Environmental Modelling & Software 25 (2010) 1640–1649

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Flood forecasting in transboundary catchments using the Open Modeling Interface

F. Fotopoulos*, C. Makropoulos, M.A. Mimikou

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece

A R T I C L E I N F O

Article history:
Received 26 February 2010
Received in revised form
28 June 2010
Accepted 30 June 2010
Available online 3 August 2010

Keywords:
3B42
3B42RT
Early warning
Evros
Flood forecasting
Open modeling interface
OpenMI
Rainfall – runoff
Remote sensing
Satellite

A B S T R A C T

Using satellite data for flood forecasting in catchments located in mid-latitudes is challenging to engineers and model developers, in no small part due to the plethora of data sets that need to be retrieved, combined, calibrated and used for simulation in real time. The differences between the various satellite rainfall data products and the continuous improvement in their quantity and quality render the development of a single software tool, able to read and process all the different data sets, particularly difficult. Even if such an endeavour was undertaken, the degree of flexibility and extensibility that such a tool would require to accommodate future versions of data sets, available in different file formats as well as different temporal and spatial resolution should not be underestimated. This paper describes the development of a flood forecasting system that addresses this issue through a modular architecture based on the use of the Open Modeling Interface (OpenMI) standard, which facilitates the interaction between a number of separate software components. It is suggested that this approach greatly simplifies programming and debugging and eliminates the need to create spatial and temporal transformation functions without significantly compromising the overall execution speed. The approach and system were tested for forecasting flood events within a particularly challenging transboundary catchment, the Evros catchment, extending between Greece, Bulgaria and Turkey. The system uses two sets of data sources, as an example (NASA's TRMM 3B42 and 3B42RT satellite data sets) to forecast flooding in the Evros catchment. Results indicate that OpenMI greatly facilitates the complex interaction of various software components and considerably increases the flexibility and extensibility of the overall system and hence its operational value and sustainability.

Software availability

Name of Software: EFLOOD, SSM, PRCOR, 3B42, 3B42RT, PRUPDATE
Developer: F. Fotopoulos
Contact address: Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece
Direct line: +30 210 651 2487
Fax: +30 210 654 8461
E-mail: fotis@technologismiki.com
Year first available: 2009
Hardware required: PC (2000 MHz or more, 1024 MB of RAM, Windows XP/Vista)
Software required: OpenMI (www.openmi.org)
Program languages: VB.NET 2008 C# 2008
Program size: 20 MB excluding data (data is over 100 GB)
Availability: CD from developer
Cost: free of charge for non-profit and research institutions

1. Introduction

Recent years have seen a considerable increase in the demand, by both water authorities and the public, of reliable flood warning systems (Hayes, 2009). Flood warning systems allow authorities to issue or rescind alert statuses for imminent flood events ranging from a few hours to several days in advance. At the core of the flood warning system, is a flood forecasting procedure, that essentially predicts stream flow using precipitation data and other relevant hydrometeorological parameters using rainfall–runoff models. Since the quality and availability of precipitation data is crucial to the success of this operation, the research community has always been looking for reliable sources for recording, processing and transmitting rainfall measurements to data processing centers, if possible, in real time (Horwood, 2004). Telemetry, and more specifically the use of satellites, although not a new idea, is one that

* Corresponding author. Tel.: +30 210 651 2487; fax: +30 210 654 8461.
E-mail address: fotis@technologismiki.com (F. Fotopoulos).

1364-8152/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2010.06.013

is now reaching operational status and arguably received significant attention (Roth, 2006; Lovejoy et al., 2008).

Currently, there exist various satellite-based rainfall products ranging from near real time to monthly averages for different spatial resolutions (grid sizes). The most popular products are those derived from the measurements of the instruments onboard the Tropical Rainfall Measurement Mission[1] satellite. These measurements are combined with similar products from other satellites thus expanding the spatial coverage beyond the tropics. A main advantage of these products is that they address almost all of the issues that usually accompany ground measurements such as insufficient data because of low density of installed raingauges (Syed et al., 2004), incompatibilities in equipment, accuracy and measurement methodology and most importantly discontinuities and lack of immediate access to data due to political boundaries that divide transboundary catchments (Hossain et al., 2007).

The importance of data is more pronounced by the realization that depending on the rainfall data product used, the same flood forecasting model can yield different results (Krzyzstofowicz, 2001; Kavetski et al., 2006). In this study we address the issue of deciding which rainfall product is best suited for flood warning, using a simplified statistical rainfall–runoff model and by taking advantage of the Open Modeling Interface (OpenMI) (Moore and Tindall, 2005) standard to switch between data products without changing the model nor the simulation architecture.

Of all the satellite-based rainfall products, the combined instrument rain calibration algorithm (Huffman et al., 2007), 3B42, was chosen for this example application. The algorithm utilizes an optimal combination of ground, visual, infrared and microwave data and outputs areal gridded rainfall with grid cell sizes equal to $0.25^\circ \times 0.25^\circ$ with 3 h temporal resolution. Due to its calibration against ground measurements, 3B42 is not available in near real time but has, on average, a two-months time lag. There is however a variation of 3B42, known as 3B42RT, which is identical to the first but lacks validation against ground measurements. Since post-processing with ground data does not take place, 3B42RT is available in near real time with an average time lag between 6 and 9 h. However, some ground stations are also available in near real time, such as the rainfall data sets from the Climate Prediction Center (CPC) (Chen et al., 2008), which are available at a 1 day temporal resolution but much coarser spatial ($0.50^\circ \times 0.50^\circ$) resolution. The possibility of calibrating 3B42RT against such ground records is therefore a desired feature that enhances the operational capabilities of flood forecasting early warning systems. The following sections describe the system architecture, key features and use of a novel, flexible, flood forecasting tool that takes advantage of the data sources mentioned to improve operational capabilities for early warning.

2. System architecture

The main challenge when dealing with diverse data sources used as input to the same system is that usually each data set has its own specifications. These specifications consist, inter alia, of the file format, spatial and temporal resolution and the unit system. Handling all the different data sets and their specification using a single software tool is often quite complex. One could either create separate modules to work with each data source or add complex routines to establish temporal and spatial compatibility between data exchange operations (Lau et al., 1999). Given that data sources are constantly evolving in quantity and quality (smaller resolution, smaller time steps, etc), the amount of different modules required to handle the different data sources and their respective versions would be large and difficult to maintain and debug. As suggested earlier this is very much the case in flood forecasting.

In our approach, instead of creating a single program to deal with the different versions of rainfall data sets, it was decided to break down the model into several smaller components. Each data set is handled by one component with the single purpose to return the rainfall height at a given time and point in space. In particular, each component is responsible for:

1. opening the data file,
2. seeking the appropriate value and
3. returning it to the caller object.

Fig. 1 outlines the component interactions that take place during a flood forecast simulation using as the data source 3B42 or 3B42RT.

The system consists of four different components and two plugins. The main component is called EFLOOD and is the core of the system. It retrieves the areal rainfall estimation for a given time period and location and computes the expected runoff by means of a statistical rainfall–runoff model.

