Architecture of Environmental Risk Modelling: for a faster and more robust response to natural disasters

Dario Rodriguez-Aseretto a,1, Christian Schaerer b, Daniele de Rigo a,c

a European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra (VA), Italy.
b Polytechnic School, National University of Asuncion, San Lorenzo, Central, Paraguay.
c Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Milano, Italy.

Abstract
Demands on the disaster response capacity of the European Union are likely to increase, as the impacts of disasters continue to grow both in size and frequency. This has resulted in intensive research on issues concerning spatially-explicit information and modelling and their multiple sources of uncertainty. Geospatial support is one of the forms of assistance frequently required by emergency response centres along with hazard forecast and event management assessment. Robust modelling of natural hazards requires dynamic simulations under an array of multiple inputs from different sources. Uncertainty is associated with meteorological forecast and calibration of the model parameters. Software uncertainty also derives from the data transformation models (D-TM) needed for predicting hazard behaviour and its consequences. On the other hand, social contributions have recently been recognized as valuable in raw-data collection and mapping efforts traditionally dominated by professional organizations. Here an architecture overview is proposed for adaptive and robust modelling of natural hazards, following the Semantic Array Programming paradigm to also include the distributed array of social contributors called Citizen Sensor in a semantically-enhanced strategy for D-TM modelling. The modelling architecture proposes a multi-criteria approach for assessing the array of potential impacts with qualitative rapid assessment methods based on a Partial Open Loop Feedback Control (POLFC) schema and complementing more traditional and accurate a-posteriori assessment. We discuss the computational aspect of environmental risk modelling using array-based parallel paradigms on High Performance Computing (HPC) platforms, in order for the implications of urgency to be introduced into the systems (Urgent-HPC).

Keywords: Geospatial, Integrated Natural Resources Modelling and Management, Semantic Array Programming, Warning System, Remote Sensing, Parallel Application, High Performance Computing, Partial Open Loop Feedback Control

1E-mail Corresponding Author: dario.rodriguez@jrc.ec.europa.eu

Cite as:
1. INTRODUCTION: CONTEXT, PITFALLS AND THE SCIENCE-POLICY INTERFACE

Europe experienced a series of particularly severe disasters in the recent years [1, 2], with worrying potential impacts of similar disasters under future projected scenarios of economy, society and climate change [3, 4]. They range from flash floods [5, 6, 7] and severe storms in Western Europe with an expected increasing intensity trend [8], large-scale floods in Central Europe [9], volcanic ash clouds [10, 11, 12] (e.g. after the Eyjafjallajkull eruption), large forest fires in Portugal and Mediterranean countries [13, 14]. Biological invasions such as emerging plant pests and diseases have the potential to further interact e.g. with wildfires [15] and to impact on ecosystem services [16] and economy with substantial uncertainties [17].

It should be underlined that these recent highlights are set in the context of systemic changes in key sectors [18, 19, 20] which overall may be expected to at least persist in the next decades. As a general trend, demands on the EU’s resilience in preparedness and disaster response capacity are likely to increase, as the impacts of disasters continue to grow both in size and frequency, even considering only the growing exposure (societal factors) [21, 22]. The aforementioned examples of disturbances are often characterised by non-local system feedbacks and off-site impacts which may connect multiple natural resources (system of systems) [23, 24, 25]. In this particular multifaceted context [26, 27, 28], landscape [29] and ecosystem dynamics show intense interactions with disturbances [30].

As a consequence, classical disciplinary and domain-specific approaches which might be perfectly suitable at local-scale may easily result in unacceptable simplifications within a broader context. A broad perspective is also vital for investigating future natural-hazard patterns at regional/continental scale and adapting preparedness planning [31, 32, 33]. The complexity and uncertainty associated with these interactions – along with the severity and variety of the involved impacts [34] – urge robust, holistic coordinated [35] and transparent approaches [36, 37]. At the same time, the very complexity itself of the control-system problems involved [38, 39, 40] may force the analysis to enter into the region of deep-uncertainty [41].

The mathematization of systems in this context as a formal control problem should be able to establish an effective science-policy interface, which is not a trivial aspect. This is easily recognised even just considering the peculiarities – which have been well known for a long time – of geospatially-aware environmental data [42] and decision support systems [43, 44, 45], their entanglement with growingly complex ICT aspects [46, 47] and their not infrequent cross-sectoral characterisation. Several pitfalls may degrade the real-world usefulness of the mathematization/implementation process. While it is relatively intuitive how a poor mathematization with a too simplistic approach might result in a failure, subtle pitfalls may lie even where an “appropriately advanced” theoretical approach is proposed.
Mathematization should resist *silo thinking* [48, 49] temptations such as academic solution-driven pressures [37, 50] to force the problem into fashionable “hot topics” of control theory: robust approximations of the real-world broad complexity may serve egregiously instead of state-of-art solutions of oversimplified problems.

