Citation for published version:

Digital Object Identifier (DOI):
10.1093/bioinformatics/btt023

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Bioinformatics

Publisher Rights Statement:
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
SBSI: an extensible distributed software infrastructure for parameter estimation in systems biology

Richard Adams1, Allan Clark1, Azusa Yamaguchi2, Neil Hanlon3, Nikos Tsorman4, Shakir Ali5, Galina Lebedeva1, Alexey Goltsov6, Anatoly Sorokin7, Ozgur E. Akman8, Carl Troein9, Andrew J. Millar1, Igor Goryanin2 and Stephen Gilmore1,2,*

1SynthSys Edinburgh, King’s Buildings, University of Edinburgh, Edinburgh, UK, 2School of Informatics, Informatics Forum, University of Edinburgh, Edinburgh, UK, 3BSkyB, Watermark House, Livingston, Scotland, 4The Systems Biology Institute, Falcon Building 5F, 5-6-9 Shirokanedai, Tokyo 108-0071, Japan, 5Horizon Digital Economy Research, Institute of Cell Biophysics RAS, Institutskaya st., 3, Pushchino, Russia, 6University of Abertay Dundee, Dundee, UK, 7Institute of Cell Biophysics RAS, Institutskaya st., 3, Pushchino, Russia, 8Centre for Systems, Dynamics and Control, University of Exeter, UK and 9Computational Biology and Biological Physics, Lund University, Söldvagan 14A, Lund, Sweden

ABSTRACT

Summary: Complex computational experiments in Systems Biology, such as fitting model parameters to experimental data, can be challenging to perform. Not only do they frequently require a high level of computational power, but the software needed to run the experiment needs to be usable by scientists with varying levels of computational expertise, and modelers need to be able to obtain up-to-date experimental data resources easily. We have developed a software suite, the Systems Biology Software Infrastructure (SBSI), to facilitate the parameter-fitting process. SBSI is a modular software suite composed of three major components: SBSINumerics, a high-performance library containing parallelized algorithms for performing parameter fitting; SBSIDispatcher, a middleware application to track experiments and submit jobs to back-end servers; and SBSIVisual, an extensible client application used to configure optimization experiments and view results. Furthermore, we have created a plugin infrastructure to enable project-specific modules to be easily installed. Plugin developers can take advantage of the existing user-interface and application framework to customize SBSI for their own uses, facilitated by SBSI’s use of standard data formats.

Availability and implementation: All SBSI binaries and source-code are freely available from http://sourceforge.net/projects/sbsi under an Apache 2 open-source license. The server-side SBSINumerics runs on any Unix-based operating system; both SBSIVisual and SBSIDispatcher are written in Java and are platform independent, allowing use on Windows, Linux and Mac OS X. The SBSI project website at http://www.sbsi.ed.ac.uk provides documentation and tutorials.

Contact: stg@inf.ed.ac.uk

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on June 28, 2012; revised on November 22, 2012; accepted on January 12, 2013

1 THE SBSI SYSTEM

SBSI is a modular software system, consisting of three major software components: (i) SBSINumerics, containing core numerical global search algorithms; (ii) SBSIDispatcher, providing access to back-end HPC machines; and (iii) SBSIVisual, a client application that provides a user-friendly consistent user interface to the tools provided by the SBSINumerics framework. Supplementary Material S1 illustrates this architecture.

Each module is developed independently, and there are no dependencies between modules, facilitating re-use of code. Therefore each module may be installed independently depending on user requirements, ranging from simple command-line access to SBSINumerics, to the full stack of client application, webserver and back-end task server.

The provision of parallel execution for parameter optimization means that SBSI has the potential to scale to higher-dimensional modelling problems than serial codes such as PottersWheel (Maiwald and Timmer, 2008) and DOTcvpSB (Hirmajer et al., 2009). In addition, unlike these optimizers, SBSI has no dependency on the MATLAB software.

SBSI offers a unique advantage over other parallel suites such as SBML-PET-MPI (Zi, 2011) in that SBSI provides an information privacy infrastructure that obfuscates the models and data that are sent to the HPC back-end. This ensures that the server receives the minimum information necessary to execute the computation, ensuring confidentiality of the model semantics. Communication between modules is secured through protocols such as https and ssh. Authentication details are encrypted using one-way hashing.

2 THE COMPONENTS OF SBSI

SBSINumerics is a C++ library that provides parallelized implementations of global optimization algorithms such as genetic algorithms, particle swarm and simulated annealing. SBSINumerics is optimized for performance: models are converted to C++ and compiled into the library before execution, and parallelization enables substantial speed-up of the search on
Our use of standard data formats such as the Systems Biology Markup Language [SBML, Hucka et al. (2003)] provides a stable basis for plugin development. Plugins can be downloaded and installed from an update site (http://www.sbsi.ed.ac.uk/update) with minimal user input. Plugins developed to date enable access to project-specific databases and editing support for standard data formats such as SBML and SED-ML (Adams et al., 2012; Adams, 2012).

3 CONCLUSIONS

The main focus of SBSI is the task of parameter estimation, which by its nature is computationally intensive. Progress of long-running jobs can be monitored by observing cost function values, and downloading intermediate result files. At the end of a parameter fitting, the user can download the fitted model, together with simulation results generated from the model, and a full log of the explored search space. Since it is beyond our resources to provide computational time for everyone wishing to use SBSI, we have endeavoured to package the software in a form that can be easily installable on other servers. SBSIDispatcher, which acts as a task manager for remotely running jobs, is deployed as a regular Java web application, and only requires configuration of database access in order to be useable on other application servers.

ACKNOWLEDGEMENT

The authors thank Kevin Stratford for consultancy on SBSSystematic.

Funding: UK Biotechnology and Biological Sciences Research Council, and Engineering and Physical Sciences Research Council (to SynthSys, a Centre for Integrative and Systems Biology) (BB/D019621/1).

Conflict of Interest: none declared.

REFERENCES


