Flexible framework for diagnosing alternative model structures through sensitivity and uncertainty analysis

S. Van Hoey[a][b]**, P. Seuntjens**[a][c][d]**, J. van der Kwast**[a]**, J.-L. de Kok**[a]**, G. Engelen**[a]**, I. Nopens**[b]

[a]*Flemish Institute of Technological Research, Environmental Modeling Unit, B-2400 Mol, Belgium*
[b]*Ghent University, BIOMATH, Departement of applied mathematics, Biometrics and Process Control, B-9000 Ghent, Belgium*
[c]*Ghent University, Departement of Soil Management, B-9000 Ghent, Belgium*
[d]*University of Antwerp, Department of Bioscience Engineering, B-2020 Antwerp, Belgium*
Email: *stijn.vanhoey@vito.be*

Abstract: To explore the significance of alternative model structures and their inadequacies, hydrological modeling frameworks that allow quick implementation and comparison of alternative structures were already developed, mostly in a lumped mode. These systems allow testing the suitability of different model components and combining them in a modular fashion. Components can be modified or added if none of the available components fulfils the problem-specific requirements. It is, however, important to highlight that there is no general superior model structure for all spatial resolutions used. So, combining this flexible handling of structural components with varying the spatial resolution is necessary to adapt the model building process to specific conditions of the system, the available data and the objectives of the study.

The next important step is to define effective strategies to diagnose and compare competitive model structures. Only then one can propose model structure improvements. There are several approaches described in the literature that help model diagnosis, like sensitivity analysis, parameter optimisation and uncertainty analysis, but these are mostly used to evaluate one specific structure or compare only a limited number of different structures and are typically not used in conjunction, but rather individually. By changing either specific processes or spatial resolution, while fixing the remainder of the model structure, rigorous testing of the model structure is possible, by addressing the effect of individual model components or spatial resolution.

We present a tool to diagnose alternative model structures and address the effect of individual model components. The tool allows improving the evaluation and selection process of appropriate model structures out of the possible combinations coming from these flexible model structures to ensure the model represents the dominant processes of the system with the required rigour. The presented strategy uses both uncertainty and sensitivity analysis in a Monte-Carlo based framework. Regional sensitivity analysis allows identifying and comparing critical parameters among the different structures for different objective functions. Uncertainty analysis quantifies output uncertainty and parameter identifiability for different likelihood functions among the different structures. Comparing the posterior distribution of the parameters with the initial sampled distribution defines how these are conditioned by the model evaluation process. Since working with flexible structures, analysis can be done on both common and non-common components and associated parameters of the different model structures in a lumped or distributed mode.

Structural components can be changed one at a time or a predefined set of model structures can be compared, using the combination of the above mentioned techniques. Selection criteria are assessed and linked to specific objectives (looking to specific flow regimes, specific objective functions or adapted metrics like flow duration curves). Moreover, it expresses the significant effect of the selected objective function(s) and the importance of using multiple evaluation criteria supporting the research question instead of only trying to reproduce the observed hydrograph.

Keywords: model structure selection, sensitivity analysis, uncertainty analysis, GLUE

1 INTRODUCTION

Different model formulations and spatial conceptualisations can be selected to model hydrological systems. Over the past two to three decades, this resulted in a wide variety of model structures for analyzing the properties and behaviour of such systems. Roughly, this variety is characterised by different levels of spatial detail and process complexity. A trade-off exists between the required level of complexity and resolution, the accepted level of uncertainty, the data-availability and the performance of the overall model. A suboptimal model structure will lead to parameter identifiability problems, increasing uncertainty and significantly reduced model predictive power. As a result, the choice of a model with respect to the appropriate complexity and spatial detail is not straightforward or transparent and is to a large extent driven by the objective of the modeling exercise.

