One of the key requirements faced by the earth and environmental science community is to improve collaboration across the many geographically-distributed research groups and enable them to share data acquired by diverse means (ground, airborne and satellite), computational and storage resources, knowledge, experimental results.

Complexity is caused by the fact that:
• very often a single instrument can serve several applications and a given application wishes to access all available instruments;
• each application needs specific time and space data sampling;
• standards and data handling practises differ in different research communities.

The Grid and emerging e-collaboration computing services are considered as powerful new instruments, enabling and strengthening remote collaboration among researchers. For example, a Grid-based infrastructure can permit alternative approaches to access and exploit large data sets. Instead of traditional remote sensing data ordering and delivering from the acquisition/storage facilities to the user sites, user specialised processing modules could be located wherever data and computing resources are available. Specific on-demand data products, based on the best user-defined parameters, could then be generated or retrieved from archives and downloaded in real-time. Such services can be available via the web to members of a virtual thematic community, using the emerging Web Services standards.

The European Space Agency at the ESRIN site (Frascati, Italy) has the mandate to acquire, process, archive and distribute data coming from the Earth observation satellites operated in Europe, organised in a fully distributed network of facilities all over Europe. The ESA historical archive already includes some 2 PetaBytes of data and the flagship ENVISAT mission, launched in 2002, is increasing this data holding by some 400 TeraBytes per year.

In recent years, ESA-ESRIN has experimented with the Grid and is currently involved in several applications and projects ranging from e-collaboration to the organisation of information in ad-hoc digital libraries, organised around the established Grid infrastructure. ESA intends to adopt the Grid computing philosophy for handling the instrument data of future Earth observation missions.

**Grid-on-Demand Specific Demonstration Applications**

Following successful demonstrations of GRID applications for Earth Science, both in studies funded by the ESA General Study Programme and through participation in EC-funded projects, in 2004 at ESRIN a schedule was defined to render mature GRID-based applications operational. In particular, the integration of the ENVISAT MERIS dedicated data processing tool (BEAM) in the so-called ‘Grid on-Demand’ activity was completed. Through Grid-on-Demand, authorised users can generate level-3 products (ie uniformly time and space remapped geophysical variables) used, for example, for the monthly mosaicking of the global chlorophyll or vegetation status derived from MERIS instrument data (see Figure 1).

Grid on-Demand supports scientific applications for future large ENVISAT e-Collaboration and Grid-on-Demand Computing for Earth Science at ESA

by Luigi Fusco, Veronica Guidetti and Joost van Bemmelen

Information and Communication Technologies (ICT) have proven to be key instruments in activities targeted at protecting the environment and its integration in sustainable development policies. The European Space Agency (ESA) considers GRID computing capabilities as powerful tools to manage large volumes of Earth observation data and provide the Earth Science community with on-demand services. ESA is currently involved in several projects requiring a Grid infrastructure to support e-collaboration and digital library applications.

![MERIS mosaic image using data from the months of May, July, October and November 2004. This image is made up of true colour images using four out of 15 MERIS spectral bands taken from Envisat (bands 2,3,5 and 7) with data combined from the selected separate orbital segments with the intention of minimizing cloud cover as much as possible by using the corresponding data flags. In total, more than 1 TeraBytes of data were processed to generate a final image of about 1 GigaBytes.](image-url)
data set access. Together with the high-performance processing capability of the Grid, it provides quick accessibility to data, computing resources and results. Figure 2 shows support for a science group interested in new algorithm development and fast validation of results.

The power of the Grid infrastructure will help to process and manage large amounts of satellite images, thus forming the basis for long term data preservation, while digital library common functionality and third party applications will allow the users to retrieve, analyze and manage the contents, the services and the virtual organisation.

**e-Collaboration**

ESA also sees the Grid as a powerful means to improve the integration of data and measurements coming from very different sources to form a Collaborative Environment. The ESA-GSP project ‘THE VOICE’ (THEmatic Vertical Organisations and Implementation of Collaborative Environments, see http://www.esa-thevoice.org) aims at building an infrastructure that allows collaboration between different groups of researchers in the Earth observation field and generates scientific prototypes in the domain of ozone data calibration and validation, in the establishment of GMES (Global Monitoring for Environment and Security) application innovative partnerships, in the agriculture and forest rural areas monitoring and the marine applications communities.

**Concluding Remarks**

As already happened with the World Wide Web, the Grid together with e-Collaboration technology is expected to have a deep impact on our life, not necessarily restricted to scientific applications. However, the extent to which Grid technology will be exploited in the future is closely connected to the adoption of common standards to allow different grids to collaborate, ie to work together.

For the Earth science community, it is important to continue and invest in activities focussing on Grid and e-collaboration. Initiatives such as ESA Grid-on-demand and projects like THE VOICE are demonstrating their relevance. These projects show that a Grid-based underlying infrastructure is a real asset for this community. It significantly improves the accessibility and usability of Earth science data, information and knowledge, and the way Earth science users collaborate.

**ERAMAS — Environmental Risk Analysis and Management System**

by Thilo Ernst, Andreas Hoheisel, Thomas Lux and Steffen Unger

The aim of the ERAMAS project is to develop a Grid-based system for analysing and managing pollutant-related environmental risks. Sophisticated simulation programs are used to forecast and evaluate the dispersion of carcinogenic and chemically toxic substances in the atmosphere, the soil and the groundwater, and to calculate the risk they pose to humans.

ERAMAS is a simulation-based analysis framework for calculating risks due to chemically toxic or carcinogenic substances being released, for example during accidents in industrial installations, the transport of dangerous goods or by terrorist attacks. It is designed to be applicable both for real-time emergency management and for risk mitigation activities such as simulation-aided studies concerning the design of approval procedures or emergency plans.

Figure 1 shows an overview of the simulation models involved in ERAMAS regarding various transportation paths for pollutants in the atmosphere and the soil. In the environmental simulation domain, this is a typical scenario; nevertheless, integrating and coupling even a small number of heterogeneous simulation models, and making them available to technically unskilled users, amounts to a very complex and time-consuming effort.

ERAMAS is being developed using the technology of the Fraunhofer Resource Grid (http://www.fhrg.fraunhofer.de), which simplifies the coupling of heterogeneously distributed software, hard-
ware and data resources (see Figure 2). The Fraunhofer Resource Grid is a Grid initiative comprising five Fraunhofer institutes and funded by the German Federal Ministry of Education and Research. Its main objectives are to develop and implement a stable and robust Grid infrastructure within the Fraunhofer Gesellschaft, to integrate available resources, and to provide internal and external users with a user-friendly interface for controlling distributed applications and services on the Grid.

ERAMAS is a system with considerable resource demands. These arise not only from the inner complexity of its components, but also from complex workflows and usage scenarios in which a substantial number of component instances need to be executed, e.g. parameter studies. Such a system cannot be expected to run on a single workstation; parallel and distributed computing techniques are obviously necessary. However, the primary advantage of using Grid technology in ERAMAS is not the performance gain from having access to additional resources, but rather the organizational advantages in building and maintaining a distributed, highly heterogeneous simulation system. The Grid is used to organize the workflow of coupled simulations and to provide uniform access to a wide variety of hardware, software and data resources. The component abstractions offered by the Fraunhofer Resource Grid make the coupling of a wide range of models and data sources very easy - detailed knowledge of the internals of the components is no longer needed.

The simulation components integrated in ERAMAS are pure command-line applications, that is, they have no graphical user interface. Specialized simulators usually originate from the research sector, where this is considered normal. However, it conflicts severely with the goals of and application scenarios envisioned for ERAMAS, which call for strong support of the less technically skilled user, such as users from on-site emergency response teams. This gap is bridged by relying on the VirtualLab platform (http://vl.nz.dlr.de/VL), which is being developed through collaboration between the German Aerospace Center (DLR) and Fraunhofer FIRST. VirtualLab contains a subsystem for dynamically generating flexible and easy-to-use Web user interfaces for command-line applications, using abstract descriptions of their input datasets. Together with its generic Web portal features (such as protected user areas that persistently store simulation runs, integrated documentation management, and Web-based administration), VirtualLab is thus able to provide a powerful Web access layer for ERAMAS.

The ERAMAS system is being developed by Fraunhofer FIRST, in collaboration with Ingenieurbüro Beger für Umweltanalyse und Forschung and the Dresdner Grundwasser Consulting GmbH in Germany. The project is funded by the Arbeitsgemeinschaft industrieller Forschungsvereinigungen Otto von Guericke (AiF) in the programme Innovationskompetenz mittelständischer Unternehmen (PRO INNO).

The funded ERAMAS project commenced in July 2002 and finished in October 2004, and the result is a demonstration prototype of the ERAMAS system. ERAMAS can be viewed as a pilot project that is introducing Grid- and Web-based e-science methods into the environmental simulation and risk management community, and is developing and deploying a dedicated platform for the purpose.

Our aim for the future is to make ERAMAS a commercially operated service that can be used in a variety of ways, e.g. for advance analysis in licensing procedures or for drawing up action plans. Potential customers include chemical companies, haulage contractors and emergency services like the fire service. Another application area is analysis under real-time conditions, whether in the case of malfunctions in industrial plants, the transport of hazardous materials or terrorist attacks.

Links:
http://www.first.fraunhofer.de/en/eramas
http://www.fhrg.fraunhofer.de
http://www.first.fraunhofer.de/en/vlab

Please contact:
Steffen Unger, Institute for Computer Architecture and Software Technology - FIRST, Fraunhofer ICT Group
E-mail: Steffen.Unger@first.fraunhofer.de
GENIE: Grid Enabled Integrated Earth System Model

by Andrew Price, Tim Lenton, Simon Cox, Paul Valdes and the GENIE team

An understanding of the astonishing and, as yet, unexplained natural variability of past climate is an essential pre-requisite to increase confidence in predictions of long-term future climate change. GENIE is a new Grid-enabled modelling framework that can compose an extensive range of Earth System Models (ESMs) for simulation over multi-millennial timescales, to study ice age cycles and long-term human induced global change. Grid technology is a key enabler for the flexible coupling of constituent models, subsequent execution of the resulting ESMs and the management of the data that they generate.

To predict the future, we must understand the past. In the case of planet Earth we do not yet fully understand the mechanisms that have driven the most fundamental change in climate over the past million years – the transitions between ice ages and warm inter-glacials. To improve understanding of the physical processes and feedbacks that are important in the Earth System, the GENIE project is creating a component framework that allows the flexible coupling of constituent models (ocean, atmosphere, land, etc.) of varying resolution (grid sizes), dimensionality (2D and 3D models) and comprehensiveness (resolved physics vs. parameterisations) to form new integrated ESMs. Through the systematic study of a hierarchy of GENIE models the project aims to determine the spatial resolution and process complexity that actually need to be included in an ESM to exhibit past Earth System behaviour.

The GENIE project is funded by the National Environment Research Council (NERC) and brings together expertise from UK and international academic institutions. The Universities of Bristol and East Anglia, the Southampton Oceanography Centre and the Centre for Ecology and Hydrology have provided mature models of major Earth System components including atmosphere, ocean, sea ice, ocean biogeochemistry, sediments, land vegetation and soil, and ice sheets. The e-Science centres at the University of Southampton and Imperial College have been engaged to provide the software infrastructure for the composition, execution and management of the integrated Earth System Models and their output on the grid. We have strong international collaborations with researchers at the Frontier Research Centre for Global Change in Japan, University of Bern in Switzerland and University of British Columbia in Vancouver.

**e-Science Challenge**

The objectives of the GENIE project are to develop a Grid-based computing framework which will allow us:
- to flexibly couple together state-of-the-art components to form a unified Earth System Model (ESM),
- to execute the resulting ESM across a computational Grid,
- to share the distributed data produced by simulation runs, and
- to provide high-level open access to the system, creating and supporting virtual organisations of Earth System modellers.

**Software**

Grid computing technology is required to ease the construction of new instances of Earth system model, automate the process of model tuning, speed up the execution of individual long integrations, enable large ensembles to be run, ease their execution, and feed and recycle data back into model development. A principle aim of the project is to ensure that the Grid is useable directly from the environment where the climate modellers are performing their work. The software deployed to meet these requirements is built upon products of the first phase of the UK e-Science programme. These include:
- Geodise Compute Toolbox
  The Geodise computational toolbox for Matlab provides a suite of Matlab functions that provide programmatic access to Globus Grid enabled compute resources. The computational toolbox uses the APIs provided by the Java CoG toolkit to allow the submission of compute jobs to Globus enabled resources, GridFTP data transfer and the management of proxy certificates. An interface to Condor resources is also provided.
- Geodise Database Toolbox
  An augmented version of the Geodise Database Toolbox has been deployed to provide a distributed data management solution for the GENIE project. The Geodise system exploits database technology to enable rich metadata to be associated with any data file, script or binary submitted to the repository for archiving. XML schemas define the structure of the metadata and are mapped into the underlying Oracle 9i database. The database system is built on open W3C compliant standards technologies and is accessed through a web services interface. Client tools are provided in Matlab and Jython which allow both programmatic and GUI access to the system.
- OptionsMatlab
  OPTIONS is a design exploration and optimisation package that has been developed in the Computational Engineering and Design Centre at the University of Southampton. This software provides a suite of sophisticated multidimensional optimisation algo-
Tuning
A key challenge to the project is to tune or re-tune the parameterisations of individual model components so that the new coupled ESMs simulate reasonable climate states. In particular, it is imperative that the fluxes passed between components are compatible if the resulting coupled model is to be stable. We have exploited the Grid enabled toolset in conjunction with the OPTIONS package to apply Response Surface Modelling techniques and Genetic Algorithms to optimise GENIE model parameters. In addition, the ensemble Kalman Filter, a data assimilation method, has also been employed. These techniques provide a comprehensive set of tools for a program of extensive model tuning which has progressed in step with model development.

Current and Future Study
We are exploiting local resources (condor pools, Beowulf clusters) and the UK National Grid Service to perform extensive studies of GENIE models. The computational Grid provides the means to perform large ensemble runs. To date, experiments have studied the stability of the thermohaline circulation to multi-parameter freshwater inputs, using typically ~1000 instantiations of the model, involving ~40 million years of model integration.

The database repository plays a central role in these studies as a resource for both steering computation and sharing of the data. Future work will involve the development of a distributed federated database system, deployment of the database on the National Grid Service data node(s) and further enhancements to the data management tools. The project will adopt the GeodiseLab Toolbox from the OMII (Open Middleware Infrastructure Institute) managed programme when this product is released to the community.

Links:
http://www.genie.ac.uk/
http://www.geodise.org/
http://www.omii.ac.uk/

Please contact:
Andrew Price,
Regional Southampton e-Science Centre,
University of Southampton, UK
Tel: +44 23 8059 8375
E-mail: a.r.price@soton.ac.uk

Chemistry GRID and its Applications for Air Pollution Forecasting
by Róbert Lovas, István Lagzi, László Kullmann and Ákos Bencsura

Computational Grid systems are becoming increasingly popular in the natural sciences. In such systems, a large number of heterogeneous computer resources are interconnected to solve complex problems. The main aim of the national research project funded by the Hungarian Ministry of Education, ‘Chemistry Grid and its Applications for Air Pollution Forecasting’, was to look at feasible applications of Grid technology in computational chemistry from a practical point of view; for example, prevention of the harmful effects of high-level ozone concentrations.

In the project, the consortium (SZTAKI; Chemical Research Institute of the Hungarian Academy of Sciences; Department of Physical Chemistry, Eötvös University; Hungarian Meteorological Service) applied new Grid technologies to provide support for a specific research area. The developed infrastructure now provides chemists with access to both Hungarian computational Grid resources, called HUNGGRID, and European-wide chemistry Grid infrastructures. The latter were established as the result of the EU-funded projects SIMBEX and EGEE.
SZTAKI has elaborated a product line: a Grid-monitoring tool called Mercury, and two integrated application development environments, called P-GRADE parallel programming environment, and P-GRADE Grid portal (see Figure 2). These tools enable the efficient and transparent parallelization of sequential applications through their high-level graphical approach and special performance debugging and analyser tools. In the framework of the project, the P-GRADE portal was developed further to provide support for the efficient execution of complex programs in various Grids, e.g., in HUNGRID. It included the dynamic execution of applications across Grid resources according to the actual state and availability conditions provided by the new information system.

Consequently, HUNGRID is not only a virtual organization within the EGEE: its new elements make it easier to use the infrastructure for solving complex problems, such as the modelling of air pollution.

The phytotoxic nature of ozone was recognized decades ago. Due to high emissions of ozone precursor substances, elevated ozone concentrations may cover large areas of Europe for shorter (episodic) or longer periods under certain meteorological conditions. These elevated concentrations can be potentially damaging to agricultural and natural vegetation. Occasional extreme concentrations may cause visible injury to vegetation, while long-term exposure, averaged over the growing season, can result in decreased productivity and crop yield.

