

Chapter 4

FlexAgg2.0: An Improved Version of the Data Aggregation Program FlexAgg

Nelson Villoria and Robert A. McDougall

This chapter describes a revised version of the command line data aggregation program FlexAgg, labeled FlexAgg2.0. This new version will be included in the GTAP Data Package starting with version 8.

4.1 What's New in FlexAgg2.0?

- FlexAgg2.0 takes advantage of GEMPACK text data files and is totally written in GEMPACK. (It does not use Awk for building aggregation mappings.)
- FlexAgg2.0 can be called from anywhere in the file system provided it has been added to the Windows path.
- Allows databases with names different to those predefined in the GTAP Data Package.
- Easy manipulation of commodity, region, and factor sets makes FlexAgg2.0 a flexible tool to accommodate and incorporate satellite datasets.
- Produces a Social Accounting Matrix (SAM) view of the GTAP Data Base.
- For ease of reference, mapping sets with description of disaggregated (original) sets are included in the sets file.
- Error messages in the aggregation process are indexed facilitating debugging of larger workflows.

4.2 System Requirements

The package has been tested on Windows XP, Vista and the 32 and 64 bit versions of Windows 7.

4.3 Preparing Aggregations

To prepare an aggregation, you need to provide the data base to be aggregated, prepare an aggregation mapping file, and run the aggregation program. Note that the basic aggregation program does not aggregate the time-series trade data; this is so that it will work with GTAP Data Bases not accompanied by such data.

4.3.1 Providing the Data Base to be Aggregated

The program aggregates a data base consisting of four files:

- [Prefix]SET.HAR : sets (Required)
- [Prefix]PAR.HAR : parameters (Required)
- [Prefix]DAT.HAR : data (Required)
- [Prefix]VOLE.HAR : energy volumes (Optional)
- [Prefix]TRADE.HAR : time series of trade data (Optional)

The prefix used in the standard GTAP Data Package is "GSD", i.e., GSDSET.HAR, GSDPAR.HAR, GSDDAT.HAR, GSDVOLE.HAR, and GSDTRADE.HAR. Upon installation of the GTAP Data Package, the files comprise a standard GTAP Data Base; you can however use FlexAgg2.0 to aggregate any standard version of the GTAP Data Base. For example, you may use the aggregation program to aggregate an already aggregated data base or perhaps for aggregating other versions of the data base. The inclusion of prefixes in FlexAgg2.0 were introduced to help with this. For example, a 10 sectors, 10 commodities, 3 factors aggregation could be prefixed 10X10X3SET.HAR, 10X10X3PAR.HAR, 10X10X3DAT.HAR. By using different prefixes, you could also have different versions of the same data base coexisting in the same folders. For example, you could have versions 6 and 7 of the data base coexisting in the same directory by using prefixes such as, say, V6 and V7.

4.3.2 Preparing an Aggregation Mapping File

To prepare an aggregation mapping file, make a copy of the file INPUT.TXT, and edit it to suit your requirements (INPUT.TXT is a GEMPACK text file, fully documented in section 6.1 of GPD-4 of the GEMPACK User Documentation Release 8.0). By using a GEMPACK text file, the user can easily change the sizes of the commodity and country sets, thus giving more flexibility to aggregate datasets with nonstandard set splits¹.

For details on layout, see the instructions included in INPUT.TXT itself. Save the edited file with a name ending in the extension .TXT.

In what follows we will use an aggregation file named MY_AGG.TXT. MY_AGG.TXT aggregates a fully disaggregated GTAP Data Base² to a $3 \times 3 \times 5$ data base. For your convenience, MY_AGG.TXT is already included in the FlexAgg2.0 directory.

4.3.3 Running the Aggregation Program

We suggest adding FlexAgg2.0 to the Windows path. If you do not have administrator privileges for changing Windows' environmental variables start the DOS-prompt (cmd.exe) and type the following command (see Figure 1)³:

```
set path=
```

¹Specific examples are the split of land into agroecological zones, the split of labor in further labor types, or the introduction of new sectors using tools such as SplitComm.