The estimated high discharge values are then compared to two pre-defined values that correspond to the alarm level and to the maximum possible flow rate that can be routed through the cross-sections of Evros river respectively. If these two values are exceeded, then EFLOOD issues either an alarm warning or a flood event. As it is fairly common for the flow rate in Evros river to exceed the alarm level but not the maximum flow rate, a false alarm (an alarm not followed by flood) may occur. False alarms should not be treated as weakness of the system but as real states during which local authorities should be in high alert. If however the system predicts flow rates greater than the maximum and in reality a flood

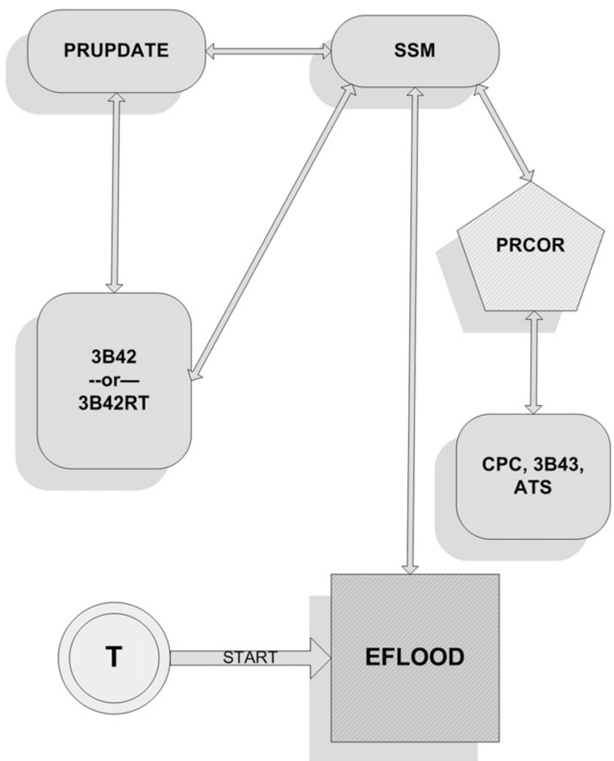

Fig. 1. Computer model interactions flow diagram.

[1] http://trmm.gsfc.nasa.gov/.

event does not occur, then a false flood event occurs which is clearly a weakness of the system.

EFLOOD does not contain any rainfall data; instead it requests rainfall data from a second component, the Satellite Simulation Model (SSM). This is a separate model that calculates the exact position of the satellites (see Fig. 2) used to produce the rainfall data products over time (in our case these satellites are TRMM, Aqua, GOES-E, GOES-W, MTSAT, Meteosat 5 and 7) and initiates the actions for retrieving and processing the areal rainfall estimate from a suitable source. For a discussion on how measurements taken from the aforementioned satellites are used to derive NASA's 3B42 & RT data sets, see Adler et al. (2000). The knowledge of the exact position of the satellites is needed to compare the rainfall products with point ground measurements. In our example application such ground measurements are retrieved from the recently installed Automatic Telemetric Stations (ATS) in the Prefecture of Eastern Thrace and Macedonia in Greece (Fotopoulos and Tsesmelis, 2006b). The ground measurements are then used to validate a (necessarily small) part of the satellites' areal rainfall estimates. The validation process consists of a calculating the difference between the ground measurements and the satellite rainfall estimates. If the two values differ by a predefined amount (i.e. 20%) an alarm is displayed prompting further investigation or action, such as adjusting the bias of satellite precipitation estimation (Boushaki et al., 2009).

Depending on the rainfall data product requested, SSM searches in the external database of that particular product for the value required. This external database is the only interchangeable component of the system. For the 3B42 satellite product, the homonymous component 3B42 is used to search a locally stored copy of 3B42 data files for the appropriate value. If the value required is not found, another component (PRUPDATE) is called to retrieve the value from the Internet. PRUPDATE connects to NASA's servers, retrieves and returns the appropriate value to SSM or if the values cannot be found, it returns a "missing value" code. The reason for maintaining locally part of the data sets is to increase the speed of the simulation. If all values were retrieved remotely, then it would take a significant amount of time for downloading them, as the total volume of data per simulation, exceeds 80 GB. Hence, only newer values that have not been stored locally are retrieved from the Internet.

If SSM succeeds in locating the desired value, it transmits the value to a model used for post-processing the values (PRCOR). PRCOR is able to correct the rainfall height estimate using ground measurements if needed. In the case of 3B42, where calibration has already taken place by NASA's algorithms, PRCOR will return the same value without any correction. If 3B42RT is used, it will combine other recorded values from products such as CPC or local rainfall measurements recorded, for example, by the Automatic Telemetric Stations (ATS) to correct the rainfall estimate.

What is important to note is that both the source data (3B42 or 3B42RT) and the reference data used for correcting the source data (CPC or 3B43) can be replaced at will. One could use for example, 3B41RT (Huffman et al., 2003) as source data and the GPCC data set (Deutscher Wetterdienst, 2009) for corrections, or use different versions of existing data sets at finer temporal or spatial resolutions. Developers can create their own components to provide the system with measured rainfall data from radars, satellites, ground stations or any combination of these, insert it in the simulation scheme and retrieve the results without having to access the source code of any of the other components.

To enable seamless data exchange between the various system components, the Flood Forecasting System presented here uses the OpenMI Standard.

3. The OpenMI Standard

The Open Modeling Interface was developed within the HarmonIT project, in 2001 (Moore, 2001) shortly after the adoption by the European Parliament and Council of the ambitious Water Framework Directive 2000/60/EC (WFD). OpenMI was the IT and Environmental community's response to the key objective of the Directive which was to achieve "good ecological status" of Europe's water bodies by 2015. The point was (and to a large extent still is) that while stakeholders readily subscribe to the aim of the Directive, they have no real means in terms of software tools and models to properly comply to the directive, due to the high level of integrated planning (and consequently integrated environmental assessment) required (Ireson et al., 2006). This integration, which is a key aspect of the WFD, calls for combining ecology, hydrology, hydraulics, chemistry, geology, social and economic sciences into one tool (see, for example, Barthel et al., 2008) which can lead to an Integrated River Basin Management Plan, necessary for monitoring

