Open-source Quality Assurance and Performance Analysis Tools

Michael Bussieck  
MBussieck@gams.com

Steven P. Dirkse  
SDirkse@gams.com

Stefan Vigerske  
SVigerske@gams.com

GAMS Development Corp • GAMS Software GmbH
Welcome/Agenda

- SQA at GAMS
- Effective Testing
- Performance Analysis Tool Paver
- Examples
<table>
<thead>
<tr>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQA at GAMS</td>
</tr>
<tr>
<td>Effective Testing</td>
</tr>
<tr>
<td>Performance Analysis Tool Paver</td>
</tr>
<tr>
<td>Examples</td>
</tr>
</tbody>
</table>
Motivation

Quality Assurance
- Essential component in most industries
- Important in most software engineering sectors

Mathematical Programming
- Less attention to quality assurance (small community)
- Specific QA issues for modeling systems (initially expensive)
- Different focus for industry and academic

industry
Focus on reliability

academia
Focus on performance
Software Quality Assurance at GAMS

- Software configuration management
- Quality control and tests of the product
- Client model testing
- Performance Analysis tools: PAVER
- Solution verification tool: Examiner
- Model converter and “encryption“ tool: Convert
Quality Test Models Library

- Include tests to verify proper behavior of the system.

- More than 600 quality test models, each containing numerous pass/fail tests:
  ```
  ...
  abort$card(delta) 'time routines have an error';
  ...
  ```

- Check basic functionality of the solver and the link:
  ```
  ...
  abort$$(( abs(cost.m-cost_m) > tol) 'bad cost.m';
  ...
  ```

- Gives developer and users assurance about the basic functionality of the link and the solver.

- Automatic generated test summaries with different level of information.
# Latest GAMS System Builds and Test Results

NOTE: The (nightly) alpha builds are internal development versions of the GAMS system. They may have known bugs, unfinished features, beta versions of third-party software, or may not function at all! Not for production use!

<table>
<thead>
<tr>
<th>nightly α System</th>
<th>Libraries</th>
<th>Build</th>
<th>Rev</th>
<th>Status and Time (UTC)</th>
<th>Initial Tests</th>
<th>Full Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thursday Inx</td>
<td>Download</td>
<td>24.2.0</td>
<td>41150</td>
<td>Test started 28Jun2013 01:43:36</td>
<td>918 runs 1 failures (q=1,s=0)</td>
<td>Report results pending</td>
</tr>
<tr>
<td>Monday leg</td>
<td>Download</td>
<td>24.2.0</td>
<td>41173</td>
<td>Test done 01Jul2013 17:15:28</td>
<td>777 runs 0 failures (q=0,s=0)</td>
<td>Report 2141 runs 1 failures (q=0,s=0,a=1) Report</td>
</tr>
<tr>
<td>Friday vs8</td>
<td>Download</td>
<td>24.2.0</td>
<td>41164</td>
<td>Test started 29Jun2013 01:06:54</td>
<td>920 runs 0 failures (q=0,s=0)</td>
<td>Report results pending</td>
</tr>
<tr>
<td>Friday wei</td>
<td>Download</td>
<td>24.2.0</td>
<td>41166</td>
<td>Test started 29Jun2013 03:23:50</td>
<td>917 runs 1 failures (q=1,s=0)</td>
<td>Report results pending</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nightly β System</th>
<th>Libraries</th>
<th>Build</th>
<th>Rev</th>
<th>Status and Time (UTC)</th>
<th>Initial Tests</th>
<th>Full Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thursday Inx</td>
<td>Download</td>
<td>24.1.2</td>
<td>41150</td>
<td>Test done 01Jul2013 18:58:07</td>
<td>897 runs 0 failures (q=0,s=0)</td>
<td>Report 16714 runs 3 failures (q=0,s=3) Report</td>
</tr>
<tr>
<td>Thursday vs8</td>
<td>Download</td>
<td>24.1.2</td>
<td>41150</td>
<td>Test done 28Jun2013 07:47:08</td>
<td>761 runs 0 failures (q=0,s=0)</td>
<td>Report 11107 runs 0 failures (q=0,s=0) Report</td>
</tr>
<tr>
<td>Friday wei</td>
<td>Download</td>
<td>24.1.2</td>
<td>41166</td>
<td>Test done 01Jul2013 13:32:22</td>
<td>896 runs 0 failures (q=0,s=0)</td>
<td>Report 16768 runs 1 failures (q=0,s=1) Report</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>alpha System</th>
<th>Libraries</th>
<th>Build</th>
<th>Rev</th>
<th>Status and Time (UTC)</th>
<th>Initial Tests</th>
<th>Full Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>20130427 aix</td>
<td>Download</td>
<td>24.1.0</td>
<td>39956</td>
<td>Test done 27Apr2013 22:18:39</td>
<td>387 runs 1 failures (q=1,s=0)</td>
<td>Report 3427 runs 5 failures (q=1,s=1,e=3) Report</td>
</tr>
<tr>
<td>20130627 deg</td>
<td>Download</td>
<td>24.2.0</td>
<td>41138</td>
<td>Test done 27Jun2013 12:50:22</td>
<td>706 runs 0 failures (q=0,s=0)</td>
<td>Report 10171 runs 0 failures (q=0,s=0) Report</td>
</tr>
<tr>
<td>20130627 leg</td>
<td>Download</td>
<td>24.2.0</td>
<td>41138</td>
<td>Test done 27Jun2013 21:15:39</td>
<td>778 runs 0 failures (q=0,s=0)</td>
<td>Report 11666 runs 0 failures (q=0,s=0) Report</td>
</tr>
</tbody>
</table>
Agenda