²For version 7 the fully disaggregated GTAP Data Base had 57 sectors, 113 regions and 5 factors

³Notice that you may need to do this each time that you restart the DOS-prompt, so in some point you may want to ask your administrator to add FlexAgg2.0 to the Windows path.

```
DOS-prompt
C:\>path=%path%;C:\FlexAgg2
C:\>cd GTAP
C:\gtap>data-agg my_agg
Processing aggregation mapping/sets information. . .
Checking that aggregated sets are well defined. . .
Processing user provided parameters . . .
Aggregating data file gsddat.har to Basedata.har. . .
Aggregating parameters file gsdpar.har to Default.prm. . .
Aggregating energy volume data file . . .
Creating GTAPView and GTAPTax files . . .
Creating SAM view of the database . . .
Done . . .
C:\gtap>cd my_agg
C:\gtap\my_agg>dir
Volume in drive C is OS
Volume Serial Number is 3A49-91AF

Directory of C:\gtap\my_agg

10/06/2011  03:29 PM    <DIR>          .
10/06/2011  03:29 PM    <DIR>          ..
10/06/2011  03:29 PM                21,137 gsdgdat.har
10/06/2011  03:29 PM                4,870 gsdgpar.har
10/06/2011  03:29 PM                20,154 gsdgsam.har
10/06/2011  03:29 PM                7,304 gsdgset.har
10/06/2011  03:29 PM                6,509 gsdgtax.har
10/06/2011  03:29 PM                26,978 gsdguview.har
10/06/2011  03:29 PM                2,680 gsdgvole.har
10/06/2011  03:29 PM                 237 info.txt
                8 File(s)      89,869 bytes
                2 Dir(s)  218,173,517,824 bytes free

C:\gtap\my_agg>
```

Figure 1: Running FlexAgg2.0 using the default method

Which assumes that the program resides in C:\FlexAgg2.0. If you have installed FlexAgg2.0 elsewhere, just be sure to substitute C:\FlexAgg2.0 with the appropriate location.

The aggregation program is DATA-AGG.BAT. To run it, make the directory with the GTAP Data Base your current directory. For example, if you *standard* GTAP Data Base resides in the directory C:\GTAP (files GSDSET.HAR, GSDPAR.HAR, GSDDAT.HAR, and GSDVOLE.HAR) first type:

```
cd C:\GTAP
```

And then invoke the program, using

```
DATA-AGG my_agg
```

Note that we omitted the extension .TXT. Also note that MY_AGG.TXT must reside in C: \GTAP.

The DOS-prompt will display the steps performed by the aggregation program. The aggregated data base is saved to a subdirectory named after the aggregation mapping file, in this case, C: \GTAP \MY_AGG. Besides the aggregated data files, the program will create report files [Prefix]gview.har and [Prefix]gtax.har containing summary information on the data base. It also creates a social accounting matrix view of the GTAP aggregation labeled [Prefix]sam.har. As shown in Figure 1, in our example these files are gsdgview.har, gsdgtax.har, gsdgsam.har. To inspect these files we recommend the GEMPACK utility ViewHAR. For later reference, FlexAgg2.0 also creates a summary file named info.TXT. This file specifies the name and location of the original datasets as well as the location of the newly aggregated datasets.

4.3.4 Running the Aggregation Program, custom_prefix

An alternative syntax is

```
data-agg my_agg custom_prefix [prefix]
```

This will look for files [prefix]dat.har, [prefix]set.har, and [prefix]par.har and produce [prefix]gdat.har, [prefix]gset.har, and [prefix]gpar.har.

This option is defined so several versions or aggregations of the GTAP Data Base can coexist in the same folder.

4.4 Useful Files and Programs in the FlexAgg2.0 Folder

- INPUT.TXT is a self-documented mapping file to assist the user in designing an aggregation.
- FlexAgg2.0 is a collection of tablo programs that are sequentially executed from a DOS-batch file (GBARUN.BAT called from DATA-AGG.BAT). GBARUN.BAT checks that the needed files are in place, creates a HAR mapping file based on the aggregation mapping file, writes the GEMPACK stored input files needed to run the tablo programs, and (if the aggregation succeeds) cleans up all the .LOG and .STI files left behind by the tablo programs by calling AGGCLEAN.BAT.