A coupled Eulerian photochemical reaction–transport model and a detailed ozone dry-deposition model were developed to investigate ozone fluxes over Hungary. The reaction-diffusion-advection equations relating to air pollution formation, transport and deposition are solved on an unstructured triangular grid. The model domain covers Central Europe including Hungary, which is located at the centre of the domain and is covered by a high-resolution nested grid. The sophisticated dry-deposition model estimates the dry-deposition velocity of ozone by calculating the aerodynamics, the quasi-laminar boundary layer and the canopy resistance. The meteorological data utilized in the model were generated by the ALADIN meso-scale limited-area numerical weather prediction model, which is used by the Hungarian Meteorological Service. The work demonstrates that the spatial distribution of ozone concentrations is a less accurate measure of the effective ozone load than the spatial distribution of ozone fluxes. The fluxes obtained show characteristic spatial patterns, which depend on soil moisture, meteorological conditions, ozone concentrations and the underlying land use (see Figure 1).

This project has demonstrated that the Grid is an efficient computer system for supporting complex collaborative work. Applications for air pollution forecasting (elaboration of smog alert response plans and a Gaussian plume simulation) have been developed and presented. The project partners have designed a collaborative application that runs on Grid to forecast air pollution in Hungary. The same application can be used to simulate earlier smog events and to analyse the efficiency of smog alert response plans and the long-term effects of various measures against air pollution.

Link:  
http://www.lpds.sztaki.hu/chemistrygrid

Please contact:  
Róbert Lovas, SZTAKI, Hungary  
E-mail: rlovas@sztaki.hu
In the Czech Republic, an operational statistical predictor of summer photo-chemical smog situations has been used since 2000. This system uses a neural network predictor and a dynamic regression model. Its inputs are the daily maxima of tropospheric ozone concentrations and temperature, measured at the ground-level stations of the Automatic Immission Monitoring system (AIM) of the Czech Hydrometeorological institute. The resulting concentration field is obtained by means of spatial interpolation. Although this system reasonably predicts daily ozone concentration maxima, there is a need for a deterministic model which would give reliable values for regions sparsely covered by AIM stations and which would reflect the air circulation. On the other hand, deterministic models often suffer from uncertainties in emission inputs; systematic bias in some locations may occur. It is also desirable to make use of any available measurements. A system with incorporated data assimilation therefore represents a natural target for modelling efforts.

The primary aim of the MEDARD project is the development of such a system for air quality forecasting. Its origins date to 2001 when the first steps were made during the EU framework V project APPETISE. The project is currently supported by the grant agency of the Academy of Sciences of the Czech Republic, within the framework of the ‘Information Society’ programme (No 1ET400300414). It involves researchers from the Institute of Computer Science (ICS) of the Czech Academy of Sciences, in collaboration with Charles University, the Czech Hydrometeorological Institute and the Institute of Meteorology at the University of Natural Resources and Applied Life Sciences in Vienna.

A suitable NWP-CTM model pair was sought for this purpose. The most natural choice for the NWP model was MM5, which is widely used in both the USA and Europe (eg the EURAD group; www.eurad.uni-koeln.de). The model was configured for the territory of the Czech Republic, it has two nested domains with resolutions 27 and 3km, the NOAH land surface model is run and the MRF scheme of the Planetary Boundary
Layer. The boundary conditions are taken from the GFS global forecast of NCEP. The CAMx model was chosen as the CTM part of the system. CAMx has pre-processors for meteorological fields from MM5, and in our configuration it runs on two nested domains derived from the above MM5 domains, and uses the SAPRC99 chemistry mechanism and EMEP emission inventories.

Upon the MM5-CAMx pair, a common presentation layer was built. The results are available at a user-friendly Web site intended for the general public. It was designed so as to provide a quick orientation to the situation, and features two alternative displays, animation and quick switching of domains and products. In the field of data assimilation, a type of ensemble filter suitable for the CAMx model has been proposed and is in the testing phase. Pilot experiments are being run, with the aim of determining how much data assimilation will improve operational forecasts. Figures 1 and 2 show an example of the output of an ensemble run: the ensemble mean of ozone concentration (Figure 1) compared to the output of a free run without data assimilation (Figure 2).

Soon after the project commenced and the first weather prediction outputs became available, demand arose from the private sector for specialized outputs like lightning activity indices, risk of icing and local wind profiles. Some products of this kind have accordingly been developed. It also emerged that public interest is biased towards weather prediction, with interest in air quality being marginal. We are attempting to attract more interest in air quality issues by presenting the outputs of the CTM as a product of comparable importance to weather forecasts.

Commencing an operational run of the air quality forecast with incorporated data assimilation, however, requires a longer time schedule. Any such system will need on-line data, not only from the country immediately involved but also, due to transport phenomena, from neighbouring countries. Unfortunately there remain large differences between individual countries in the availability of on-line measurements of pollutant concentrations.

A big effort has to be made in designing a system for the downloading and validation of these on-line measurements. Together with the enormous computing-power demands of assimilation methods, the development of an operational data assimilating system for air quality prediction remains a significant challenge.

---

**OASI: Integrated Monitoring and Decision Support for Environmental Systems**

by Roberto Mastropietro, Lorenzo Sommaruga and Andrea E. Rizzoli

OASI is a system for regularly and automatically collecting environmental data from a network of hundreds of sensors distributed throughout the Canton of Ticino in southern Switzerland. It has been in successful operation for several months.

Canton Ticino is the southernmost part of Switzerland and it lies right on one of the most important European transport axis, connecting Italy with Central Europe. It is a sub alpine region, with remarkable natural beauties, but its environment is subject to many pressures, among which road emissions play a major role. In 2001 the Swiss government decided to tackle the problem of the growing traffic on the main national highways by funding initiatives aimed at making available the necessary data and tools to understand the impact of traffic on the environment, in order to enable policy makers to make appropriate decisions.

Within this context, the Land Management Department of Canton Ticino launched a project to develop an integrated monitoring and decision support system to collect, analyse and process multi-domain environmental data. The project outcome has been named OASI (Osservatorio Ambientale Svizzera Italiana), and was developed by SUPSI/DTI, the Department of Innovative Technologies at the University of Applied Science of Southern Switzerland.

OASI regularly and automatically collects environmental data from a network of hundreds of sensors distributed all over Canton Ticino. Then, the collected data automatically undergo a statistical, intersite and interdomain validation process before being made available to the end-users. In this process, past data, already validated, are used to detect anomalies in the measurements which have been recently collected.

Scientists can access the data repository via the OASI application in order to perform integrated data analyses. Users can select locations, parameters, time intervals and diagram types. Multiple curves can be displayed and compared in a single diagram. Data belonging to different domains can be shown on a single screen, for comparison purposes, as shown in Figure 1.

The OASI software system has a scalable architecture that allows additional
data domains, as diverse as basin levels and landslides monitoring indicators, to be seamlessly added to the set of monitored data. This is achieved thanks to its 3-tier architecture composed of a data layer, a logical layer and a presentation layer — and to a flexible database design.

Data from different domains are collected in separate databases in the data layer, but they can be transparently integrated for validation and analysis in the logical layer. The software architecture therefore easily supports the integration of multiple databases, and it allows the deployment of the system in distributed configurations making the solution flexible and scalable. The system allows the definition of user roles and data ownership. An organization may or may not allow its data to be seen or modified by users belonging to other organizations.

One of the most challenging aspects of the project is related to the amount of measurements being collected and the size of the databases, which results in interesting storage management and system performance issues solved using special design and advanced techniques made available by the database management system (Andretta et al. 2004).

Finally, the presentation layer is designed to serve both the needs of scientists and researchers as well the general public that wants to be informed on the state of the environment.

The OASI system has been running in production mode for a few months and it has raised the interest of other Cantons that are now also integrating their data in the system for validation and analysis purposes.

Further developments of the OASI project will push towards the provision of an even more open and mobile access to data, supporting interoperability among different client-server nodes, easy extensibility for integrating any kind of client device into the system, both for regular users and system administrators. In particular, within the OASI context, Web services technology will be effectively exploited for integrating the needs for dissemination of analytical data about environment, such as air, noise, traffic, etc., and the needs of different users having accessibility requirements for their devices, being distributed and heterogeneous systems, remote and mobile clients (Arauco and Sommaruga, 2004).

Link:
Oasi website: http://www.ti.ch/oasi
Please contact:
Roberto Mastropietro
University of Applied Science of Southern Switzerland (SUPSI)
Tel: +41 91 6108578
E-mail: roberto.mastropietro@supsi.ch

Figure 1: A screenshot of the OASI workspace.

Data Assimilation and Air Pollution Forecasting: the Berlin Case

by German Ariel Torres, Steffen Unger, Torsten Asselmeyer-Maluga, Vivien Mallet, Denis Quéo, Bruno Sportisse, Isabelle Herlin and Jean-Paul Berroir

In a joint French-German project, a three-dimensional chemistry-transport model called Polair3D has been improved in order to achieve better air pollution forecasts.

The first objective of this project was to prepare the Polair3D model for forecasting the air pollution in the Berlin-Brandenburg area. Polair3D is a 3D chemistry-transport model developed by Cerea (ENPC - EdF R&D), and is used for air quality forecasting on both continental (Europe) and regional scales (urban air pollution in Paris, Lille and Marseille). The second objective is to improve the quality of the forecasts by developing a sequential data assimilation.
framework adapted to Polair3D. Experiments are run by assimilating data provided by Berlin's air pollution monitoring network Blume.

Running Polair3D on a new application site requires a set of input data to be prepared. This comprises:
- the meteorological fields, generated by the mesoscale meteorological model MM5, which was developed by the Pennsylvania State University and the National Center for Atmospheric Research
- the emission inventory, based on EMEP European emission inventory and on the CityDelta project for urban emissions in the Berlin-Brandenburg area
- the boundary conditions, obtained by nesting Berlin regional runs within Polair3D runs at a European scale, now routinely performed.

The methodology is systematic, and can be applied to regional air pollution forecasting with Polair3D in any European city for which a CityDelta emission inventory is available.

Following this phase, sequential data assimilation is used to adjust model inputs and parameters, in order to minimize the difference between forecasts and measurements. This process is performed each time a new measurement is acquired, hence the name 'sequential'. Data assimilation therefore helps to improve the quality of air pollution forecasts provided by the model.

One of the main advantages of sequential data assimilation is that it does not require the availability of an adjoint model, and can be implemented as a post-processing toolbox that can be plugged into any model with little effort. However, it carries a considerable computational cost, since it requires several direct runs for each new measurement.

During the project, two types of Kalman data assimilation procedure were implemented: a rank-reduced Kalman filter, which reduces the dimension of the problem, and an ensemble Kalman filter, where statistics of model errors are generated by performing many forward runs of the model. These procedures have been validated in case studies, and their use for the Berlin case is under development, since it requires the parallelization of the Kalman filter. We are currently carrying out an evaluation of Polair3D results by performing a set of forecasts for the test period April-September 2001. Forecast error statistics are then computed by comparing the forecasts with measured in situ data. The project will be completed by quantifying the improvement brought by sequential data assimilation relative to direct runs without assimilation. A comparison with variational data assimilation methods is also possible.

This work has been carried out in the framework of a French-German project, funded under the Procope programme, and supported by an ERCIM post-doctoral fellowship. It involves the Clime team, a common INRIA and ENPC project located in the Paris area, and Fraunhofer FIRST, Berlin. Research was carried out at the three sites - INRIA, ENPC and Fraunhofer - during the period 2003-2004.

The project is continuing through cooperation with the University of Córdoba in Argentina. Polair3D will be used for air quality forecasts in this area, which will require the preparation of an emission inventory. Sequential data assimilation will be used to improve knowledge of the emissions. During this project, the feasibility of assimilating satellite measurements of chemical concentrations will be studied.

Please contact:
Vivien Mallet,
Clime Team (ENPC/INRIA), Ecole Nationale des Ponts et Chaussées, France
E-mail: mallet@cerea.enpc.fr
There is an increasing need for local and regional government authorities to incorporate computerized systems to assist in evaluating the current status of the physical environment in their area and in assessing the impact of proposed future developments. In general such authorities already have extensive environmental sampling programs (e.g., water characteristics at designated places on streams, rivers and lakes; air quality at specific locations; noise levels along certain roads) whose outputs must be incorporated in the new systems. Also to be incorporated are such data as population densities, patterns of transport usage including car ownership, agricultural practices and patterns, industrial development and so on. The computerized systems must include relevant models of various aspects of the environment (e.g., associations between water quality and population density, agricultural production and phosphate/nitrate levels, traffic and air quality, local weather patterns, social capital and car usage), and must be capable of incorporating new data sources and models in the future. The utility of such systems depends crucially on the quality of their input data, which involves maximising the information yield from available measurements. The particular focus of this article is on extracting environmental information from airborne hyperspectral measurements.

Pigments are chemical compounds that reflect only certain wavelengths of visible light, which makes them appear "colourful". More important is the ability of pigments to absorb certain wavelengths. In particular, Chlorophylls are greenish pigments of which the most important for photosynthesis is Chlorophyll a. In general, since each pigment reacts with only a narrow range of the spectrum, there is usually a need to produce several kinds of pigments, each of a different colour, to capture more of the sun's energy. Typically, each hyperspectral sample consists of 512 measurements, collected by means of an airborne 'Portable Multispectral Imaging Spectrometer', of the irradiance reflectance at equally spaced intervals in the visible spectrum. Flight campaigns have been carried out over several inland lakes and adjacent lands, with many samples collected during each flight.

Traditionally, broad band ratio techniques, which would make use of just a subset of the available measured reflectances, have been used to convert multispectral measurements into estimates of indices of vegetative biomass or photosynthetic activity. An important part of the present work is the development of full spectral vegetation indices, with the aim of extracting the maximum amount of information from measurements. Figure 1 provides an example of environmental monitoring of Suspended matter content, and Dissolved Organic Matter and other pigments for the water industry, and remote sensing (airborne) application of hyperspectral techniques for water quality and land productivity quantification and mapping.

Spectral Signatures is a University College Dublin campus company which for several years has specialised in extracting environmental information on water quality and land productivity by means of optical instruments that have been designed and built in-house. Capabilities include in-vivo and non-contact sampling of water Chlorophyll content, and Suspended matter and Dissolved Organic Matter and other pigments for the water industry, and remote sensing (airborne) application of hyperspectral techniques for water quality and land productivity quantification and mapping.
Getting the Most out of Earth Observation

by Martin Juckes

The imperative to understand our environment becomes ever more urgent as global change brings new uncertainties. At the same time, rapid technological advances are making it possible to observe the global environment from space in ever finer detail. In response to the challenges of handling an ever greater volume and diversity of data, the Multiple Instrument Stratospheric Tracer Assimilation (MISTA) project at Rutherford Appleton Laboratory has developed a new algorithm for extracting optimal information from satellite measurements. Thanks to gains in efficiency, a task normally done on supercomputers, can now be carried out easily on a PC.

Photons leaving the Earth's atmosphere carry the signature of the emitting gasses. Space instruments orbiting the Earth use a variety of filtering and detecting technologies to measure these photons and thence provide information which can be used to determine the state and composition of the atmosphere.

Imaging devices on geostationary satellites, measuring visible light reflected from clouds or solid surfaces, can give global coverage every 15 minutes or so. But this luxury is not available for those who want to study atmospheric composition in detail.

Our climate is significantly modulated by ozone, water vapour and methane in the stratosphere, all of which are present in concentrations of only 1 to 10 parts per billion (10^9) by volume. Detection of the weak signal emitted by these gases requires satellites flying much closer to the Earth. Europe's environmental monitoring research satellite, ENVISAT, launched in 2003 at a cost of one billion Euro, is typical of this class of satellites, flying at around 800km altitude and orbiting the Earth every 100 minutes. The MISTA project has so far focused on results from one instrument on ENVISAT: the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS). Of the three instruments on ENVISAT which were designed to detect stratospheric gasses MIPAS has been the most successful to date.

The greater detail which can be obtained by instruments flying closer to the Earth is gained at the expense of losing the regular global coverage provided by the geostationary instruments. MIPAS provides a string of observations along the satellite orbital path. As the Earth rotates under the orbit a picture of the global atmosphere can be built up. However, many of the features which are observed evolve significantly over the course of a day because of physical processes, many of which are well understood.

A wide range of methods have been developed to combine the prior knowledge we have of the physical processes with the information stream coming from the measurements. The problem can be formulated in a Bayesian framework, combining observational information with prior information expressed in a computational approximation to the relevant physical laws. So far so simple, at least to those involved in this area of research. The novelty in this project is in the way it tackles the numerical challenges thrown up by the Bayesian framework. Many other research institutes working on the same problem are exploiting existing forecast models. These are computational representations of the physical laws designed to predict the future state of the system from a prescribed initial state. The MISTA project and others use a new algorithm for extracting optimal information from satellite measurements. Thanks to gains in efficiency, a task normally done on supercomputers, can now be carried out easily on a PC.
A project has started from scratch and developed code specifically designed for data analysis. This allows the full exploitation of the elliptical structure of the problem, a structure which follows naturally from a standard Bayesian formulation. A multigrid relaxation algorithm has been implemented. This type of algorithm is commonly used for the solution of elliptical partial differential equations, but has not been applied in this context before.

A particular advantage of the new algorithm is that it fully exploits observations both prior to and after an analysis time. Information propagates both forwards and backwards in time within the analysis system, with a typical half-life of around 2 days. This means that images such as the figure, which shows the ozone distribution in the stratosphere (at around 30km altitude) on July 6th, 2003, can be significantly more accurate than anything produced by a forecast system which, by its nature, can only exploit observations prior to the analysis.