Other long-lasting academic claims are “towards” fully automated scientific workflows in computational science, maybe including self-healing and self-adapting capabilities of the computational models implementing the mathematization. These kinds of claims might easily prompt some irony [51] among experienced practitioners in wide-scale transdisciplinary modelling for environment (WSTMe, [52]) as a never-ending research Pandora’s box with doubtful net advantages [53]. Complex, highly uncertain and sensitive problems for policy and society, as WSTMe problems typically are, will possibly never be suitable for full automation: even in this family of problems, “humans will always be part of the computational process” [54] also for vital accountability aspects [55].

While a certain level of autonomic computing [56] capabilities might be essential for the evolvability and robustness of WSTMe (in particular, perhaps, a higher level of semantic awareness in computational models and a self-adapting ability to scale up to the multiple dimensions of the arrays of data/parameters; see next section), here the potential pitfall is the illusion of *fully automating* WSTMe. The domain of applicability of this puristic academic *silo* – although promising for relatively simple, well-defined (and not too policy-sensitive) case studies – might be intrinsically too narrow for climbing up to deal with the *wicked problems* typical of complex environmental systems [57, 58, 59].

The discussed pitfalls might deserve a brief summary. First, perhaps, is the risk of “solving the wrong problem precisely” [60] by neglecting key sources of uncertainty – e.g. unsuitable to be modelled within the “warmly supported” solution of a given research group. During emergency operations, the risks of providing a “myopic decision support” should be emphasised; i.e. suggesting inappropriate actions [61] – e.g. inaction or missing precaution – due to the potential overwhelming lack of information [62] or the oversimplification/underestimation of potential chains of impacts due to the lack of computational resources for a decent (perhaps even qualitative and approximate) rapid assessment of them.

Overcoming these pitfalls is still an open issue. Here, we would like to contribute to the debate by proposing the integrated use of some mitigation approaches. We focus on some general aspects of the *modelling architecture* for the computational science support, in order for emergency-operators, decision-makers, stakeholders and citizens to be involved in a participatory [63] information and decision support system which assimilates uncertainty and precaution [64, 57]. Since no silver bullet seems to be available for mitigating the intrinsic wide-extent of complexity and uncertainty in environmental risk modelling, an array of approaches is
integrated and the computational aspects are explicitly connected with the supervision and distributed interaction of human expertise. This follows the idea that the boundary between classical control-theory management strategies for natural resources and hazards (driven by automatic control problem formulations – “minimize the risk score function”) and scenario modelling under deep-uncertainty (by e.g. merely supporting emergency-operators, decision-makers and risk-assessors with understandable information – “sorry, no such thing as a risk score function can be precisely defined”) is fuzzy. Both modelling and management aspects may be computationally intensive and their integration is a transdisciplinary problem (integrated natural resources modelling and management, INRMM [65]).

2. ENVIRONMENTAL RISK MODELLING - ARCHITECTURE

Figure 1 illustrates a general modelling conceptualization where the interactions among natural hazard behaviour, related transdisciplinary impacts, risk management and control strategies are taken into account. The special focus on the many sources of uncertainty [36] leads to a robust semantically-enhanced modelling architecture based on the paradigm of Semantic Array Programming (SemAP) [66, 67], with an emphasis on the array of input, intermediate and output data/parameters and the array of data-transformation modules (D-TM) dealing with them.

Arrays of hazard models \( h_j^\zeta(\cdot) \), dynamic information forecasts \( X^\zeta(\cdot) \) (i.e. meteorology) and static parametrisation \( \theta^\zeta(\cdot) \) (i.e. spatial distribution of land cover) are considered. Their multiplicity derives from the many sources on uncertainty \( \zeta = \{\zeta_f, \zeta_X, \zeta_\theta\} \) which affect their estimation (or implementation, for the D-TM software modules \( f_i^\zeta(\cdot) \) which are the building blocks of the hazard models \( h_j^\zeta(\cdot) \)).