When the aggregation is unsuccessful, FlexAgg2.0 stops and refers the user to a GEMPACK-generated log file documenting the problem (more details in the subsection below).

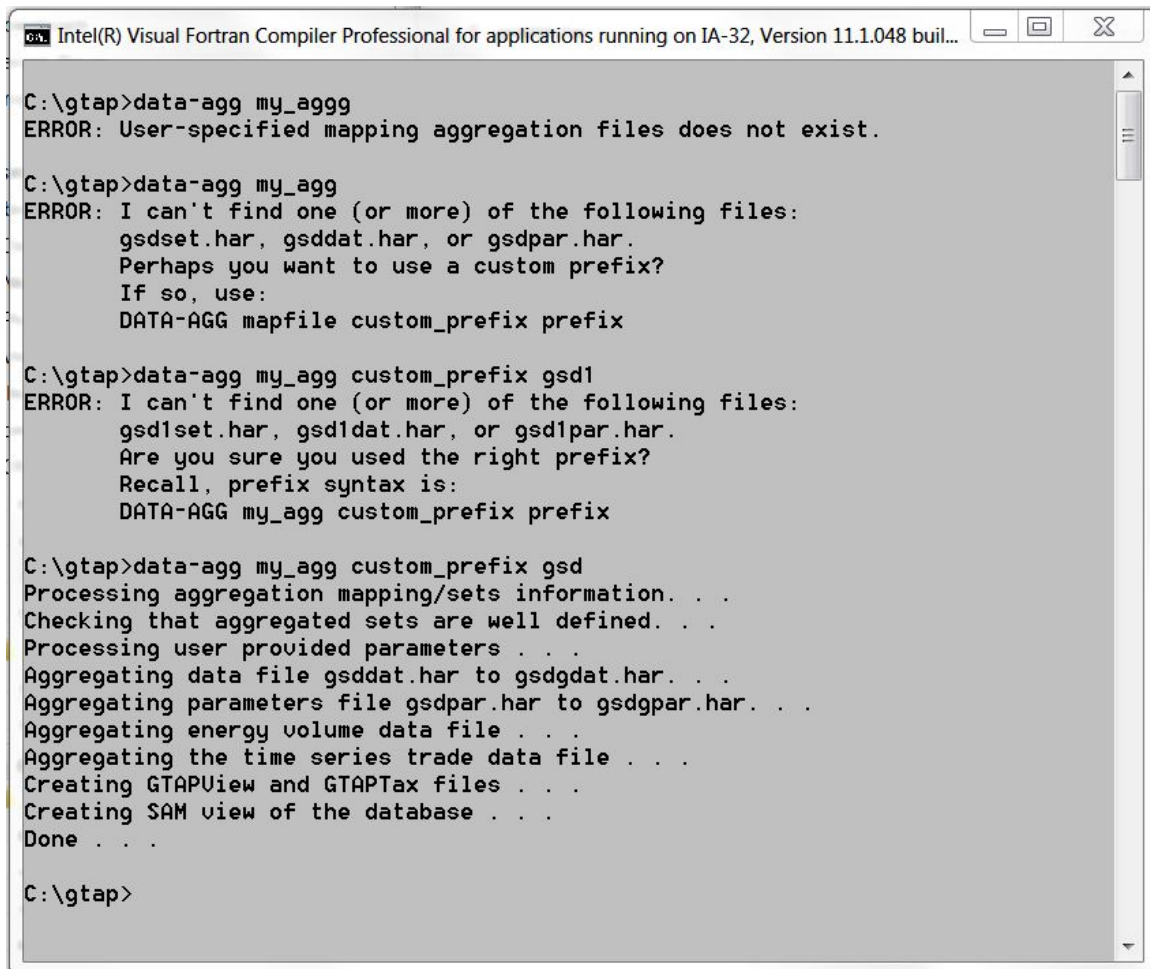
The tablo programs run from GBARUN.BAT are:

- AGGCHECK (AXS, AXT, EXE), checks that the aggregated sets are well defined.
- AGGDAT (AXS, AXT, EXE), aggregates the GTAP Data Base.
- AGGPARG (AXS, AXT, EXE), aggregates the GTAP parameters.
- AGGVOLUME (AXS, AXT, EXE), aggregates the data on energy volumes.
- AGGTRADE (AXS, AXT, EXE), aggregates the time series trade data.
- SAMVIEW (AXS, AXT, EXE), creates a social accounting matrix from the aggregated data base.
- GTAPVIEW (AXS, AXT, EXE), creates alternatives views of the data base (economic flows and tax rates).

