Americas Conference on Information Systems (AMCIS)

AMCIS 2009 Proceedings

Association for Information Systems *Year* 2009

A Loosely-Coupled Collaborative Integrated Environmental Modelling Framework

Isaac Wong* David C.-L. Lam†
William G. Booty‡ Phil Fong**

*Environment Canada, isaac.wong@ec.gc.ca
†Environment Canada, david.lan@ec.gc.ca
‡Environment Canada, bill.booty@ec.gc.ca
**Environment Canada, phil.fong@ec.gc.ca

A Loosely-Coupled Collaborative Integrated Environmental Modelling Framework

Isaac Wong
Environment Canada
Isaac.Wong@ec.gc.ca

David C.-L. Lam
Environment Canada
David.Lam@ec.gc.ca

William G. Booty
Environment Canada
Bill.Booty@ec.gc.ca

Phil Fong
Environment Canada
Phil.Fong@ec.gc.ca

ABSTRACT

Integration of environmental models requires full support of the modelling community. When a large number of models are integrated, it requires consistency within scale, datasets, and model to model interactions to minimize the uncertainty among the models. The integrated environmental modelling (IEM) framework is a necessary approach to integrate multiple environmental models for a particular study. When modellers cannot afford considerable amount of time to get involved with full and tightly-integrated IEM or an IEM has very short time frame to complete, then a loosely-coupled collaborative IEM environment can provide the benefits of the integrated approach while minimizing the effort of each individual modeller. However, such a framework will require setting rules that all participants must adhere to. These rules address the issues of model inputs and model to model interaction. The framework should also provide value-added functionality to make the IEM framework more transparent and applicable.

Keywords

integrated modelling, environmental models, collaborative modelling framework

INTRODUCTION

In order to assist decision makers and to support complex environmental issues, environmental scientists and modellers are often asked to investigate and understand the inter-relationship of the ecosystem across multi-media such as air, water and soil, and across multi-disciplines such as the environment and economics. No one model can handle such complex conditions alone and it is a forgone conclusion that integrated environmental modeling (IEM) is required to provide a better understanding of the impact to the ecosystem. The US Environmental Protection Agency (EPA) and the European Union (EU) are working on concepts to make the implementation of IEM much smoother. In this paper, we examine two areas that affect the uncertainty analysis of IEM the most: the data input for each model and the model to model data exchange and interaction. In an ideal world, input data is QA/QC and we have perfect data for the IEM. However, this is not the case in the real world situation. For example, in Canada, the water quality data can be sporadic and the quality of the GIS (Geographic Information System) layers can also be problematic. In particular, the temporal and spatial scaling issues sometimes make an IEM exercise difficult. The boundary between two models within the IEM may have different spatial and temporal scales. The reconciliation to make the models co-exist within the IEM requires careful consideration and more importantly, uncertainty needs to be addressed.

There are many configurations of IEM. The US EPA has implemented the Framework for Risk Analysis of Multi-Media Environmental Systems (FRAMES) (Babendreier and Castleton, 2005) and in that configuration data is well defined and model exchanges are also defined in a prescribed framework. Since the model developers and the model input data are all from within the same organization, the tight integration works nicely. In addition, FRAMES also addresses model uncertainty for the model integration. However, there is a major limitation in the FRAMES configuration: it allows only one way flow of data within the model integration, i.e. the models are linked sequentially in one direction. There is no feedback to prior models in the model chain and two-way calibration of models is not allowed in FRAMES. At the other end of the spectrum, the European Union created an IEM framework called OpenMI (Moore and Tindall, 2005). It allows the model integration among models with relative minor modifications to the original model source codes. However, the OpenMI framework does not handle any modelling uncertainty nor input checking. The onus lies with the modeller since it merely provides a mechanism for linking models together. One other limitation of the OpenMI framework is that it requires the modeller to have access to the source code and in the cases of proprietary or legacy models, this will pose a major obstacle for the OpenMI implementation. IEM is recognized as one of the important tools for science assessment and it has been used in several varieties of projects in Canada. There are both tightly and loosely-coupled IEM; the type used is dependent on the availability of models, data and partnership. An example of a tightly-coupled IEM is the Mackenzie GEWEX Study (MAGS) which is a major Canadian scientific initiative with the goal of closing the water and energy budgets within the Mackenzie River basin. A major component of MAGS was a tightly-coupled IEM that involves coupling atmospheric and hydrological models (Soulis, Kouwen, Pietroniro, Seglenieks, Snelgrove, Pellerin, Shaw and Martz, 2005). The Laboratory Information Management System (LIMS) also moves towards the field of interoperability among different data models to provide integral connections to laboratory data, instruments, analysis and reports (Kasner, 2007).