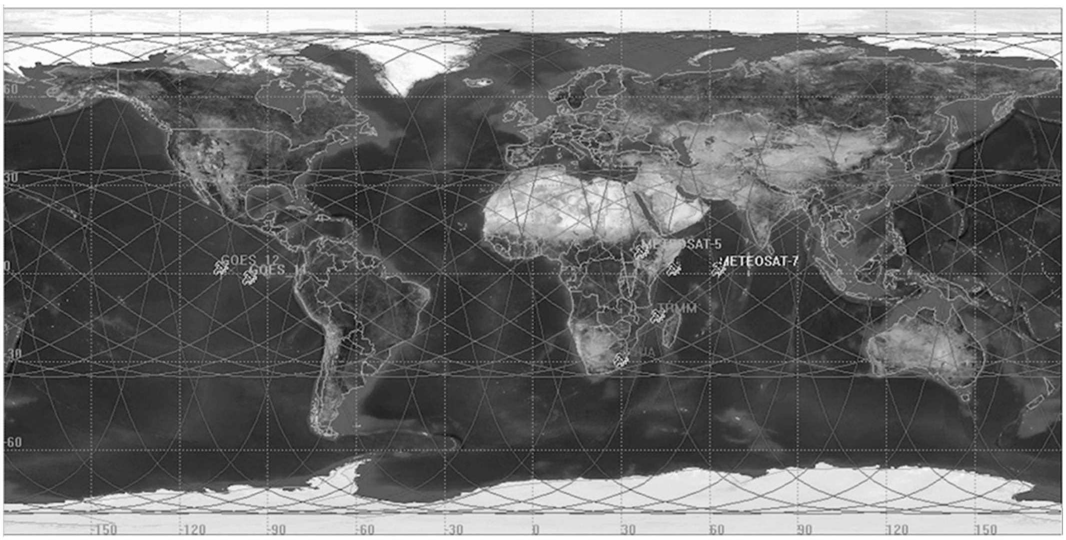

Fig. 2. Constellation orbits in 24 h.

and ultimately achieving "good ecological status". Furthermore, a Directive on Flood Risk came into force in November 2007. The 2007/60/EC directive aims to reduce and manage the risks that flooding poses to human health, the environment, cultural heritage and economic activity. It requires Member States to establish flood risk management plans focused on prevention, protection and preparedness by 2015. This adds a new level of complexity to the requirements of integrated modeling and makes the need for standards facilitating modeling linking all the more pressing.

Instead of taking the route of building and constantly upgrading/customizing complex integrated models, OpenMI offers the possibility to combine slightly modified versions of existing models (including commercial and academic models and tools (Makropoulos et al., 2009)) and achieve the same if not better results. The OpenMI Standard dictates the way models can be linked to other models and exchange information in real (run) time (Gregersen et al., 2007), i.e. without using external files. In simple terms, if a model is *compliant* with the OpenMI standard, it can be linked to other compliant models, thus forming an "ensemble-model", capable of simulating complex physical processes and the interactions that take place between them, even if none of the connected models can simulate them by itself. It is also possible to replace one model from this ensemble with another compliant model that offers improved simulation of certain physical processes or that has more functions (Argent, 2005). This addresses the issue of having to constantly re-code popular programs in order to link them: these programs consist of millions of code lines, they are written in different programming languages, utilize a variety of visualization techniques and are often impossible to modify due to incompatibilities between programming languages (Gregersen et al., 2005; Goodall et al., 2008).

To facilitate model migration to OpenMI, a wrapper was introduced (Sinding et al., 2005), which is essentially a collection of functions that handle the data exchange between the model's computational engine and other computational engines. The wrapper must be configured specifically for each model and programmed in a modern programming language. Using wrappers simplifies model migration since the model's computational engine remains intact while the necessary functions for implementing the migration remain external to it. To facilitate this process the model's computational engine must not be coupled with a graphical user interface. If it is, then one must separate the user interface from the computational core and create a dynamic link library (DLL) containing exclusively the model's engine.

Therefore, the two mandatory rules that every developer has to follow when migrating or programming a new model, in order to comply with the OpenMI Standard is to separate the graphical user interface from the computational core and to include the computational core in a DLL that exposes its methods to the Operating System. The exposed methods must be at a minimum those needed for data exchange. For example, a water budget model can return the computed discharge at a given time by using the function GetRainfall(*time) but it could also use GetEvaporation(*time) and SetEvaporation(*time), to let the end user override the built-in evaporation routine and replace it with another. The more exposed functions are offered by a DLL, the more flexibility the end user (modeler) can exploit.

4. Implementation of OpenMI in the flood forecasting system

The implementation of the flow diagram illustrated in Fig. 1 was performed via the OpenMI Configuration Editor v1.4.0 (OCE). The two different scenarios proposed are shown separately using the OCE in Fig. 3 (3B42) and 4 (3B42RT). The end user can drag and drop new (compliant) models on the Editor and re-arrange them using the interface to achieve a clear linking picture.

All connections have a specific direction from the source component (the component asking for something e.g. the value of rainfall at some specified location at a specified time step) to the target component (the component that will reply to the demand). An additional module, termed Trigger (shown as Oatc.OpenMI.-Gui.Trigger in Figs. 3 and 4) should also be added to act as the "run button" for the integrated model.

In the case of the Flood forecasting system developed here, the objective of the trigger is to initialize the simulation by asking EFLOOD for a result, without really knowing what kind of result is expected. EFLOOD starts a timed loop asking at each time step SSM for satellite rainfall data. SSM calculates the exact position of the TRMM satellites for each time step, to determine the value of rainfall measured by the ATS ground stations which will be used to validate the data retrieved from the satellite products. However, SSM cannot return the satellite rainfall data by itself as it has no knowledge of how to access and process the satellite data sets. So in turn, depending on the configuration, calls the appropriate satellite rainfall database (3B42 or 3B42RT) requesting the data values. The database searches locally for the values that correspond to the specific time step and if found, they are sent to SSM. If not, then the database calls PRUPDATE, which connects to the data provider's servers and tries to retrieve the value remotely. If successful, the remotely retrieved value is returned to SSM and if not, an error code is returned, ordering the system to hibernate for a predefined amount of time (i.e. 30 min) after which it will try again to obtain the missing value. When the locally or remotely obtained value has returned to SSM, it is bundled with several flags. Flags are integer numbers that can be used to trace back the value to its origin (local or remote) and dictate how the system must process the value prior to using it. If the processing value is set to true, the SSM sends the value to PRCOR, a separate component designed to calibrate satellite data products against ground measurements. PRCOR then uses near real-time ground data sets (such as CPC or ATS) to calibrate the input value which is then sent back to SSM. Finally, SSM returns the corrected value to EFLOOD and the latter uses to perform the rainfall-runoff simulation by accumulating rainfall and using historical pairs of accumulated rainfall and observed runoff to estimate the discharge at the points of interest.

Since data exchanged have no geo-location information, an ID based exchange is applied. ID based connections are established using string references. For example, when SSM asks 3B42 for rainfall data, it does so using two parameters, time and location. While time is a double precision number representing Julian Centuries, location is a unique string identifying the subcatchment's outlet. Each outlet is a collection of grid cells which are combined together in order to compute the total rainfall height over the subcatchment. Connection between different components is achieved by selecting the same outlet. An example of connecting CPC and PRCOR components is given in Fig. 5. In this figure, four outlets of the Case Study area are available (Ardas, Delta, Evros and Tundza). By ticking the "id delta" in both models a link is established specifying which catchment's CPC areal rainfall will be used and which catchment's ground parameters will be used by PRCOR to correct inbound satellite data.