Link:
http://home.badc.rl.ac.uk/mjuckes/mista/

Please contact:
Martin Juckes, British Atmospheric Data Centre Research, Space Science and Technology Department, Rutherford Appleton Laboratory, CCLRC.
Tel: +44 1235 445124
E-mail: M.N.Juckes@rl.ac.uk

Ozone concentration at 30km altitude (parts per billion by volume), 6 July 2003.

Meso-Meteorological Modelling for Air Pollution Applications in Mid-Latitudes

by Jose-Luis Palau, Goka Pérez-Landa and Millán M. Millán

Any approach to meteorological modelling needs to take into account a variety of processes that interact synergistically at different scales (mainly in mid-latitude and complex terrain areas). Because of their potential to resolve regional and local atmospheric circulations, meso-meteorological models represent interesting tools. They are highly complex and their adaptation to a particular region requires an in-depth understanding of the limits of parameterization applicability.

Over the past few years there has been a growing need to simulate meteorological fields for complex situations at finer spatial resolutions. This has been partly stimulated by scientific and technological advances (eg in dynamics, computational methods and facilities) and partly by policy pressures requiring more detailed assessments of air pollution on urban to regional scales. As a consequence, complex dynamic models have increasingly been used in Europe and the USA for meteorological and air pollution applications. Models developed for short- or long-range applications, however, are not always appropriate in flow conditions involving intermediate mesoscale features and processes (of the order of 1-1000 km). This is because parameterizations, working hypotheses and configurations need to be different for differently scaled models (as stated in the explanatory memorandum for the implementation of the European Concerted Research Action designated COST Action 728: ‘Enhancing mesoscale meteorological modelling capabilities for air pollution and dispersion applications’).

In this context, our group is interested in situations of complex atmospheric flows for which mesoscale models are necessary (eg sea breezes, valleys and layered flows). It is generally acknowledged that current models are far from perfect; our goal is therefore to identify the gaps in our knowledge. We are performing this research within the following European frameworks: the Cluster of European Air Quality Research (CLEAR) project FUMAPEX (integrated systems for Forecasting Urban Meteorology, Air pollution and Population EXposure); the Network of Excellence ACCENT (‘Atmospheric Composition Change: A European Network’); and the ‘CarboEurope-IP, Assessment of the European Terrestrial Carbon Balance’.

Meteorological fields applied to air quality models (from global to local scales) may contain significant uncertainties that adversely affect simulations. There are a large number of meteorological variables needed for ‘generic’ air quality models, including horizontal and vertical wind components, temperature, water vapour mixing ratio, cloud fraction and liquid-water content, precipitation (rain/snow), solar actinic flux, sea-level pressure, boundary layer depth, turbulence intensity, and surface fluxes for heat, moisture and momentum. In addition to these variables, the forecasting of air quality in mid-latitudes (as in Mediterranean coastal regions) is also highly sensitive to the fact that non-local (mesoscale) effects strongly determine flows at urban scales (as reiterative experimental results have evidenced).
The aforementioned atmospheric state variables are insufficient under meteorological conditions marked by non-local dynamic effects (as, for example, the compensatory subsidences over the Mediterranean coasts associated with the orographic injections resulting from the coupling of sea breezes and inland orographic upslope winds). In this sense, the meteorological models must be configured in such a way that they are able to reproduce the mesoscale dynamics at these subtropical latitudes. For instance, the interaction between different scales must be reproduced: land-use, soil moisture, sea-surface temperature, grid nesting, domain configuration, and horizontal and vertical resolution are key magnitudes/parameters in describing wind flows for air-pollution forecasting purposes and must be set up properly (see Figure 1).

In recent years, different research projects around the world have demonstrated that the pollutant ‘exchange rate’ between different regions and under different meteorological (and climatic) conditions is driven by interactions and forcings between different meteorological scales reaching distances of thousands of kilometres (see Figure 2).

For a believable evaluation of the impact of anthropogenic emissions from urban to global scales, one must therefore implement within numerical models all these spatial and temporal interactions, together with feedback from climate, regional air quality and transport (at local, regional and global scales). Moreover, in order to address certain issues (eg how different meteorological scales contribute to the Long-Range Transport - LRT - of regional pollution), it is necessary to resolve some scientific questions related to these meteorological interactions. In this sense, there are indications that the formation and distribution of photo-oxidants in urban plumes, at regional or continental scales, in the boundary layer and in the free troposphere, are all linked together. There are a number of EC research projects relevant to this, including MECAPIP - Meso-meteorological cycles of air pollution in the Iberian Peninsula; RECAPMA - Regional cycles of air pollution in the west central Mediterranean area; and SECAP - South European cycles of air pollution.

Figure 1: Meteorological (top) and dispersive (bottom) simulations performed by running RAMS v.4.3.0 and HYPACT v. 1.2.0 on a Mediterranean coastal site for three days in July.

Upper figures: Time evolution of the vertical distribution of the simulated wind fields and boundary layer height simulated on a Mediterranean coastal site using the same resolution grid. This analysis of the meteorological data shows important differences in the mesoscale model outputs between two different meteorological approaches. On the right, the high resolution effects are included thanks to the two-way option between grids (resolving 1.5km meteorological features) and on the left, the model uses the resolution of the grid (13.5 km) without any feedback from the inner domains.

Lower figures: To check the implications of the two different meteorological approaches in the simulation of a point-source plume in a Mediterranean coastal area, both model outputs were employed to run two respective Lagrangian dispersion simulations. As shown, the vertical distribution of the simulated SO2 concentration emitted from a power plant (black vertical line) is strikingly different at identical simulated times. Units are g/m³.

Figure 2: Feedbacks and interactions/forcings governing the transport of anthropogenic pollutants emitted at urban/local scales.

Experimental data and complementary modelling results from these projects have established links between atmospheric circulations from local, through regional, to sub-continental scales, particularly in summer and in the central and western Mediterranean basins.

Link: http://www.gva.es/ceam

Please contact:
Jose Luis Palau, Fundacion CEAM, Spain
Tel: +34 96 131 82 27
E-mail: jlp@confluencia.net
Air Pollution Modelling in Complex Terrain: ‘Els Ports-Maestrat’ Regional-Scale Study

by Jose-Luis Palau, Goka Pérez-Landa and Millán M. Millán

Characterization of atmospheric pollutant dispersion (advection and turbulent diffusion) requires a detailed description of the wind and turbulence fields, especially on complex terrain under summer conditions. Gaussian regulatory models frequently incorporate simplifications that are not valid under these circumstances. Software-engineering development and technological advances within the hardware industry have permitted the use of high-performance computing systems that are able to supply the computing power necessary for high-resolution meteorological and Lagrangian simulations.

At present, the numerical simulation of atmospheric dynamics is a necessary tool for meteorological diagnosis, analysis and forecasting. At the beginning of the seventies, global models based on primitive equations represented an important development in the characterization of synoptic-scale atmospheric processes. Afterwards, the need to analyse lower-scale phenomena, together with the availability of increased computing power, resulted in the development of regional models (or limited-area models), which were capable of resolving mesoscale atmospheric features. In the Western Mediterranean Basin, characterized by strong mesoscale circulations, these regional models (executed with high-resolution meshes) are required to solve the atmospheric circulations/forcings.

During the last decade, the role of traditional supercomputers and workstations has been taken over by PCs. The huge PC market has benefited from the development of hardware technology and now shows an excellent price/performance ratio when compared with workstations. The growth of free software has facilitated the development of very productive software engineering capable of efficiently interconnecting computers and parallelizing the demands of computing power. Consequently, PC clusters are now able to respond to high-performance computing needs. This has been an important advance for most atmospheric research groups, giving them more computing power for mesoscale modelling purposes and at a lower cost.

One of our scientific objectives is to incorporate this new technology into the air pollution research being undertaken at our institution (at both national and international levels). One of our national research projects - the ‘Els Ports-El Maestrat’ project - is being used as a pilot case to fit these new technologies/methodologies to complex terrain areas with strong thermal and orographic forcings within the lower troposphere.

The ‘Els Ports-El Maestrat’ field campaigns are sponsored by the government of the Autonomous Community of Valencia, Spain (specifically the ‘Conselleria de Territori i Habitatge’ and the ‘Conselleria de Cultura, Educació i Sport’), and have been conducted on the south-western border of the Ebro basin (Northeast Iberian Peninsula) since November 1994. One of the objectives of these field campaigns is to monitor (both aloft and on the ground) the plume of sulfur dioxide (SO2) emitted from the 343m-tall chimney of the Andorra power plant located in Teruel (Spain). The ‘Els
Particulate matter (PM) is a mixture of particles that can adversely affect human health, damage materials and form atmospheric haze that degrades visibility. PM is usually divided up into different classes based on size, ranging from total suspended matter (TSP) to PM$_{10}$ (particles less than 10 microns in aerodynamic diameter) to PM$_{2.5}$ (particles less than 2.5 microns). In general, the smallest particles pose the highest human health risks. PM exposure can affect breathing, aggravate existing respiratory and cardiovascular disease, alter the body's defense systems against foreign materials, and damage lung tissue, contributing to cancer and premature death. Particulate matter includes dust, dirt, soot, smoke and liquid droplets directly emitted into the air by sources such as factories, power plants, cars,

Simulated dispersion results are generally checked against measurements of tracer-pollutant surface concentrations, with the dispersion analysis limited to the impact areas. The availability of measurements aloft enables us to verify the patterns of advection and turbulent diffusion that govern air pollution dynamics in the area. This is followed by an analysis of the cause-effect relation between the emission source and the ground-level concentration.

The mesoscale model uses a nested-grid configuration with five domains (100x100 grids spaced at 108, 36, 12, 4 and 1.3 km, respectively) centred over the power plant. The model predicts the wind field and turbulence parameters. The LPD model solves the turbulent wind components via the Markov process, which takes into account wind velocity variances and the three Langrangian autocorrelations. To solve the inhomogeneous turbulence in this complex terrain, the time step used is considered to be a function of the Langrangian time scale.

From the point of view of turbulent dispersion (whole-body advection + differential advection), the coupled models were able to reproduce the typical stationary-period advection, classical dispersion scenarios as experimentally characterized with the COSPEC. However, a significant temporal delay was detected between the simulation and experimental measurements of the plume dispersion (see Figures 1 and 2).

From the point of view of turbulent dispersion (differential advection + turbulent diffusion), there is a significant discrepancy during the transition between dispersion scenarios (see Figure 2), between the experimental and modelled values of the horizontal distribution of plume concentration (sy, defined from the transversal axis to the average transport direction). This stands in contrast to the situation during stationary periods (see Figure 1). In the former situation, with no defined transport direction, classical dispersion parameters lose their physical meaning.

In conclusion, using an adequate configuration of these two models (MM5 and FLEXPART) and with the methodology shown, it is possible to simulate/characterize the main meso-meteorological and dispersive features in complex terrain under very strong insolation conditions (where Gaussian regulatory models fail when applied to air pollution sources).

**2-Days Ahead PM$_{10}$ Prediction in Milan with Lazy Learning**

by Giorgio Corani and Stefano Barazzetta

The lazy learning algorithm is used to predict PM$_{10}$ air pollution levels in Milan, providing forecasts for the next two days with reasonable accuracy. In addition to the traditional data acquired by the air quality monitoring network, specific micrometeorological variables are algorithmically estimated and then used as input for our model.
construction activity, fires and natural windblown dust.

The ‘Air Sentinel’ project, developed by the ‘Agenzia Milanese per la Mobilità e l’Ambiente’ (Agency for Mobility and the Environment, Milan, Italy), aims at publishing forecasts of pollutant concentrations in Milan. One-day statistical linear predictors for different measuring stations of PM10 have already been shown to provide satisfactory performances. These models compute a prediction of the daily PM10 average for the current day at 9 a.m.. Evaluation of these predictors via cross-validation on a yearly scale shows a true/predicted correlation higher than 0.85, and a mean absolute error of about 10mg/m3 (out of a yearly PM10 average of about 43mg/m3).

We have now developed a two-day predictor; ie, at 9 a.m. this model predicts the daily PM10 average for the following day. We use the lazy learning (LL) prediction approach by Bontempi et al. LL has been shown to be viable for nonlinear time series prediction and, in particular, also for air pollution prediction. The strengths of the lazy learning approach are its predictive accuracy, its fast design (the development of a LL model is much quicker than a neural network), the easy model update procedure, and the readability of the model structure. The most relevant contribution to LL development and diffusion in recent years has been probably done by the Machine Learning group working at the University of Bruxelles, which continuously produces LL algorithmic enhancements and applications, and also releases the LL implementation as open-source code.

The yearly average of PM10 in Milan is substantially stable (about 45 mg/m3) since the beginning of monitoring in 1998 and, just to give an idea of the severity of the phenomenon, the PM10 daily average exceeds the limit (50mg/m3) about 100 times every year. PM10 concentrations in Milan follow a typical weekly and yearly pattern: in particular, winter concentrations are about as twice as high as summer ones, both because of unfavourable dispersion conditions and of the additional emissions due to residential heating. Sunday concentrations are about 25% lower than the other days of the week, because of the reduction in traffic volumes.

Our PM10 prediction application requires a careful investigation of the suitability of several variables; in addition to the classical data available from the air quality network (such as pollutant concentrations, wind speed and direction, temperature, atmospheric pressure etc., measured at the ground level), we also consider micrometeorological variables (such as mixing layer height, Monin-Obukhov length, etc.). These variables are algorithmically estimated and make it possible to characterize the dispersion conditions.

### Results and Discussion

We presents the results obtained by cross-validating the model. At each run, a year of data is used as testing set, in order to assess the ability of the model to predict previously unseen data. We compare the performances of the base model (whose inputs are PM10 (ie, autoregressive term), SO2, temperature and atmospheric pressure), with those of the model with micrometeorological data. The probability of detecting over-threshold values and the false alarm rate refer to a threshold value set at 50mg/m3.

Provided that predictions of the model are available as a range of values, rather than as a crisp number, the forecast accuracy can be considered satisfactory, especially if micrometeorological data are used. Figure 1 provides a simulation sample for January 2003.

### Links

- Website of the Machine Learning group at the University of Bruxelles: [http://www.ulb.ac.be/di/mlg/](http://www.ulb.ac.be/di/mlg/)
- Repository of papers presented at the 9th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (Garmisch, 2004). From among them, one can find the paper by Barazzetta and Corani dealing with 1-day ahead prediction of PM10: [http://www.harmo.org/conferences/Proceedings_/Garmisch/Garmisch_proceedings.asp](http://www.harmo.org/conferences/Proceedings_/Garmisch/Garmisch_proceedings.asp)

Please contact: Giorgio Corani, Stefano Barazzetta Agenzia Mobilità e Ambiente S.r.l., Italy E-mail: giorgio.corani@ama-mi.it, stefano.barazzetta@ama-mi.it

<table>
<thead>
<tr>
<th>Cross-validation year</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>True/predicted correlation</td>
<td>0.744</td>
<td>0.689</td>
<td>0.872</td>
<td>0.717</td>
<td>0.578</td>
<td>0.680</td>
</tr>
<tr>
<td>Mean abs. Err (mg/m3)</td>
<td>12.84</td>
<td>13.64</td>
<td>14.13</td>
<td>13.10</td>
<td>16.10</td>
<td>14.01</td>
</tr>
<tr>
<td>Mean error (mg/m3)</td>
<td>0.59</td>
<td>0.45</td>
<td>0.20</td>
<td>0.20</td>
<td>0.28</td>
<td>0.27</td>
</tr>
<tr>
<td>Prob. detection over-threshold values (%)</td>
<td>75.0%</td>
<td>70.3%</td>
<td>70.0%</td>
<td>65.0%</td>
<td>66.1%</td>
<td>69.32%</td>
</tr>
<tr>
<td>False alarm rate (%)</td>
<td>27.9%</td>
<td>31.1%</td>
<td>32.4%</td>
<td>34.5%</td>
<td>34.0%</td>
<td>32.00%</td>
</tr>
</tbody>
</table>

Table 1: Cross validation performances of the base model.

<table>
<thead>
<tr>
<th>Cross-validation year</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>True/predicted correlation</td>
<td>0.804</td>
<td>0.744</td>
<td>0.679</td>
<td>0.792</td>
<td>0.642</td>
<td>0.722</td>
</tr>
<tr>
<td>Mean abs. Err (mg/m3)</td>
<td>11.88</td>
<td>12.06</td>
<td>14.21</td>
<td>12.31</td>
<td>15.27</td>
<td>13.91</td>
</tr>
<tr>
<td>Mean error (mg/m3)</td>
<td>0.804</td>
<td>0.744</td>
<td>0.679</td>
<td>0.792</td>
<td>0.642</td>
<td>0.722</td>
</tr>
<tr>
<td>Prob. detection over-threshold values (%)</td>
<td>73.0%</td>
<td>79.50%</td>
<td>70.95%</td>
<td>68.37%</td>
<td>64.91%</td>
<td>71.50%</td>
</tr>
<tr>
<td>False alarm rate (%)</td>
<td>22.22%</td>
<td>26.47%</td>
<td>31.43%</td>
<td>29.20%</td>
<td>33.33%</td>
<td>28.33%</td>
</tr>
</tbody>
</table>

Table 2: Cross validation performances of the model enriched with micrometeorological indicators.

