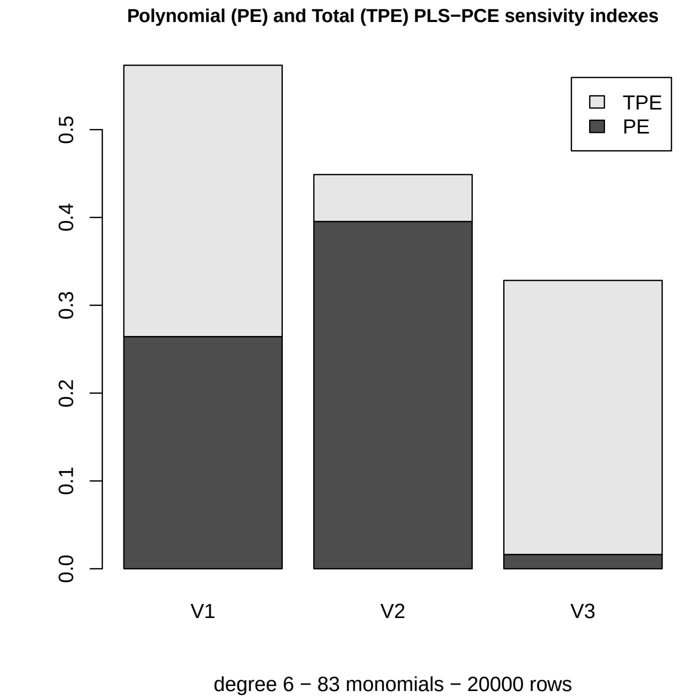
```
Explanation-prediction level of the response (Q2 and Q2cum)
        Q2  Q2cum
c1  0.3042 0.3042
c2  0.6010 0.7224
c3  0.3292 0.8138
c4  0.5824 0.9222
c5  0.3631 0.9505
c6  0.2860 0.9646
c7  0.2819 0.9746
c8  0.0931 0.9770
c9  0.0629 0.9784
c10 0.1076 0.9807
c11 0.0621 0.9819
c12 0.0544 0.9829
c13 0.0300 0.9834
c14 0.0330 0.9840
c15 0.0096 0.9841
c16 0.0313 0.9846
c17 0.0162 0.9849
c18 0.0153 0.9851
c19 0.0108 0.9853
c20 0.0091 0.9854
c21 0.0093 0.9855
c22 0.0075 0.9856
c23 0.0005 0.9857
c24 0.0000 0.9857
c25 0.0000 0.9857
c26 0.0000 0.9857
c27 0.0000 0.9857
c28 0.0000 0.9857
c29 0.0000 0.9857
c30 0.0000 0.9857

Optimal number of components:  23

Explanation level of the optimal number of components
       R2    %R2 %R2cumulated
c23 1e-04 0.0054      99.9825

Explanation-prediction level of the optimal number of components
       Q2  Q2cum
c23 5e-04 0.9857

Root Mean Square Prediction of the optimal number of components
     rmsep
c23 0.1176

PLS-PCE sensivity indexes
       LE     PE    TPE
V1 0.1109 0.2643 0.5733
```

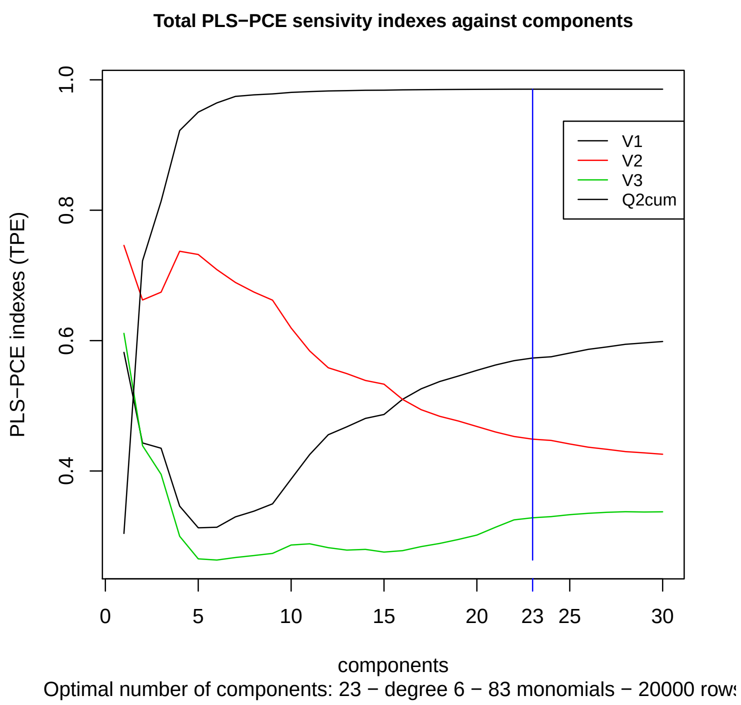
```
V2 0.0003 0.3954 0.4488
V3 0.0122 0.0162 0.3283

%PLS-PCE sensivity indexes
        LE      PE     TPE
V1 89.8935 39.1033 42.4535
V2  0.2203 58.4972 33.2371
V3  9.8862  2.3995 24.3094
```