Despite the high variety in complexity in developed models, the majority of hydrological model structures applied in current research are conceptual models, with a fixed structure based on a certain understanding of the dominant processes in the system (Wagener et al., 2001). These conceptual models commonly consist of a number of soil water reservoirs and routing routines representing various runoff processes. As parameter values can not be related to physical or measurable properties, observed data is necessary to inversely identify the parameter values.

Yet, different combinations of parameters and model structures yield similar results in terms of a defined objective function (Beven and Binley, 1992). This lack of identifiability is expressed as 'equifinality', accepting the existence of multiple behavioral parameter sets and model structures. This affects the quantification of the uncertainty, as Krueger et al. (2010) expresses that focusing on a single model structure is likely to result in modeling bias and underestimation of model uncertainty. Identifying the most appropriate model structure for a given problem and quantifying the uncertainty is therefore more than fitting the parameters of a fixed model structure to an observed hydrograph. It is an iterative learning process of rejecting unbehavioural model structures (Fenicia et al., 2008), based on a combination of diagnostics (Gupta et al., 2008). Wagener and Wheater (1999) define an appropriate model structure as a function of the modeling objectives, the characteristics of the hydrological system under investigation and the available data.

Since determining a priori which conceptual model structure is most appropriate for a given situation remains a challenging problem in hydrology (Clark et al., 2008) and recognising that no model structure is suitable for all modeling tasks, using an adaptable and flexible model structure framework in conjunction with a decision framework is preferable. Rather than attempting to find one general model structure capable to perform well over widely differing characteristics (Andréassian et al., 2009; Linsley, 1982; Sten, 1990), it gives the possibility to adapt the model structure to the specific conditions and research questions. Still, the focus is not to generate an extensive number of model structures, but rather to assist the expert in retrieving the information necessary to discriminate between concurrent model structures, each representing a hypothesis of the underlying processes. To take previously mentioned uncertainty and identifiability issues into account, this paper presents a framework to compose a set of lumped or distributed conceptual model structures and evaluate them in a sensitivity and uncertainty framework.

2 FLEXIBLE MODEL STRUCTURE

Modular modeling approaches allow to create environmental models from basic components (Argent, 2005), which makes composing model structures less time-intensive. Modular approaches to design model structures exists to construct hydrological models (Leavesley et al., 2002; Clark et al., 2008; Wagener and Wheater, 1999), but also for ecosystem and ecological modeling (Voinov et al., 2004; Villa, 2001), wastewater simulation (Vanhooren et al., 2003) and general spatial models (Argent, 2005; Wesseling et al., 1996). The construction of these models can be done with an explicit coupling framework connecting components in an user interface (Vanhooren et al., 2003), which increases user accessibility. Although, it is also done by a model language approach (Wesseling et al., 1996; Kraft et al., 2010), where functions and building blocks are represented by coded definitions. The latter approach of using scripting tools got the advantage of being easily extended and at the same time it can be used as a 'glue' to external models or components (Kraft et al., 2010).

In hydrology, the model structures that can be implemented in most flexible model environments like in
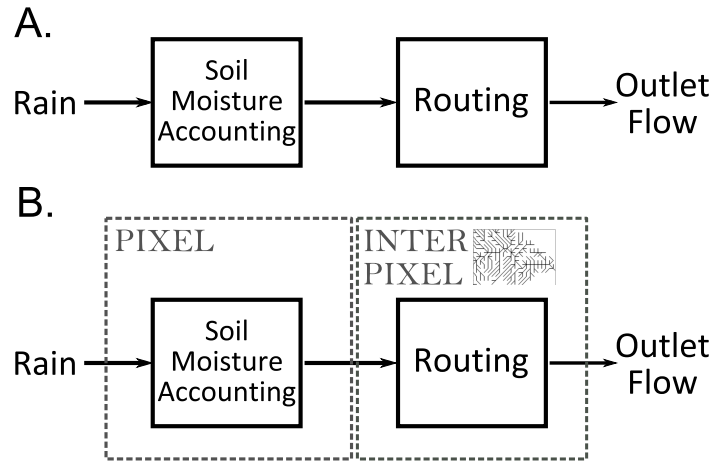
Figure 1. Simplified general outlay of (A.) Lumped hydrological model and (B.) Distributed grid-based hydrological model