- SQA at GAMS
- Effective Testing
- Performance Analysis Tool Paver
- Examples
Effective Testing

- Test cases
  - Widely available collection of standardized test instances
- Data collection tools
  - Automatic collection of solution and statistics
  - Capture test environment setting (hardware, software)
- Data analysis tools
  - Standard quality and performance measurements

- Give the tools in the hand of the user.
  Make your own benchmarks!
Welcome to the GAMS World

This is the home page of the GAMS World, a web site aiming to bridge the gap between academia and industry by providing highly focused forums and dissemination services in specialized areas of mathematical programming.

Substantial progress was made in the 1980s and 1990s with the development of algebra based modeling systems, algorithms, and computer codes to solve large and complex mathematical programs. The application of these tools, however, was less than expected. The abstraction, expression, and translation of real world problems into reliable and effective operational systems requires highly specialized and domains specific knowledge. The process of acquisition and dissemination of this knowledge is complex and poorly understood and the number of "good modelers" is much less than we all hoped for. Similarly, the process of transforming a new algorithm into a reliable and effective solution system is a slow and expensive process and there are few "good implementers". This web site hopes to address some of these problems by helping with the collection and dissemination of domain specific information and knowledge that is outside the established channels because of its content or form.

For example, model structures and results get published in commercial and academic papers but it is virtually impossible to reproduce any of those results or lift model components and data from one study to be used in some other study. Algorithm implementers face a similar dilemma when trying to get their hands on real world data models and data to test and refine their systems. This web site offers a few, well focused and maintained services to help with the dissemination of problems and solutions.

The GAMS World Google discussion group is associated with GAMS World.
Welcome to the Performance World!

Performance World is a forum for discussion and dissemination of information and tools about all aspects of performance testing of solvers for mathematical programming problems. This world has been established in response to user demands for independent and reproducible performance results.