Although a tightly-coupled IEM is desirable, sometimes it may not be feasible to implement because of the huge effort required. There are many factors that hinder this type of IEM. Firstly, the availability of people time may not be readily available due to other priorities and commitments. The discrepancy in people time and effort will be an issue during the development phase. Secondly, modellers may come from different geographical locations. Even with today's advanced technologies, this may introduce road blocks to development when frequent face to face visits are critical to the success of the development. Thirdly, modellers may not have access to the source code of some models. They are merely users of the models and these models sometimes are proprietary with very little chance to make them available in the open-source environment. Without the access to the source codes, it is extremely difficult for those models to engage in a tightly-coupled IEM framework. However, these difficulties can be overcome by relaxing the tightly-coupled IEM to a form of loosely-coupled collaborative IEM. The resulting IEM can be viewed as a collaboration support system to serve the wide range of interdisciplinary data and models. This system is complex, distributed, open and dynamic in nature but it also requires human commitment and buy-in.

The main characteristics of a loosely-coupled collaborative IEM are fully participated, standards-based, modelling community-based, consensus and transparency. The idea is to share consistent data and model results in a common framework. This approach works best for the case in which scientists work on their own models and data in their own area but intend to collaborate with other research scientists in other areas, and yet are unwilling to spend too much time and effort to tightly integrate all models. One example is the acid rain integrated assessment modelling system (IAM), which was developed to investigate multi-media and multi-disciplinary acid rain assessment. The system integrates various air quality, water quality, wildlife, forestry, fish and economic models and data (Lam, Puckett, Wong, Moran, Fenech, Jefferies, Olson, Whelpdale, McNicol, Mariam and Minns, 1998) to assess the damages of rivers and lakes, forest soil and wildlife due to acid rain. It also provides scenario gaming facilities to answer "what if" questions based on future emission reductions from legislation. IAM further provides an optimal solution to indicate where and how much emissions need to be reduced while minimizing the impact on the economy. Another example is the land-water integration decision support system (DSS) (Wong, Fong, Booty, Neilsen, Benoy and Swayne, 2008). The agricultural industry accounts for eight percent of the gross domestic product (GDP) in Canada. Because of its significant contribution to the economy, its impact to the environment needs to be understood. If we understand how the non-point source pollutants from agricultural practices interact with the ecosystem, we can design a suite of best management practices (BMP) to minimize the impact while keeping the agricultural sector productive. The idea of using multiple non-point source pollutant models that complement the strength of each other is the backbone of the DSS. In this case, the first model, a time-dependent model, predicts a time series of loadings of the pollutants at various focal points in the watershed. The second model, an event-driven model, would use information from the first model to trace the location of the major pollutant sources. Various BMPs are investigated for these locations and the optimal solution such as the length of a buffer strip is estimated based on the integrated model results. More recently, this approach has been extended to study the integration of the watershed-lake interaction for the Lake Winnipeg basin. In this case, various watershed and lake models will be used to assess the eutrophication problem due to excessive nutrients in Lake Winnipeg. The BMP of land use management will affect the phosphorus loadings in the lake. Hence, the integration of watershed-lake interaction is necessary to analyze the complete situation.