Wagener and Wheater (1999) are spatially lumped representations and can be summarized by the combination of a soil moisture accounting module and a routing module (Figure 1 A), where different options can be selected for both parts. Bai et al. (2009) uses a modular modeling structure of three modules: Soil moisture accounting, actual evapotranspiration and routing, with different options for the three components. The Framework for Understanding Structural Errors (FUSE) (Clark et al., 2008) combines modeling options from well-known hydrological models to build up new equally plausible model structures, where the model components can be evaluated in isolation.

When only focusing on the dominant processes representing the system (parsimonious modeling), lumped models are prefered. Nevertheless, distributed models are taking advantage of spatially distributed forcing of the routing and transport processes (Tang et al., 2007) (Figure 1 B) and therefore can be useful, if sufficient data is available, despite their potential for overparameterization. Spatial development of flexible model structures need a computational system that couples and coordinates modules in a simulation together with a GIS tool to perform the spatial analysis within the simulation environment. Wesseling et al. (1996) developed the dynamical modeling language PCRaster, which can be used to construct spatio-temporal models and can be called from the Python programming language, which provides a powerful and flexible environment for creating different model structures.

3   DIAGNOSING MODEL STRUCTURE

Flexible model structures with interchangeable components give the opportunity to easily construct different model structures. To really get the benefit from this approach, effective strategies to diagnose and compare these model structures in terms of performance, uncertainty, identifiability and complexity are necessary. Bai et al. (2009) use an ensemble of 4 signatures, representing different timescales, in a fuzzy evaluation scheme to select model structures and assess the necessary level of complexity. Clark et al. (2008) use the Shuffled Complex Evolution optimization algorithm (SCE) to test if all model stuctures perform equally well if provided with an optimal parameter set. Lee et al. (2005) are using the distance between the optimal values of two objective functions as an evaluation of how a model structure can simultaneously meet two different modeling objectives. The Monte-Carlo Analysis Toolbox (MCAT) (Wagener and Wheater, 1999) includes a number of analysis methods to evaluate the results of Monte Carlo parameter sampling experiments or model optimisation methods. Vache and McDonnell (2006) are using a rejectionist framework to evaluate model structures using residence time data.

In general, methods are mostly based on an exploration of the parameter spaces coming from the different structures. Further, multiple measures giving objective guidance as to whether a selected structure is suitable or not need to be defined (Wagener et al., 2003) to reject inadequate model structures. The

view of rejecting models that are non-behaviourable is also the idea behind the Generalised Likelihood Uncertainty Estimation (GLUE) method introduced by Beven and Binley (1992), based on the Regional Sensitivity Analysis (RSA) Spear and Hornberger (1980) and applied in this framework.

3.1 Regional Sensitivity Analysis (RSA) and Generalised Likelihood Uncertainty Analysis (GLUE)

Regional Sensitivity Analysis (RSA) is a simple, qualitative way to estimate the model sensitivity based on Monte Carlo simulations. Sets of parameters are decomposed in a behavioural group and a non-behavioural group. Frequencies of occurrence of the parameter values are accumulated for both groups of parameter sets separately and plotted against the parameter values. The difference between both distributions reflects the impact of the parameter, defining a sensitive parameter as a parameter introducing a large difference between the distributions. Freer et al. (1996) extended this by dividing the behavioural parameter sets into 10 equally sized groups based on a sorted model performance measure and comparing the cumulative distributions of these ten sampled sub-ranges. To interpret the qualitative sensitivity of the parameter to a specific performance measure, the degree of dispersion of the ten lines represents the sensitivity of the parameter. Figure 2 shows an output for a sensitive (2(a)) and an insensitive parameter (2(b)), represented by the spread of the different distributions.