- The subdirectory `.\src` contains the source of all the programs described above. It also contains `COMPILE.BAT`, which will attempt to compile all the `.TAB` files, generated the executable files and copy them to the `FlexAgg2.0` folder.
- `MODHAR.EXE`, a GEMPACK program used to read the aggregation mapping file (GEMPACK text file) and create the aggregation mapping `GMAP.HAR`.

4.5 Error Messages in FlexAgg2.0

FlexAgg2.0 first ensures that both the aggregation mapping file and the required files of the GTAP Data Base (i.e., `*data.har`, `*set.har`, and `*.par.har`) exist in the user-defined directory. The error messages produced when one of these conditions are not met are shown in Figure 2.



```

Intel(R) Visual Fortran Compiler Professional for applications running on IA-32, Version 11.1.048 build...
C:\gtap>data-agg my_aggg
ERROR: User-specified mapping aggregation files does not exist.

C:\gtap>data-agg my_agg
ERROR: I can't find one (or more) of the following files:
      gsdset.har, gsddat.har, or gsdpar.har.
Perhaps you want to use a custom prefix?
If so, use:
DATA-AGG mapfile custom_prefix prefix

C:\gtap>data-agg my_agg custom_prefix gsd1
ERROR: I can't find one (or more) of the following files:
      gsd1set.har, gsd1dat.har, or gsd1par.har.
Are you sure you used the right prefix?
Recall, prefix syntax is:
DATA-AGG my_agg custom_prefix prefix

C:\gtap>data-agg my_agg custom_prefix gsd
Processing aggregation mapping/sets information. . . .
Checking that aggregated sets are well defined. . . .
Processing user provided parameters . . . .
Aggregating data file gsddat.har to gsdgdat.har. . . .
Aggregating parameters file gsdpar.har to gsdgpar.har. . . .
Aggregating energy volume data file . . . .
Aggregating the time series trade data file . . . .
Creating GTAPView and GTAPTax files . . . .
Creating SAM view of the database . . . .
Done . . . .

C:\gtap>

```

Figure 2: Error messages displayed by FlexAgg2.0 when the aggregation mapping file or required datasets are not in the user-specified directory. The last step shows a successful aggregation.

FlexAgg2.0 also exploits the log files generated by the GEMPACK programs to facilitate debugging. Whenever one of the FlexAgg2.0 programs cannot complete its task, the program tells the user to look at log file. In most cases the log files are informative enough for debugging any

error. FlexAgg2.0 also assigns a code to each error. This code is the value of the DOS variable %errorlevel% that in turn can be used to identify problems in FlexAgg2.0 when it is used as part of a larger workflow using DOS batch files. Table 1 shows a complete list of the errors messages included in FlexAgg2.0 as well as the log file that documents the program execution.

Table 1: Error messages and exit codes

Code	Error	Log File
1	Need to specify a mapping file	
2	User-specified mapping aggregation files does not exist.	
3	Required data files are not in the directory	
4	Required data files with specified prefix are not in the directory	
5	error in the aggregation mapping file	gmap.log
6	aggregated and disaggregated sets are inconsistent	aggcheck.log
7	can't convert mapping file to HAR format	gmap.log
8	can't aggregate data file	aggdat.log
9	can't aggregate parameters file	aggpar.log
10	can't aggregate energy volume data file	aggvole.log
11	can't aggregate the time series data file	aggtrade.log
12	can't create gtapview file	gtapview.log
13	can't create SAM view file	samview.log

4.6 Legalese

As always, to the extent permitted by applicable law, THE GTAP DATA PACKAGE COMES WITH ABSOLUTELY NO WARRANTY.

4.7 Contact Information

Questions or comments about malfunctioning of FlexAgg2.0 can be send to Rob McDougall (rmcdougall@purdue.edu) or Nelson Villoria (nvillori@purdue.edu).