MOMC User's Guide

1. Description

MOMC is a FORTRAN77 program that computes a discrete approximation of the front of Pareto-optimal (PO) selection designs in case of one or more quality and one or more diversity objectives. The program can handle both single and multi-stage selection designs. The limitations of the program are listed below. For full details of the program see De Corte, Sacket & Lievens (2023).

2. Requirements to Run the Program

The program runs under Windows 10 and Windows 11. To run the program a recent 64 bits of Cygwin must be installed on the computer. To download Cygwin, go to the webpage of Cygwin: https://www.cygwin.com

After installing Cygwin update the installation (i.e., go back to the Cygwin webpage) by adding the libgfortran5 package to your installation. If the latter package is not in your installation, the program does not run!

3. Installing the Program

Go to the web site http://users.ugent.be/∼wdecorte/software.html and download the program (i.e., the momc.exe file). Copy the program to a directory that Cygwin has access to.

3. Running the Program

Prepare the input file for the selection situation. Details on this are given below. Copy the input file to the same directory that holds the program. Run Cygwin from your PC. This opens a command window under Cygwin. If necessary change to the directory where both the program and the input file are stored. Assuming that you did not change the name of the program (i.e., momc.exe) and that the name of your input file is momc.i, type the following command at the prompt:

./momc < momc.i

Note that the name momc.i can be replaced with an arbitrarily chosen name for the input file.

During execution the program displays information on the progress of the computations in the command window. Completion may take seconds to a few minutes, but also hours in high density 5-objective situations.

The full output of the program is on a separate file with name as given in the input file. The latter file contains a list with the PO systems obtained by the new hybrid method and the 2D subspace procedures. The file is ready for use in R scripts to perform further analyses (e.g., filtering preferred PO solutions, drawing the PO front, etc.). An example of such an R script can be downloaded from the website.

It is recommended to run the program twice. The first time with the value of MINREQ (see below) as chosen by the user. The second time with (the or) one of the values sugggested for MINREQ after the first run.

4. The Input File

Important: in preparing the input file, use a simple text editor such as Notepad, Wordpad or any other standard ASCII producing editor. DO NOT USE TEXT PROCESSING PROGRAMS SUCH AS MS-WORD. Here is a sample input file for the MOMC program.

```
  9   4 10   0
  2   4   1   3   15   1
  1 2 3 9
  1 2 4
  6
 1.00  0.34  0.57  0.17  0.13  0.20  0.12  0.25  0.45
 0.34  1.00  0.11  0.13  0.12  0.09  0.28  0.55  0.10
 0.57  0.11  1.00  0.11  0.58  0.28  0.12  0.19  0.16
 0.17  0.13  0.11  1.00  0.11  0.18  0.16  0.01  0.03
 0.13  0.12  0.58  0.11  1.00  0.03  0.24  0.37  0.36
 0.20  0.09  0.28  0.18  0.03  1.00  0.11  0.24  0.34
 0.12  0.28  0.12  0.16  0.24  0.11  1.00  0.59  0.07
 0.25  0.55  0.19  0.01  0.37  0.24  0.59  1.00 -0.06
 0.45  0.10  0.16  0.03  0.36  0.34  0.07 -0.06  1.00
 .60 .22 .03 .15
 .00   .00   .00   .00   .00   .00   .00   .00   .00
 .08   .23   .05 -.49 -.20 -.54   .03   .42 -.83
-.01   .21   .16 -.24   .10 -.38   .17   .28 -.72
-.31   .31 -.06   .04 -.35 -.11   .02 -.05   .06
  2   2
  1   4
  2   3
  0.0  0.15
  0.6  0.3
  250
 momof
```

We now comment each of the input lines

- # 1: **NVT, NGT, NRE, ISH**
  - NVT (integer): the total number of predictors and criteria. In the example NVT = 9. NVT < 11
  - NGT (integer): the total number of applicant groups (including the majority group). In the example NGT = 4. NGT < 7

  - NRE (integer): the number of times all nonlinear programs are solved. In the example NRE = 10. Recommended values are between 5 and 20.
  - ISH (integer): ISH has a value of either 0, meaning full hybrid and subspace solutions are computed, or 1, meaning only subspace solutions are computed.