Figure 1: A simulation sample for January 2003.
Chaos was identified in meteorological systems in the 1960s by E.N. Lorenz. All solution curves of Lorenz’s equations eventually become confined to a bounded region of space, yet two different solutions with only slightly perturbed initial conditions diverge from one another at an exponential rate. This divergence has dire consequences for any attempt to simulate the weather or climate with a computer, since errors that are necessarily made in the process of representing the original, analytical model in discrete form may be seen as perturbations of the solution. These unavoidable errors will, under the influence of chaos, cause the simulated solution to diverge from the true solution exponentially. With current numerical weather prediction technology, model and observation errors overwhelm the solution in roughly ten days.

On the other hand, the errors incurred in a numerical computer simulation are not random but systematic, taking the form of phase errors, numerical damping and so forth. For well-designed methods, the simulated solution may, while diverging exponentially from the true trajectory, still remain close to or ‘shadow’ some other solution satisfying a modified initial state, modified parameters, or even a perturbed model. In such cases, the solution may still be meaningful when properly interpreted. In other words, one may still be accurately simulating some weather; it is just unlikely that it will be the actual observed weather.

To deal with chaos, modern predictions of weather and climate often utilize an ensemble of simulations: a whole series of runs in each of which the initial state of the system is slightly perturbed. This yields a distribution of weather scenarios that can be subjected to statistical methods, and allows one to make statements like: “In 73% of all trials, rain was predicted for Amsterdam next Thursday”.

When carrying out simulations on long-time intervals in which the numerical map is iterated, say, a hundred thousand to a hundred million times, the qualitative differences between algorithms become amplified. It is therefore crucial to employ methods having properties similar to those of the physical model being simulated (such as energy and mass conservation) or having similar mathematical structure (such as symmetries).

Research at CWI, in collaboration with Potsdam University, focuses on preservation of Hamiltonian structure in simulations of atmospheric models. Hamiltonian structure lies at the root of classical mechanical systems, and can be used to unveil many of the conservation laws present in such systems. Since viscosity can be neglected in large-scale atmospheric flows, the equations of motion are conservative, and from the Hamiltonian structure one can derive conservation laws of mass, energy and vorticity. By ensuring that the computational algorithm also adheres to this structure we can automatically guarantee that these conservation laws will also be inherited by the simulated solution. The preservation of Hamiltonian structure also has important implications for ensemble simulations. One can think of an ensemble of initial conditions as a set

---

Better Weather Forecasts with Use of Conservation Laws

by Jason Frank

Research at CWI aims to ensure structure in the midst of chaos. In long-time simulations of weather and climate, the Hamiltonian Particle Mesh method preserves Hamiltonian structure, and thereby energy, mass and vorticity. This technique could provide better quality solutions for weather forecasts.

Figure 1: Simulation of a 2D non-hydrostatic flow over a mountain ridge, computed using the HPM method. (Source: CWI.)
of points in space, each tracing out a solution. Initially the points are close together: for example, they form a ball. The ball is stretched and deformed as each point follows its solution trajectory. However, for Hamiltonian systems, the volume of the ball remains constant. This property is also inherited by the numerical method, and ensures that the members of an ensemble of simulations will each explore different possible scenarios, as opposed to being contracted to a smaller number of extreme solutions.

Preserving Hamiltonian structure using standard, purely grid-based computational techniques is difficult if not impossible. Instead we follow an approach in which a fluid is represented by a finite number of particles representing fluid masses, which satisfy classical mechanical equations of motion. For computational expediency, the potential field in which the particles move is represented on a grid. This is all carefully formulated to preserve Hamiltonian structure. The result is the Hamiltonian Particle-Mesh (HPM) method.

The HPM method has been under development since 2001. It has been used successfully for rotating planar and spherical shallow water equations, two-layer flows, and most recently, 2D vertical non-hydrostatic flow over topography. Figure 1 shows a simulation of a 2D non-hydrostatic flow over a mountain ridge using the HPM method, with the wind field (arrows) and potential temperature (color contours) indicated. Interest in non-hydrostatic models has grown in recent years as increased computational power has allowed higher grid resolutions to be attained. Current plans in this project address the issues surrounding 3D simulations.

Modelling Ecological Health using AI Techniques

by Martin Paysley and Bill Walley

The conservation of the natural environment is crucial to our long term survival. Scientists are now turning to artificial intelligence as a basis for better understanding of our surroundings.

The Centre for Intelligent Environmental Systems (CIES), Staffordshire University, UK, specialises in the application of artificial intelligence (AI) to problems affecting the natural environment. Projects to date have concentrated on the development of intelligent systems for the biological monitoring of river quality, with the Centre’s expertise in this field growing out of the pioneering work carried out by Bill Walley and Bert Hawkes in the early 1990s.

Biological monitoring of river quality has grown in importance over the past few decades and the approach now underpins legislation recently introduced in the EU, the Water Framework Directive (WFD). According to the WFD all freshwater bodies should be classified in terms of their ecology, with the target of reaching at least ‘good’ ecological status by 2015. The objective for CIES in the last few years has been to develop robust models for the interpretation of biological and environmental variables in terms of water chemistry and vice versa. Diagnosis is a key function; that is, determining likely pressures such as organic pollution, excess nutrients or acidity, from biological and environmental data. Prediction is also required; that is, forecasting the biological response to changes in pressures, perhaps as a result of modifications to land use or the treatment of waste water. Both of these functions are of fundamental importance for meeting the requirements of the WFD.

The methods used by CIES stem from the belief that experts use two complementary mental processes when diagnosis or prediction is required, namely probabilistic reasoning based upon their scientific knowledge; and pattern recognition based upon their experience of previous cases. Consequently, the group has followed two lines of AI research in parallel - probabilistic reasoning based on Bayesian methods and pattern recognition based on neural networks and information theory.

The robust and holistic nature of Bayesian belief networks (BBNs) makes them well-suited to modelling complex systems, like river ecology, where the interaction between the elements is probabilistic in nature. This uncertainty is represented explicitly in the model by probability distributions that encode cause-effect relationships. CIES has used BBN technology to develop, in collaboration with the Environment Agency for England and Wales, a model of river ecology called RPBBN (River Pollution Bayesian Belief Network). The strength of this approach has been that the model can be used to diagnose the pressures that affect a site and/or predict the ecological changes resulting from a programme of remedial measures.

The group has also developed its own pattern recognition system, MIR-Max (Mutual Information and Regression Maximisation), based upon information theory. This is now central to its pattern recognition systems and forms the basis of an application called RPDS (River
Pollution Diagnostic System (OPDS) is a prototype computerised diagnostic tool designed to help the Environment Agency to determine the type of oil product involved, then to “fingerprint” it to identify its source and to provide solid evidence for legal proceedings against the polluter. This is based on detailed matching of patterns in the gas chromatograms of the pollutant and the original product, after allowing for the effects of weathering. The pattern matching techniques of the OPDS can assist scientists in cases where the chromatograms are particularly complex and matches are hard to achieve.

Link: http://www.cies.staffs.ac.uk.

Please contact:
Martin Paisley
E-mail: m.f.paisley@staffs.ac.uk
David Trigg
E-mail: d.j.trigg@staffs.ac.uk
Rob Buxton
E-mail: r.buxton@staffs.ac.uk
Bill Walley
E-mail: wjwalley@globalnet.co.uk.
Tel: +44 (0) 1785 353510

Mathematical Models for the Simulation of Environmental Flows: From the Strait of Gibraltar to the Aznalcollar Disaster

by Carlos Parés, Jorge Macías, and Manuel J. Castro

The realistic simulation of environmental flows is a subject of major economic, social, human and scientific interest. Human interaction with natural flows, through adding pollutants or being responsible for disasters such as the dam-break of Aznalcollar, feeds the necessity for numerical modelling and prediction.

The DAMFLOW project is looking at the development of efficient techniques for the numerical solution of hydrodynamic flows, with a particular focus on environmental applications. Among the project’s aims is the development of robust and reliable numerical tools with low computational cost, which can predict and simulate hazards or emergency situations such as floods or oil spills. These numerical models may also assist, in some particular cases, in understanding the physical nature of the problem and the key processes involved.

Currently, a set of fascinating problems has attracted our attention. One of these is the two-layer water masses exchange through the Strait of Gibraltar. The confinement of the flow by a strait can give rise to profound dynamic consequences including choking or hydraulic control, a process similar to that by which a dam regulates the flow from a reservoir. The funnelling geometry can lead to enhanced tidal modulation and increased velocities, giving rise to local instabilities, mixing, internal bores, jumps and other striking hydraulic and fine-scale phenomena. In short, sea straits represent choke points which are observationally and dynamically strategic and which contain a range of interesting processes. The flow configuration in the Strait of Gibraltar is characterized by two counter-currents: at the
surface the less saline water of the Atlantic flows eastward, spreading into the Mediterranean and, at depth, the waters of the Mediterranean flow westward toward the Atlantic Ocean. This situation means that it is possible and useful to use two-layer models to represent the region’s dynamics and to better understand the key processes involved. As result of the DAMFLOW project, a set of finite-volume two-layer shallow water models have been developed, comprehensively tested, and used to numerically simulate the dynamics of the Strait of Gibraltar. One-dimensional tidal modelling of the Strait (see Figure 1) revealed a complicated pattern of time-dependent hydraulic fluctuations involving changing interfacial levels, moving control points and reversal of the layer flows at different stages of the tide, in good agreement with observations. Two-dimensional modelling (see Figure 2) produces a more complex picture of tidal dynamics for the interpretation of which several numerical and graphical tools have also been developed.

A second environmental problem that deserved our attention was the simulation of the Aznalcollar dam-break. On 25 April 1998, a massive spillage of tailings occurred from a pyrites mine lagoon at Aznalcollar (Seville, Spain) and damaged the environmentally sensitive Doñana region. This caused an estimated economic cost of 300 million euros and an unquantifiable cost in terms of environmental damage. Our project has focused on reproducing the Aznalcollar dam-break conditions in order to assess our model’s performance. First, a digital model of the affected area was made from GIS data. A two-dimensional one-layer finite volume model was used (see Figure 3), which showed robust behaviour and produced good results; propagation speeds, the flooded area (see Figure 4) and draught sections were in good agreement with the available data. The model had previously been validated against test cases with analytical solutions and laboratory data. The results provided by the model have given us confidence that this numerical tool can produce and provide a full description of a potential flood wave and flooded area in other scenarios of potential risk.

At the present stage of the project:
• robust 1D and 2D finite-volume one- and two-layer solvers have been implemented, taking into account realistic topographical data
• an exhaustive model validation has been undertaken
• rigorous comparisons have been made with observed data
• input data for model problems were taken from observed measurements
• sensibility studies have been performed on various physical parameters such as friction or density ratio in oceanographic applications
• fortnightly and monthly signals have been investigated in model variables (Strait of Gibraltar).

Numerical modelling and simulation require, besides algorithmic research and implementation, the use of ‘interfaces’ and other technological developments. In the framework of this project we have also undertaken the following tasks:
• development of pre-processing, post-processing, analysis and visualization tools
• parallelization of the numerical schemes on a PC cluster, providing good efficiency
• use of mesh adaptation techniques.

This research is partially funded by the C.I.C.Y.T. (Comisión Interministerial de Ciencia y Tecnología), project BFM2003-07530-C02-02, which commenced in December 2003 and will end in December 2006. The research team involved in this project belongs to the Differential Equations, Numerical Analysis and Applications group at the University of Málaga. Close collaborations exist with the Numerical Analysis in Fluid Mechanics group at Seville University led by Tomás Chacón, and other research groups at the Universities of Santiago de Compostela, Zaragoza, Cádiz, I.E.O. (Spanish Institute of Oceanography), CITEEC (University of La Coruña) and Université de Corse (France).

Link: http://www.damflow.com

Please contact:
Carlos Paredes, Universidad de Málaga, Spain
Tel: +34 9 5213 2017
E-mail: pares@anamat.cie.uma.es

by Fadi El Dabaghi

The AWIIS is a Spatial Decision Support System for rational planning, operation and management involving water resources. It provides the user with a customized overview or partial view depending on the user profile and the use context. This is achieved through dynamic ontology, using the concept of relevance feedback and a thesaurus based on substitutes of documents.

The management of water resources is a highly complex and tedious task that involves multidisciplinary domains including data acquisition and treatment, multi-processes and therefore multi-models, numerical modelling, optimization, data warehousing, and the analysis of exploitation, socio-economic, environmental and legal issues. In addition, given water’s vital role in human life, the continuous decrease in its availability due to over-exploitation and pollution, and its scarcity in arid areas, there is a clear need for better design and optimal use of hydro-systems. In this context, a rational analysis must be based on an approach that considers all related causes and effects and systematically evaluates the various alternatives. These include preserving water for irrigation, managing watersheds, building dams to alleviate the impact of floods and improve the water supply, simulating floods for risk prevention, and oxygenating lakes to eliminate eutrophication. For this, we have developed a Web-Integrated Information System (WIIS) dedicated to the management and modelling of water resources. This system is an infrastructure of development: perennial, open, modular, robust, effective and convivial, offering a complete, dynamic and configurable environment.

The WIIS architecture consists in two levels: the first contains general modules such as viewers, GIS, grid generators, a data warehouse and simulators, while the other is domain-specific and includes all the relevant modules tackling numerical modelling for the water-related physical application under study, ie hydrology, hydraulics, hydrodynamics and floods. This architecture is based on integration principles linking software components with heterogeneous data sources. Within this objective, we confront an important conflict. On the one hand, there is a considerable need for flexibility in the simulation of water phenomena and the representation and handling of input/output data, which are characterized by the absence of fixed and rigid structures. On the other hand, these data, exchanged within the WIIS, are a priori unstructured, heterogeneous and distributive, which means they are not easily accessible or exploitable. Within this framework, we looked at the definition of a model for the representation and structuring of the data using XML techniques. Accents were put on both the customization of the WIIS and its adaptation according to the user profile and the context of use.

In the first stage, we led a pseudo-XMLization of all the data accessed or exchanged within the WIIS through an Information Indexation and Research System (IIRS). This IIRS operates on a thesaurus based on Substitutes of Documents under XML structure of the data warehouse, characterized by strongly heterogeneous and distributed contents. The second stage aimed at setting up an Adaptive Web-Integrated Information System (AWIIS), which is based on the WIIS and is robust and effective in terms of integrated tools and data. This AWIIS must provide a convivial, adaptive interface, rich in information, and possibly with an overview or various partial views according to the situation. This customization is closely dependent on the user profile and the context of use. The AWIIS will also integrate a PIRS (Personalized Information Research System), which generalizes on the IIRS. It should be dynamic and evolutionary, use the concept of relevance feedback, and its driving idea is the combination...
Numerical Modelling and Analysis of Water Free Surface Flows

by Fadi El Dabaghi

A number of environmental engineering applications related to water resources involve unsteady free surface flows. A full three-dimensional model based on Navier-Stokes equations can give good descriptions of the physical features of certain phenomena, including lake eutrophication, transport of pollutants, floods and watersheds. However these models are characterized by significant computational cost. We aim to reduce this through the use of two-dimensional models or appropriate coupling models of different dimensions.

The models developed in this context can be classified into two categories: a two-phase flow model based on Navier-Stokes equations, and a shallow-water flow model.

Two-phase flow models have been used to simulate the remedial aeration used to combat eutrophication effects in lakes. A water reservoir is generally considered eutrophized when the concentration of dissolved oxygen drops below 3 mg/L. The idea behind remedial aeration is to inject compressed air into the bottom of the reservoir in order to stir up and oxygenate the water. The numerical simulation of the resulting flow by conventional models such as the two-fluids or Lagrangian models leads to many difficulties. This is mainly due to the models’ complexity and the need for a fine grid in order to achieve a good representation of the effect of the bubbles.

These difficulties limit interest in these classical models and lead us to suggest some cheap and realistic alternatives. We consider a one-phase flow model based on velocity-pressure semi-compressible Navier-Stokes equations. Here, bubble dynamics are taken into account, firstly through boundary conditions related to the air injection velocity at the aerator position, and secondly by introducing correction terms representing the forces applied by the bubbles to the water. In the same framework, a more general two-fluids (air/water) model with a moving water free surface has been developed. We use realistic assumptions regarding the wind velocity and the atmospheric pressure with a convection equation describing the void fraction function of the water, which determines the wet domain. This allows us to treat the free surface effects of the lake in a more suitable manner, and consequently improve the dynamic results of the bubbles’ aeration effects.

The second category of model is the shallow-water model: these describe river flooding, tidal fluctuations, bay and estuary flows, breaking waves on shallow beaches etc. They are derived from 3D incompressible Navier-Stokes equations by depth-averaging of the continuum mass and momentum balances. These models, known also as Saint-Venant, involve fluid-domain geometries characterized by their complexity and variability, and the large-scale computational aspects and fine grid necessary for realistic simulations.

To overcome these problems, we have used a priori and a posteriori error analysis. Geometric error indicators have been derived to improve the numerical simulations models by adaptive mesh techniques; these ensure an optimal quality solution at a required precision for a given computational cost. Moreover, the error indicators accurately.

