informatics mathematics

Activity Report 2015

Project-Team CLIME

Coupling environmental data and simulation models for software integration

RESEARCH CENTER
Paris - Rocquencourt

THEME
Earth, Environmental and Energy Sciences

Table of contents

Project-Team CLIME

Creation of the Project-Team: 2005 September 01

Keywords:

Computer Science and Digital Science:

3.3. - Data and knowledge analysis
5.3. - Image processing and analysis
5.9. - Signal processing
6.2.1. - Numerical analysis of PDE and ODE
6.2.3. - Probabilistic methods
6.4. - Automatic control

Other Research Topics and Application Domains:

3.3. - Geosciences
3.4. - Risks
4.2. - Renewable energy production

1. Members

Research Scientists

Isabelle Herlin [Team leader, Inria, Senior Researcher, HdR]
Vivien Mallet [Inria, Researcher]

Faculty Members

Julien Brajard [Univ. Paris VI, Associate Professor, on delegation]
Étienne Huot [Univ. Versailles, Associate Professor]

Engineer

Nicolas Claude [Inria]

PhD Students

Paul Baudin [Inria, until Sep 2015]
Yann Lepoittevin [Inria]
Jean Thorey [EDF]
Raphaël Ventura [Inria]

Post-Doctoral Fellow

Sylvain Girard [IRSN, until Feb 2015]

Administrative Assistant

Nathalie Gaudechoux [Inria]

Others

Jeremy Lefort [Inria, internship, from May 2015]
Illyyne Saffar [Inria, internship, from Mar 2015 until Aug 2015]
Haykel Snoussi [Inria, internship, from Mar 2015 until Aug 2015]
Betissam Tazarine [Inria, internship, from Mar 2015 until Aug 2015]
Dominique Béréziat [Univ. Paris VI, Associate Professor, external collaborator, HdR]

2. Overall Objectives

2.1. Clime in short

The international politic, economic and scientific contexts are pointing out the role that is played by models and observation systems for forecasting and evaluating environmental risks.

The complexity of environmental phenomena as well as the operational objectives of risk mitigation necessitate an intensive interweaving between geophysical models, data processing, simulation, visualization and database tools.

For illustration purpose, we observe that this situation is met in the domain of atmospheric pollution, whose modeling is gaining an ever-increasing significance and impact, either at local (air quality), regional (transboundary pollution) or global scale (greenhouse effect). In this domain, numerical modeling systems are used for operational forecasts (short or long term), detailed case studies, impact studies for industrial sites, as well as coupled modeling, such as pollution and health or pollution and economy. All these scientific subjects strongly require linking/coupling the models with all available data either of physical origin (e.g., models outputs), coming from raw observations (satellite acquisitions and/or information measured in situ by an observation network) or obtained by processing and analysis of these observations (e.g., chemical concentrations retrieved by inversion of a radiative transfer model).

Clime has been created for studying these questions with researchers specialized in data assimilation and image processing.

Clime carries out research activities in three main areas:

- Data assimilation methods: inverse modeling, network design, ensemble methods, uncertainties estimation, uncertainties propagation.
- Image assimilation: assimilation of structures in environmental forecasting models, study of ill-posed image processing problems with data assimilation technics, definition of dynamic models from images, reduction of models.
- Development of integrated chains for data/models/outputs (system architecture, workflow, database, visualization).

3. Research Program

3.1. Data assimilation and inverse modeling

This activity is one major concern of environmental sciences. It matches up the setting and the use of data assimilation methods, for instance variational methods (such as the 4D-Var method). An emerging issue lies in the propagation of uncertainties by models, notably through ensemble forecasting methods.

Although modeling is not part of the scientific objectives of Clime, the project-team has complete access to air quality models through collaborations with École des Ponts ParisTech and EDF R&D: the models from Polyphemus (pollution forecasting from local to regional scales) and Code_Saturne (urban scale). In regard to other modeling domains, such as oceanography and meteorology, Clime accesses models through co-operation with LOCEAN (Laboratoire d'OCEANographie et du climat, UPMC) and Météo-France.

The research activities of Clime tackle scientific issues such as:

- Within a family of models (differing by their physical formulations and numerical approximations), which is the optimal model for a given set of observations?
- How to reduce dimensionality of problems by Galerkin projection of equations on subspaces? How to define these subspaces in order to keep the main properties of systems?

- How to assess the quality of a forecast and its uncertainty? How do data quality, missing data, data obtained from sub-optimal locations, affect the forecast? How to better include information on uncertainties (of data, of models) within the data assimilation system?
- How to make a forecast (and a better forecast!) by using several models corresponding to different physical formulations? It also raises the question: how should data be assimilated in this context?
- Which observational network should be set up to perform a better forecast, while taking into account additional criteria such as observation cost? What are the optimal location, type and mode of deployment of sensors? How should trajectories of mobile sensors be operated, while the studied phenomenon is evolving in time? This issue is usually referred as "network design".

3.2. Satellite acquisitions and image assimilation

In geosciences, the issue of coupling data, in particular satellite acquisitions, and models is extensively studied for meteorology, oceanography, chemistry-transport and land surface models. However, satellite images are mostly assimilated on a point-wise basis. Three major approaches arise if taking into account the spatial structures, whose displacement is visualized on image sequences:

- Image approach. Image assimilation allows the extraction of features from image sequences, for instance motion field or structures' trajectory. A model of the dynamics is considered (obtained by simplification of a geophysical model such as Navier-Stokes equations). An observation operator is defined to express the links between the model state and the pixel values or some image features. In the simplest case, the pixel value corresponds to one coordinate of the model state and the observation operator is reduced to a projection. However, in most cases, this operator is highly complex, implicit and non-linear. Data assimilation techniques are developed to control the initial state or the whole assimilation window. Image assimilation is also applied to learn reduced models from image data and estimate a reliable and small-size reconstruction of the dynamics, which is observed on the sequence.
- Model approach. Image assimilation is used to control an environmental model and obtain improved forecasts. In order to take into account the spatial and temporal coherency of structures, specific image characteristics are considered and dedicated norms and observation error covariances are defined.
- Correcting a model. Another topic, mainly described for meteorology in the literature, concerns the location of structures. How to force the existence and to correct the location of structures in the model state using image information? Most of the operational meteorological forecasting institutes, such as Météo-France (in France), UK-met (in United Kingdom), KNMI (in Netherlands), ZAMG (in Austria) and Met-No (in Norway), study this issue because operational forecasters often modify their forecasts based on visual comparisons between the model outputs and the structures displayed on satellite images.