(a) Sensitive Parameter   (b) Insensitive Parameter
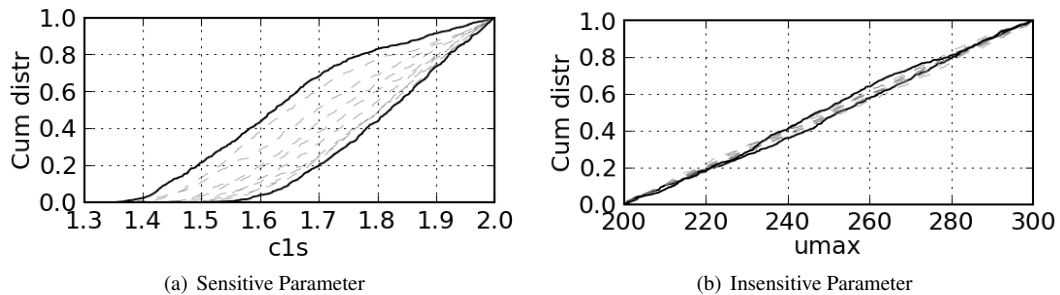
Figure 2. Typical output of a Regional Sensitivity Analysis. A parameter is considered more sensitive towards a model performance measure when the cumulative distribution functions of the sampled sub-ranges vary more.

The Generalised Likelihood Uncertainty Analysis (GLUE) is well described in Beven and Binley (1992) and Beven and Freer (2001) and is responding to the principle of equifinality, saying that different model parameterizations and model structures can give similar good model results. The methodology basically selects behavioural models based on a predefined likelihood measure and uses the output of these behavioural models to assess the uncertainty.

The method is applicable to both model parameterizations and model structures, however most applications in literature are only focusing on differences in parameterizations. In the presented framework, the use of different model structures is integrated in the workflow. Besides the use of combining all these structures to asses the output uncertainty of the ensemble of structure as done by Krueger et al. (2010), information can also be extracted from the rejected and accepted structures and the effect of individual components can be assessed.

3.2 Objective Functions and evaluation metrics

The importance of using multiple criteria is stressed in Gupta et al. (1998), as it better explores the information contained in the data. When just using one objective function, like Nash-sutcliffe efficiency, only a general agreement between the modelled output and the measurements is verified. A combination of different measures of information supporting the research question is preferable. Different objective functions can be combined in a multi-objective calibration approach (Efstratiadis and Koutsoyiannis, 2010),
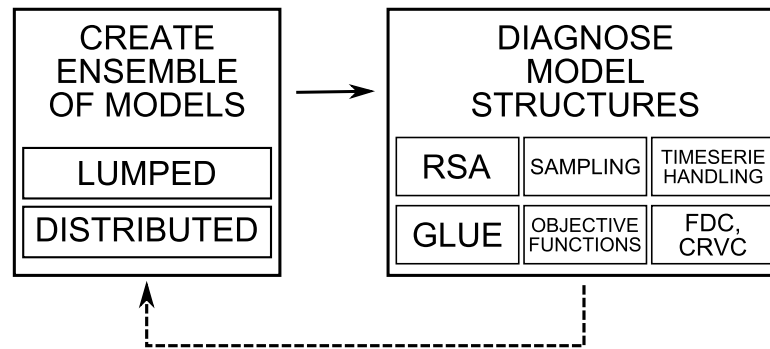
Figure 3. Simplified general schematic of the general model structure diagnosing framework.

but also evaluation criteria coming from auxiliary data (Vache and McDonnell, 2006), identifiability of the parameters (Wagener and Kollat, 2007), temporal clustering (de Vos et al., 2010), evaluation of different periods of the hydrograph (Wagener et al., 2001) or from subflow filtering (Willems, 2009) can be used. One must take into account that effective testing requires data of a sufficient long period to prevent acceptance of inadequate models (Linsley, 1982). While it is accepted that no model structure is suitable for all tasks, the evaluation of model structures is mostly a trade-off and is highly dependent of the research question. The presented framework supports the user in selecting a set of evaluation criteria and verifing the performance of the different model structures to these criteria.