This work was supported by the WADI Euro-Mediterranean Project and French bilateral cooperation programmes (CMIFM, CMEP and PLATON). The WADI project was managed by ERCIM through a consortium consisting of INRIA-France, FORTH/IACM-Greece, EMI/ONEP-Morocco, ENP-Algeria, Calabria University-Italy and ESIB/CREEN/EUCLID-Lebanon.

Please contact:
Fadi El Dabaghi, INRIA, France
Tel:+33 1 3963 5343.
E-mail: fadi.el_dabaghi@inria.fr
predict in which parts of the fluid domain the flow model may be considered 1D, 2D or 3D, and can couple the resulting multi-dimensional models. In addition, we have integrated a conceptual hydrological model based on the HEC-HMS: this model computes run-off volume by subtracting the volume of water that is intercepted, infiltrated, stored, evaporated, or transpired from the precipitation in the catchment area. The result of this step, a flood hydrogram, forms the boundary conditions for the numerical models presented above (St-Venant) in simulating the flood.

From the numerical point of view, the approximation of the models mentioned above is based on the characteristics method for the time discretization of the advection terms. This method leads to an upwind scheme, which has the double advantage of being physically closed to convection, and being an explicit scheme unconditionally stable in the finite element context, thereby allowing the use of large and reasonable time steps. At each time level, we have to solve a quasi-Stokes problem, approximated by $P_1/P_1$ mixed finite elements for the velocity-height St-Venant formulation, or by $(P_1 + \text{bubble } /P_1)$ mixed finite elements for the velocity-pressure Navier-Stokes formulation. This ensures the discrete LBB condition necessary in this context. For both formulations, the a priori error estimates are verified numerically on many academic test cases before the models are validated on real cases such rivers in flood or crossed 2D lake sections. Moreover, given that the real-time and large-scale computing requirements are very important for this kind of simulation, we used the HPCN facilities; this permitted the execution of developed codes, particularly on real test cases requiring finer and more complex meshes.

This work was supported by many Euro-Mediterranean projects (WADI, ESIMEAU and CruCID) and French bilateral cooperation programmes (CMIFM, CMEP and PLATON). The EU projects were managed by ERCIM through a partnership constituted by INRIA-France, FORTH/IACM-Greece, EMI/ONEP-Morocco, ENP-Algeria, ENIT-Tunisia, RIKS-Holland, Calabria University-Italy and ESIB/CREEN/EUCLID-Lebanon. Please contact: Fadi El Dabaghi, INRIA, France Tel: +33 1 3963 5343 E-mail: fadi.el_dabaghi@inria.fr

SACADEAU: A Decision-Aid System to Improve Stream-Water Quality

by Marie-Odile Cordier

Water quality is a critical environmental issue. In this project, we use a pesticide transfer model to simulate effects on the quality of stream-water. Since a large number of parameters are involved in pollution phenomena, modelling, simulation and machine learning are useful techniques for acquiring knowledge in this poorly understood domain.

The objective of the SACADEAU is to build a decision-aid tool to help specialists in charge of catchment area management to preserve stream-water quality. This is done by coupling a qualitative transfer model (simulating pesticide transfer through the catchment area) with a qualitative management model (simulating farmers’ decisions concerning weeding strategies and herbicide application). This has two main advantages: it allows the impact of recommendations to be evaluated by simulating high-level scenarios, and it allows simulation results to be analysed by using machine learning and data mining techniques to discover discriminating variables and to acquire knowledge in this poorly understood domain.

The experimentation site (the Fremeur catchment area) is located in Brittany, France and covers about seventeen square kilometres.

A Transfer Model Coupled with Three Input Models

The transfer model simulates river contamination by pesticides. It models pesticide transfer through a catchment area and simulates the daily river contamination that this causes. This phenomenon depends on numerous
parameters, including human activities, climate, soil type and catchment area topology. Since these parameters are complex and difficult to formalize, we created three sub-models to describe them (see Figure).

These are as follows:

- A decision model, which models farmers’ strategies. This provides herbicide application characteristics (date, substance, quantity) and agricultural interventions (soil preparation, seeding date, weeding dates) according to predefined farmers’ strategies and weather conditions.
- A climate model, which provides daily weather data such as the temperature and the quantity of rainwater.
- A spatial model, which is in charge of the spatial distribution of agricultural activities, according to the catchment area topology.

Using the outputs of these three sub-models, a biophysical transfer model determines pesticide transfer from application locations, through the catchment area, to the river. The model takes into consideration all the possible ways in which rainwater can flow through the catchment area (run-off and leaching).

A High-Level Language

In order to achieve qualitative results, a high-level language for inputs and outputs of the model is required. The aim is to describe qualitatively, via a scenario, numerical inputs and outputs of the model. This can be seen as the process of discretization of quantitative data. This process is fundamental if we want to construct comprehensive results for decision-makers. The initial step was to gather a set of scenarios suggested by experts; for example, “What happens if a post-emergence weeding strategy rather than a pre-emergence strategy is applied on all plots close to the river?” or “What is the impact of pesticide application dates on stream-water quality?”

To simulate a scenario, a methodology was defined that consists in generating a large set of instances of the scenario. These instances are then simulated and the results generalized to give a qualitative description (in response to a qualitative question) using machine-learning techniques. For example, given the question above concerning the impact of application dates, a response could be: “Concentration peaks appear when pesticide application dates are close (less than two days) to significant showers (quantity > 10mm)”.

Learning from Simulation Results

The global model generates pesticide quantities and concentrations according to the parameters mentioned above. The set of simulation inputs and outputs is called an example in machine-learning vocabulary, and the set of examples is a set of instances of a scenario we have simulated. We used ICL (http://www.cs.kuleuven.ac.be/~wimv/ICL/), inductive logic programming software, to learn a set of rules summarizing the examples. These rules can be formatted according to the task being addressed. For example:

- qualitatively predicting water pollution
- identifying which variables play an important role in water pollution
- characterizing important risks (eg whether pollution is due to too many concentration peaks, or to a constant and significant source of pollution).

First results have been obtained with a simplified model, and according to the experts, they show both expected and surprising relationships. For instance, concerning stream-water quality, a post-emergence weeding strategy did not show better results than a pre-emergence strategy. This was unexpected, and a discussion on the impact of weeding strategies was generated, with some possible explanations given by experts.

Future work is three-fold, and involves validating the model, refining the high-level language, and providing recommendations to experts based on relationships discovered through the model.

The SACADEAU project is in its third year of development. Contributing members are M. O. Cordier, V. Masson, A. Salleb (IRISA/Univ. Rennes 1, Rennes), C. Gascuel-Odoux, F. Tortrat, P. Aurousseau, R. Trepos (INRA-ENSAR/UMR SAS, Rennes), F. Garcia (INRA/BIA, Castanet Tolosan), B. Chanomordic (INRA/LASB, Montpellier), M. Falchier, D. Heddadj and L. Lebouille (Chambre d’agriculture de Bretagne). SACADEAU is funded by Conseil Général du Morbihan and INRA.

Link:
http://www.irisa.fr/dream

Please contact:
Marie-Odile Cordier, Université Rennes 1/IRISA, France
Tel : +33 2 9984 7100
E-mail: Marie-Odile.Cordier@irisa.fr
Developing an Environmental Modelling Framework for Integrated Assessment of EU Agricultural Policies

by Andrea E. Rizzoli, Carlo Lepori, Roberto Mastropietro and Lorenzo Sommaruga

‘SEAMLESS’, System for Environmental and Agricultural Modelling; Linking European Science and Society, is an environmental modelling project aimed at creating a framework, named SEAMFRAME, which is specifically targeted for integrated assessment in the agricultural domain. The goal is to deliver a software toolkit to facilitate the comparison and integration of the many different models in existence today.

European agriculture is currently undergoing radical changes due to economical pressures imposed by the enlargement of the EU, the revision of WTO agreements, and changes in farm support payments. Such changes interact with the physical and natural environment (eg, climate change, loss of biodiversity). Meanwhile, society demands a green and clean landscape, and farming communities in rural areas are faced with continuous technological innovation.

Scientists and modelers are therefore confronted with the increasing need to deliver scientific results, which can be used by policy makers to assess new policies, at the regional and national scale, that can facilitate agriculture’s contribution to sustainable development. The process of evaluation of a policy with respect to the three components of sustainable development (environment, economy, society) requires what has been defined ‘integrated assessment’ (Parker et al. 2002).

Yet, the obstacles to performing integrated assessment studies are substantial, given the dimension and the complexity of the problem. Integration requires that researchers from different disciplines merge and compare their ideas, often scaling up and down the dimension of the problem under study. For instance, in the case of the integrated assessment of agricultural policies, the effect of a subsidy for a given product has a different impact in the various regions, and the related change in farm activity has a very local effect on the environment.

Science, through innovation and research, has progressively built a vast body of knowledge specific to various sectors, which can now be considered as components of an integrated system. For instance, for a given crop growth process, there may be tens of different formulations solving the same modelling problem. Thus, the main issue becomes to make this knowledge accessible and re-usable. In fact, computer models are monolithic, difficult to reuse in a different context, and strongly tied to the modelling domain and scale of resolution. For this reason, many groups have been researching the problem of model integration and modelling frameworks to allow a ‘seamless’ and transparent model integration and re-use across different scientific domains and scales (eg, OpenMI (http://www.harmonit.org/), TIME (http://www.toolkit.net.au/), OMS (http://oms.ars.usda.gov/) just to cite a few).

A modelling framework is basically a software environment that supports the development of models, but it also provides a set of facilities to set up model runs, data visualisation and analysis, documentation, archival and discovery of models. A modelling framework also supports the packaging of end-user applications, which can be effectively used to support decision making. An environmental modelling framework adds domain-specific knowledge that enables solving environmental-related issues, for instance, it considers spatial modelling.

A new Integrated Project “SEAMLESS: System for Environmental and Agricultural Modelling; Linking European Science and Society” aims to develop an environmental modelling framework, named SEAMFRAME, which is specifically targeted for integrated assessment in the agricultural domain. This project was defined as a result of a research call from the European Commission within the EU 6th Framework Programme (Global Change and Ecosystems).
SEAMFRAME will capitalise on previous research in the development of environmental modelling frameworks. Moreover, it will support a ‘declarative’ approach to modelling (Muetzelfeldt and Massheder, 2003). A clear-cut separation between model representation (equations) and model manipulation (data processing, integration routines) will allow the effective re-use of the model in different contexts. The model equations will be semantically denoted, by means of Semantic Web technologies such as RDF and Ontologies. This will open the model structure to automated processing, leading to the possibility of searching models according to their specifications — and model linking and composition will be greatly facilitated.

Finally, another distinctive feature of SEAMFRAME will be the component-oriented approach of its software architecture. Declarative models will be deployed as software components, which can be manipulated by tools such as calibrators, simulators, optimisation routines, which are themselves software components. The use of introspection (Rahman et al. 2004) will allow tools to adapt to the published interface of models, thus granting the required flexibility and extensibility of the framework.

Thirty research institutions from thirteen European countries, including several new member states, are involved in the project. These institutions bring together a vast amount of knowledge and expertise from economic, environmental, agronomic, social and information technology disciplines. The project also includes co-operation with an African and an American research institute. The total budget is 15 million Euros. In 18 months the first prototype should be available and in four years the system should be fully operational. The project is coordinated from Wageningen University (The Netherlands), while the development of SEAMFRAME is coordinated by the Swiss institute IDSIA, part of USI (University of Lugano) and SUPSI (the University of Applied Sciences of Southern Switzerland).

Image Processing for Forest Monitoring

by Josiane Zerubia and Paul-Henry Cournede

Aerial and satellite imagery have a key role to play in forestry management. The increasing availability of data and their high spatial resolution, which is now submetric, allow automatic tools to be developed that can analyse and monitor forests by accurately evaluating the vegetation resources. The Ariana research group at INRIA Sophia Antipolis, with its strong experience in remote-sensing image analysis and feature extraction, is working on this topic as part of the joint research effort Mode de Vie.

Digitized aerial photographs and satellite images of forests represent convenient data for developing computerized assessments of forestry resources. Such automatic tools are useful for a number of reasons, including the help with which they provide forest managers in classifying species. Such work is currently done by specialists, using difficult image analysis combined with ground verifications. Some tools already exist for this purpose, and use texture information and classification based on parameters such as covariance matrices. However, few take advantage of high data resolution. Nowadays, it is possible to study forests on the scale of individual trees, by resolving the extraction of tree crowns. This is one of the aims of the Ariana research group, which in the last year has adapted its knowledge in stochastic geometry (object processes) to forest-resource evaluation. This will allow the automatic assessment of economically and environmentally important parameters such as the number of tree crowns, the distribution of their diameters or the stem density.

Our approach consists in modelling the forestry images as realizations of a marked point process of trees. This stochastic framework aims at finding the best configuration of an unknown number of geometrical objects in the image, with respect to a probability density defined a priori. This density takes into account both the data, in order to fit the objects to the feature we want to extract, and the interactions between these objects, to favour or penalize some arrangements. In the case of tree-crown extraction, we modelled the trees as ellipses, defined by the position of their centre, their orientation, and their major and minor axes. After simulation, we obtain a collection of objects and have access to several statistics such as the number of trees in the stand, their position and their diameter. If different species coexist, we can add a mark to the objects to describe their type, and obtain a classification during the extraction. Some tests have been performed on digitized aerial photographs of stands of poplars, courtesy of the French Forest Inventory (IFN). These photographs are taken in the infrared, which enhances the chlorophyll matter of vegetation. In future investigations, we will be studying texture parameters in order to distinguish different species during the simulation, and lower-level information to obtain a pre-segmentation of the image before the extraction process.

Another application relevant to forest monitoring has been developed by Ariana in collaboration with Alcatel...
Space, and addresses forest-fire detection on satellite images, using the random fields theory. It consists in modelling the image as a realization of a Gaussian field, in order to extract rare events like potential fires that could grow and imply serious damage. For this purpose, the thermal infrared channel (TIR) is selected, since fires show up as peaks of intensity at these wavelengths.

Of further interest is the involvement of Ariana in the ARC joint research effort Mode de Vie, in collaboration with MAS Laboratory (Ecole Centrale Paris), the Digiplante research group (INRIA Rocquencourt, CIRAD), and LIAMA (Sino-French Laboratory of Informatics and Automation, Academy of Sciences, Beijing, China). The purpose of this joint action is to link the assessment of forestry resources with the dynamical models of plant growth developed in Digiplante and LIAMA. The ultimate target would be to develop a complete tool for vegetation monitoring, which could both evaluate the present biomass and predict its evolution.

Figure 1: (left) In the original image (courtesy of IFN), the plantation of poplars on the left-hand side can be isolated thanks to a Gabor filter; (right) the result of the extraction process by stochastic geometry (Guillaume Perrin).

GreenLab: A Dynamical Model of Plant Growth for Environmental Applications

by Paul-Henry Cournède and Philippe de Reffye

Some of the most critical issues for sustainable development and the protection of the environment concern the rationalization of agriculture and forest exploitation. For this purpose, a research project at INRIA, DigiPlant, aims at developing both mathematical models of plant growth and related computing tools. Over the years, an important Sino-European collaboration has emerged, involving research institutes and universities, mathematicians and biologists.

The cultivated areas of Europe, including agricultural land and exploitation forests, have a strong impact on global environmental conditions. Erosion, resource impoverishment due to over-exploitation, and pollution by fertilizers or pesticides are crucial problems that agronomy and forestry hope to solve through harmonious cultivation modes and exploitation strategies. For this purpose, they must take into account production needs on one hand and the environment on the other; that is to say, both quantitative and qualitative criteria. In this context, mathematical models of plant growth describing interactions between the architecture of the plant and its physiological functioning have a key role to play. They allow the exchanges (of water, carbon, minerals etc) between plants and their natural environment to be quantified.

GreenLab is just such a functional-structural model, and is the result of a long dialogue between botanists, physiologists and mathematicians. Derived from the AMAP model developed in the 1990s at CIRAD, GreenLab’s new formulation was introduced at LIAMA (Beijing) in 2000. Today, the model is studied and improved upon through the DigiPlant project, which is run by a joint team of researchers from INRIA, CIRAD and Ecole Centrale Paris, and hosted by INRIA. Some very close partnerships exist with LIAMA, China Agriculture University, Wageningen University, Orsay University and INRA.

A number of choices have been made in order to simplify biological knowledge and write the dynamical equations of growth. Organogenetic growth cycles are defined, and an automaton describing the evolution of the buds determines the plant architecture. The botanical concept of physiological age, defining a typology
of the axes, allows a powerful factorization of the plant structure. Fresh biomass production is computed from transpiration and is then distributed among expanding organs (including buds) according to their demands (sinks). The available biomass determines the rules of the organogenetic automaton and the organ sizes, and plant architecture has a strong influence on photosynthesis. Thus, there is a very strong coupling between the growth equations.

Stresses are also taken into account, especially for environmental resources. In particular, we have studied the competition between plants for light and water in plantations or crop fields. Plant density is also an important control parameter of the model.

The introduced mathematical formalism had a significant impact on the model analysis. First, the simulation software is strictly derived from the mathematical formalism. Thanks to the structural factorization, the computation time grows linearly with the chronological age of the plant (in comparison with a bud-by-bud simulation whose computing time grows exponentially with the plant’s chronological age). Moreover, the mathematical study of the system allowed theoretical results on plant viability and yield prediction to be obtained. The best prospects, however, rely on our ability to clearly formulate optimal control problems and to fit the theoretical models to real plants. These two points are the keys to applications in agronomy and forestry.