Furthermore, during emergency modelling support the lack of timely and accurate monitoring systems over large spatial extents (e.g. at the continental scale) may imply a noticeable level of uncertainty to affect possibly even the location of natural hazards (geoparsing [68] uncertainty). This peculiar information gap may be mitigated by integrating remote sensing (e.g. satellite imagery) with a distributed array of social contributors (Citizen Sensor [69, 70, 71]), exploiting mobile applications (Apps) and online social networks [72]. Remote sensing and the Citizen Sensor are here designed to cooperate by complementing accurate (but often less timely) geospatial information with distributed alert notifications from citizens, which might be timely but not necessarily accurate. Their safe integration implies the supervision of human expertise, even if the task may be supported by automatic tools [73]. Assessing the evolution in the timespan \( U^k = [t_{\text{begin}}, t_{\text{end}}] \) of a certain hazard event for the associated array of impacts \( C^{k,t} \) may be also complex (e.g. [41, 74, 75]). In particular, the array of impacts is often irreducible to a unidimensional quantity (e.g. monetary cost) [76, 77].
Figure 1 - Modular architecture for environmental risk modelling. Based on Urgent HPC, it follows the Semantic Array Programming paradigm (image adapted from [36, 52]) integrating as inputs remote sensing, meteo data and the Citizen Sensor.

The analysis of non-trivial systems subject to environmental risk and natural resources management may naturally lead to multi-objective (multi criteria) control problems, which might benefit from advanced machine learning techniques for...
mitigating the involved huge computational costs [78]. Indeed, the multiplicity of modelling dimensions (states; controls; uncertainty-driven arrays of parameters and scenarios; arrays of D-TM modules to account for software uncertainty) may easily lead to an exponential increase of the required computational processes (the so called “curse of dimensionality”). A viable mitigation strategy might be offered by HPC tools (such as Urgent HPC [79, 80, 81]) in order to sample high-dimensional modelling space with a proper method.

Box 1 – In a nutshell.

Context Demands on the EU’s resilience in preparedness and disaster response capacity are likely to increase, as the impacts of disasters continue to grow.

- Classical disciplinary and domain-specific approaches which might be perfectly suitable at local-scale may result in unacceptable simplifications in a broader context.

Pitfalls Mathematization of systems in this context as a formal control problem should be able to establish an effective science-policy interface. Academic silo thinking should stop advertising solution-driven oversimplification to fit control theory “hot topics”.

- Although in this family of problems “humans will always be part of the computational process” (despite any academic potential illusion of fashionable full automation),

- evolvability (for adapting models to new emerging needs and knowledge) and robustness (for supporting uncertainty-aware decision processes) would still need
  - a higher level of semantic awareness in computational models and
  - a self-adapting ability to scale up to the multiple dimensions of the arrays of data/parameters.

Multiplicity: uncertainty and complexity In this context, the boundary between classical control-theory management strategies for natural resources and hazards and scenario modelling under deep-uncertainty is fuzzy (INRMM).

- A key aspect of soundness relies on explicitly considering the multiple dimensions of the problem and the array of uncertainties involved.

- As no silver bullet seems to be available for reliably attacking this amount of uncertainty and complexity, an integration of methods is proposed.

Mitigating with an integrated approach Array programming is well-suited for easily managing a multiplicity of arrays of hazard models, dynamic input information, static parametrisation and the distribute array of social contributions (Citizen Sensor).

- Array-based abstract – thus better scalable – modularisation of the data-transformations (D-TM), and a semantically-enhanced design of the D-TM structure and interactions (Semantic Array Programming) is proposed to consider also the array of uncertainties (data, modelling, geoparsing, software uncertainty) and the array of criteria to assess the potential impacts associated with the hazard scenarios.

- The unevenly available information during an emergency event may be efficiently exploited by means of a POLFC schema.

- Its demanding computations may become affordable during an emergency event with an appropriate array-based parallelisation strategy within Urgent-HPC.
SemAP can simplify WSTMe modelling of nontrivial static [82, 83] and dynamic [41, 75, 84] geospatial quantities. Under the SemAP paradigm, the generic $i$-th D-TM module $Y_i = f_i(\theta_i, X_i)$ is subject to the semantic checks $\text{sem}_i$ as pre-, post-conditions and invariants on the inputs $\theta_i, X_i$, outputs $Y_i$ and the D-TM itself $f_i(\cdot)$. The control problem is associated with the unevenly available dynamic updates of field measurements and other data related to an on-going hazard emergency. An Emergency Manager may thus be interested in assessing the best control strategy $u^t(\cdot)$ given a set of impacts and their associated costs as they can be approximately estimated (rapid assessment) with the currently available data. This data-driven approach can be implemented as Partial Open Loop Feedback Control (POLFC) approach [85] for minimizing the overall costs associated with the natural hazard event, from the time $t \in \mathcal{U}^t$ onwards:

$$u^t(\cdot) = \arg \min_{u \in \mathcal{U}^{t_{\text{end}}}} \left[ C^{1,t}, C^{2,t}, \ldots C^{k,t}, \ldots C^{n,t} \right]$$

where the $k$-th cost $C^{k,t}$ is linked to the corresponding impact assessment criterion. This POLFC schema within the SemAP paradigm may be considered a semantically-enhanced dynamic data-driven application system (DDDAS) [41, 75, 84]. Finally, the Emergency Manager may communicate the updated scenarios of the emergency evolution (by means of geospatial maps and other executive summary information) in order for decision-makers and stakeholders to be able to assess the updated multi-criteria pattern of costs and the preferred control options. This critical communication constitutes the science-policy interface and must be as supportive as possible. It is designed to exploit web map services (WMS) [86, 87] (on top of the underpinning free software for WSTMe, e.g. [88]) which may be accessed in a normal browser or with specific Apps for smart-phones [73].

3. CONCLUDING REMARKS

NSF Cyberinfrastructure Council report reads: While hardware performance has been growing exponentially - with gate density doubling every 18 months, storage capacity every 12 months, and network capability every 9 months - it has become clear that increasingly capable hardware is not the only requirement for computation-enabled discovery. Sophisticated software, visualization tools, middleware and scientific applications created and used by interdisciplinary teams are critical to turning flops, bytes and bits into scientific breakthroughs [89]. Trans-disciplinary environmental problems such as the ones dealing with complexity and deep-uncertainty in supporting natural-hazard emergency might appear as seemingly intractable [90]. Nevertheless, approximate rapid-assessment based on computationally intensive modelling may offer a new perspective at least able to support emergency operations and decision-making with qualitative or semi-
quantitative scenarios. Even a partial approximate but timely investigation on the potential interactions of the many sources of uncertainty might help emergency managers and decision-makers to base control strategies on the best available – although typically incomplete – sound scientific information. In this context, a key aspect of soundness relies on explicitly considering the multiple dimensions of the problem and the array of uncertainties involved. As no silver bullet seems to be available for reliably attacking this amount of uncertainty and complexity, an integration of methods is proposed, inspired by their promising synergy. Array programming is perfectly suited for easily managing a multiplicity of arrays of hazard models, dynamic input information, static parametrisation and the distribute array of social contributions (Citizen Sensor). The transdisciplinary nature of complex natural hazards – their need for an unpredictably broad and multifaceted readiness to robust scalability – may benefit (1) from a disciplined abstract modularisation of the data-transformations which compose the models (D-TM), and (2) from a semantically-enhanced design of the D-TM structure and interactions. These two aspects define the Semantic Array Programming (SemAP, [66, 67]) paradigm whose application – extended to geospatial aspects [52] – is proposed to consider also the array of uncertainties (data, modelling, geoparsing, software uncertainty) and the array of criteria to assess the potential impacts associated with the hazard scenarios. The unevenly available information during an emergency event may be efficiently exploited by means of a partial open loop feedback control (POLFC, [85]) schema, already successfully tested in this integrated approach [41, 75, 84] as a promising evolution of adaptive data-driven strategies [91]. Its demanding computations may become affordable during an emergency event with an appropriate array-based parallelisation strategy within Urgent-HPC.

References


forest fire progression map using MODIS active fire observation and Citizens sensor. IEEE
Earthzine 7(2). Submitted.
77] GASPARATOS, A., 2010. Embedded value systems in sustainability assessment tools and
DDDAS for forest fire behavior prediction. Lecture Notes in Computer Science 5545: 489-499.
ization. Lecture Notes in Computer Science 5544: 227-236.
behaviour prediction in Europe. Procedia Computer Science 18: 1861-1870.
85] CASTELLETTI, A., et al, On-Line design of water reservoir policies based on inflow
prediction. IFAC-PapersOnLine 17: 14540-14545.
the European Forest Data Centre. Geophysical Research Abstracts 15: 12101+.
Science Foundation.
prediction. In: Computational Science and Engineering, 2008. CSE ’08. 11th IEEE Inter-
national Conference on. IEEE, pp. 275-282.