ISSUES OF LOOSELY-COUPLED COLLABORATIVE IEM

Upon closer examination of IEM examples, there are several issues we need to address for a loosely-coupled collaborative IEM. Each model is operated by domain experts in their field, i.e. the watershed experts will use selected watershed models and the lake modellers will apply the selected lake models for the assessment. It is a challenge to integrate various models from various sources and personnel. The models may have to integrate together in a certain configuration. Figure 1 shows a simple model network configuration of three models (A, B and C) where model C receives inputs from models A and B. If for some reason the modeller for model A decides to withhold the results, there will be a delay in getting the model A results to model C. Or, if Modeller B is waiting for a key dataset from a data holder in order to run the model, this will introduce holdups. Without the results of Model C, the system fails to provide the necessary information for impact assessment. Therefore, it is important for the IEM to have all key personnel, including modellers and data holders, to “buy-in” to avoid any pitfalls and failure.

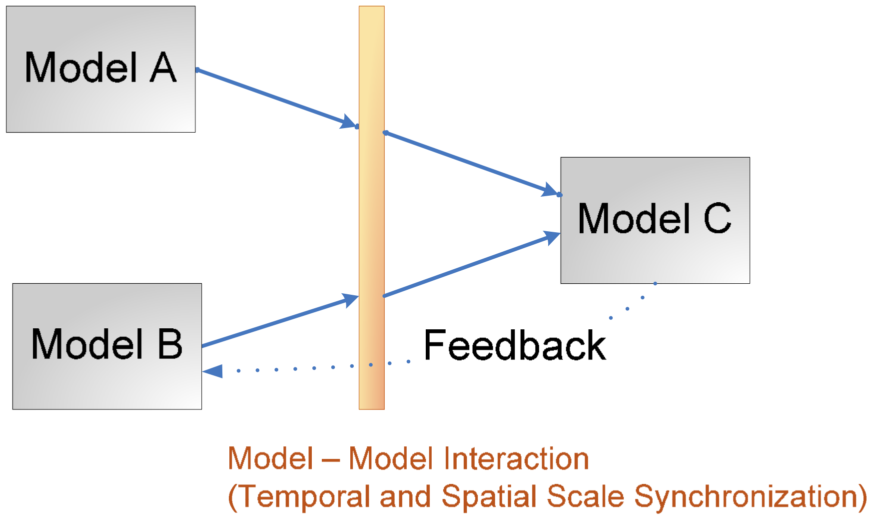

Figure 1. Integrated environmental modelling framework

The modelling data quality is a major issue in a loosely-coupled collaborative IEM. The IEM needs to ensure the absolute best quality of input data is used for all facets of modelling to minimize errors and reduce uncertainties about data, errors in individual model results and measured data, model-model interaction and the current science knowledge. Since it is loosely-coupled, multiple models may need similar input data but data may come from various sources. This may introduce inconsistencies as well as incompatibilities in the IEM. In some cases, data should not be used from certain sources because their laboratory analytical methods are significantly different from others. In other cases, data from various sources should be combined and integrated as a single dataset for modelling purposes. In general, successful calibration and validation of the models requires large amounts of high quality data.

The model to model data exchange is also an issue. The IEM needs to ensure the model results from Model A are indeed the correct inputs for Model C. This includes consistency of the variables in terms of the same units, and temporal and spatial scales.

Another issue is the timeliness of the data. Since modellers run their models from time to time, new model results will be generated periodically. Also, new input data may also become available periodically. With each new update of new model results and data, it requires all stakeholders within the IEM to know and revise their own results. This ripple effect trickles down within the IEM until all the information and results are updated.

To address these issues surrounding the loosely-coupled collaborative IEM, an intelligent modelling data assurance framework is proposed to ensure the quality and minimize uncertainties of the IEM. The uncertainty measurement can be quantified by agreeable statistical approaches and/or qualified by artificial intelligence approaches. In the next section, we will describe the components of this framework to ensure quality performance, i.e. to minimize uncertainties.

THE INTELLIGENT LOOSELY-COUPLED COLLABORATIVE INTEGRATED MODELLING FRAMEWORK

Initial Buy-In

In the beginning, all modellers and data holders are required to work closely together to define the configuration of the system. The IEM developer collects all necessary information about the IEM and proposes a workable configuration of the IEM. This process iterates until every stakeholder in the IEM is completely satisfied and comfortable with the proposed solution so that one is capable of delivering one’s own portion of work within the IEM. This step is critical because it ensures the IEM has the right configuration to deliver what it proposes to deliver. More importantly, the early involvement of all stakeholders and the transparency of the process make it easier to receive the buy-in. In most cases, the success of the IEM depends on human involvement and commitment.