- # 2: **NF, NP, NT, NG, NS, ND**
  - NF (integer): the number of selection stages. In the example NF = 2. NF < 6, but NF > 3 will substantially increase the running time of the program.
  - NP (integer): the number of actually used predictors. In the example NP = 4.
  - NT (integer): the number of quality goals in the actual selection
  - NG (integer): the total number of applicant groups (including the majority group) in the actual selection situation. Here NG = 3.
  - NS (integer): NS controls the density of the discrete approximation of the full PO front (i.e., in case that ISH = 0) as computed by the SBG and the NBI methods. Here NS = 15. If ISH = 0 then NS > 3 and NS less than NS_{max}, where NS_{max} decreases from 50 to 20 and to 10 for the total number of objectives (equal to NT + NG - 1) increasing from 3 to 4 and to 5. If ISH = 1, NS must not exceed 1000. Also, the total number of objectives must be greater than 2 and less or equal to 5 (when ISH = 0) or less or equal to 7 (when ISH = 1)
  - ND (integer): ND specifies the chosen metric for the quality and diversity objectives. if ND equals 1 then quality corresponds to the expected criterion performance of the selected applicants and diversity corresponds to the AIR of the minority groups, if ND equals 2 then quality corresponds to the validity of the selection composite and diversity corresponds to the AIR of the minority groups. Note that ND equal to 2 is only permitted if NF equals 1.

- # 3: **IDP(I), I=1,NP**.
  - IDP(I), I=1,NP (integers): The line contains NP values and each value indicates the row (and column) of the correlation matrix PCT (see below) that corresponds to one of the actually used predictors. So, in this case, the actually used predictors 1, 2 and 3 correspond to the first three rows (columns) of the matrix PCT, whereas predictor 4 corresponds to the 9th row (column) of PCT.

- # 4: **IDG(I), I=1,NG**.

  - IDG(I), I=1,NG (integers): The line contains NG values and the values indicate the rows of the effect size matrix EF (see below) that correspond to the groups included in the selection. The program assumes that the first group is the majority group and that the group is always included in the actual selection. The numbers 2 and 4 in the line correspond to the actual minority groups one and two (i.e., minority group one is the group with effect sizes detailed in row 2 of the EF matrix, whereas minority group 2 is the group with effect sizes detailed in row 4 of the EF matrix.

- # 5: **IDT(I), I=1,NT**.
  - IDT(I), I=1,NT (integers): The line contains NT values and the values indicate the rows (columns) of the correlation matrix PCT (see below) that correspond to the quality criteria of the actual selection

- #6.1–#6.NVT: **PCT** matrix.
  - PCT (decimal point numbers): In total NVT lines that specify the elements of the correlation matrix PCT

- # 7: **PRG(I), I=1,NGT**.
  - PRG(I), I=1,NGT (decimal point numbers): Vector of length NGT with elements that specify the proportional representation of the different application groups. The proportions must sum to 1. The program does not check this! If NG is smaller than NGT, the program automatically adapts the given proportions

- # 8.1–#8.NGT: **EF** matrix.
  - EF (decimal point numbers). In total NGT lines that specify the effect sizes of the NVT predictors and criteria for the NGT applicant groups. The program assumes that the first group is the majority group such that all elements in the first row of the ef matrix are all equal to zero.

- #9: **NPS(I), I=1,NF**
  - NPS(I), I=1,NF (integers). The NF elements of NPS specify the number of predictors used in the first, the second a.s.o stage until the number used in the final NFth stage.

- # 10.1–#10.NF: **IPS** matrix.