3.3. Software chains for environmental applications

An objective of Clime is to participate in the design and creation of software chains for impact assessment and environmental crisis management. Such software chains bring together static or dynamic databases, data assimilation systems, forecast models, processing methods for environmental data and images, complex visualization tools, scientific workflows, ...

Clime is currently building, in partnership with École des Ponts ParisTech and EDF R&D, such a system for air pollution modeling: Polyphemus (see the web site http://cerea.enpc.fr/polyphemus/), whose architecture is specified to satisfy data requirements (e.g., various raw data natures and sources, data preprocessing) and to support different uses of an air quality model (e.g., forecasting, data assimilation, ensemble runs).

4. Application Domains

4.1. Introduction

One application domain of the project-team is atmospheric chemistry. We take part to the development (in partnership with École des Ponts ParisTech and EDF R&D) of the air quality modeling system Polyphemus, which includes several numerical models (Gaussian models, Lagrangian model, two 3D Eulerian models including Polair3D) and their adjoints, and different high level methods: ensemble forecast, sequential and variational data assimilation algorithms.

The activity on assimilation of satellite data is mainly carried out for meteorology and oceanography, but it is planned to extend it to agriculture. This research is addressed in cooperation with external partners who provide the numerical models. Concerning oceanography, the aim is to assess ocean surface circulation, by assimilating fronts and vortices displayed on image acquisitions. Concerning meteorology, the focus is on correcting the location of structures related to high-impact weather events (cyclones, convective storms, etc.) by assimilating image features.

4.2. Air quality

Air quality modeling implies studying the interactions between meteorology and atmospheric chemistry in the various phases of matter, which leads to the development of highly complex models. The different usages of these models comprise operational forecasting, case studies, impact studies, etc., with both societal (e.g., public information on pollution) and economical impacts (e.g., impact studies for dangerous industrial sites). Models lack some appropriate data, for instance better emissions, to perform an accurate forecast and data assimilation techniques are recognized as a major key point for improving the forecasts' quality.

In this context, Clime is interested in various problems, the following being the crucial ones:

- The development of ensemble forecast methods for estimating the quality of the prediction, in relation with the quality of the model and the observations. The ensemble methods allow sensitivity analysis with respect to the model's parameters so as to identify physical and chemical processes, whose modeling must be improved.
- The development of methodologies for sequential aggregation of ensemble simulations. What ensembles should be generated for that purpose, how spatialized forecasts can be generated with aggregation, how can the different approaches be coupled with data assimilation?
- The definition of second-order data assimilation methods for the design of optimal observation networks. The two main objectives are: management of combinations of sensor types and deployment modes and dynamic management of mobile sensors' trajectories.
- How to estimate the emission rate of an accidental release of a pollutant, using observations and a dispersion model (from the near-field to the continental scale)? How to optimally predict the evolution of a plume? Hence, how to help people in charge of risk evaluation for the population?
- The assimilation of satellite measurements of troposphere chemistry.

The activities of Clime in air quality are supported by the development, in partnership with École des Ponts ParisTech and EDF R&D, of the Polyphemus air quality modeling system. This system has a modular design, which makes it easier to manage high level applications such as inverse modeling, data assimilation and ensemble forecast.

4.3. Oceanography

The capacity of performing a high quality forecast of the state of the ocean, from the regional to the global scales, is of major interest. Such a forecast can only be obtained by systematically coupling numerical models and observations (in situ and satellite data). In this context, being able to assimilate image structures becomes a key point of current research. Examples of such image structures are:

- surface velocity;
- trajectories, obtained either by tracking features or by integrating the velocity field;
- spatial objects, such as fronts, eddies or filaments.

Image models of these structures are developed and take into account the underlying physical processes. Image acquisitions are assimilated with these models to derive pseudo-observations of state variables, which are further assimilated in ocean forecasting models.

4.4. Meteorology

Meteorological forecasting constitutes a major applicative challenge for image assimilation. Although satellite data are operationally assimilated with models, this is mainly done on an independent pixel basis: the observed radiance at one position is linked to the state variables via a radiative transfer model, that plays the role of an observation operator. Indeed, because of their limited spatial and temporal resolutions, numerical weather forecast models fail to exploit image structures, such as precursors of high impact weather:

- cyclogenesis related to the intrusion of dry stratospheric air in the troposphere (a precursor of cyclones),
- convective systems (supercells) leading to heavy winter time storms,
- low-level temperature inversion leading to fog and ice formation, etc.

To date, there is no available method for operational assimilation of such data, which are characterized by a strong coherence in space and time. Meteorologists have developed qualitative Conceptual Models (CMs), for describing the high impact weathers and their signature on images, and tools to detect these CMs on image data. The result of this detection is used for correcting the numerical models, for instance by modifying the initialization. The aim is therefore to develop a methodological framework allowing to assimilate the detected CMs with numerical forecast models. This is a challenging issue given the considerable impact of the related meteorological events.

5. New Software and Platforms

5.1. Data assimilation library: Verdandi

Participants: Nicolas Claude, Vivien Mallet, Gautier Bureau [M3DISIM], Dominique Chapelle [M3DISIM], Sébastien Gilles [M3DISIM], Philippe Moireau [M3DISIM].

The leading idea is to develop a data assimilation library (see the web site http://verdandi.sourceforge.net/) intended to be generic, at least for high-dimensional systems. Data assimilation methods, developed and used by several teams at Inria, are generic enough to be coded independently of the system to which they are applied. Therefore these methods can be put together in a library aiming at:

- making easier the application of methods to a great number of problems,
- making the developments perennial and sharing them,
- improving the broadcast of data assimilation works.