4 FRAMEWORK LAYOUT AND METHODOLOGY

The presented framework implemented in Python programming language provides a set of tools to develop model structures and diagnose structural differences. It consists out of a model structural development section and an analysing section as presented in Figure 3. Both can also be used independently, so analyzing external models as well as extending the framework with other evaluating methods is possible.

An ensemble of model structures can be created by combining predefined coded model components, using PCRasterPython for the spatial component and based on the work of (Wesseling et al., 1996). The Python libraries Numpy/Scipy, Matplotlib, scikits.timeseries are used for numeric calculation, visualisation and timeseries-handling respectively. In lumped mode, a similar approach as in Clark et al. (2008) and Wagener and Wheater (1999) is used with a number of soil moisture and routing components to select from and combined in a way to support the isolated evaluation of the components. In distributed mode, a water balance model is calculated for each cell similar to the lumped mode and routing is represented by its intercell connectivity (Figure 1 B), each with a predefined set of options for the model structure composition.

The diagnoses of the ensemble of model structures is supported by a set of analysis methods and supporting modules. Evaluation metrics can be selected from the set of implemented objective functions or derived from the information given by Flow Duration Curves (FDC) or Cumulative Runoff Volume Curves (CRVC). A combination with the timeseries handling functionality, namely dividing the hydrograph in specific time periods, gives the possibility to apply the objective functions on specific time spans (e.g. seasonal evaluation) to respond to specific research purposes.

The sampling module supports the Monte Carlo based analysis by taking samples for the different parameters of the ensemble of model structures, using a Latin Hypercube sampling or a quasirandom sampling approach. In distributed mode, the sampled values are used to relatively change the distributed parameter values altogether. While the framework brings the possibility of formulating highly parameterized models, it must be taken into account that sampling from high dimension spaces is difficult and defining the sampling distributions should be done carefully. For lower dimensional problems quasirandom sampling is advantageous, but for high-dimensional problems both methods have restrictions (Mauntz and Kucherenko, 2007). The output of the RSA and the GLUE analysis gives information about the sensitiv-

ity, interaction and identifiability of the parameters and about the output uncertainty. Evaluation of the uncertainty for the ensemble of behavioural structures as in Krueger et al. (2010) as well as making an assessment of the effect of changing one specific component on the rest of the model structural parameters is possible by fixing the remainder of the model.

5 CONCLUSIONS

A set of scripted modules and tools is presented to make a rigourous assessment of the relation between model structure and model performance. This flexible model approach responds to the fact that it is unlikely that a single model structure provides the best streamflow simulation for multiple catchments or for the same catchment when exploring different objectives. As such, model identification can be seen as a combined effort of parameterization and structure identification, supporting an iterative approach to modeling, considering the model as a set of hypotheses that evolves progressively, adapting to the catchment characteristics and to the data (Fenicia et al., 2008).

By making use of the scripting language Python instead of a GUI-software package, the framework is flexible in use and easy to extend. The framework provides the possibility of doing a rigorous analysis of the model structural underlying hypotheses instead of an automated way to define the optimal model structure. Although the model components are defined to construct hydrological models, the evaluation methods can be used to investigate other dynamical models. The combination of a modular model approach and a diagnostic environment has got a lot of potential as it supports testing of different research questions. Extending the model components as well as adding new analysing methods is possible and form part of further development.

REFERENCES

Andréassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M.-H. Ramos, and a. Valéry (2009). HESS Opinions: Crash tests for a standardized evaluation of hydrological models. *Hydrology and Earth System Sciences 13*(10), 1757–1764.

Argent, R. M. (2005). A case study of environmental modelling and simulation using transplantable components. *Environmental Modelling & Software 20*, 1514–1523.

Bai, Y., T. Wagener, and P. Reed (2009). A top-down framework for watershed model evaluation and selection under uncertainty. *Environmental Modelling and Software 24*(8), 901–916.