- IPS (integers). In total NF lines. The elements in the first line indicate the predictors used in the first stage, the elements in the second line indicate the predictors used in the second stage, a.s.o. Be careful to note that the indication refers to the NG numbers listed in Line #3 of the input. So, in this example, the numbers 1 and 4 in the line 10.1 indicate that the predictors used in the first stage correspond to the variable in the first and the nineth row (column) of the PCT correlation matrix; whereas the predictors used in the second stage correspond to the variables in the first and the second row (column) of the PCT correlation matrix.

- # 11: **WMI, FSR**
  - WMI (decimal point number): lower bound for the predictor weights. The program enforces that the sum of the predictor weights in each stage equals 1.
  - FSR (decimal point number): the final selection rate

- #12.1–#12.(NF-1)
  - In total (NF -1) lines, the first line indicates the upper and lower bound (decimal point numbers) on the retention rate after the first selection stage, the second line indicates the upper and lower bound on the retention rate after the second selection stage, and so on. However note that the lines 12.1-12.(NF-1) **must be omitted if nf equals 1**. Also, take care that the upper bound is at least equal to the lower bound for each intermediate stage and that the upper bound of the retention rate in stage (x+1) is at most equal to the lower bound on the retention rate in the preceding stage x. Again, the program does not check this!.

- #13: **MINREQ**
  - MINREQ (integer): specifies the number of required PO solutions when using the E-NBI method. MINREQ should not exceed 300, 200 and 100 if NF equals 3, 4 or 5. At the end of the first run, the program suggests one or more suitable values for MINREQ. Using one of these values in the second run typically results in a more even representation of the PO front.

- #14: **NAME**
  - A character string specifying the name (at most 9 characters) of the file holding the list of computed PO-systems.

5. Program Output

The program results in an output file with name as given in the input file (i.e., the file with the name "momof" in the example). The format of the file is such that it can be read as a data frame in the R programming environment. Thus, the first line specifies the header line and the following lines each specify a solution. Each of these lines contains the following elements:

- **num**: the number of the system;
- **inn**: a value of 1 if the system is PO else a value of 0
- **typ**: identifies the solution set the system is part of, with value 1 or 2 if the system is obtained by the SBG method; a value of 3 if the systems is obtained by the Das & Dennis NBI method; a value of 11 if the system is obtained by using the new hybrid method; a value of 4 if the system is obtained by the subspace procedure using average quality and minimum diversity as aggregated quality and diversity goals; a value of 5 if the system is obtained by the subspace procedure using average quality and average diversity as aggregated quality and diversity goals; and a value of 6 if the system is obtained by the subspace procedure using average quality and average selection rate as aggregated quality and diversity goals. The solution set of the new hybrid method consists of the solutions with typ = 2 and typ = 11.
- **QUA1-QUA(NT)**: the value of the NT quality trade-off components of the system;
- **DIV1-DIV(NG-1)**: the value of the NG-1 diversity trade-off components of the system;
- **SRG1-SRG(NG)**: selection rates for the NG groups
- **SRS1-SRS(NF)**: retention (selection) rates for the NF stages
- **PWE1-PWE(NPC)**: weights of the NPC predictors (i.e., weights for the predictors that are used in combination with at least one other predictor in the same stage).

Note that the value of the trade-off components is reported in the metric as chosen in Line 2 of the input file for the quality and diversity objectives. In the present example Line 2 specifies using the expected performance on the quality criterion of the selected applicants as metric for gauging quality, and using the AIR metric for gauging the two diversity objectives. So, the three values for the first PO system (i.e., 0.501, 0.369 and

1.041 indicate that the expected performance of the selected candidates on the quality criterion equals 0.501 and that the AIR for the first and the second minority group equals 0.369 and 1.041, respectively. The expected performance value is standardized with respect to the majority group metric.

The website provides several examples of input and corresponding output files.

6. Acknowledgement

When reporting results obtained by the present program, due reference should be made to De Corte, Sackett & Lievens (2023)

7. References

De Corte, W., Sackett, P. & Lievens (2023). Designing Pareto-optimal Selection Systems for Multiple Minority Subgroups and Multiple Criteria. *Journal of Applied Psychology* (accepted)