5. The case study area

The catchment of the Evros River has been selected as a case study for the flood forecasting system. Evros is one of the most important transboundary basins in the Balkans, because of its great impact on the economies of Bulgaria, Greece and Turkey and the frequent occurrence of significant flood incidents. It should be

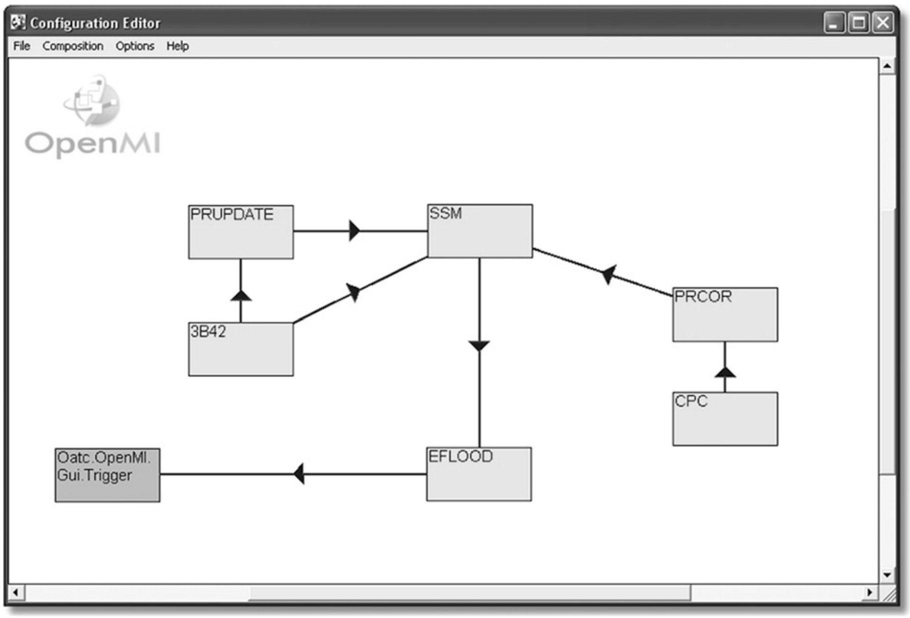

Fig. 3. Model connections required for 3B42 simulation.

noted that due to the large size of the basin, no flash floods occur, hence the proposed flood forecasting system (which could not have predicted flash floods) is appropriate.

The Evros Basin is located in southeastern Europe and is shared by Bulgaria, Greece and Turkey, between north latitudes 41 and 43 and east longitudes 24 and 28. The basin borders to the north with Danube, to the east and south with the Black, Marmara and Aegean Seas and to the west with River Nestos. The river has its source in the mountains of Bulgaria to the east of Sofia and flows east and southeast to the city of Edirne in Turkey. Here, the river changes course abruptly, flows generally southward for a distance of approximately 215 km and empties into the Aegean Sea near Enez, Turkey. Some 20 km to the northwest of Edirne, the river crosses the Bulgarian border and for the remainder of its course forms a physical boundary between Greece and Turkey. Since 1923 it also acts as the political boundary between the two countries, however the riverbed shifted in the past 90 years and its deviation from the position it had in 1923 is non-negligible. At the delta the drainage area of the basin is 53,000 km^2 of which about 66% or 34,980 km^2 is in Bulgaria, about 27.5% or 14,575 km^2 is in Turkey and the remaining 6.5% or 3445 km^2 is in Greece (Fotopoulos and Tsesmelis, 2006a).

Within Greece and Turkey four major tributaries join the main stream (see Fig. 6): the Ardas from the west and the Tundza from the north at Edirne, the Erithropotamos from the west at

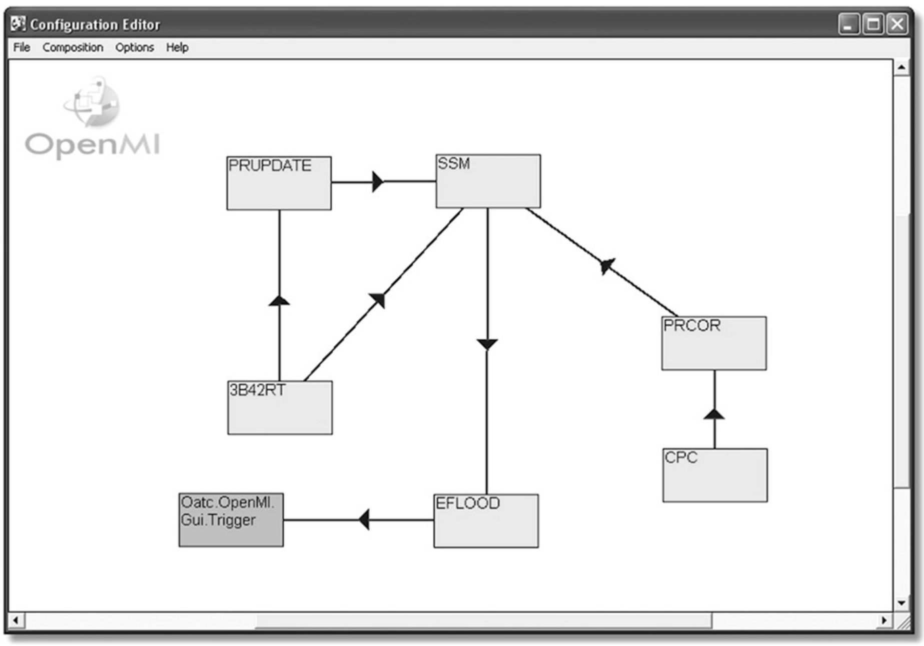

Fig. 4. Model connections required for 3B42RT simulation.

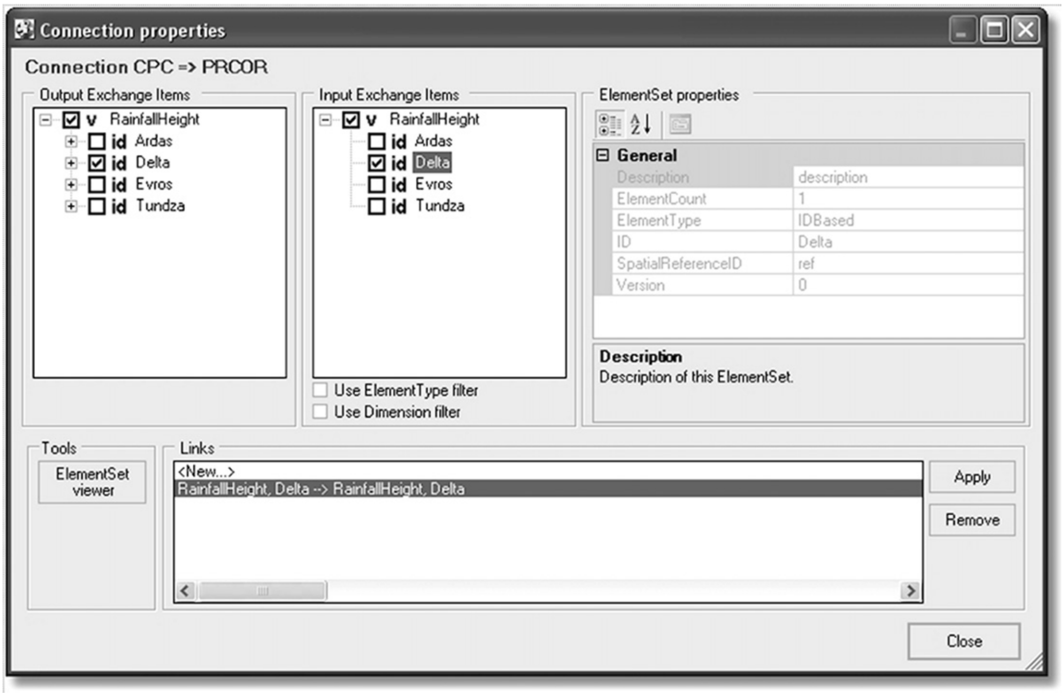

Fig. 5. ID Based connection between PRCOR and CPC.