Handling Modelling Input

Getting data at source is becoming more and more the norm for any successful IEM, where the benefits to the IEM of basing inputs on up-to-date accurate data are clearly recognized. Although it is desirable to keep data at source, sometimes it may not be feasible. For example, a key dataset may not be available due to the network security protocol of a participant within his/her organization. Expert rules to handle exceptions should be defined and reinforced at all times. Model to model data exchange can also be handled in the same fashion. A user should be able to receive the best data available. We propose an intelligent hybrid web-based system to facilitate the activities within the IEM framework.

An Intelligent Hybrid Web-based System

A central web-based system is established and serves as an intelligent agent among various data sources and model results. Figure 2 illustrates the key components of this system. The system records the sources of all the inputs using standards-based meta-data such as FGDC (Federal Geographic Data Committee, 1998) or ISO (International Organization for Standardization, 2003) -19115 meta-data standards to make the data more consistent for use among the models. For each data source, it records key meta-data about the dataset. This includes sampling method, laboratory analytical method, time duration, and units of each variable. In addition, it should record the access method and the update frequency. The system also records meta-data on the sampling site. This includes the name and location of the site, contact information and time duration. For spatial data, the system should record the scale, bounding box, the contact information and the access method. Likewise, the system should record the meta-data about the input and output criteria of each model. Specifically, the temporal and spatial scales of the model inputs and outputs must be registered. The model to model exchange at the boundary will need this information. The use of meta-data is important to address the data and model results compatibility issue.