The hidden parameters of the model are very few (for example, twelve for an annual plant like maize). Agronomic data on maize, sunflowers, tomato, rice and wheat have been fitted with very good accuracy, and fittings on different plants of the same type have shown that the parameters are very stable. Some more complex plants like coffee, cotton and young beech trees are currently being studied.

We have developed tools for a variety of objectives:
- Optimization and control of the cultivation modes: in the case of limited resources, there is an optimal strategy of fertilizing and watering during plant growth. Likewise, controlling plant density or partial forest clearings can be beneficial. In this way, we can improve water resources and land management and reduce pollution by fertilizers.
- Control of plant sanitation and pesticides treatment: by coupling the plant-growth model and insect-population dynamics, we can control the use of pesticides and thus reduce costs and pollution.
- Selection of crop variety: we are currently working with geneticists, in order to prove that the plant genes directly determine the physiological parameters of the GreenLab model. In this way, we expect to propose better strategies for crop selection.
- Virtual simulation and visualization of plantations: computer graphics techniques allow the results of numerical simulations to be visualized. This is very important in urbanism or landscaping for predicting the long-term evolution of projects.

The results of this research seem to show that in the near future, new tools of prediction, optimization and control could be effectively used in agriculture and forest exploitation on a large scale, and would drastically improve the management of the environment.

Links:
http://www.inria.fr/recherche/equipes/digiplante.en.html
http://www.mas.ecp.fr
http://amap.cirad.fr

Please contact:
Paul-Henry Courmède, MAS - Ecole Centrale Paris, France
Tel: +33 141 131 289
E-mail: courmede@ecp.fr
Philippe de Reffye, INRIA, AMAP - CIRAD
Tel: +33 139 635 774
E-mail: Philippe.de_Reffye@inria.fr
Lossless Compression of Meteorological Data

by Rodrigo Iza-Teran and Rudolph Lorentz

Weather forecasts are getting better and better. Why? The models include more physical processes and resolution has been improved. The consequence: more data is generated by every weather forecast. Much more data. Very much more data. What does one do with all this data? Answer: compress it.

In the Department for Numerical Software (NUSO) at the Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), a new group has been founded to look at the compression of data resulting from numerical simulations. One of the fields in which it has been active is the compression of meteorological data. The motivation was the drastically increased amounts of meteorological data needing to be stored. This is a consequence of improved weather models and of the higher resolution with which forecasts are being made. However the data is also being used for new purposes: direct sales, data-mining and reanalyses. All of these tend to increase the amount of data that is archived for longer periods of time.

The initial impetus was provided by the plans of the German Weather Service (Deutscher Wetterdienst) to switch from their Local Model (LM) to the Local Model-Europe (LME). This new model not only covers a larger area but also has a higher vertical resolution. The amount of data they plan to archive can be seen in the graph below. Medium-term storage requirements, ie for about a year, are planned to be just short of 4 Petabytes. That is 4,000,000,000,000,000 bytes.

Compression

Compression involves changing the form of data in a file, so that the compressed file takes up less space than the original file. There are two types of compression: lossy and lossless. If the compression is lossy, one cannot retrieve the original file from the compressed file. The advantage of this approach is that the data can be compressed to a much greater degree. Lossy compression is typically used to compress graphic and video files, especially in the context of the Internet. Typical lossy compression programs are JPEG and MPEG, which can reduce the size of a file by factors of ten to fifty.

On the other hand, if lossless compression is used, the original file can be retrieved exactly from the compressed file. Lossless compression is typical for text files, but also for files containing sensitive numerical data, eg medical or meteorological data. All Zip utilities perform lossless compression, with compression factors of around 1.5 to 3.

Meteorological data is typically stored in the GRIB format. This format, gridded data in binary form, is an international standard for the storage and exchange of meteorological data. The usual procedure is to first put the data obtained from a weather forecast into the GRIB format. This is lossy compression. Afterwards, however, any compression is required to be lossless. The new group established at SCAI has developed a program, GRIBzip, for the lossless compression of meteorological data stored in the GRIB1 format (ie GRIB Version 1). As an example, if the data is formatted in GRIB1 with 16-bit precision, then the GRIB files produced by a typical weather forecast can be reduced in size losslessly by, on average, a factor of three. Extrapolating this to the example of the German Weather Service, storage would be required for only 1.5 Petabytes of data instead of 4 Petabytes.

Archiving data in a compressed form has another advantage that may not be immediately apparent. Normally the archiving system consists of hardware separate from the other computers, meaning the connection to the archiving system can be a bottleneck. By using compressed data, the bandwidth of the connection is effectively increased by the same factor as the data has been compressed. In our example, the bandwidth would be effectively increased by a factor of three. A patent for the techniques used in the program has been applied for.

Future Activities

The programs developed in SCAI can compress GRIB data on topologically rectangular grids. Other types of grids, such as a triangular grid covering the whole globe, or grids which become sparser towards the poles, are also allowed by the GRIB format. We are planning programs able to losslessly compress this type of data. In addition, spectral data is sometimes stored in the GRIB format, and its compression is also desirable.

The group has also developed programs for compressing data produced by crash simulations (Pamcrash, LS-Dyna) in the automobile industry. Since this data is defined on irregular grids, the techniques used are quite different from those for data on regular grids. Using the expertise gained from these extreme cases, the group intends to compress data resulting from other types of simulations.

Link: http://www.scai.fraunhofer.de/nuso.0.html?&L=1

Please contact: Rudolph Lorentz, Institute for Algorithms and Scientific Computing - SCAI, Fraunhofer ICT Group
Tel: +49 (0) 2241 143 480
E-mail: Rudolph.Lorentz@scai.fraunhofer.de

Planned archive size, German Weather Service.
Air-Quality Information for Multi-Channel Services: On the Air with Mobile Phones

by Gertraud Peinel and Thomas Rose

The APNEE project (Air Pollution Network for Early warning and information exchange in Europe) has designed customizable services providing information on air quality for citizens. These services draw on various information channels, including mobile technology, interactive portals for the Internet and street panels.

Based on such high-quality dissemination services, even environmental information becomes an attractive product once it is perceived as providing citizens with indicators of their levels of comfort. The APNEE information services are based on the following concepts:

- public sector information is promoted as catalytic content for new and compelling applications
- citizens require information on air quality in a user-friendly form
- EU directives are calling for new ways to disseminate information on air quality to citizens.

Authorities and scientists have large amounts of data with perfect means of visualization and well-researched models for calculation, forecasts and interpolations. Normally however, normal citizens are not targeted as potential consumers of these data. Yet citizens are now beginning to call for timely and high-quality environmental information for reasons of comfort and health care, especially when they are individually affected by respiratory diseases. While general information is available via Internet servers and TV video text, some things are still lacking:

- classification of measurement values by user-friendly indexes (ie ‘good’ or ‘bad’ instead of, say, 77 mg)
- information regarding health impact and what action to take
- appropriate channels through which to reach citizens on the move (ie anytime and anywhere).

To date, most warnings of adverse environmental conditions have made use of broadcast channels. However, radio or TV inevitably reaches a much larger audience than that section of the population actually at risk. National and regional authorities have found it difficult to ‘narrowcast’ specific environmental warnings or advice only to those living within a specific area. It is this capability that APNEE and APNEE-TU have implemented and tested by utilizing modern IT and telecommunication technology at locations across Europe.

Project APNEE commenced in January 1999 as an RTD project in the Fifth Framework Programme of the European Commission, IST Programme, Key Action 1 (Systems and Services for the Citizen), Action Line I.5.1. It concluded successfully in December 2001, and a take-up action named APNEE-TU (4/2003 – 3/2004) adapted and tested the APNEE systems at additional user sites in Europe, as well as employing new technologies (eg handhelds, smart phones, PDAs and new mobile protocols like GPRS) for the dissemination of air pollution information.

Field trials for the evaluation of these innovations in APNEE and APNEE-TU took place at nine test sites (Oslo, Greenland, Athens, Thessalonica, Marseilles, Canary Islands, Madrid, Andalusia, and the whole of Germany), and the projects included 21 partners.
from research, government and the IT and telecommunications sectors. The test sites represent a diverse selection in terms of geography, pollution problems and cultural settings. In addition, their experience and development of air monitoring stations were different.

We designed and implemented the following ‘business collaboration concept’, which has proven successful:

• in each APNEE region an authority provides and authorizes the measured air pollution data
• research institutes and universities operate and control models for the forecasting of air pollution
• technological partners design and realize the Internet and WAP portals as well as street panel interfaces
• mobile and Internet information portal providers integrate the APNEE solution into their portals
• telecommunication companies distribute the messages through SMS, MMS and WAP, as well as via smart phones and PDAs.

APNEE is, first of all, the architecture of a dissemination platform, built around a harmonized environmental database scheme. In this respect, APNEE supports the harmonization of air-quality management systems. This eases the exchange of data (eg among authorities), allows further standardization of methods and directives and simplifies integration into other services and service providers. In addition, APNEE provides a reference implementation of components, required to set up an APNEE infrastructure at a new user site. APNEE is mostly based on royalty-free open-source software that has a very good developer community for support. It can be installed on a standard PC and is easily administrated once the set-up has been completed. The installation and operational costs of APNEE are very low, allowing cities and communities to set up an information portal for air-quality (or any other environment-related) information, even on a low budget.

APNEE created a set of reference core modules, including a common database scheme, service triggers, and regional server applications. These core modules are considered to be the heart of the system; however, they may be bypassed or not implemented in cases where only the electronic services are of interest for installation and operational usage, provided that there is a database and a pull-and-push scheme offered via alternative software infrastructures. Thus, the project’s e-services pallet currently includes:

• pull services: WAP, J2ME, PDA, Internet-based, GIS information services, voice server
• push services: SMS, e-mail, newsletter
• street panels.

These services may be applied individually, in combination, or in parallel with other pre-existing services, where they exist.

The complementarity and in particular the situateness of services have proven decisive in reaching citizens, that is, in advising people of episodes that might cause harm to their health. MMS services sound attractive from a technology point of view and WAP services allow for more detailed information when on the move. Nevertheless, SMS services have proven sufficient in several field trials.

APNEE has been built upon the use of complementary communication channels. Having successfully completed the field trials, we now know which channels are best suited for various kinds of environmental episode. The information dissemination platform is distinguished by its customizability: messages can be customized to the location of users, their preferences or the types of content that affect them. This facility means APNEE excels in providing early warning of hazardous events. Recipients can be targeted in a focused fashion, with the type of message also able to be tailored.

The project was recently presented as a success story among IST projects by Commissioner Viviane Reding during a press conference in Brussels. Future applications of APNEE will be early warning services. We are currently working on risk management concepts and services for cities and urban regions.

APNEE was supported by the European Commission DG XIII under the 5th Framework Pro-gramme, IST-1999-11517. Partners included FAW Ulm, Germany (coordinator), Airmaraix, France, Aristotle University Thessaloniki, Greece, Ayuntamiento de Madrid, Spain, Expertel Consulting, France, NILU, Norway, NORGIT AS, Norway, Seksjon for kontroll og utvåking i Grenland, Norway, SICE, Spain, Telefonica I+D, Spain, Universidad Politecnica de Madrid, Spain. APNEE-TU was supported by the EC under contract IST-2001-34154. Additional partners are the Storm Weather Center and National Road Protection Agency, Norway, ITC Canarias and Andalusia Network, Spain, OMPEPT and Siemens, Greece, and UMEG, Fraunhofer FIT (new coordinator), and t-info GmbH, Germany.

Links:
Project Web site: http://www.apnee.org
Spain: http://panda.lma.fi.upm.es/andalusia
http://panda.lma.fi.upm.es/canary/
http://atmosfera.lma.fi.upm.es:8080/
regional-eng/servlet/regional/template/Indext.vm
Greece:
http://www.apnee.gr (temporarily out of order)
Norway:
http://www.luftkvalitet.info
Germany:
http://www.t-info.de
http://www.t-zones.de
France:
http://www.aimaraix.com

Please contact:
Gertraud Peinel, Thomas Rose,
Institute for Applied Information Technology - FIT, Fraunhofer ICT Group
Tel: +49 22 4114 2432 (~2798)
E-mail: gertraud.peinel@fit.fraunhofer.de, thomas.rose@fit.fraunhofer.de

Please contact:
Gertraud Peinel, Thomas Rose,
Institute for Applied Information Technology - FIT, Fraunhofer ICT Group
Tel: +49 22 4114 2432 (~2798)
E-mail: gertraud.peinel@fit.fraunhofer.de, thomas.rose@fit.fraunhofer.de

Please contact:
Gertraud Peinel, Thomas Rose,
Institute for Applied Information Technology - FIT, Fraunhofer ICT Group
Tel: +49 22 4114 2432 (~2798)
E-mail: gertraud.peinel@fit.fraunhofer.de, thomas.rose@fit.fraunhofer.de
Articles in this Section

47 Building a Bridge for Communication between Patients, Family Doctors, and Specialists
by Matteo Padletti, Loriano Galeotti and Carlo Marchesi

48 Connecting Remote Tools: Do it by yourSELF!
by María Alpuente and Salvador Lucas

50 Google Teaches Computers the Meaning of Words
by Rudi Cilibrasi and Paul Vitányi

51 Reconstruction, Modelling and Motion Analysis of the Human Knee based on MR Images
by Gábor Renner and György Szántó

53 Computer Controlled Cognitive Diagnostics and Rehabilitation Method for Stroke Patients
by Cecília Sik Lányi, Julianna Szabó, Attila Páll, Ilona Pataky

54 Web-based NeuroRadiological Information Retrieval System
by Sándor Dominich, Júlia Góth and Tamás Kiezer

55 ARGO: A System for Accessible Navigation in the World Wide Web
by Stavroula Ntoa and Constantine Stephanidis

56 BRICKS: A Digital Library Management System for Cultural Heritage
by Carlo Meghini and Thomas Risse

58 Student Programming Project becomes Multi-Million Dollar Company
by Nancy Bazilchuk

59 Software Engineering Institute goes International in Software Process Research
by Mario Fusani
Building a Bridge for Communication between Patients, Family Doctors, and Specialists

by Matteo Paoletti, Loriano Galeotti and Carlo Marchesi

According to recent reports of the World Health Organization, the next two decades will see dramatic changes in the health needs of the world’s populations with chronic diseases, mental illness and injuries as the leading causes of disability. Increases in the senior population ‘confined’ within the home are also expected. ICT technologies must tackle this challenge by providing the means for fast communication and consulting services between the chronically ill, the general practitioner and the hospital specialist.

The objective of the AMICUS project, launched in 2003 at the BIM Lab, University of Florence, is to design a personalized communication system that can improve the quality of daily life and medical care for the chronically ill, at the same time helping to reduce the number and duration of hospital recoveries. AMICUS aims at contributing to the reduction of health costs and at providing easy access to medical assistance. This is in line with a recent position statement from WHO that strongly encourages the development of devices able to give patients an active, conscious role in the management of their disease: actions are needed that emphasize collaborative goal setting, patient skill building to overcome barriers, self-monitoring, personalized feedback, and systematic links to community resources.

AMICUS proposes a number of innovative solutions all aimed at increasing the control of patients over their health. First of all, patients should be encouraged to play an active role in monitoring their disease through the application of custom-designed procedures embedded in portable (possibly wearable) devices. In order to obtain the necessary medical and technical specifics for the implementation of devices of this type, AMICUS is studying the development of an instrument which will monitor the patient’s vital signals and derive patterns of evolution over time. Although such devices are already potentially marketable, in our opinion a rigorous assessment phase is first needed in order to evaluate the level of knowledge and the experience necessary for a correct employment. Fully reliable systems are not yet available and a too-early introduction on the market could be dangerous. The AMICUS project will build a prototype system, known as COMPASS, which should offer wide possibilities of in itinere intervention (hardware and software changes) as a result of experimentation and field trials. The system has a split architecture and consists of two units: one is a wearable Blue Tooth biosignal transmitter whose range covers the patient home area and the other is a PC work-station, which receives data from the patients and classifies events, submitting the information to the health care network, when necessary.

A second objective of the project is the definition of an accurate personalization of system parameters based on contextual information about the patient’s case history and current dynamic conditions. A preliminary tuning is obtained on the patient’s first admission, through a set of multivariate analysis techniques that
Connecting Remote Tools: Do it by yourSELF!
by Maria Alpuente and Salvador Lucas

Fifty years of research in computer science have seen a huge number of programming languages, theories, techniques, and program analysis tools come into existence. The World-Wide Web makes it possible to gain access to resources in a number of ways, ranging from remote downloads followed by local executions, to remote execution via WWW services. Unfortunately however, few of the existing systems and tools are effectively connectable, even if they address similar problems or rely on similar theoretical bases.

As recently noticed by T. Hoare, the quest for a verifying compiler is a classic, but still urgent problem for both the software industry and the computer science community. Of course, program verification, debugging and analysis (especially when oriented towards improving program efficiency) will be essential components of such a tool. Moreover, the effective development of such a system will require an incremental and cooperative effort from work teams all around the world. Thanks to the Internet, the physical distance separating those teams is becoming less and less important. Unfortunately, many existing systems and tools are not really suitable for working together, even if they address closely related problems or rely on similar theoretical bases. The Internet and middleware technology, however, provide numerous possibilities for removing integration barriers and dramatically improving the reusability of previous theoretical results and development efforts.