Didimotichon and the Ergene from the east near Balabancik. The relative size of each of these tributaries together with the extent of each of the drainage areas within the three countries is broken down in Table 1. The flood plain extends along both banks of the river from the Bulgarian border to the sea. Of the 82,000 ha of land in the plain about half (40,000 ha) lies in Greece and the remaining half (40,000 ha) in Turkey (Harza Engineering Company, 1953).

During the past 10 years, five significant flood events took place (Brackenridge et al., 2008). For each event, begin and end date, the affected areas in the Greek side, the severity class and the magnitude are given in Table 2. The severity class is an integer number from 1 to 3 depending of the return period of the event. A flood event occurring at a return period (T) of less than 20 years has a severity class equal to 1. For T greater than 100 years the severity

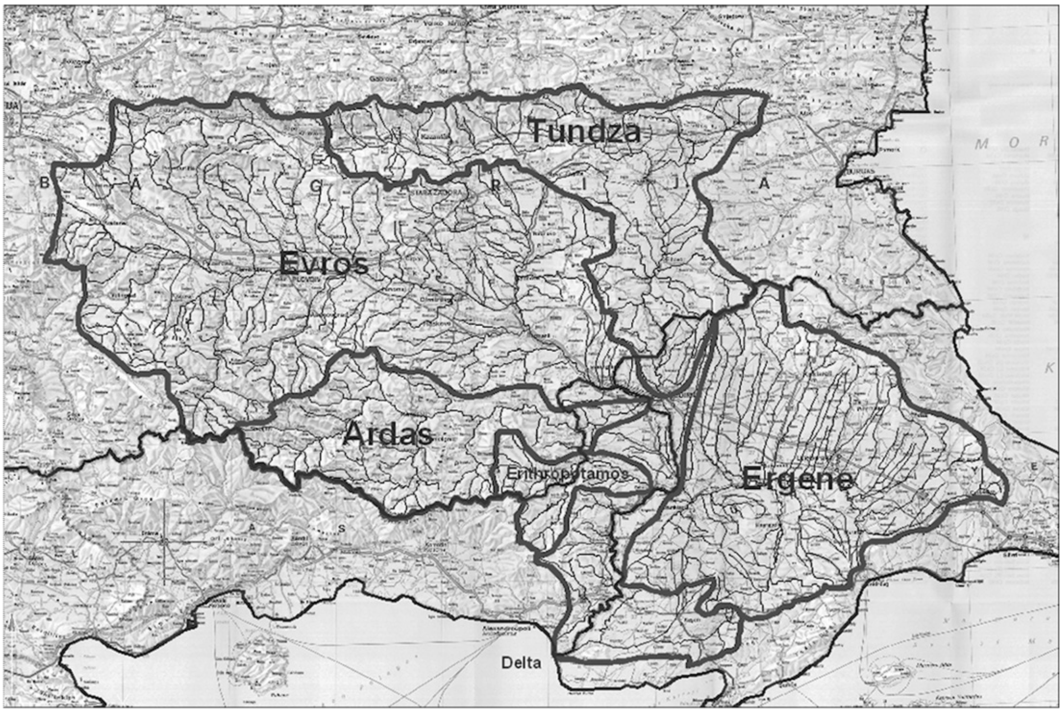

Fig. 6. Evros catchment and major tributaries.

Table 1
Drainage area for each tributary per country (km^2).

Drainage area (km^2)				
Tributary	Bulgaria	Greece	Turkey	Total
Ardas	5250	350	–	5600
Tundza	7790	–	710	8500
Erithropotamos	670	830	–	1500
Ergene	–	–	11,000	11,000

class is equal to 3 and for T less than 100 years and greater than 20 is equal to 2. The flood magnitude is computed using the following equation (1):

$$M = \frac{d \cdot s \cdot \sqrt{A}}{100} \tag{1}$$

Where: M is the flood event magnitude (dimensionless), d the duration of flood event in days, s, the severity class (dimensionless) and A the flooded area in km^2.

The Greek and Turkish authorities have empirically established an alarm system based on the flow depths at Kipoi Bridge and at the monitoring station located downstream of Edirne respectively. These flow depths can be replaced by flow rates when the riverbed geometry at the two locations is considered. It has been found, that the alarm flow rates are equal to 1150 m^3/s and 800 m^3/s for Kipoi and Edirne respectively (Fotopoulos and Tsesmelis, 2006b). These values correspond to approximately 85% of the total capacity of the cross-sections at the two locations. The aim of the flood forecasting simulation is to successfully forecast the exceedance of these two alarm flow rates.

In order to achieve such a prognosis, a statistical rainfall-runoff model was created using historical values of rainfall and recorded flow rates downstream of Edirne. The model accepts daily areal rainfall and daily flow rates. First, it is calibrated using 85% of all available historical data. The historical flow rates are converted to daily volumes. The daily volumes and the equivalent daily areal rainfall heights are accumulated over a variable number of days, from 1 to 180 (approximately 6 months), therefore creating 180 time series. The pairs of accumulated rainfall and volume undergo a statistical analysis, to assess the correlation between the two for each one of the 180 time series separately. The remaining 15% of the historical data are reserved to validate the results of the analysis. In Fig. 7 the results of the analysis using rainfall from CPC data sets are presented. In both cases, Edirne and Kipoi bridge, the optimal number of days is 121 and 120 days respectively (approximately 4 months) and the correlation coefficients are 0.581 and 0.597 using 85% of the historical data and 0.559 and 0.588 using the remaining 15% of the historical data. Then the same analysis was repeated using all of the historical data, yielding practically the same correlation coefficients (0.577 and 0.595). The exact same procedure was applied for rainfall derived from 3B42, 3B42RT and 3B42RTc data sets.

Table 2
Major flood events recorded from 2000 to 2009.