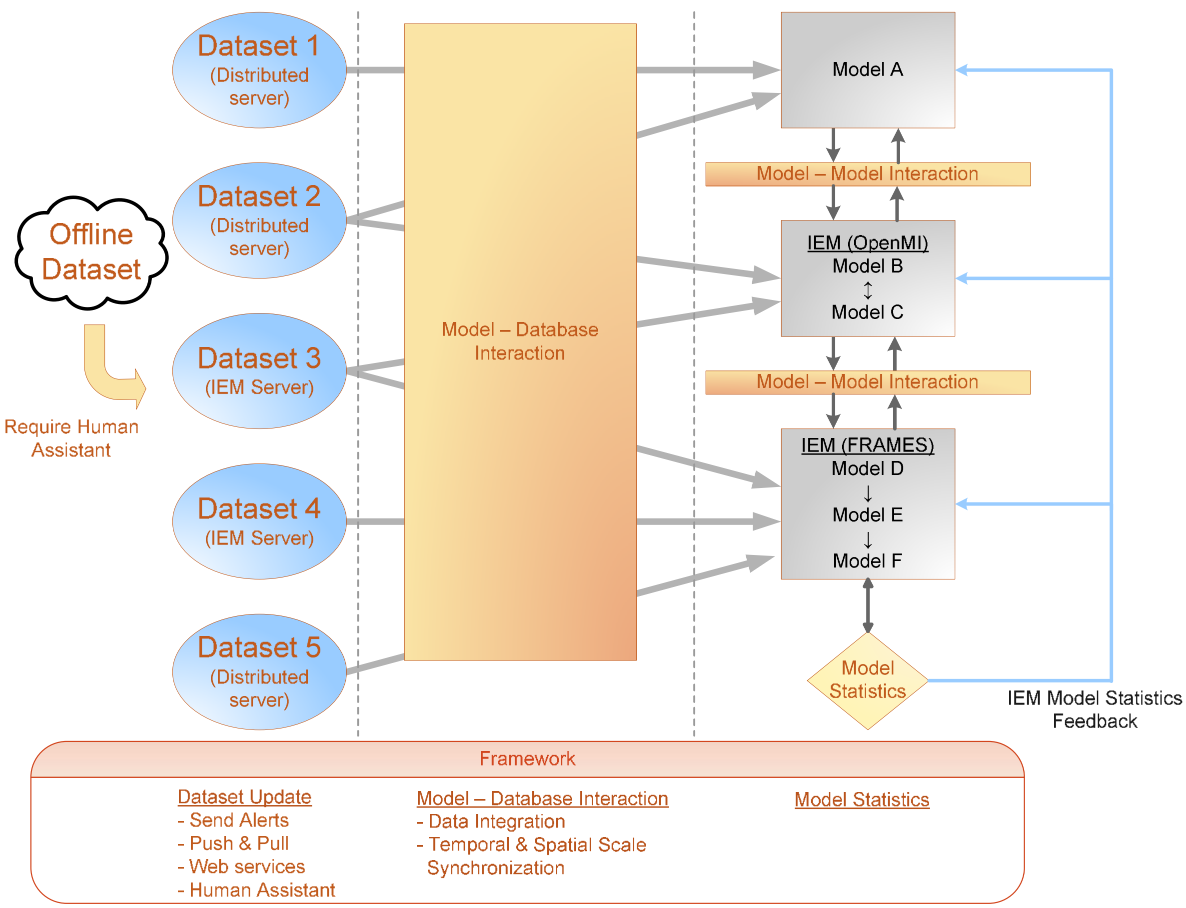

Figure 2. The schematic diagram of the loosely-coupled collaborative IEM framework

IEM Configuration Change

The ability of adding, replacing and deleting models within the IEM is critical to the success of the system. Sometimes a model identified in the initial configuration might turn out to be inappropriate or a new modeller may join and bring in his/her own model. The IEM configuration should be flexible and adopt a modular approach. When this happens, the IEM developer will revise the IEM configuration and requests all stakeholders to review and validate the configuration change.

Value-Added Functionality

The system should provide additional value-added functionality to allow the users a consistent approach to access the data and model results. One functionality is the spatial upscaling and downscaling of model results to a suitable resolution acceptable for other models within the IEM. A significant functionality is the integration of sampling datasets. With the aid of meta-data standards and some prescribed expert rules to combine datasets, the system is able to examine all data sources and make the necessary integration as an option to the users. It is important to note that the system will provide measures of uncertainty arising from the integration of datasets from multiple sources (as opposed to using a single data source). The third functionality is the update alert. If a new set of data or model results is available, the system should actively alert the IEM users to the change of data and models. The users should act on the change as soon as possible. The IEM will go through its update of model results until completion. The overall model statistics of the IEM are provided by the system to measure the goodness of fit. By allowing the users a platform to integrate all the IEM model results and allowing all the stakeholders to review the overall model statistics of the IEM and its uncertainty analysis, it allows for a collaborative modelling environment among modellers. One modeller may decide to re-run a model in the hopes of improving the overall model statistics after reviewing all the results. This will trigger a ripple effect and many models will also be re-run. Since most IEM systems are multi-way calibration, this activity will produce better model statistics for the overall IEM. The intelligent system should keep track of all the history in case the modeller decides to roll back to certain previous states. This happens when the modeller discovers some mistakes in some of the later runs. When this happens, there is no need to re-run the models.

IEM Within IEM

With the added functionality, the intelligent model data assurance system strives to ensure all the model users in the IEM access the same data, to integrate the datasets in a consistent way, to remove the burden of scaling and downloading problems and more importantly, to provide transparency within the IEM. It is important to keep in mind that this system does not actually run the models. It merely provides a platform for data and model results sharing and model statistics enhancement. The actual model run still resides in the hands of each individual modeller. Therefore, it is perfectly fine for a modeller to have two of his own models and set up his own IEM using other technologies such as OpenMI or FRAMES.

Semantic links in IEM

The IEM adopts an approach based on the concept of ontologies to represent the relationships among data and models. The use of ontology coupled with modelling meta-data standards not only improves the quality of the IEM, but also reduces the complexity of handling data and model results. Building a semantic enriched framework based on ontology produces an intelligent agent utilizing the knowledge base to detect any discrepancy or problem within the IEM (Rizzoli, Donatelli, Athanasiadis, Villa, Muetzelfeldt and Huber, 2005).