An object-oriented language (C++) has been chosen for the core of the library. A high-level interface to Python is automatically built. The design study raised many questions, related to high dimensional scientific computing, the limits of the object contents and their interfaces. The chosen object-oriented design is mainly based on three class hierarchies: the methods, the observation managers and the models. Several base facilities have also been included, for message exchanges between the objects, output saves, logging capabilities, computing with sparse matrices.

In 2015, version 1.7 was released. We introduced an implementation of nudging. A level-set observation manager was added. Further tests were included. We added the option to build Verdandi as a library.

5.2. Image processing library: Heimdali

Participants: Isabelle Herlin, Dominique Béréziat [UPMC], David Froger [SED].

The initial aim of the image processing library Heimdali was to develop a library based on standard and open source tools, and mostly dedicated to satellite acquisitions.

The leading idea of the library is to allow the following issues:

- making easier the sharing and development of image assimilation softwares. For that purpose, the installation is easily achieved with the package manager Conda.
- developing generic tools for image processing and image assimilation based on ITK (Insight Segmentation and Registration Toolkit http://www.itk.org).
- in reverse, providing tools to ITK and contribute to the ITK community.

The main components of Heimdali concern:

- the pre/post processing of image sequences,
- the image assimilation with numerical models,
- the visualization of image sequences.

In 2015, additional functions were introduced in the library in order to allow more pre/post processing tools.

5.3. Polyphemus

Participants: Vivien Mallet, Sylvain Doré [CEREA], Karine Sartelet [CEREA], Yelva Roustan [CEREA].

Polyphemus (see the web site http://cerea.enpc.fr/polyphemus/) is a modeling system for air quality. As such, it is designed to yield up-to-date simulations in a reliable framework: data assimilation, ensemble forecasts and daily forecasts. Its completeness makes it suitable for use in many applications: photochemistry, aerosols, radionuclides, etc. It is able to handle simulations from local to continental scales, with several physical models. It is divided into three main parts:

- libraries that gather data processing tools (SeldonData), physical parameterizations (AtmoData) and post-processing abilities (AtmoPy);
- programs for physical pre-processing and chemistry-transport models (Polair3D, Castor, two Gaussian models, a Lagrangian model);
- model drivers and observation modules for model coupling, ensemble forecasting and data assimilation.

Fig. 1 depicts a typical result produced by Polyphemus.

Clime is involved in the overall design of the system and in the development of advanced methods in model coupling, data assimilation and uncertainty quantification (through model drivers and post-processing).

In 2015, version 1.9 was released, with all previous developments on the interface between Verdandi and Polyphemus. The other improvements were developed by CEREA on aerosol modeling.

6. New Results

6.1. Simulation, observation and state estimation for analysis and forecast

The objective of Clime is the merging of simulation and observations, with data assimilation methods, for state estimation in environmental applications. However, this aim previously requires, as seen in some of the next subsection, to collect the observations and carry out the simulations.

6.1.1. Assimilation of drifter data in the East Mediterranean Sea

Participants: Julien Brajard, Milad Fakhri [CNRS, Lebanon], Daniel Hayes [Oceanography Centre, Cyprus], Leila Issa [Lebanese American University, Lebanon], Laurent Mortier [LOCEAN], Pierre-Marie Poulain [Oceanography Institute of Trieste, Italy].

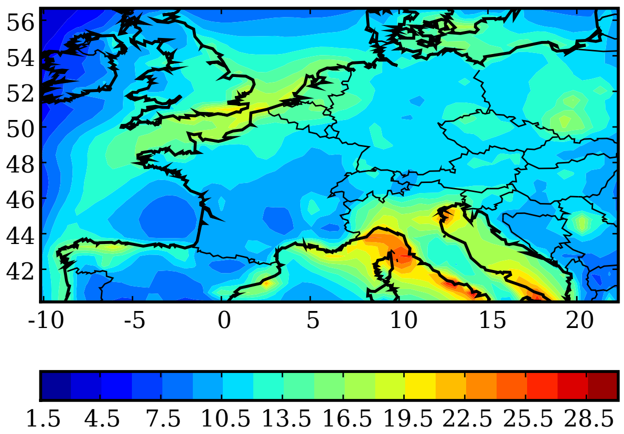


Figure 1. Map of the relative standard deviation (or spread, %) of an ensemble built with Polyphemus (ozone simulations, $\mu g\ m^{-3}$). The standard deviations are averaged over the summer of 2001. They provide an estimation of the simulation uncertainties.

Surface velocity fields of the ocean in the Eastern Levantine Mediterranean are estimated by blending altimetry and surface drifters data. The method is based on a variational assimilation approach for which the velocity is corrected by matching real drifters positions with those predicted by a simple advection model, while taking into account the wind effect. The velocity correction is done in a time-continuous fashion by assimilating at once a whole trajectory of drifters using a sliding time window. A divergence-free regularization term is added to the cost function mnimized during the assimilation process in order to constrain the velocity field. First results show that with few drifters, the method improves the estimation of the surface velocity: an eddy between the Lebanese coast and Cyprus is better assessed and the values of velocities along the Lebanese coast are more accurate.

6.1.2. Traffic simulation

Participants: Vivien Mallet, Vincent Aguiléra [CEREMA], Ruiwei Chen [CEREA].

The ANR project ESTIMAIR aims at propagating uncertainties in the complete simulation chain of air quality at urban scale. A key step in the chain lies in traffic assignment and the computation of the corresponding emissions. We take part to the simulation of traffic in the streets of Clermont-Ferrand metropolitan area, with the dynamic traffic assignment model LADTA. The simulations are evaluated against observations from loop counters and also against the simulations of the reference static model VISUM.

From the traffic assignment, the emissions are computed for nitrogen dioxide and particulate matter, using COPERT IV formulae. Preliminary work shows large uncertainties in the emissions due to the fleet composition.

6.1.3. Observation of noise pollution

Participants: Vivien Mallet, Raphaël Ventura, Valérie Issarny [MiMove], Pierre-Guillaume Raverdy [Ambientic], Fadwa Rebhi [MiMove].