Begin date	End date	Affected areas	Severity class	Magnitude
17.01.03	03.03.03	Riverside areas	1	7.4
17.02.05	24.03.05	Pithio, Sofiko, Didimotichon, Lavara, Poros	2	6.3
02.01.06	20.01.06	Lavara, Kissario, Amorio, Tichero Thimaria, Psathades, Pithio, Trigono	1	2.0
09.03.06	25.03.06	Soufli, Tichero	1	4.0
16.11.07	02.12.07	Sofiko, Thourio, Pithio, Petrades Didimotichon, N. Vyssa	1	6.0

An alarm is issued when the difference in predicted total volumes for two consecutive days exceeds the volumes that correspond to the alarm flow rates (1150 m^3/s at Kipoi bridge and 800 m^3/s at the monitoring station downstream Edirne). If an alarm is issued, it does not necessary follow that a flood event will occur, as depending on the rainfall, the flow depths can gradually decline and drop below the alarm level without ever exceeding the maximum capacity along the river. In this case, a “false alarm” is triggered. However, if maximum capacity is exceeded, then the number of days between the first issue date and the actual occurrence of the flood event are defined as the “warning days”. The number of warning days may be negative if the actual occurrence of the flood event predates the first issue date. An appropriate warning would give several positive warning days to the public, otherwise it would lose much of its value towards civil protection. Clearly, negative warning days are completely useless for operational purposes. If on the other hand maximum capacity is not exceeded contrary to the system’s prediction, then a “false flood event” occurs.

The calibration procedure aims to identify the optimal number of days one has to accumulate in order to minimize the number “false flood events” and maximize the number of successfully predicted flood events.

Data is scarce for the Evros catchment mainly due to political issues. Until recently, only Greece was a member of the European Union and all relevant Directives and policies, including the WFD, could only be applied to the Greek part of the River. Bulgaria has recently joined the European Union and soon the unilateral management of the water resources for Evros is bound to change. Turkey however is still not a member of the European Union and hence EU Directives can only be observed on voluntary basis by the Turkish water resources authorities. Even if one could bypass the (complex) political issues and collect data from all three countries, one would find that the ground network’s density varies significantly from place to place, as does the measuring methodology, accuracy and temporal resolution. Combining rainfall measurements from the three countries for the purposes of flood forecasting would be an extremely hard task with uncertain results and certainly not possible in real time.

This lack of rainfall data led to the adoption of a satellite measurements solution for the flood forecasting system discussed here. NASA TRMM data sets are published in real time (with a lag of no more than 9 h) and cover the whole Evros catchment. For the purposes of this study, two data sets were chosen, 3B42RT and 3B42. The first set, 3B42RT (2002–2008), is a real time 3-h areal gridded precipitation product, given over a grid of 0.25° × 0.25°, while the second (2000–2008) is similar to the first with the addition of post calibration using several ground stations. Since post-processing takes place, this set is published at a later time (usually 1–2 months after the measurements take place) and cannot be used in real time.

There are two important issues with the aforementioned precipitation data sets that need to be taken into serious consideration: In mid-latitudes, where light rain (<1 mm/h) is responsible for 85% of the total rainfall height and very light rain (<0.1 mm/h) accounts for 35% of the total height, satellite estimated precipitation tends to significantly underestimate the real precipitation height (Kidd, 2007). Moreover the instruments onboard the satellites, the precipitation radar and the passive microwave imager, cannot separate rainfall from snowfall. Although these issues are not a problem in the tropics where the satellite areal rainfall estimates can be treated as very accurate, in mid-latitudes they severely diminish the accuracy of the estimates. And while the processing algorithms used by NASA try to compensate for these problems, validating the satellite data sets prior to using them is

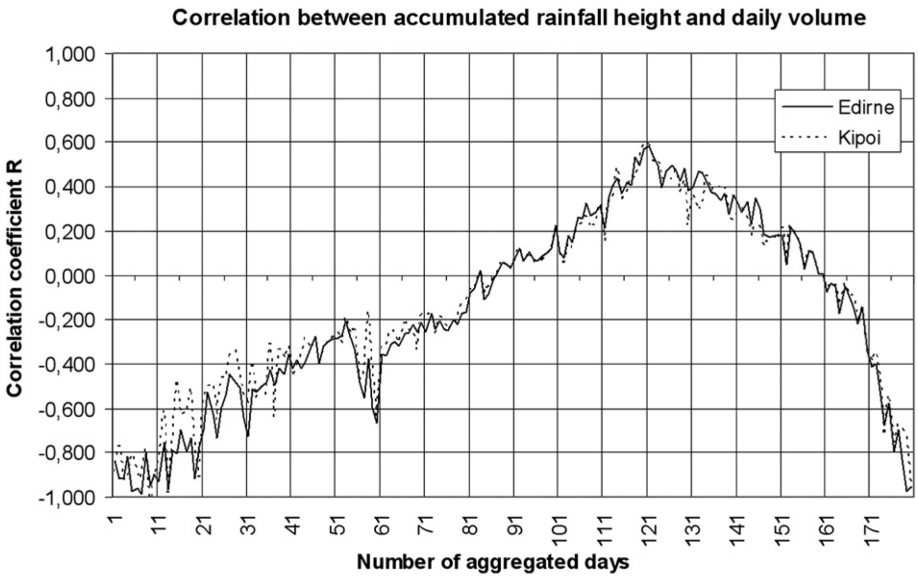

Fig. 7. Impact of number of aggregated days on correlation coefficient between accumulated rainfall height and daily volume (CPC).

a necessary step to account for any systematic bias. To validate the satellite data sets over the Evros catchment, the Climate Prediction Center's (CPC) areal rainfall estimations were used. The CPC grid cells have a spatial resolution of 0.50°×°0.50° and the areal estimates are computed on a daily basis. The CPC areal data sets were also used in combination with level measurements in Edirne and Kipoi bridge to calibrate the rainfall–runoff model.

6. Results

Tables 3–5 summarize the simulation results for flood forecasting downstream of Edirne and at the river's delta, using four different rainfall data sets: CPC, 3B42, 3B42RT and corrected 3B42RT (3B42RTc). The results obtained using the CPC's areal rainfall estimates compiled by ground observations only, were found to predict all actual flood events at both locations, with great accuracy. Reduced but acceptable prediction is obtained when the 3B42 data set is used. However the same does not hold true for the 3B42RT data whose use results in a disappointing simulation, which not only fails to predict 2 out of 5 events, but when it does predict a flood event, it fails to issue an appropriate warning. On the other hand, the performance of the corrected 3B42RTc data set is significantly enhanced. The simulation using this dataset predicts all five flood events (as does the simulation using CPC and 3B42), and gives fairly accurate warning, even during the '07 flood, where both CPC and 3B42 failed to issue an alarm. It should be noted that in the Evros basin, floods last for several weeks.

The interactions that took place using OpenMI to enable the use of the system are more complex than those of other applications of the standard (e.g. Makropoulos et al., 2009; Safiolea et al., 2009) that usually involve up to 3 components. Using several separate software components instead of a single large component, such as the ones used in this work, makes it easier to locate and debug the source code. To ensure the correctness of the simulation, the components were merged with the 3B42 database and a single standalone software program, without the use of OpenMI, was created for benchmarking, comparison purposes and to provide proof that the whole system behaves as expected. The results obtained from the standalone program were then compared to those derived from the OpenMI linked application to ensure that the links between the models work properly.

A series of benchmark tests were also conducted to assess potential reduction in performance when using a fragmented

Table 3
System performance depending on rainfall source (values in days).