Beven, K. and J. Freer (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology 249*(1-4), 11–29.

Beven, K. J. and A. M. Binley (1992). The Future of Distributed Models: Model Calibration and Uncertainty Prediction,. *Hydrological Processes 6*(3), 279–298.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research 44*, 14.

de Vos, N. J., T. H. M. Rientjes, and H. V. Gupta (2010). Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering. *Hydrological Processes 24*(20), 2840–2850.

Efstratiadis, A. and D. Koutsoyiannis (2010). One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal 55*(1), 58–78.

Fenicia, F., J. J. McDonnell, and H. H. G. Savenije (2008). Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research 44*(6), 13.

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2008). Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research 44*(1), 13.

Freer, J., K. Beven, and B. Ambroise (1996). Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach. *Water Resources Research 32*(7), 2161.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research 34*(4), 751.

Gupta, H. V., T. Wagener, and Y. Q. Liu (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes 22*(18), 3802–3813.

Kraft, P., S. Multsch, K. B. Vaché, H.-G. Frede, and L. Breuer (2010). Using Python as a coupling platform for integrated catchment models. *Advances in Geosciences 27*, 51–56.

Krueger, T., J. Freer, J. N. Quinton, C. J. A. Macleod, G. S. Bilotta, R. E. Brazier, P. Butler, and P. M. Haygarth (2010). Ensemble evaluation of hydrological model hypotheses. *Water Resources Research 46*(7).

Leavesley, G. H., S. L. Markstrom, P. J. Restrepo, and R. J. Viger (2002). A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling. *Hydrological processes 16*, 173–187.

Lee, H., N. McIntyre, H. Wheater, and A. Young (2005). Selection of conceptual models for regionalisation of the rainfall-runoff relationship. *Journal of Hydrology 312*(1-4), 125–147.

Linsley, R. (1982). *Rainfall-runoff models - an overview*. Littetown, CO, USA: Water Resources Publications.

Mauntz, W. and S. Kucherenko (2007). Application of global sensitivity indices for measuring the effectiveness of quasi-Monte Carlo methods. *Monte Carlo Methods and Simulation*.

Spear, R. and G. Hornberger (1980). Eutrophication in Peel Inlet, II. Identification of Critical Uncertainties via Generalised Sensitivity Analysis. *Water Research*, 43–49.

Sten, B. (1990). Principles and Confidence in Hydrological Modelling. *Nordic Hydrology 22*(2), 123–136.

Tang, Y., P. Reed, K. V. Werkhoven, and T. Wagener (2007). Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research 43*(6), 1–14.

Vache, K. B. and J. J. McDonnell (2006). A process-based rejectionist framework for evaluating catchment runoff model structure. *Water Resources Research 42*(2).

Vanhooren, H., J. Meirlaen, A. Y, F. Claeys, H. Vangheluwe, and P. A. Vanrolleghem (2003). Modeling biological wastewater treatment. *Journal of Hydroinformatics 5*, 27–50.

Villa, F. (2001). Integrating modelling architecture: a declarative framework for multi-paradigm, multi-scale ecological modelling. *Ecological Modelling 137*(1), 23–42.

Voinov, A., C. Fitz, R. Boumans, and R. Costanza (2004). Modular ecosystem modeling. *Environmental Modelling & Software 19*, 285–304.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences 5*(1), 13–26.

Wagener, T. and J. Kollat (2007). Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. *Environmental Modelling Software 22*(7), 1021–1033.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes 17*(2), 455–476.

Wagener, T. and H. S. Wheater (1999). A Generic Framework for the Identification of Parsimonious Rainfall-Runoff Models. *Environmental Engineering*, 434–439.

Wesseling, C., G. Karssenberg, W. van Deursen, and P. Burrough (1996). Integrating dynamic environmental models in GIS: the development of a dynamic modelling language. *Transactions in GIS 1*, 40–48.

Willems, P. (2009). A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling & Software 24*(3), 311–321.