We can further motivate this point of view with two concrete examples extracted from our personal experience: model checking and termination analysis. Model checking is becoming a standard technique for automated software verification. Its success has made the term relatively popular for describing other verification methods. An example is the automatic analysis of general correctness properties of concurrent systems that do not require a particular representation with a property language (eg deadlock absence, dead code and arithmetic errors). Moreover, advances in model checking position it as an important contributor in the future development of the verifying compiler.

Still, model checking is not commonly used in, for example, object-oriented programming, where ‘de facto’ standard modelling methods (eg UML) and programming languages (eg C, C++ and Java) are common practice. The current pattern in verification tools for these languages essentially consists of automatic model extraction from the source code to the input language of existing, efficient model-checking tools like...
SPIN. However, the use of these tools for verifying parts of a program currently being developed, say, Java is not easy. First, the syntax of Java programs does not correspond to the input language of any of the above-mentioned tools (e.g., Promela). Some translation is therefore required, but hardly covers a subset of the language. Second, there is no easy way of calling one of these tools from the Java interpreter. Third, we lack methods to analyse the results, especially counter-examples for non-deterministic executions. This is partly due to the different interfaces of the tools. For instance, in contrast to the usual GUI of most Java interpreters, SPIN has a command-based interface accepting text from the standard input (stdin), and writes its results on the standard output (stdout). Fourth, the tools can be written in different programming languages, which may make the eventual embedding of one tool into another significantly more difficult. Finally, there is no documented API to gain external access to the functionalities of the applications.

As a different (although related) motivating example, we consider the termination analysis of programs written in programming languages such as CafeOBJ, Elan, Erlang, Maude and OBJ, whose operational principle is based on reducing expressions according to the well-known paradigm of term rewriting. Proofs of termination of term-rewriting systems can be used to prove termination of such programs. A number of termination tools exist which can, in fact, be used to address this problem: for instance, the tool MU-TERM can be used to prove termination of simple Maude programs (see Figure 1). Termination of rewriting has also recently been related to certain classes of standard verification problems that can be reduced to termination problems.

Unfortunately, however, it is not easy to connect independently developed analysis tools (like MU-TERM) to a practical environment such as the Maude interpreter.

The problems are analogous to those enumerated above.

Interoperability (i.e., making it possible for a program on one system to access programs and data on another system) is a general problem in software engineering, and a number of solutions (namely, middleware systems) have been devised to solve it. The developers of (formal) software tools should consider such solutions to be an essential part of their design and development. Tackling this problem seriously would also be a positive step for the research community in formal methods and declarative languages. These areas are often used (or should be used) to design and implement analysis tools. By isolating the various aspects of a complex tool (e.g., GUI, numeric computation, constraint solving, program transformation and manipulation etc) into different modules possibly written in different languages, we gain the flexibility to use the most appropriate language. Tightly coupled support techniques such as RPC, CORBA and COM have not undergone sufficient experimentation and development in this setting. The XML WWW services (or just WWW services) also provide a flexible architecture for achieving interoperability of loosely coupled systems that are developed in different programming languages. Again, few efforts have been made to conciliate the design and development of software tools with this technology. We claim that, when considering the design and use of software systems and analysis tools, not only correctness and efficiency must be systematically considered but also interoperability across platforms, applications and programming languages.

Starting this year, we plan to develop these ideas into the research project SELF (Software Engineering and Lightweight Formalisms). This will include research on:

• Suitable XML-like formats for expressing the most important features of currently used (families of) programming languages. A related effort has been recently addressed for developing the CLI/CTS/CLS/CLR framework, which forms the basis of Microsoft’s .NET platform.

• Suitable XML sub-languages for expressing analysis/verification requirements to existing tools.

• Middleware translators and interfaces from existing programming languages to the formalisms or lower-level languages which underlie the program analysis tools.

• The inclusion of existing verification and analysis tools into the SOAP/WS/WSDL/UDDI framework of XML WWW services in order to gain systematic access to their functionality.

Links:
SELF: http://self.lcc.uma.es/
ELP group: http://www.dsic.upv.es/users/elp/elp.html

Please contact:
Salvador Lucas, Universidad Politècnica de Valencia/SparCIM
E-mail: slucas@dsic.upv.es
http://www.dsic.upv.es/~slucas
Google Teaches Computers the Meaning of Words

by Rudi Cilibrasi and Paul Vitányi

Computers can learn the meaning of words with the help of the Google search engine. CWI researchers Rudi Cilibrasi and Paul Vitányi found a way to use the World Wide Web as a massive database from which to extract similarity of meaning between words. The approach is novel in its unrestricted problem domain, simplicity of implementation, and manifestly ontological underpinnings.

To make computers more intelligent one would like to represent the meaning of words and phrases in computerdigestable form. Long-term and labor-intensive efforts like the Cyc project (Cyc Corporation) and the WordNet project (Princeton University) try to establish semantic relations between common objects, or, more precisely, names for those objects. The idea is to create a semantic web of such vast proportions that rudimentary intelligence and knowledge about the real world spontaneously emerges. This comes at the great cost of designing structures capable of manipulating knowledge, and entering high quality contents in these structures by knowledgeable human experts. While the efforts are long-running and large scale, the overall information entered is minute compared to what is available on the World Wide Web.

The rise of the World Wide Web has enticed millions of users to type in trillions of characters to create billions of web pages of on average low quality contents. The sheer mass of the information available about almost every conceivable topic makes it likely that extremes will cancel and the majority or average is meaningful in a low-quality approximate sense. We devise a general method to tap the amorphous low-grade knowledge available for free on the World Wide Web, typed in by local users aiming at personal gratification of diverse objectives, and yet globally achieving what is effectively the largest semantic electronic database in the world. Moreover, this database is available for all by using search engines like Google.

We developed a method that uses only the name of an object and obtains knowledge about the semantic (meaning) similarity of objects by tapping and distilling the great mass of available information on the web. Intuitively, the approach is as follows. The meaning of a word can often be derived from words in the context in which it occurs. Two related words will be likely to give more hits — web pages where they both occurred — than two unrelated words. For instance, the combined terms ‘head’ and ‘hat’ will give more hits in a Google search than ‘head’ and ‘banana’.

The Google search engine indexes around ten billion pages on the web today. Each such page can be viewed as a set of index terms. A search for a particular index term, say “horse”, returns a certain number of hits, say...
Reconstruction, Modelling and Motion Analysis of the Human Knee Based on Magnetic Resonance Images

by Gábor Renner and György Szántó

New computer methods and programs have been developed at SZTAKI in cooperation with the Orthopaedic Department of the Semmelweis University and the CT/MR Laboratory of the International Medical Center, Budapest. They support the three-dimensional visualization, reconstruction and modelling of the tissues and organs of the human body. These methods were used in examining the knee joint in several situations and have provided us with a better understanding of the three-dimensional representation of the cartilage surfaces in the knee joint. The knee joint is a very delicate and frequently damaged component of the human motion system. A wide range of imaging techniques (MR, CT) is available to clinical practitioners investigating the injured parts. These methods offer invaluable support to the orthopaedic surgeon by making visible the three-dimensional visualization, reconstruction and modelling of the tissues and organs of the human body. These methods were used in examining the knee joint in several situations and have provided us with a better understanding of the three-dimensional representation of the cartilage surfaces in the knee joint.

When analysing articular motion, the determination of the contact points (regions) of the cartilage surfaces in three-dimensional (3D) space is especially important. Delineation of the contacting surfaces is extremely difficult due firstly to the similar grey-scale representation of the synovial fluid between the opposite surfaces and the hyalin, and secondly to the fact that the two surfaces are partly covering each other. We have developed computer methods for the image analysis, the 3D reconstruction, and for building geometric models of the knee joint based on MR/CT images. An iso-surface raytracing method with a Phong illumination model is used to create surfaces from volume grid models according to a predefined value of grey-scale threshold.

Surface points extracted from the iso-surface models represent only rough information on the shape. Accurate geometrical models are required for a detailed analysis of the shape and func-
tionality of the knee structures, especially when studying the contact properties of proximal or touching cartilage surfaces.

Spatial geometrical models are created using a set of contours extracted from MR images. We have developed active contour and fast marching methods for the detection of tissue contours in the sequence of MR images relevant to the knee joint (bone, cartilage). Our algorithms and programs based on these two methods work fast and reliably for bone contours; however, they require several improvements for cartilage contours.

By using contours extracted from 2D images and suitable transformations, a digitized representation (point set) of functional surfaces of the knee joint can be constructed. We have developed several methods and programs, which allow us to visualize and evaluate the morphology of those surfaces (eg noise elimination, decimation and triangulation). Various geometrical properties (eg lines of intersection, curvature distributions and feature points) can be calculated, visualized and evaluated.

The study of the knee in motion is important both from anatomical and pathological points of view. This requires the alignment of images and spatial models corresponding to different flexions of the knee in a common coordinate system (registration). Different types of registration methods (such as iterative closest point methods and methods based on anatomical markers or feature points) have been developed for the accurate and fast registration of knee structures.

Our experiments show that continuous and smooth representations of the active surfaces are needed for detailed biological or medical investigations (especially for estimating the size and shape of the contacting regions of cartilages).

Continuous surface fit starts with the topological ordering of points, which is done by creating a triangulation over the surface points. A good triangulation must meet a number of requirements: it must be topologically correct, must eliminate outlier points, triangles must have comparable side lengths and angles, their size must reflect the curvatures of the surface, and so forth. Algorithms and computer programs have been developed for the automatic generation of triangular structures. The mathematical description of surfaces with a high degree of continuity is performed by parametric surfaces commonly used in computer graphics and CAD (eg Bézier, B-spline and NURBS surfaces). The mathematical representation of continuous surfaces is computed by minimizing a functional. This contains the squared distances of data points to the continuous surface, as well as geometrical quantities reflecting continuity and smoothness.

Contact properties of cartilage surfaces (their shape and extension) can be best evaluated by a distance function defined between the two surfaces. This is the length of the shortest path from a surface point to the other surface. We have developed a procedure to evaluate the distance function for the continuous surface models of the cartilage. Surfaces can be colour-coded according to the distance function, which provides an extremely efficient tool for evaluating the shape and size of the contact regions. Figure 3 shows the colour-coded distances for the two contacting cartilage surfaces (femur and tibia).

The above methods provide the basis for a prototype system to be used in the surgical treatment of orthopaedic patients. The complex nature of the problems in this area means that close cooperation is required within a multidisciplinary team. The project, which has run from 2002-2005, has been greatly supported by the National Research and Development Programme of the Ministry of Education of Hungary.

Please contact:
Gábor Renner, SZTAKI, Hungary
Tel: +36 1 279 6152
E-mail: renner@sztaki.hu
Szántó György, SZTAKI, Hungary
Tel: +36 1 279 6187
E-mail: szanto@sztaki.hu
Computer Controlled Cognitive Diagnostics and Rehabilitation Method for Stroke Patients

by Cecilia Sik Lányi, Julianna Szabó, Attila Páll, Ilona Pataky

The HELEN (HELp Neuropsychology) computer-controlled diagnosis and therapeutic system was developed at the Colour and Multimedia Laboratory at the University of Veszprém, Hungary during the period 2002-2004. It was prepared to support the diagnosis and therapy of stroke patients, to keep records of the rehabilitation process and to help the therapeutic process. The system has been successfully tested at the National Centre of Brain Vein Diseases.

The system we have developed provides help for neuropsychological clinicians by supporting them in setting up the anamnesis, providing tests and training material for rehabilitation work and keeping track of patients’ progress.

The system contains a form for holding patients’ personal data, numerous tests and tutorial tasks, and a database in which the rehabilitation process is archived. The single tasks contain decision-making situations, and the system provides assistance in these cases if the patient seeks it. In case of failure the system provides the possibility of a new trial. The method used by the patient to solve the problems is stored, as this represents important information for the neuropsychologist. To enable the patient to find his or her best solution to the problem, we have developed interactive procedures that permit the solution to be reached using different cognitive approaches. It was important that the different routes taken by patients should not influence the test results, which are expressed in points, but should guide the rehabilitation process.

During the diagnostic phase of the rehabilitation, the program keeps track of the patient’s problem-solving strategy and uses this in selecting the best rehabilitation plan. In order to do this, we use syndrome analysis, meaning our system is not merely a collection of computerized tests, but has been built in such a way that it uses the solution provided by the patient to select the next task. The system contains tasks that relate to perception, memory and attention, including writing, reading and counting tasks. Further on, a new aspect in our solution is that through modern technical support one can change the difficulty of the task by using colouring, changing the local or global stroke intensity, or adding noise. The program also numerically presents the efficiency of the task solution. This means that diagnostic and therapeutic programs are not only used according to the individual needs, but also supply exact results and comparative data. One can therefore investigate cognitive and behaviour deficits in such a way as to activate the intact functions of the brain, thereby allowing the patient to control the course of the procedures. The pace of the rehabilitation is thus influenced by the cooperation of the patient, and by this involvement one hopes that the rehabilitation speed will also increase. In addition, the main menu of the HELEN program is constructed in order to allow new tests and practicing tasks to be added at any time.

The Specification of the Logical Structure of the HELEN Software

The following data are stored in the database of the software:
- personal data
- left- or right-handedness (this has to be known for the best strategy of certain tasks)
- data from the investigation, eg test results stored in the different sub-menus of the program, their solution strategy, drawings prepared by the patient, audio recordings etc.

The tasks of the HELEN interactive and therapeutic test and teaching software are shown in the table.

<table>
<thead>
<tr>
<th>Tasks to be used via the Internet</th>
<th>Therapeutic tests</th>
<th>Therapeutic tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many pictures do you see?</td>
<td>Picture with noisy background</td>
<td>Ordering of sentences</td>
</tr>
<tr>
<td>How many black and white checked drawings do you see?</td>
<td>Stories similar to the Binet pictures</td>
<td>Logical cards</td>
</tr>
<tr>
<td>How many cubes do you see?</td>
<td>Ordering according to form and colour</td>
<td>Clock</td>
</tr>
<tr>
<td>Where?</td>
<td>Recognition of emotions</td>
<td>Blind map</td>
</tr>
<tr>
<td>Memory game</td>
<td>Identify!</td>
<td>Story</td>
</tr>
<tr>
<td>Concealed pictures</td>
<td>Ordering of pictures</td>
<td></td>
</tr>
<tr>
<td>Puzzle game</td>
<td>Recognition of faces</td>
<td></td>
</tr>
</tbody>
</table>

Table: Tasks of the HELEN interactive and therapeutic test and teaching software.

Figure 1: The “Where?” tasks.

Figure 2: How many pictures do you see?
Steps of the Solutions of the Applied Research
The HELEN system contains several interactive computer-controlled rehabilitation programs. Using these programs the neurologist can reveal individual changes in the mental capacity of the patient, and can get new information on the cognitive functioning of the patient’s brain. The program can be used also in group therapies.

The information stored in the database helps the neurologist to follow the individual rehabilitation path of the patient. It can also provide data on the typical behaviour of the patient, how he or she recognizes something, and give further insight into the sequencing of perception.

In the future we intend to supplement the system with further sub-tasks, so as to increase the variability of tests the neurologist can use. Based on user feedback, we intend to include stories that are more interesting and understandable for the stroke patients, so that they become more interested and involved in the rehabilitation process.

The rapid progress in the image-processing capabilities of modern computers enables the use of more sophisticated pictures, the introduction of animation, the use of virtual reality worlds etc. We are of the opinion that patients would appreciate more realistic presentations, stimulating them to become more involved and thus engage their mental capacity. All this could lead to quicker and more complete rehabilitations. Naturally this is at the present moment still only speculation; nevertheless, we feel it to be worth exploring, since it could help a non-negligible percentage of the now still inactive population to re-integrate themselves into normal life.

The development of the system was supported by the National Research and Development Fund, Hungary, Grant no. NKFP 2/052/2001

Please contact: Cecília Sik Lányi, University of Veszprém, Hungary Tel: +36 88 624601 E-mail: lanyi@almos.vein.hu

NeuRadIR: A Web-Based NeuroRadiological Information Retrieval System
by Sándor Dominich, Júlia Góth and Tamás Kiezer

The Cost-Effective Health Preservation Consortium Project was formed to develop novel, Web-based Computer Tomograph (CT) image retrieval techniques for use in medical practice, research and teaching. The project was carried out in the Department of Computer Science, Faculty of Information Technology, University of Veszprém, within the Center of Information Retrieval between 2001 and 2004. It was sponsored by the National Research and Development Fund of the Széchenyi Plan, Hungary.

Medical images, like CT images, are increasingly important in healthcare, therapy treatment and medical research. The research results consist in applying information retrieval methods to human brain CT images and reporting retrieval. This represents a solution to the problem. Radiologists have of finding cases that match exactly or partially or are similar to the information they need, depending on the user’s viewpoint. To satisfy these different requirements, three different information retrieval techniques were adopted in one system. The first is employed when the user requires cases that exactly match their information need. A Boolean retrieval technique using AND queries was used for this purpose. When the user wants to retrieve partially matching cases, the retrieval is performed by the Hyperbolic Information Retrieval (HIR) algorithm. Lastly, in order to find similar cases the Interaction Information Retrieval (I2R) method was implemented.