Data set [number of aggregated days]	Successful predictions	False alarms	Missed floods	False flood events
Edirne				
CPC [121]	3092	258	18	125
3B42 [93]	3047	192	46	106
3B42RT [144]	2276	565	52	369
3B42RTc [156]	2383	72	49	15
Kipoi bridge				
CPC [120]	3099	334	8	98
3B42 [112]	2896	336	34	199
3B42RT [140]	2263	195	52	108
3B42RTc [156]	2405	50	49	9

Table 4
Flood forecasting results downstream Edirne.

No	Flood (actual dates)		Data set	Alarm start	Alarm end	Warning (days)
	Start	End				
1	17.01.03	03.03.03	CPC	29.09.02	06.03.03	111
			3B42	30.08.02	05.03.03	140
			3B42RT	25.01.03	07.04.03	−8
			3B42RTc	27.10.03	31.03.03	82
2	17.02.05	24.03.05	CPC	04.02.05	30.04.05	13
			3B42	15.02.05	30.04.05	2
			3B42RT	30.12.04	29.03.05	49
			3B42RTc	14.02.05	13.04.05	3
3	02.01.06	20.01.06	CPC	18.12.05	29.01.06	15
			3B42	05.01.06	13.01.06	−2
			3B42RT	03.12.05	04.03.06	31
			3B42RTc	11.12.05	09.01.06	23
4	09.03.06	25.03.06	CPC	24.02.06	16.04.06	13
			3B42	06.03.06	27.05.06	4
			3B42RT	–	–	–
			3B42RTc	15.02.06	20.04.06	23
5	16.11.07	02.12.07	CPC	17.11.07	17.02.08	−1
			3B42	17.11.06	07.01.08	−1
			3B42RT	–	–	–
			3B42RTc	03.11.07	09.01.08	13

Table 5
Flood forecasting results at Evros Delta.

No	Flood (actual dates)		Data set	Alarm start	Alarm end	Warning (days)
	Start	End				
1	17.01.03	03.03.03	CPC	30.09.02	07.03.03	110
			3B42	23.09.02	03.03.03	116
			3B42RT	24.01.03	25.02.03	−7
			3B42RTc	29.09.02	08.04.03	110
2	17.02.05	24.03.05	CPC	04.02.05	16.06.05	13
			3B42	03.02.05	21.06.05	14
			3B42RT	02.01.05	14.03.05	46
			3B42RTc	11.02.05	22.04.05	6
3	02.01.06	20.01.06	CPC	27.11.05	30.01.06	36
			3B42	23.11.05	15.01.06	40
			3B42RT	28.11.05	11.02.06	35
			3B42RTc	14.12.05	11.01.06	19
4	09.03.06	25.03.06	CPC	24.02.06	26.04.06	13
			3B42	24.02.06	08.07.06	13
			3B42RT	–	–	–
			3B42RTc	04.03.06	22.04.06	5
5	16.11.07	02.12.07	CPC	17.11.07	19.02.08	−1
			3B42	16.11.07	17.02.08	0
			3B42RT	–	–	–
			3B42RTc	22.10.07	09.02.08	25

system linked via OpenMI instead of an equivalent standalone program. Three tasks were devised and timed using hardcoded timers embedded in the programs' source codes. Due to the way Windows XP handle physical and virtual memory, it was found that the second time any of the settings were timed, the results were different. This difference was attributed (partially) to the fact that dynamic link libraries are memory resident and persistent. Therefore, the following benchmarking scheme was adopted:

- A notebook running Windows XP, Intel Core 2 Duo T6400 2.00 GHz processor with 4 GB Ram and with no other applications installed was used.
- OpenMI configuration editor v1.4.0 was installed.
- All primary source data (either satellite or ground rainfall data sets) were stored in an external USB hard disk drive with 1 TB capacity.
- Each time a single benchmark was concluded the system was restarted.
- To account for unexpected actions that occasionally take place during a Windows session that can interfere with the benchmark (such as random hard disk drive activity), three identical measurements were taken for each task. The final outcome reported here is the average of the three measurements.

Table 6 presents the benchmark results with an accuracy of 1/10 of a second. In the fourth column of the table, the difference in seconds for the conclusion of each task is shown. The negative sign indicates that OpenMI takes longer than the standalone program for the same task. An interesting finding is that while for simulating a single year the difference in speed is around 14.7%, this number does not remain constant but gradually decreases. Thus, for the whole period (slightly more than 9 years), the difference in performance is down to 14.36%. This could be attributed to the fact that more memory is consumed when all data is stored in a single program than when the same amount of data is stored in different programs.

Table 6
OpenMI vs. Standalone Program benchmark test.

Task	OpenMI enabled ensemble model	Standalone model	Difference
1 year simulation	63.2	55.1	−8.1
5 years simulation	290.6	253.6	−37.0
Complete simulation	575.7	503.4	−72.3

7. Conclusions

In this study, the use of OpenMI was demonstrated facilitating the building of a component-based flood forecasting system, using a variety of rainfall data sets, primarily based on satellite products. The system comprises of a series of OpenMI compliant components that are then linked to perform repetitive complex tasks supporting the forecasting of flood events. The system was tested in forecasting recorded flood events in the Evros transboundary catchment, which would not have been possible with more “traditional” flood forecasting systems due to issues with data availability and reliability. It was found that while more reliable, corrected data sets (such as the 3B42 data and the CPC data) performed adequately, these were not available at near real time. To overcome this obstacle 3B42RT data, that are published in near real time, were used, but had to be corrected, within the proposed system using CPC ground measurements. The corrected data set, 3B42RTc, performed even better than 3B42 in predicting the historical flood events. It is concluded that utilizing OpenMI was particularly helpful in developing and testing the system, since it allowed for a seamless change of the data source provider and a plug and play functionality for the core system components. Instead of having a single large component to perform all necessary tasks, several smaller ones were programmed, making it easier to test and debug the source code. Using wrappers to achieve compatibility with OpenMI allowed the main source code to remain intact as all functions responsible for data exchange are external. The only drawback identified while using OpenMI was a slight sacrifice in terms of performance as indicated by benchmark tests that were undertaken. It is suggested however, that given the constant improvement in computer speed and memory capacity and the efficiency of their operating systems, this difference in performance could soon be negligible.

On the other hand the benefits of OpenMI to integrated modeling, are substantial, not only for linking different simulation models between them, but, as demonstrated in this work, also for linking models with different data sources, both locally and over the internet. This approach provides an example of a possible way forward to issues raised in the ongoing debate about emerging technologies to support environmental modeling (incl. component-based models and the use of web-based data (see for example Argent, 2004)). It is also timely in view of relevant thinking in the US (for example with the proposed development of a Community Hydrologic Modeling Platform (CHyMP) by the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI)) where issues of data-model integration using OpenMI are being discussed (Maidment et al., 2009).

It is believed that as the ideas of distributed, component-based computing evolve (Villa et al., 2009), such applications of OpenMI will grow in number and significance.

References

Adler, R.F., Huffman, G.J., Bolvin, D.T., Curtis, S., Nelkin, E.J., 2000. Tropical rainfall distributions determined using TRMM combined with other satellite and rain gauge information. Journal of Applied Meteorology 39, 2007–2023.

Argent, R.M., 2004. An overview of model integration for environmental applications—components, frameworks and semantics. Environmental Modelling and Software 19 (3), 219–234.