Overall performance highly depends on problem formulation, solver, and tuning parameters. Our performance tools are designed to serve the different needs of our user community. One user may be interested in finding the most reliable way to solve a proprietary or classified model. On the other hand, an academic researcher may be interested in testing a new algorithm against a set of existing test problems and competing approaches. The main features are:

- Uniform access to a comprehensive set of established and new test problems
- Automation tools for collecting performance measurements
- Tools for analyzing and visualizing test results

What's New:

- Experimental PAVER 2.0 Server is on-line.
- The paper PAVER 2.0: An Open Source Environment for Automated Performance Analysis of Benchmarking Data is available.
- Several new libraries (Fixed Cost Network Flow and the Princeton NLP collection) have been added to the Performance Libraries.
- A collection of quadratically constrained programs (QCP) have been added to the LINLIB set of models.
- The paper A Server for Automated Performance Analysis and Benchmarking of Optimization Software is available which includes an NLP benchmark using PAVER. See the results on (all models or a subset of models).
- Presentations on benchmarking and performance testing from the INFORMS 2002, San Jose conference have been added to the related links section.
- Try our online PAVER Server for automated performance analysis and visualization, batch file creation and model translation.
- New tools for analyzing non-convex or discrete models (quality of solution information).
- MINLP type models from the MINLP World have been added to the PerformanceLib.

Performance World is featured by GAMS World
Performance Libraries

- Performance tests require public test libraries
- Shared test libraries allow reproducible results

The following libraries are available:

1. AMPLBookLib - LP, MIP, and NLP
2. FCNetLib - MIP
3. GLOBALLib - NLP
4. LINLib - LP, MIP, and QCP
5. MacMOOPLib - LP, NLP
6. MIPLIB 2010 - MIP
7. MPLLib - LP, NLP
8. PrincetonLib - NLP
9. Selected Continuous Global Optimization Lib - NLP and DNLP
10. XPRESSLib - LP, NLP

MacMOOP - A Collection of Multiobjective Optimization Testproblems

MacMOOP is a collection of multiobjective optimization testproblems from Sven Leyffer. See the MacMOOP Collection. The GAMS models were translated by Andre Savistky as is from the original AMPL source.

The raw GAMS models were converted into GAMS scalar format using the CONVERT facility. For models with multiple solves, only the first solve instance is translated.

Download

Download MacMOOPLib.zip

Total number of models: 14
Original Source: My AMPL Collection of Multiobjective Optimization Testproblems (Sven Leyffer)

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>#Eqns</th>
<th>#Vars</th>
<th>#NZ</th>
<th>#NNZ</th>
<th>Bestknown Objective</th>
<th>GAMS Scalar</th>
<th>AMPL Source</th>
<th>GAMS Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc-comp</td>
<td>NLP</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>17.96428571</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>ex001</td>
<td>NLP</td>
<td>4</td>
<td>6</td>
<td>21</td>
<td>11</td>
<td>0.55508075</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>ex002</td>
<td>NLP</td>
<td>3</td>
<td>6</td>
<td>13</td>
<td>3</td>
<td>8.00000000</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>ex003</td>
<td>NLP</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>7.25098758</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>ex004</td>
<td>LP</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>2.00000000</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>ex005</td>
<td>NLP</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>-4.00000000</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>hs05x</td>
<td>NLP</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td>5</td>
<td>0.14049587</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>liswem</td>
<td>NLP</td>
<td>3</td>
<td>9</td>
<td>13</td>
<td>7</td>
<td>-3.41233322</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>molpg_1</td>
<td>NLP</td>
<td>9</td>
<td>9</td>
<td>51</td>
<td>0</td>
<td>-26.01668874</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>molpg_2</td>
<td>NLP</td>
<td>17</td>
<td>13</td>
<td>67</td>
<td>0</td>
<td>-13.37500000</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>molpg_3</td>
<td>NLP</td>
<td>15</td>
<td>11</td>
<td>55</td>
<td>0</td>
<td>-13.40000000</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>moqp_1</td>
<td>NLP</td>
<td>14</td>
<td>24</td>
<td>102</td>
<td>60</td>
<td>-57.29327326</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>moqp_2</td>
<td>NLP</td>
<td>13</td>
<td>24</td>
<td>102</td>
<td>60</td>
<td>-13.16755651</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
<tr>
<td>moqp_3</td>
<td>NLP</td>
<td>14</td>
<td>24</td>
<td>101</td>
<td>60</td>
<td>-24.54193065</td>
<td>gms</td>
<td>ampl</td>
<td>raw gams</td>
</tr>
</tbody>
</table>
MINLP written by GAMS Convert