Exposure to noise pollution is highly variable in space. As a consequence, it is very difficult to determine individual exposure using only numerical simulations of noise levels. Together with the MiMove Inria project-team, we take part to the SoundCity project that aims at collecting noise observations from smartphones and better evaluating the individual exposure. We assist MiMove in the development of an Android application that automatically senses noise along the day and collects the data (when the user agrees) for the improvement of simulated noise maps. Clime especially contributes to the calibration of the application. Comparisons between the measurements of smartphones and a sound meter allow us to estimate the bias of the main smartphones available on the market.

The SoundCity application was launched in July 2015 with Bernard Jomier, deputy mayor responsible for health, disability, and relations with Paris public hospital system, during a press conference organized by Paris City. The application received a positive coverage in the media, so that the application gained about 2500 users. About one million observations are collected every four days and ongoing work tries to process these data to correct Paris noise maps.

6.1.4. Evaluation of fire models

Participants: Jérémy Lefort, Vivien Mallet, Jean-Baptiste Filippi [CNRS].

In the field of forest fires risk management, important challenges exist in terms of people and goods preservation. Answering to strong needs from different actors (firefighters, foresters), researchers focus their efforts to develop operational decision support system tools that may forecast wildfire behavior. This requires the evaluation of model performance.

We carry out the evaluation of several fire propagation models based on over 500 real fires. We use the data as they would be available in operational conditions, so as to avoid any tuning that would be incompatible with real-time forecasting. The study shows significant performance difference between the models, despite the poor data quality.

6.2. Image assimilation

Sequences of images, such as satellite acquisitions, display structures evolving in time. This information is recognized of major interest by forecasters (meteorologists, oceanographers, etc.) in order to improve the information provided by numerical models. However, the satellite images are mostly assimilated in geophysical models on a point-wise basis, discarding the space-time coherence visualized by the evolution of structures such as clouds. Assimilating in an optimal way image data is of major interest and this issue should be considered in two ways:

- from the model's viewpoint, the location of structures on the observations is used to control the state vector.
- from the image's viewpoint, a model of the dynamics and structures is built from the observations.

6.2.1. Model error and motion estimation

Participants: Isabelle Herlin, Dominique Béréziat [UPMC].

Data assimilation technics are used to retrieve motion from image sequences. These methods require a model of the underlying dynamics, displayed by the evolution of image data. In order to quantify the approximation linked to the chosen dynamic model, an error term is included in the evolution equation of motion and a weak formulation of 4D-Var data assimilation is designed. The cost function to be minimized depends simultaneously on the initial motion field, at the beginning of the studied temporal window, and on the error value at each time step. The result allows to assess the model error and analyze its impact on motion estimation. The approach is used to estimate geophysical forces (gravity, Coriolis, diffusion) from images in order to better assess the surface dynamics and forecast the displacement of structures like oilspill.

6.2.2. Tracking of structures from an image sequence

Participants: Isabelle Herlin, Yann Lepoittevin, Dominique Béréziat [UPMC].

The research concerns an approach to estimate velocity on an image sequence and simultaneously segment and track a given structure. It relies on the underlying dynamics' equations of the studied physical system. A data assimilation method is designed to solve evolution equations of image brightness, those of motion's dynamics. The method is for instance applied on meteorological satellite data, in order to track tropical clouds on image sequences and estimate their motion, as seen on Fig. 2.

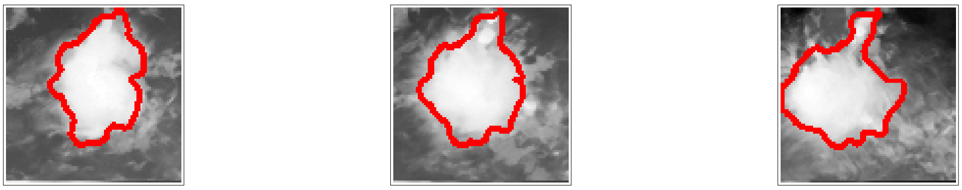


Figure 2. Tracking a tropical cloud. Frames 3, 9, 18 of the sequence.

Data assimilation is performed either with a 4D-Var variational approach or with a Kalman ensemble method. In the last case, the initial ensemble is obtained from a set of optical flow methods of the literature with various parameters values.

Various ways for representing the structures are studied and compared.

- For the variational approach, we consider: 1- a distance map modeling the tracked structures, which is added to the state vector, 2- anisotropic regularization terms, which are added to the cost function minimized during the assimilation process, 3- covariances between pixels, which are included in the background error covariance matrix.
- For the filtering approach, we focus either on domain decomposition or on explicit localization, which are both related to the displayed structures.

6.2.3. Applying POD on a model output dabase for defining a reduced motion model

Participants: Isabelle Herlin, Etienne Huot.

Dimension reduction may be obtained by determining a small size reduced basis computed by Proper Orthogonal Decomposition (POD) of a motion fields database and applying the Galerkin projection. This database is constructed for characterizing accurately the surface circulation of the studied area, so that linear combinations of the basis elements obtained by POD accurately describe the motion function observed on satellite image sequences. The database includes the geostrophic motion fields obtained from Sea Level Anomaly reanalysis maps that are available from the MyOcean European project website (http://marine.copernicus.eu/). Fig. 3 displays such SLA maps and the associated motion fields.

Image assimilation with the POD reduced model allows estimating motion as displayed on Fig. 4.

6.2.4. Rain nowcasting from radar image acquisitions

Participants: Isabelle Herlin, Yann Lepoittevin.

This research concerns the design of an operational method for rainfall nowcasting that aims at mitigating flash floods. The nowcasting method is composed of two main components:

- a data assimilation method, based on radar images, estimates the state of the atmosphere: this is the estimation phase.
- a forecast method uses this estimation to extrapolate the state of the atmosphere in the future: this is the forecast phase.

The method is transfered to the industrial company Weather Measures.