Push and Pull of Data and Model Results Inquiry

In the ideal world, when a new set of data or model results is available, the system should receive alerts about this. However, there is no guarantee that this would happen. Even if the alert system is in place, the intelligent system should periodically pull data out of the data source to ensure the link remains intact. In the case of offline data, it is essential for the IEM developer to check with the data holder or modellers to assist them to upload and/or download the information. Active management will minimize this discrepancy.

CONCLUSIONS AND RECOMMENDATIONS

When a large suite of environmental models is required to be integrated, a loosely-coupled collaborative IEM environment is more suitable. The advantage is that the modellers do not have to spend a considerable amount of time in IEM but still receive most of the benefits of the IEM, i.e. the transparency and interaction among models. There are also cases that only an IEM can address. These cases tend to be complex environmental issues in nature.

Even though the collaborative IEM framework is loosely-coupled, there are many issues to be addressed. These issues include the update of data and model results; the addition of new data and models; the upscaling and downscaling issues of data and model results and finally the human issue. In addition to addressing these issues, the intelligent system should provide certain value-added functionality to ensure the IEM is consistent throughout. The functionality includes integration of datasets; a platform to review all the model results and overall model statistics; and an alert system to notify the users of any changes in data or model results. The success of the system is dependent on active management and the full support of stakeholders. Also, more formal specifications of the IEM architecture and behaviour using software engineering standards such as UML (Unified Modelling Language) or MDA (Model Driven architecture) will be required to develop a successful and robust system.

Although parallel computing is not in the scope of this IEM, many models within the IEM require parallel computing technology to make it time efficient and thus enhance the IEM.

REFERENCES

1. Babendreier, J. E. and Castleton, K. J. (2005) Investigating uncertainty and sensitivity in integrated, multimedia environmental models: tools for FRAMES-3MRA, *Environmental Modelling and Software*, 20, 1043-1055.
2. Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998), Federal Geographic Data Committee, Washington, D.C.
3. International Organization for Standardization (2003) 19115 Geographic information – metadata, International Organization for Standardization, Geneva, Switzerland.
4. Lam, D. C. L., Puckett, K. J., Wong, I., Moran, M. D., Fenech, G., Jefferies, D. S., Olson, M. P., Whelpdale, D. M., McNicol, D., Mariam, Y. K. G. and Minns, C. K. (1998) An integrated acid rain assessment model for Canada: from source emission to ecological impact, *Water Quality Research Journal of Canada*, 33, 1–17.
5. Kasner, R. (2007). Determining LIMS functionality, cost, and ROI: system architecture strengths and limitations, PharmaTech.com, http://pharmtech.findpharma.com/pharmtech/Article/Determining-LIMS-Functionality-Cost-and-ROI-System/ArticleStandard/Article/detail/469677?ref=25.
6. Moore, R. V. and Tindall, C. I. (2005) An overview of the open modelling interface and environment (the OpenMI), *Environmental Science and Policy*, 8, 279-286.
7. Rizzoli, A.E., Donatelli, M., Athanasiadis, L. Villa, F., Muetzelfeldt, R. and Huber, D. (2005) Semantic links in integrated modelling frameworks, in Zerger, A. and Argent, R.M. (Eds.) *MODSIM 2005 International Congress on Modelling and Simulation*. *Modelling and Simulation Society of Australia and New Zealand*, December 2005, 704 - 710, http://www.mssanz.org.au/modsim05/papers/rizzoli.pdf.
8. Soulis, E. D., Kouwen, N., Pietroniro, A., Seglenieks, F. R., Snelgrove, K. R., Pellerin, P., Shaw, D. W. and Martz, L. W. (2005) A framework for hydrological modelling in MAGS, in Spence, C., Pomeroy, J. W. and Pietroniro, A. (Eds.) *Prediction in Ungauged Basins: Approaches for Canada's Cold Regions, Canadian Water Resources Association*, 119-138.
9. Wong, I., Fong, P., Booty, W. G. Neilsen, C., Benoy, G. and Swayne, D. A. (2008) The land and later integration decision support system, *AMCIS 2008 Proceedings*, August 14 - 17, 2008, Toronto, Canada, Paper 176.