Variables

\[ b_1, b_2, i_3, i_4, i_5, i_6, i_7, i_8, x_9; \]

Binary Variables \( b_1, b_2; \)

Integer Variables \( i_3, i_4, i_5, i_6, i_7, i_8; \)

Equations

\[ c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, \]
\[ e_{11}, e_{12}, e_{13}; \]

\[ c_1 \cdot 0.1 \cdot b_1 + 0.2 \cdot b_2 + i_3 + i_4 - x_9 = 0; \]
\[ e_2 \cdot 460 \cdot i_5 + 570 \cdot i_7 - L - 1900; \]
\[ e_3 \cdot 460 \cdot i_6 + 570 \cdot i_8 = L = 1900; \]
\[ e_4 \cdot 460 \cdot i_5 + 570 \cdot i_7 - G = 1700; \]
\[ e_5 \cdot 460 \cdot i_6 + 570 \cdot i_8 = G = 1700; \]
\[ e_6 \cdot i_5 + i_7 - L - 5; \]
\[ e_7 \cdot i_6 + i_8 - L - 5; \]
\[ e_8 \cdot i_3 + i_4 - L - 5; \]
\[ e_9 \cdot i_3 \cdot 17 + i_4 \cdot 18 - G - 7; \]
\[ e_{10} \cdot b_1 - i_3 - L - 0; \]
\[ e_{11} \cdot b_2 - i_4 - L - 0; \]
\[ e_{12} \cdot - 15 \cdot b_1 + i_3 = L = 0; \]
\[ e_{13} \cdot - 15 \cdot b_2 + i_4 = L = 0; \]

* sol non default bounds

\[ i_3 \text{. up } = -15; \]
\[ i_4 \text{. up } = -15; \]
\[ i_5 \text{. up } = -5; \]
\[ i_6 \text{. up } = -5; \]
\[ i_7 \text{. up } = -5; \]
\[ i_8 \text{. up } = -5; \]

Model m / all /;
Solve m using MINLP minimizing x9;
PAVER - GAMS Model Translation Web Submission Tool (GMS2XX)

The PAVER GAMS model translation web-submission tool runs the GAMS/CONVERT "solver" to translate GAMS models into the following supported languages:

- AlphaECP
- AMPL
- AmpINLC
- BARON
- CoinFML
- CplexLP
- CplexMPS
- Dict
- FixedMPS
- GAMS (scalar)
- Jacobian
- Lago
- Lgo
- LindoMPI
- LINGO
- MINOPT
- NLP2MCP
- ViennaDag
Data Collection – GAMS Trace Files

* Trace Record Definition
  * GamsSolve
  * InputFileName, ModelType, SolverName, OptionFile, Direction, NumberOfEquations,
  * NumberOfVariables, NumberOfDiscreteVariables, NumberOfNonZeros,
  * NumberOfNonlinearNonZeros, ModelStatus, SolverStatus, ObjectiveValue,
  * ObjectiveValueEstimate, SolverTime, ETSolver, NumberOfIterations, NumberOfNodes

30n20b8, MIP, SCIP, 1, 0, 577, 18381, 11098, 109709, 0, 1, 1, 302, 302, 186.80, 189.833, 464659, 466
acc-tight5, MIP, SCIP, 1, 0, 3053, 1340, 1339, 16136, 0, 1, 1, 0, 0, 366.28, 367.651, 1788064, 1971
aflow40b, MIP, SCIP, 1, 0, 1443, 2729, 1364, 8148, 0, 1, 1, 1168, 1168, 1411.99, 1425.472, 5232401, 3773
Timo Berthold: Measuring the impact of primal heuristics

\[ P(T) := \int_{t=0}^{T} p(t), \]