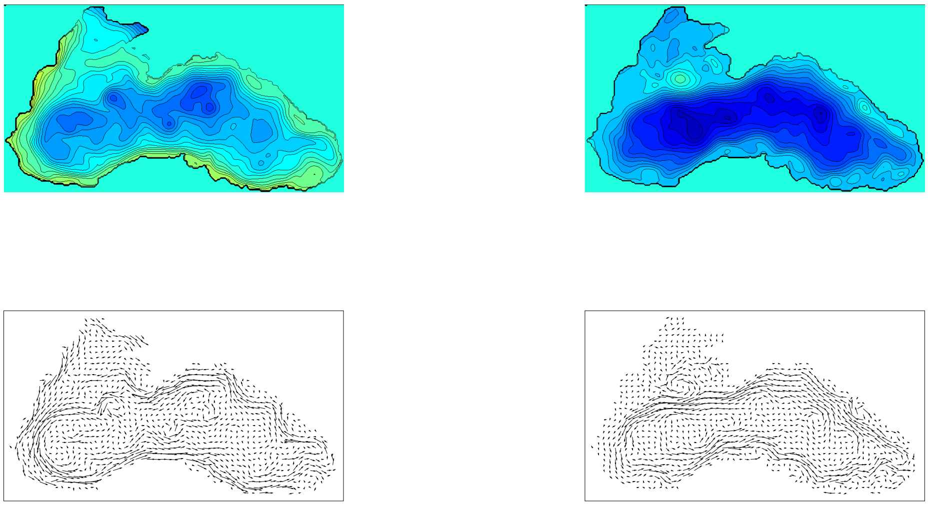


Figure 3. Top: reanalysis of SLA. Bottom: geostrophic motion.

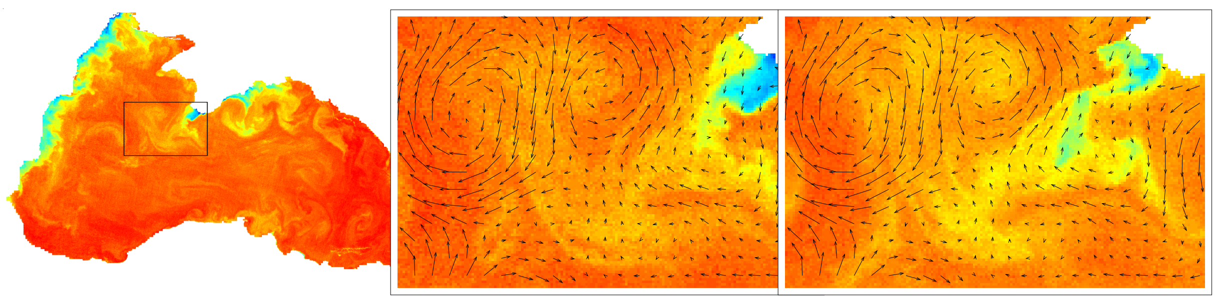


Figure 4. Zoom on a region of interest and motion estimation superposed on two consecutive images.

Current research concerns the use of object components in the state vector in order to get an improved motion estimation and a better localization of endangered regions. Assimilation of pluviometers measures in the nowcasting system is also investigated.

6.3. Uncertainty quantification and risk assessment

The uncertainty quantification of environmental models raises a number of problems due to:

- the dimension of the inputs, which can easily be 10^5-10^8 at every time step;
- the dimension of the state vector, which is usually 10^5-10^7;
- the high computational cost required when integrating the model in time.

While uncertainty quantification is a very active field in general, its implementation and development for geosciences requires specific approaches that are investigated in Clime. The project-team tries to determine the best strategies for the generation of ensembles of simulations. In particular, this requires addressing the generation of large multimodel ensembles and the issue of dimension reduction and cost reduction. The dimension reduction consists in projecting the inputs and the state vector to low-dimensional subspaces. The cost reduction is carried out by emulation, i.e., the replacement of costly components with fast surrogates.

6.3.1. Application of sequential aggregation to meteorology

Participants: Paul Baudin, Vivien Mallet, Gilles Stoltz [CNRS].

Nowadays, it is standard procedure to generate an ensemble of simulations for a meteorological forecast. Usually, meteorological centers produce a single forecast, out of the ensemble forecasts, computing the ensemble mean (where every model receives an equal weight). It is however possible to apply aggregation methods. When new observations are available, the meteorological centers also compute analyses. Therefore, we can apply the ensemble forecast of analyses, which consists in weighting the ensemble of forecasts to better forecast the forthcoming analyses. Before any forecast, the weights are updated with past observations and past forecasts. The performance of the aggregated forecast is guaranteed, in the long run, to perform at least as well as any linear combination of the forecasts with constant weights.

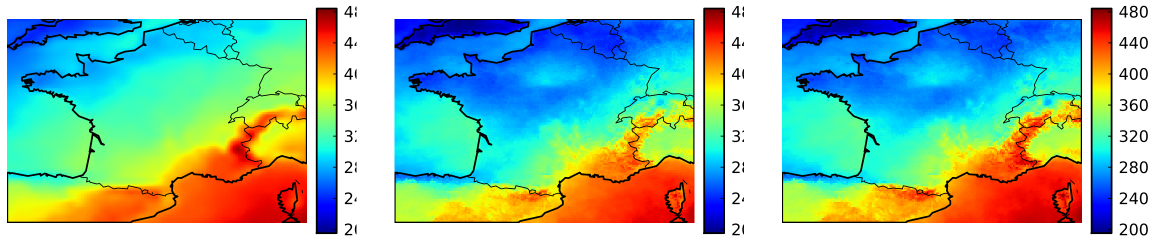
Ensembles of forecasts for mean sea level pressure, from the THORPEX Interactive Grand Global Ensemble, are aggregated with a forecast error decreased by 18% compared to the best individual forecast. The approach is also proved to be efficient for wind speed. The contribution of the ensembles (from different meteorological centers) to the performance increase are evaluated.

6.3.2. Sequential aggregation with uncertainty quantification and application to photovoltaics production

Participants: Paul Baudin, Vivien Mallet, Jean Thorey, Christophe Chaussin [EDF R&D], Gilles Stoltz [CNRS].

We study the aggregation of ensembles of solar radiations and photovoltaic productions. The aggregated forecasts show a 20% error decrease compared to the individual forecasts. They are also able to retrieve finer spatial patterns than the ones found in the individual forecasts (see Figure 5).