The advantage of using all three strategies is that they complement each other: the Boolean search retrieves every case exactly matching the query, the hyperbolic search retrieves every case having a partial match to the query, while the interaction search returns the most closely associated cases based on the query. The last two techniques were developed in the Center of Information Retrieval (CIR).

In medical practice, most image retrieval systems are designed to help experienced physicians with diagnostic tasks, and they require that users have prior knowledge of the field. This means they are not suitable for educational purposes. The results of our research from the application viewpoint will enhance the quality of both specialist consultation and medical education. General practitioners may confirm a diagnosis or explore possible treatment plans by consulting the CT retrieval system over the Web. Medical students may want to relate images with diagnoses, or to see images corresponding to different pathological cases such as lesion, bleed or stroke. Our system enables users (both radiologists and general practitioners) to make use of medical text and image databases over the Web, both in order to facilitate health preservation and to assist diagnosis and patient care. An assessment by radiologists found the application to be both relevant and effective.

System Description
The NeuRadIR (NeuroRadiological Information Retrieval System) appli-
tion (see Figure 1) consists of computer program modules written in several languages, as well as related documentation. Communication between the Web server and the search program is based on the CGI protocol. The CT Base Editor makes it possible to create and modify the database containing the images, corresponding reports and index files. The Search Module is used on-line on the Web, and consists of user interfaces and search programs. The query is formally analysed and interconnected with the database. A controlled vocabulary was created based on both textual reports and standard specialist queries.

The title page of the ’Search Screen’ contains a header at the top, whilst the area underneath is divided into two halves: the left half contains the terms of the controlled vocabulary, whilst the right half contains the search buttons and the query. Medical terms can be selected from the vocabulary or freely entered from the keyboard in the query line. The user selects one of the searching strategies by clicking on the appropriate button. This returns a hit list, and by clicking on any hit both textual information and CT images are displayed.

The NeuRadIR system has been demonstrated on several national and international forums and can be accessed from the project homepage. In the future, we plan to extend the application by making use of feature extraction modules to index images.

Link: http://www.dcs.vein.hu/CIR/neuradir/app

Please contact: Sándor Dominich, Júlia Göth, Tamás Kiezer, Department of Computer Science, University of Veszprém, Veszprém, Hungary
Tel: +36 88 624711
E-mail: dominich@dcs.vein.hu, goth@dcs.vein.hu, kiezer@dcs.vein.hu

ARGO: A System for Accessible Navigation in the World Wide Web

by Stavroula Ntoa and Constantine Stephanidis

ARGO is a system for accessible navigation in the World Wide Web, which addresses the requirements of diverse target user groups, including people with disabilities. ARGO supports visual and non-visual interaction in order to satisfy the requirements of blind users, users with vision problems, and users with mobility impairments of the upper limbs. ARGO has been developed by ICS-FORTH in the context of the EQUAL PROKLISI Project.

In the context of the ‘Community Initiative EQUAL’ of the European Commission, the Project PROKLISI – ‘Identifying and combating simple and multiple discrimination faced by people with disabilities in the labour market’, coordinated by the Hellenic National Confederation of People with Disabilities, aims to identify and combat the difficulties that people with disabilities and their families face in accessing and remaining in the labour market. One of the goals of the PROKLISI project is to exploit the potential that new technologies provide in order to develop innovative systems and tools aimed at ensuring access to information and communication to groups that are usually discriminated in the employment market.

Under this perspective, a system for accessible navigation in the World Wide Web has been developed, called ARGO, which addresses the requirements of diverse target user groups, including people with disabilities. The system supports visual and non-visual interaction in order to satisfy the requirements of:

• blind users and users with vision problems
• users with mobility impairments of upper limbs.

The ARGO system operates on Microsoft Windows XP, supports all typical functionalities of a browser application, and additionally includes:

• innovative mechanisms for the representation of links and the structure of web pages
• innovative mechanisms for accelerating user interaction with the system
• facilities for finding words and phrases in web pages
• speech synthesis software support
BRICKS (Building Resources for Integrated Cultural Knowledge Services) is an Integrated Project of the 6th Framework Programme (IST 507457) within the research and technological development programme ‘Integrating and Strengthening the European Research Area (2002-2006)’. BRICKS began in January 2004 and has a duration of 42 months. The BRICKS Consortium consists of 24 partners; 7 from the academia, and the rest equally distributed between users and industry, mostly SMEs. The target audience is very broad and heterogeneous and involves cultural heritage and educational institutions, research community,

BRICKS: A Digital Library Management System for Cultural Heritage

by Carlo Meghini and Thomas Risse

The aim of the BRICKS project is to design and develop an open, user- and service-oriented infrastructure to share knowledge and resources in the Cultural Heritage domain. BRICKS will provide cultural heritage institutions and users the possibility to provide or share their content and services with other users.

In order to address the needs of the various target user groups, ARGO employs different operation modes (visual and non-visual interaction), as well as alternative input and output devices.

The system operates in three different modes:

- non-visual mode, appropriate for blind and users with severe vision impairments
- visual mode enhanced with scanning technique, appropriate for users with severe motor impairments of the upper limbs
- visual mode without assistive interaction techniques or input and output devices, appropriate for able-bodied users.

In non-visual interaction, user input is provided through the keyboard, while the system output is provided through the use of synthetic speech and, when necessary, warning sounds. All visual interface elements are organised in a non-visual hierarchical tree structure. By using the appropriate keyboard input commands, the user can navigate in the hierarchy and interact with the currently active interface element.

In visual interaction with scanning, the user provides input through three switches and receives the system output through the screen. All the interactive interface elements of the application are automatically scanned (highlighted), providing the user with the opportunity to interact with each one of them. Thus, there is only one active interface element for each dialogue step, indicated by a coloured border that highlights it. The user may interact with the indicated element, or move the scanning dialogue to the next interactive interface element, by pressing the appropriate switch.

Users with less severe impairments, such as low vision or limited dexterity, can also be assisted by the system, by using the simple visual interface of ARGO in combination with the accessibility features of the operating system (e.g., magnifier, mouse keys, sticky keys, etc).

Evaluations with usability experts and users that took place in the usability laboratory of ICS-FORTH indicated that the system is usable and can be useful to people with disability.

According to the project requirements, the ARGO system has been designed in order to operate as an information point in public places. Therefore, specifications have been provided regarding the setting of the information system and the arrangement of the assistive devices for full accessibility. To facilitate blind users in locating the system and learning how to use it without any assistance, a camera is attached to the system, and when motion is detected, instructions of use are announced.

The ARGO system is currently installed and in use in two information kiosks in the municipalities of Cholargos (Attiki district) and Neapoli (Thessaloniki district).

Future work includes evaluating ARGO with real users in the municipalities where it is currently available, and improving the system according to the evaluation results.

Please contact:
Stavroula Ntoa, ICS-FORTH
Tel: +30 2810 391744
E-mail: stant@ics.forth.gr
industry, and citizens. The project is coordinated by Engineering SpA, Italy.

**Requirements**
The BRICKS infrastructure will use the Internet as a backbone and has been designed to enable:

- Expandability, which means the ability to acquire new services, new content, or new users, without any interruption of service
- Graduality of Engagement, which means offering a wide spectrum of solutions to the content and service providers that want to become members of BRICKS
- Scalability
- Availability
- Interoperability.

In addition, the user community requires that an institution can become a member of a BRICKS installation with minimal investments, and that the maintenance costs of the infrastructure, and in consequence the running costs of members are minimal. BRICKS membership will also be flexible; parties can join or leave the system at any point in time without administrative overheads.

**Approach**
It has been decided that the BRICKS architecture will be decentralised, based on a peer-to-peer (P2P) paradigm, i.e. no central server will be employed. Every member institution of a BRICKS installation is a node (a BNode in the BRICKS jargon) of the distributed architecture. BNodes communicate and use available resources for content and metadata management. Any BNode only has direct knowledge of a subset of other BNodes in the system. However, if a BNode wants to reach a member with which it is not in direct communication, it will forward a request to some of its known neighbour BNodes; these will deliver the request to the final destination or forward again to other nodes.

**The bBricks of BRICKS**
The figure shows the architecture of a BNode. The components (bricks, in the BRICKS jargon) making up the BRICKS architecture are Web Services, and are divided into 3 broad categories:

- Fundamental bricks: these are required on a BNode for its proper functioning and to maintain membership in the BRICKS community. In addition to the P2P layer, other fundamental services are: Decentralised XML Storage (providing data ubiquity on the P2P architecture), Service Registration and Discovery, and Index Manager.
- Core bricks: these are needed if a local BNode installation wants to provide its local users with access to the BRICKS. They include: User Management, Authentication and Authorisation, and Search and Browse.
- Basic bricks: these are optional, and are deployed on the BNode only in case of need (i.e. the Content Management brick will be deployed only if the institution running the BNode exposes content to the BRICKS installation). They include: Content Management, Metadata Management, Accounting, IPR Protection, Annotation Management, and Service Composition.

**Pilot Applications**
In order to render BRICKS operational in the Cultural Heritage domain, the project will develop 4 applications (Pillars, in the BRICKS jargon):

- Archaeological Site, which includes four scenarios: Cultural Landscape Discoverer (sharing knowledge about cultural landscapes), Finds Identifier (identification of objects brought to a museum by members of the public), Landscapes Reconstructed (reconstruction of knowledge about cultural landscapes), and Pompeii and Roma (access to visual artefacts).
- The European Museum Forum, which aims at introducing a digital application process for the European Museum of the Year Award.
- Living Memory, which aims at developing a collaborative environment targeted at the general public, and allowing the creation of cultural contents and their sharing of them among the rest of the community.
- Scriptorium, which will facilitate the work of historians and archive professionals, universities, cultural research centres, libraries, history professors and teachers, and others, by the definition of a new way to exploit and manage Distributed Digital Texts and historical documents. The specific scenario will be the creation of a critical edition of an ancient manuscript.

Link:
BRICKS Community: http://www.brickscommunity.org/

Please contact:
Carlo Meghini, ISTI-CNR, Italy
Tel: +39 050 315 2893
E-mail: carlo.meghini@isti.cnr.it

Thomas Risse, Integrated Publication and Information Systems Institute -IPSI, Fraunhofer ICT Group
Tel: +49 (0) 6151 / 869-906
E-mail: risse@ipsi.fraunhofer.de
In March 2005, the two graduates — now entrepreneurs — moved their rapidly growing business, Norwegian Market Monitor AS, to new headquarters in Trondheim, Norway.

The program developed by Petter Fornass and Tore Steinkjer, who studied for their Master's degrees together at NTNU, uses a high accuracy wrapper to crawl the Internet and collect information about the odds that have been set for 25 different sports across the globe. About 300 sites are regularly checked for information. This technology has become Betradar.com, which monitors odds and is updated every few minutes.

"Since we are collecting numbers, everything has to be 100 percent correct," Fornass says. More than 100 bookmakers around the globe now use this web-crawling technology. Because even small differences in odds can result in huge losses for bookmakers, bookmakers can use Betradar.com to check to make sure the odds they have set for various sporting events are in line with odds set elsewhere on the planet.

In March 2005, the pair moved their rapidly growing business out of the Innovasjonssenter Gloshaugen to new offices on the Trondheim waterfront. Among the products developed by the group is Sports Fixtures, which reports starting times and dates for hockey, soccer and tennis matches and other sports, with daily updates on more than 10 sports one week in advance. Results are delivered 2-3 times per day.

Another key service is called Odds Radar, which detects the odds and starting times from more than 140 bookmakers. Bookmakers can view 8000 odds and 2000 starting times of their competitors each day, and are provided a comprehensive monitoring service of more than 200 Internet sources for odds each day. The monitoring module shows the odds key that was used in setting odds and the odds history and date changes for each bookmaker detected as a part of the program. The service allows bookmakers to compare their odds and starting times with the market average, with alerts of critical differences by e-mail or SMS. The program also detects incorrect results, errors from typos or matches that have been missed or left out, and presents the official schedules from federations or tournament organizers if available on the Internet.

The company employs 27 people, with 10 employees in Norway, 13 employees in Germany, three in Switzerland and one in Czechoslovakia, and more than 100 clients spread across the globe.

Fornass and Steinkjer launched their company from their offices at the Gloshaugen campus of NTNU in 2000. The launch was a little more unusual than most: instead of participating in the Venture Cup competition in Norway, the business plan competition that helps Scandinavian students, researchers and others to take their business idea from concept to actual start-up, Fornass and Steinkjer decided to look for venture capital on their own "and found it. It

Petter Fornass is now CEO of Norwegian Market Monitor AS, which owns and operates Betradar.com. His interest in developing an information collection technology began with a simple desire to finance expenses associated with his master's degree in computer science at NTNU. He developed a web-crawling program with Steinkjer that allowed them to check and compare hundreds of websites for betting odds. The program enabled the pair to find differences in odds for the same sporting event. With this information, they could then place sure bets that guaranteed a win.

But as Fornass and Steinkjer discovered, placing bets and playing the odds can be cash-intensive, even if you are always guaranteed a win.

They were also far more interested in the programming challenges of their project rather than the financial benefit in and of itself.

Fornass and Steinkjer launched their company from their offices at the Gloshaugen campus of NTNU in 2000. The launch was a little more unusual than most: instead of participating in the Venture Cup competition in Norway, the business plan competition that helps Scandinavian students, researchers and others to take their business idea from concept to actual start-up, Fornass and Steinkjer decided to look for venture capital on their own "and found it. It

Petter Fornass, co-founder of Norwegian Market Monitor AS.
Software Engineering Institute goes International in Software Process Research

by Mario Fusani

In industrial prototyping and production, the shifting of the object of study from product to process is more than a century old and has been at the basis of significant product improvement. In software technology, a similar effort has been attempted for a couple of decades, both with empirical software engineering approaches and with mathematical models, yet the results in software-dependent products and services are not as bright as it could be expected.

On the empirical side, corporate and public standards for software development have flourished in recent years. It has been shown that standards impact directly on industry (even the software industry) much more than solutions proposed in specialised literature. For this very reason, standards should embody applicable results of scientific research, whereas frequently they do not. A consequence is that, despite an increasing tendency of developers to follow recognised international norms, there is growing concern about the ability of empirical software development processes to respond to the emerging needs of the global information society. On the formal methods side, indicative is the fact that the application of mathematical models (together with the instruments that can be derived) by software process managers is still a somewhat rare occurrence. So the big questions are:

• what will be the successful paradigm(s) for the software process in the near future?
• what requirements must the software satisfy?
• how will these requirements be influenced by other (social, political, economical, geographical) factors?

At the Software Engineering Institute (SEI), this concern has been taken on-board as an exciting challenge (the bigger the problem, the happier the researcher, one could say). ‘Exploring What’s Next in Software Process Research’ has been the leading concept of a recent initiative, sponsored by SEI and by several private corporations and has led to the constitution, in August 2004, of the International Process Research Consortium (IPRC), a world-wide association of researchers. Six European countries are joining SEI in this effort, Italy being represented by a researcher from ISTI-CNR.

The main goal of IPRC is to prepare a Roadmap — for software process research for the next 5-10 years. The aim will be to provide indications on how the technology challenges of the near future should be addressed by the software industry. However, software is not the only issue of interest for IPRC researchers. Here below we summarise the main activities:

• a number of working groups have been set-up to investigate distinct areas of interest
• six workshops will be held between August 2004 to August 2006 in which ideas of the IPRC member researchers will be proposed and debated; working groups will continue their activity in the meantime
• different nationalities, cultures and needs are represented in the research teams, to avoid localisation and privileged solutions
• experts in non-technical disciplines, such as economy, psychology and organizational science are invited to lecture to the technical team
• experts (from SEI) in working group behaviour act as facilitators, catalysing reactions among researchers
• various interaction mechanisms are activated by the facilitators to obtain maximum benefit from the collaboration;

and the current IPRC approaches and contents:

• technical and non-technical discussion has been solicited from the members in an unbiased way, allowing much degree in divergence of points of view and perspectives; convergence is expected as ideas and activity mature around elements of the roadmap
• widely-disparate content has been deployed so far in the form of statements in order to establish relationships and priorities: needs, hot/warm/cold research topics, state-of-art and trends in software process solutions, human factors, known and mysterious forces pushing forward / retarding progress
• a scenario-based approach is being explored to analyse the impact of a future multi-dimensional ‘trends space’ in any of the software process elements; the lack-of-crystal-ball uncertainty is amended by considering each trend in two opposing directions: this implies much work if completeness is the goal, but knowledge can be increased, even with partial analyses.

The outcomes of the two workshops held so far are considered very promising, thanks to the huge amount of material examined and the ideas generated by the discussions. Although, much work is needed before there can be convergence towards concrete proposals, we already have the impression that in addition to the Roadmap other results and new lines of research will emerge from the work of the Consortium.

Link:
http://www.sei.cmu.edu
http://www.sei.cmu.edu/iprc/

Please contact:
Mario Fusani, ISTI-CNR, Italy
Tel: +39 050 315 2916
E-mail: Mario.Fusani@isti.cnr.it