4 Plots

```
> plot(ret, pce) #apply method 'plot' on the returned object
```




Polynomial (PE) and Total (TPE) PLS−PCE sensivity indexes
TPE
PE
0.0
0.1
0.2
0.3
0.4
0.5
V1
V2
V3
degree 6 − 83 monomials − 20000 rows


Total PLS−PCE sensivity indexes against components
V1
V2
V3
Q2cum
PLS−PCE indexes (TPE)
0.4
0.6
0.8
1.0
0
5
10
15
20
23
25
30
components
Optimal number of components: 23 − degree 6 − 83 monomials − 20000 rows

5 Monomials Selection

When there are many monomials, some of them are often nonsignificants. We select the most significant ones by using the option `forward` of the function `polyLeg`: only the monomials for which the R^2 of the linear regression is the greatest are kept.

Here, only the 50 more significant monomials are kept. The calculations are then processed as usual by function `calcPLSPCE`.

```
> pcef <- polyLeg(X, Y, degree, forward=50)
> print(pcef)
```

```
Total number of monomials: 83
Number of selected monomials: 50
Number of inputs:  3
Polynomial degree:  6
Number of rows: 20000
```

```
> retf <- calcPLSPCE(pcef, nc=25)
> print(retf)
```

```
Explanation level of the response (R2, its percentage and cumulated percentage)
        R2     %R2 %R2cumulated
c1  0.3537 36.0389      36.0389
c2  0.3629 36.9716      73.0105
c3  0.1404 14.3094      87.3199
c4  0.0755  7.6881      95.0080
c5  0.0184  1.8743      96.8823
c6  0.0080  0.8141      97.6964
c7  0.0024  0.2469      97.9433
c8  0.0038  0.3888      98.3321
c9  0.0018  0.1870      98.5191
c10 0.0030  0.3036      98.8227
c11 0.0014  0.1460      98.9687
c12 0.0018  0.1806      99.1494
c13 0.0020  0.2015      99.3509
c14 0.0009  0.0909      99.4418
c15 0.0010  0.1021      99.5438
c16 0.0013  0.1351      99.6789
c17 0.0008  0.0819      99.7608
c18 0.0006  0.0570      99.8178
c19 0.0002  0.0194      99.8372
c20 0.0005  0.0526      99.8898
c21 0.0003  0.0303      99.9201
c22 0.0003  0.0304      99.9505
c23 0.0003  0.0294      99.9799
c24 0.0001  0.0127      99.9926
c25 0.0001  0.0074     100.0000

Explanation-prediction level of the response (Q2 and Q2cum)
        Q2  Q2cum
```

```
c1  0.3535 0.3535
c2  0.5613 0.7164
c3  0.4954 0.8569
c4  0.5275 0.9324
c5  0.2720 0.9508
c6  0.1621 0.9588
c7  0.0582 0.9612
c8  0.0978 0.9650
c9  0.0516 0.9668
c10 0.0888 0.9697
c11 0.0460 0.9711
c12 0.0599 0.9728
c13 0.0715 0.9748
c14 0.0336 0.9756
c15 0.0392 0.9766
c16 0.0550 0.9779
c17 0.0342 0.9786
c18 0.0238 0.9791
c19 0.0062 0.9793
c20 0.0223 0.9797
c21 0.0115 0.9800
c22 0.0116 0.9802
c23 0.0110 0.9804
c24 0.0023 0.9805
c25 0.0000 0.9805

Optimal number of components:  24

Explanation level of the optimal number of components
       R2    %R2 %R2cumulated
c24 1e-04 0.0127      99.9926

Explanation-prediction level of the optimal number of components
        Q2  Q2cum
c24 0.0023 0.9805

Root Mean Square Prediction of the optimal number of components
     rmsep
c24 0.1366

PLS-PCE sensivity indexes
       LE     PE    TPE
V1 0.1458 0.1594 0.7526
V2 0.0097 0.2000 0.5706
V3 0.0095 0.0101 0.6322

%PLS-PCE sensivity indexes
        LE      PE     TPE
V1 88.3481 43.1479 38.4864
V2  5.9044 54.1247 29.1812
```

```
V3  5.7475  2.7273 32.3324
```