An important issue is the estimation of the uncertainties associated with the aggregated forecasts. We devise a new approach to predict a probability density function or a cumulative distribution function instead of a single aggregated forecast. In practice, the aggregation procedure aims at forecasting the cumulative distribution function of the observations which is simply a Heaviside function centered at the observed value. Our forecast is the weighted empirical cumulative distribution function based on the ensemble of forecasts. The method guarantees that, in the long run, the forecast cumulative distribution function has a Continuous Ranked Probability Score (CRPS) at least as good as the best weighted empirical cumulative function with weights constant in time.



(a) (b) (c)

Figure 5. Yearly average of the map of downward shortwave solar radiation in Wm^{-2}*, for an ensemble mean (a), for our aggregated forecasts (b) and observed (c).*

The CRPS is a classical score to evaluate the probabilistic forecasts. However, applying the CRPS on weighted empirical distribution functions (derived from the weighted ensemble) introduces a bias because of which minimizing the CRPS does not produce the optimal weights. Thus, we propose an unbiased version of the CRPS which relies on clusters of members and is strictly proper.

6.3.3. Sensitivity analysis in the dispersion of radionuclides

Participants: Sylvain Girard, Vivien Mallet, Irène Korsakissok [IRSN].

We carry out a sensitivity analysis of the dispersion of radionuclides during Fukushima disaster. We considered the dispersion at regional scale, with the Eulerian transport model Polair3D from Polyphemus. Simulations of the atmospheric dispersion of radionuclides involve large uncertainties originating from the limited knowledge of meteorological input data, composition, amount and timing of emissions and some model parameters. We studied the relative influence of each uncertain input on several outputs. In practice, we used the variance-based sensitivity analysis method of Sobol. This method requires a large number of model evaluations which are not achievable directly due to the high computational cost of the model. To circumvent this issue, we built a mathematical approximation of the model using Gaussian process emulation.

In previous studies, the uncertainties in the meteorological forecasts were crudely modeled by homogeneous and constant perturbations on the fields. Hence, we started investigating the use of ensembles of meteorological forecasts instead of just one base meterological forecast. Including such ensembles allows to better represent the directions along which meteorological uncertainties should lie.

6.3.4. Fire risk assessment

Participants: Jérémy Lefort, Vivien Mallet, Jean-Baptiste Filippi [CNRS].

During days with extreme weather conditions, every wildland fire must be fought within minutes of its occurrence. This means that sufficient firefighting force is available at the right place and at the right time. In practice, firefighters wait at different critical locations, so that they can act quickly. For efficient preventive positioning of the firefighters, forecasting the risks of ignition of large fires is essential. This requires to predict where a fire may start, to estimate its potential size, to evaluate fighting scenarios and to anticipate which urban or protected areas may be under threat.

We designed a surrogate propagation model based on Gaussian process emulation of the model ForeFire. This surrogate model is fast enough to be run all over a region with high fire risk, e.g., Corsica. It can even be used for Monte Carlo simulations, with perturbations in the meteorological conditions and vegetation state, over Corsica. It is then possible to generate a risk map that identifies all the locations where a new fire can grow large.

6.3.5. Ensemble variational data assimilation

Participants: Julien Brajard, Isabelle Herlin, Marc Bocquet [CEREA], Jérôme Sirven [LOCEAN], Olivier Talagrand [LMD, ENS], Sylvie Thiria [LOCEAN].

The general objective of ensemble data assimilation is to produce an ensemble of analysis from observations and a numerical model which is representative of the uncertainty of the system. In a bayesian framework, the ensemble represents a sampling of the state vector probability distribution conditioned to the available knowledge of the system, denoted the a-posteriori probability distribution.

Ensemble variational data assimilation (EnsVar) consists in producing such an ensemble, by perturbating N times the observations according to their error law, and run a standard variationnal assimilation for each perturbation. An ensemble of N members is then produced. In the case of linear models, there is a theoretical guarantee that this ensemble is a sampling of the a-posteriori probability. But there is no theoretical result in the non-linear case.

Numerical experiments using non-linear numerical models suggest that the conclusion reached for linear models still stands for non-linear toy models.

The objective of this work is to study the ability of EnsVar to produce "good" ensemble (i.e. that sampled the a posteriori probablility) on a more realistic model: a shallow-water model. Some statistical properties of the ensemble are presented, and the sensitivity to the main features of the assimilation system (number, distribution of observations, size of the assimilation window, ...) are also studied.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- Clime is partner with IRSN (http://www.irsn.fr/), the French national institute for radioprotection and nuclear safety, for uncertainty estimation of dispersion simulations. The collaboration aims at better estimating emission sources, at improving operational forecasts for crisis situations and at estimating the reliability of forecasts. The work is derived at large scale (continental scale) and small scale (a few kilometers around a nuclear power plant).
- Clime takes part to an Inria innovation lab with the group SETH (Numtech http://www.numtech.fr/). The objective is to (1) transfer Clime work in data assimilation, ensemble forecasting and uncertainty estimation, with application to urban air quality, (2) identify the specific problems encountered at urban scale in order to determine new research directions.

8. Partnerships and Cooperations

8.1. National Initiatives

- The ANR project Estimair aims at quantifying the uncertainties of air quality simulations at urban scale. The propagation of uncertainties requires the use of model reduction and emulation. A key uncertainty source lies in the traffic emissions, which will be generated using a dynamic traffic assignment model. Ensembles of traffic assignments will be calibrated and used in the uncertainty quantification. Estimair is led by Clime.
- The IPSL project "AVES" (Ensemble Variational Assimilation applied to a shallow-water model) aims at estimating the quality of an ensemble produced by a variational ensemble algorithm applied on a shallow-water numerical model. A focus is made on the bayesian properties of the ensemble, i.e. its capacity to sample the a-posteriori probability law of the model state.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

Program: e-Infrastructures

Project acronym: EoCoE

Project title: Energy oriented Centre of Excellence for computer applications

Duration: 3 years

Coordinator: CEA (Commissariat à l'énergie atomique et aux énergies alternatives)

Other partners: Forschungzentrum Jülich GMBH and 11 other partners. Inria is third-linked party of CEA.