Argent, R.M., 2005. A case study of environmental modelling and simulation using transplantable components. Environmental Modelling and Software 20 (12), 1514–1523. doi:10.1016/j.envsoft.2004.08.016.

Barthel, R., Janisch, S., Schwarz, N., Trifkovic, A., Nickel, D., Schulz, C., Mauser, W., 2008. An integrated modelling framework for simulating regional-scale actor responses to global change in the water domain. Environmental Modelling and Software 23 (9), 1095–1121.

Brackenridge, G.R., Anderson, E., Caquard, S., 2008. Global active archive of large floods 1985–2007, Dartmouth Flood Observatory (DFO), Hanover, USA. www.dartmouth.edu/~floods/Archives/ (last accessed July 2009.).

Boushaki, F., Hsu, K.-L., Sorooshian, S., Park, G.-H., Mahani, S., Wei, S.I., 2009. Bias adjustment of satellite precipitation estimation using ground-based measurement; a case study evaluation over the southwestern United States. Journal of Hydrometeorology 10 (5), 1231–1242.

Chen, M., Shi, W., Xie, P., Silva, V.B.S., Kousky, V.E., Higgins, R.W., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. Journal Geophysical Research 113 (D04110). doi:10.1029/2007JD009132.

Deutscher Wetterdienst, 2009. Global Precipitation Climate Center. Federal Ministry of Transport, Building and Urban Affairs, Climate and Environment. GPCC Web Site. URL: http://gpcc.dwd.de/ (last accessed May 2009).

Fotopoulos, F., Tsesmelis, G., 2006a. Monitoring of the characteristics (quantity & quality) of the rivers' Evros (Maritsa), Ardas and Erythropotamos (Bjala). In: Proc. Interreg III-A Greece. Bulgaria Steering Committee, Thessaloniki.

Fotopoulos F., Tsesmelis, G., 2006b. Early Warning System for Floods in Evros River. In: Proc. Conf. for Meric Flood Protection, 22 September 2006, Ankara, Turkey.

Goodall, J.L., Robinson, B.F., Shatnawi, F.M., Castronova, A.M., 2008. Linking Hydrologic Models and Data: The OpenMI Approach. American Geophysical Union, San Francisco, USA. Fall Meeting, December 15–19, 2008.

Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P., Blind, M., 2005. OpenMI: the essential concepts and their implications for legacy software. Advances in Geosciences 4, 37–44.

Gregersen, J.P., Gijsbers, P.J.A., Westen, S.J.P., 2007. OpenMI: open modeling interface. Journal of Hydroinformatics 9 (3), 175–191.

Harza Engineering Company, 1953. Master Plan for Meric-Evros River Development, Chicago, USA.

Hayes, J., 2009. Strategic planning for better decisions. In: Proc. 8th Biennial Conference and Exposition. National Hydrologic Warning Council, Vail, Colorado May 18–21, 2009.

Horwood, R., 2004. The importance of real-time data during an operational river flood. In: Proc. Northeast Regional Operational Workshop, November 2004, Albany, New York.

Hossain, F., Katiyar, N., Wolf, A., Hong, Y., 2007. The emerging role of satellite rainfall data in improving the hydro-political situation of flood monitoring in the under-developed regions of the world. Natural Hazards 43, 199–210. doi:10.1007/s11069-006-9094-x.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F., Wolff, D.B., 2007. The TRMM multi-satellite precipitation analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. Journal of Hydrometeorology 8, 38–55.

Huffman, G.J., Adler, R.F., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., 2003. Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time. In: Combined Preprints CD-ROM, 83rd AMS Annual Meeting, Poster P4.11 in: 12th Conf. on Sat. Meteor. and Oceanog., 9–13 February 2003, Long Beach, CA, 6 pp.

Ireson, A., Makropoulos, C., Maksimovic, C., 2006. Water resources modelling under data scarcity: coupling MIKE BASIN and ASM groundwater model as a tool for the Water Framework Directive Implementation. Water Resources Management 20 (4), 567–590.

Kavetski, D., Kuczera, G., Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 2 application. Water Resources Research 42 (W03408). doi:10.1029/2005WR004376.

Kidd, C., 2007. How low can you go? Retrieval of light precipitation in mid-latitudes. In: Proc. 3rd IPWG Workshop on Precipitation Measurement, EUMETSAT.

Krzyzstofowicz, R., 2001. The case for probabilistic forecasting in hydrology. Journal of Hydrology 249, 2–9.

Lau, L., Young, R.A., McKeon, G., Syktus, J., Duncalfe, F., Graham, N., McGregor, J., 1999. Downscaling global information for regional benefit: coupling spatial models at varying space and time scales. Environmental Modelling and Software 14 (6), 519–529. doi:10.1016/S1364-8152(99)00016-X.

Lovejoy, S., Schertzer, D., Allaire, V.C., 2008. The remarkable wide range spatial scaling of TRMM precipitation. Atmospheric Research 90, 10–32. doi:10.1016/j.atmosres.2008.02.016.

Maidment, D., Goodall, J., Ames, D., 2009. Model/Data Integration: Linking the CHyMP with CUAHSI Data Services, CUAHSI CHyMP Workshop, MemphisTN, 31 March 2009 Available online: http://www.cuahsi.org/chymp-agenda-20090331.html.

Makropoulos, C., Safiolea, E., Efstratiades, A., Oikonomidou, E., Kaffes, V., Papathanasiou, C., Mimikou, M., 2009. Multi-Reservoir Management with OpenMI, 11th International Conference on Environmental Science and Technology, September 3rd–5th, Chania, Crete.

Moore, R.V., Tindall, C.I., 2005. An overview of the open modelling interface and environment (the OpenMI). Environmental Science and Policy 8 (3), 279–286.

Moore, R.V., 2001. Description of Work for the HarmonIT PROJECT Agreed during Contract Negotiation, IT Frameworks (HarmonIT), Proposal Number: EVK1 2001-00037. www.HarmonIT.com.

Roth, G., 2006. The role of remotely sensed precipitation datasets in flood forecasting. Geophysical Research Abstracts 8 SRef-ID: 1607–7962/gra/EGU06-A-04029.

Safiolea, R., Makropoulos, C., Mimikou, M., 2009. Benefits and challenges in integrated water resources modeling using OpenMI: the case of the Pinios River basin, Greece. In: CCWI Conference, Sheffiled, September 2009.

Sinding, P., Gijsbers, P.J.A., Westen, S.J.P., Gregersen, J.B., 2005. OpenMI Document Series: Part F – org.OpenMI.Utilities Technical Documentation. HarmonIT European Programme, Butford Technical Publishing Ltd, Pershore, UK. September 2005.

Syed, T.H., Lakshmi, V., Paleologos, E., Lohmann, D., Mitchell, K., Famiglietti, J., 2004. Analysis of process controls in land surface hydrological cycle over the continental United States. Journal of Geophysical Research 109 (D22105). doi:10.1029/2004JD004640.

Villa, F., Athanasiadis, I., Rizzoli, A., 2009. Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. Environmental Modelling and Software 24 (5), 577–587.