*solvetrace file SCIP.miptrace: ID = SCIP 3.0.1*
* fields are lineNum, seriesID, node, seconds, bestFound, bestBound*

1, S, 1, 0, 260614197.6, 216717059.8
2, T, 3, 1.1205, 260614197.6, 217028062.2
...
63, E, 2550, 38.3884, 220249516.8, 217928729.7
*solvetrace file closed*
Agenda

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SQA at GAMS</td>
<td></td>
</tr>
<tr>
<td>Effective Testing</td>
<td></td>
</tr>
<tr>
<td>Performance Analysis Tool Paver</td>
<td></td>
</tr>
<tr>
<td>Examples</td>
<td></td>
</tr>
</tbody>
</table>
PAVER Server

- PAVER server (Performance Analysis and Visualization for Effortless Reproducibility)

- 2013: Complete rewrite PAVER 2.0 (Python):
  - New functionality:
    - consistency checks
    - comparison against solution database
    - many more metrics
  - Easily extendable
  - Open source (COIN-OR project)
PAVER 2 – Web Submission

The PAVER web-submission tool runs the PAVER performance analysis tool. See the README for details.

Users can input their data in the form of trace files, which are in the .trc format.

If you find this tool useful, please consider citing the paper by M. Bussieck, S. Dirkse, S. Vigerske (2013). PAVER 2.0: A new performance analysis tool for MPECs. MSc Thesis. Department of Business Informatics, University of Heidelberg. (online access)

Note: there is a maximum total file size limit of 1Mb.

Submit trace files:
- Trace 1: Browse...
- Trace 2: Browse...
- Trace 3: Browse...
- Trace 4: No file selected.
- Trace 5: No file selected.
- Trace 6: No file selected.
- Trace 7: No file selected.
- Trace 8: No file selected.

Submit solution files (optional):
- MINLPLib
- GlobalLib
- LinLib
- MIPLIB 2010
- Solutions 1: No file selected.

Optional Settings:
- Relative Tolerance on Bounds: 1e-6
- Absolute Tolerance on Bounds: 0.0001
- (Primal) Feasibility Tolerance: 2e-6
- Optimality (Dual Feasibility) Tolerance: 2e-6
- Reference Solver (Name):
- Shift for Time (s): 10
- Shift for Number of Nodes: 100
- Minimal Time: 1
- Time in case of failure:
- Number of Nodes in case of failure:
- (Relative) Gap Tolerance: 1e-6
- Threshold for being relatively faster: 0.1
- Threshold for relatively better obj. value: 0.1
- Regard Dual Bounds (if available): checked
- Number of ticks (points): 40
- Extended Performance Profiles:
- Include vrt. best solver: checked
- Option file name is runtime:

Run PAVER

For questions or comments please contact the PAVER 2 author.
PAVER 2 – Output

Analysis Results
Your data was successfully submitted to the PAVER - Performance Analysis

- http://www.gamsworld.org/performance/paver2/analyzer

You can also download the results at
- http://www.gamsworld.org/performance/paver2/analyzer

Submit data:

Date/Time: Mon Jul 1 16:02:46 EDT 2013

Log file

Solver Runs
- Thorin
- Gimli
- Bombur
- Balin
- Bifur

Solving Data

Statistics (Counts and Means)

Performance Profiles

Graph showing performance profiles for SolverTime with 0.0 <= Gap <= 1e-06
## Agenda

<table>
<thead>
<tr>
<th>SQA at GAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective Testing</td>
</tr>
<tr>
<td>Performance Analysis Tool Paver</td>
</tr>
<tr>
<td>Examples</td>
</tr>
</tbody>
</table>
PAVER Example

- 5 MIP Solvers
  - Thorin, Gimli, Bombur, Balin, Bifur
  - Models from MIPLIB 2010 (Benchmark set)
  - Time limit: 1 hour

```
python src/paver/paver.py \
  balin.trc bifur.trc bombur.trc gimli.trc thorin.trc \
  solu/miplib2010.solu \
  --failtime 3600 --refsolver Gimli --writehtml mip
```