Abstract: the aim of the project is to establish an Energy Oriented Centre of Excellence for computing applications, (EoCoE). EoCoE (pronounce "Echo") will use the prodigious potential offered by the ever-growing computing infrastructure to foster and accelerate the European transition to a reliable and low carbon energy supply. To achieve this goal, we believe that the present revolution in hardware technology calls for a similar paradigm change in the way application codes are designed. EoCoE will assist the energy transition via targeted support to four renewable energy pillars: Meteo, Materials, Water and Fusion, each with a heavy reliance on numerical modelling. These four pillars will be anchored within a strong transversal multidisciplinary basis providing high-end expertise in applied mathematics and HPC.

8.2.2. Collaborations with Major European Organizations

Partner: ERCIM working group "Environmental Modeling".

The working group gathers laboratories of ERCIM working on developing models, processing environmental data or data assimilation.

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

Partner: Marine Hydrophysical Institute, Sevastopol.

The collaboration concerns the study of the Black Sea surface circulation and the issue of image assimilation in forecasting models.

Partner: IBM Research, Dublin, Ireland

The collaboration addresses the assimilation of classical observations as well as images, with application to geophysics. New assimilation methods are developed, mainly based on minimax filtering.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. Member of the organizing committees

- Vivien Mallet: Aristote seminar on "Le calcul scientifique pour la ville intelligente" (scientific computing for smart cities), IFSTTAR, Champs-sur-Marne, March 2015.

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committees

- Isabelle Herlin: ISPRS workshop on Image Sequence Analysis for Object Extraction (ISPRS Geospatial Week 2015).

9.1.2.2. Reviewer

- Isabelle Herlin: International Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), International Conference on Image Processing (ICIP).

9.1.3. Journal

9.1.3.1. Reviewer - Reviewing activities

- Julien Brajard: IEEE Transactions on Geoscience and Remote Sensing (TGRS).
- Isabelle Herlin: IEEE Transactions on Geoscience and Remote Sensing (TGRS).
- Vivien Mallet: Annals of Nuclear Energy. Environmental Science & Technology. Quaterly Journal of the Royal Meteorological Society.

9.1.4. Leadership within the scientific community

- Isabelle Herlin is a member of the Scientific Council of CSFRS (High Council for Strategic Education and Research in France).

9.1.5. Scientific expertise

- Isabelle Herlin participates to the scientific expertise on satellite observations for agriculture for CVT Allenvi.
- Isabelle Herlin is a member of the hiring committee for researchers at Inria - Rennes Bretagne Atlantique and for research directors at Inria.
- Isabelle Herlin is a member of the program committee of DIGITEO, french research cluster in science and technology of information.
- Isabelle Herlin is a member of the program committee of the interdisciplinarity mission at CNRS.
- Isabelle Herlin is a member of the Evaluation Committee at Inria.

9.1.6. Research administration

- Isabelle Herlin is a member of the Center Committee for Inria-Paris research center.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master OACOS/WAPE: Vivien Mallet; Introduction to Data Assimilation for Geophysics; 10.5 hours; M2; UPMC, X, ENS, ENSTA ParisTech, École des Ponts ParisTech; France.

Master SGE and 3rd-year class at École des Ponts ParisTech: Vivien Mallet; Air quality modeling; 9h; M2; Universities Paris Diderot- Paris 7, Paris 12 and École des Ponts ParisTech, France.

Training: Vivien Mallet; Uncertainty Quantification: Ensembles and Data Assimilation – Application to Climate and Geosciences; 5.25 hours; CERFACS; France.

9.2.2. Supervision

- HdR : Isabelle Herlin, “Images and Dynamics”, University Paris 6, June 2015.
- PhD : Paul Baudin, “Sequential prediction with ensemble aggregation: application to meteorological prediction with uncertainties”, University Paris-Sud, November 2015.

- PhD : Yann Lepoittevin, “Estimation of motion from observed structures in image sequences, University Paris 6, December 2015.
- PhD in progress: Ruiwei Chen, “Quantification d’incertitude en simulation des émissions du trafic routier”, November 2014, Vivien Mallet.
- PhD in progress : Pacôme Eberhart, “Génération automatique de codes performants et fiables pour l’assimilation de données”, September 2013, Fabienne Jezequel, Pierre Fortin and Julien Brajard
- PhD in progress : Jean Thorey, “Prévision d’ensemble du rayonnement solaire pour la production photovoltaïque du parc EDF”, November 2013, Vivien Mallet.
- PhD in progress: Raphaël Ventura, “Simulation numérique de la ville par couplage entre la modélisation et l’observation”, September 2014, Vivien Mallet.

9.3. Popularization

- Isabelle Herlin made a talk on the careers in image processing at GdR ISIS in November 2015.
- Vivien Mallet took part to a one-day introduction to Inria research at Sénat, in February 2015.
- Vivien Mallet and Raphaël Ventura took part to the UrbanDirtLab organized during the Futur-en-Seine digital festival, in June 2015.
- Vivien Mallet was auditioned by "l’Office parlementaire d’évaluation des choix scientifiques et technologiques" ("Assemblée nationale" and "Sénat") on big data and agriculture, in July 2015.

10. Bibliography

Major publications by the team in recent years

[1] M. Bocquet, P. Sakov. *An iterative ensemble Kalman smoother*, in "Quarterly Journal of the Royal Meteorological Society", October 2013 [*DOI :* 10.1002/qj.2236], http://hal.inria.fr/hal-00918488

[2] D. Béréziat, I. Herlin. *Solving ill-posed Image Processing problems using Data Assimilation*, in "Numerical Algorithms", February 2011, vol. 56, n^{o} 2, pp. 219-252 [*DOI :* 10.1007/s11075-010-9383-z], http://hal.inria.fr/inria-00538510

[3] D. Garaud, V. Mallet. *Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: Application to air quality*, in "Journal of Geophysical Research", October 2011, vol. 116 [*DOI :* 10.1029/2011JD015780], http://hal.inria.fr/hal-00655771

[4] I. Herlin, D. Béréziat, N. Mercier, S. Zhuk. *Divergence-Free Motion Estimation*, in "ECCV 2012 - European Conference on Computer Vision", Florence, Italie, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (editors), Lecture Notes in Computer Science, Springer, October 2012, vol. 7575, pp. 15-27 [*DOI :* 10.1007/978-3-642-33765-9_2], http://hal.inria.fr/hal-00742021

[5] M. R. Koohkan, M. Bocquet. *Accounting for representativeness errors in the inversion of atmospheric constituent emissions: application to the retrieval of regional carbon monoxide fluxes*, in "Tellus B", July 2012, vol. 64, n^{o} 19047 [*DOI :* 10.3402/tellusb.v64i0.19047], http://hal.inria.fr/hal-00741930

[6] G. K. Korotaev, E. Huot, F.-X. Le Dimet, I. Herlin, S. Stanichny, D. Solovyev, L. Wu. *Retrieving ocean surface current by 4-D variational assimilation of sea surface temperature images*, in "Remote Sensing of Environment", April 2008, vol. 112, n^{o} 4, pp. 1464-1475, Remote Sensing Data Assimilation Special Issue [*DOI :* 10.1016/j.rse.2007.04.020], http://hal.inria.fr/hal-00283896

[7] V. MALLET. *Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation*, in "Journal of Geophysical Research", December 2010, vol. 115 [*DOI :* 10.1029/2010JD014259], http://hal.inria.fr/inria-00547903

[8] B. SPORTISSE. *Pollution atmosphérique. Des processus à la modélisation*, Ingénierie et développement durable, Springer-Verlag France, 2008, 350 p. , http://hal.inria.fr/inria-00581172

[9] V. WINIAREK, M. BOCQUET, O. SAUNIER, A. MATHIEU. *Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant*, in "Journal of Geophysical Research Atmospheres", March 2012, vol. 117 [*DOI :* 10.1029/2011JD016932], http://hal.inria.fr/hal-00704999

[10] L. WU, V. MALLET, M. BOCQUET, B. SPORTISSE. *A comparison study of data assimilation algorithms for ozone forecasts*, in "Journal of Geophysical Research", October 2008, vol. 113 [*DOI :* 10.1029/2008JD009991], http://hal.inria.fr/inria-00582376

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] P. BAUDIN. *Sequential prediction with ensemble aggregation: application to meteorological prediction with uncertainties* , Université Paris 11, November 2015, https://hal.inria.fr/tel-01239436

[12] I. HERLIN. *Images and dynamics*, Université Paris 6, June 2015, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01245112

[13] Y. LEPOITTEVIN. *Estimation of motion from observed structures in image sequences*, Université Paris 6, December 2015, https://hal.inria.fr/tel-01241514

Articles in International Peer-Reviewed Journals

[14] D. BÉRÉZIAT, I. HERLIN. *Coupling dynamic equations and satellite images for modelling ocean surface circulation*, in "Communications in Computer and Information Science", 2015, vol. 550, pp. 191-205 [*DOI :* 10.1007/978-3-319-25117-2_12], https://hal.inria.fr/hal-01245369

[15] P. EBERHART, J. BRAJARD, P. FORTIN, F. JÉZÉQUEL. *High Performance Numerical Validation using Stochastic Arithmetic* , in "Reliable Computing", December 2015, vol. 21, pp. 35-52, https://hal.inria.fr/hal-01254446

[16] J. THOREY, V. MALLET, C. CHAUSSIN, L. DESCAMPS, P. BLANC. *Ensemble forecast of solar radiation using TIGGE weather forecasts and HelioClim database*, in "Solar Energy", October 2015, vol. 120, pp. 232–243 [*DOI :* 10.1016/J.SOLENER.2015.06.049], https://hal-mines-paristech.archives-ouvertes.fr/hal-01184650

[17] S. ZHUK, F. JASON, I. HERLIN, R. SHORTEN. *Data assimilation for linear parabolic equations: minimax projection method*, in "SIAM Journal on Scientific Computing", 2015, vol. 37, n^{o} 3, pp. A1174-A1196 [*DOI :* 10.1137/13094709X], https://hal.inria.fr/hal-01174081

International Conferences with Proceedings

[18] S. HACHEM, V. MALLET, V. RAPHAËL, P.-G. RAVERDY, A. PATHAK, V. ISSARNY, R. BHATIA. *Monitoring Noise Pollution Using The Urban Civics Middleware*, in "IEEE BigDataService 2015", San Francisco, United States, March 2015, https://hal.inria.fr/hal-01109321

[19] I. HERLIN, E. HUOT. *Dimensionality reduction on ocean model's outputs: Application to motion estimation on satellite images*, in "IGARSS - IEEE International Geoscience and Remote Sensing Symposium", Milan, Italy, July 2015, https://hal.inria.fr/hal-01174027

[20] Y. LEPOITTEVIN, I. HERLIN, D. BÉRÉZIAT. *An Image-Based Ensemble Kalman Filter for Motion Estimation*, in "VISAPP - International Conference on Computer Vision Theory and Applications", Berlin, Germany, Springer Verlag, March 2015, pp. 437-445, https://hal.inria.fr/hal-01095360

[21] Y. LEPOITTEVIN, I. HERLIN. *Assimilation of radar reflectivity for rainfall nowcasting*, in "IGARSS - IEEE International Geoscience and Remote Sensing Symposium", Milan, Italy, July 2015, https://hal.inria.fr/hal-01174046

[22] Y. LEPOITTEVIN, I. HERLIN. *Modeling high rainfall regions for flash flood nowcasting*, in "MultiTemp - IEEE Workshop on the Analysis of Multitemporal Remote Sensing Images", Annecy, France, July 2015, https://hal.inria.fr/hal-01171005

[23] Y. LEPOITTEVIN, I. HERLIN. *Regularization terms for motion estimation. Links with spatial correlations*, in "VISAPP - International Conference on Computer Vision Theory and Applications", Rome, Italy, Springer Verlag, February 2016, https://hal.inria.fr/hal-01235718

Other Publications

[24] L. ISSA, J. BRAJARD, M. FAKHRI, L. MORTIER, P.-M. POULAIN. *Modeling Surface Currents in the Eastern Levantine Mediterranean*, April 2015, vol. 17, EGU General Assembly Conference, Poster, https://hal.inria.fr/hal-01191755