H13-167

MODEL REDUCTION VIA PRINCIPAL COMPONENT TRUNCATION FOR THE OPTIMAL DESIGN OF ATMOSPHERIC MONITORING NETWORKS

Olivier Saunier[1,2], Marc Bocquet[2,3], Anne Mathieu[1] and Rachid Abida[2,3]

[1]Institut for Radiological Protection and Nuclear Safety (IRSN), Fontenay-aux-roses, France
[2]CEREA, Joint Laboratory Ecole des Ponts ParisTech/EDF R&D, Champs-sur-Marne, France
[3]INRIA, Paris Rocquencourt Research Centre, France

Abstract: IRSN is planning to renovate the French nuclear monitoring network. This renovated network should be able to forecast accurately the accidental plume by using only measures of this network. In this presentation, a numerically efficient methodology for the optimal design of monitoring networks is proposed. In this method, a large database of dispersion accidents over one year of meteorology and from 20 French nuclear sites is built and a cost function measures the ability of a potential network to provide measurements in order to reconstruct any accidental plume from the database.

We introduce methods based on principal component analysis to optimally reduce this database and consequently decrease significantly CPU time. Then, the reduced optimisation method is applied to suggest an optimal strategy for the sequential deployment of the network. Finally, we propose the set-up of networks which take into account foreign potential radiological sources in Europe and French density population.

Key words: *network design, Atmospheric dispersion, principal component analysis, optimisation*

INTRODUCTION

IRSN is planning to renovate the French nuclear monitoring network. It should be able to measure activity concentrations of various aerosol radionuclides. Abida *et al.* (2008) have developed a method allowing to obtain an optimal spatial design. This optimal network is established by minimising a cost function which measures the discrepancy between the extrapolated map obtained from observations given by a network, and the simulation concentration map from a database of dispersion accidents over one year of meteorology and from 20 French nuclear sites.

However, in the optimisation of such cost functions with respect to networks, most of the computational time is spent in the evaluation of the function, especially if the accidents database is large. Consequently, it becomes difficult to deal with several computations to answer to essential questions like for example:

- Can foreign nuclear power plants be taken into account in the design?
- Can one consider moderately suboptimal sequential deployment of the network?
- Can French population be taken into account in the design?

Answering these questions is computationally very expensive. That is why the first objective of this work is to propose a reduction of a database to obtain a more important flexibility of the method. The second objective consists in providing solutions to the questions mentioned above.

In the first section, the different steps of the method are presented and the reduction of the database is described. In the second section, several computations are performed: the first computation is devoted to the definition of an optimal deployment strategy of the network. The foreign potential radiological sources in Europe and French density population are taken into account in a second and a third computation respectively. The influence of these parameters on the shape of the optimal network is consistently discussed.

METHODOLOGY

The methodology is based on the work of Abida *et al.* (2008) where a network of about 100 nuclear aerosols automatic stations was considered. The objective of the network is to monitor 20 civil nuclear sites over France (19 power plants and one recycling facility) of similar radiological signature, from the point of view of aerial dispersion. In this method, the construction of a close to optimal monitoring network is based on the following schedule:

- Construct a database of accidents
- Extrapolate the observations to an activity concentration field over the whole domain
- Define a cost function which evaluates the performance of a network and measures the discrepancy between the extrapolation of the measurements and the concentration field reference from the database
- Truncation of the cost function in order to reduce CPU time by using principal component analysis [PCA]
- Optimise the network using simulated annealing to minimize the cost function

Construction of database

The database is obtained by performing simulations of the potential accidents computed with the Eulerian model POLAIRD3D (Quélo *et al.*, 2007), part of the Polyphemus platform. The meteorological fields are computed on the 0.25° x 0.25° resolution by using the mesoscale atmospheric model MM5 (Anthes and Warner, 1978), forced by the National Center for Environmental Prediction (NCEP) analysis. One year of meteorology was obtained for the year 2004. The source term that is used corresponds to a leak in the primary system of a nuclear power plant which leads to a swift meltdown of the core. It lasts 7 days, so do the simulations (168h). The computations have been performed only for iodine 131. In order to get a good variability in the meteorological conditions, one accident was generated every 21h, which entails a shift of 3h every day. Consequently, the database has 8360 accidents.

Interpolation technique

The interpolation scheme allows to establish maps of activity concentrations on the basis of the measurements obtained from monitoring stations. In this study, the closest point approach is used: an activity concentration on an arbitrary point on the domain is given by the closest monitoring station measurement This technique is very simple and its cost is weak in comparison with other methods like kriging techniques for example. Moreover, Abida *et al.* (2008) have demonstrated that the interpolation scheme should little affect the optimisation of the network which is sensitive to the cost function

Cost function

The cost function allows to evaluate the performance of a given network by a scalar. The mathematical objective used in the design is that observations are to be used to estimate the activity concentrations over France. Both potential locations and estimation points were defined by the cells of a grid, whose resolution is 0.25° x 0.25°.The mathematical criterion defined to evaluate a network is given by a Hölder norm:

$$J_\alpha = \left(\frac{1}{N} \sum_{i=1}^{q} \sum_{k=1}^{n} \left| \overline{c_k^i} - c_k^i \right|^\alpha \right)^{\frac{1}{\alpha}} \quad (1)$$

which measures the departure of the estimation activity concentration field $\bar{c}$ to the reference field c, for many situations indexed by $i = 1, \ldots, q$, such as dispersion accidents $k = 1, \ldots, n$, runs on all space and time grid cells of the domain. Exponent α allows to tune the importance of high concentrations in the design. The optimisation of the cost function (1) is performed by using a simulated annealing method with a swap procedure allowing to keep the number of stations constant.

Truncation of the cost function

Since the sum in (1) has N= q x n = 1.4 x 10^9, most of the computational time is spent in evaluation of the cost function. This computation was parallelised on a cluster of 40 cores but the CPU time remains considerable in the situation where we have several computations to perform. Consequently, a reduction of the database has been carried out. However instead of reducing bluntly the database using PCA or a variant, we shall demonstrate its usefulness at the level of the cost-function. Indeed, the optimisation of the network is a nonlinear process (unrelated with the fact that the dispersion physics could be linear). Therefore a reduction of the database would have an impact on the design which is difficult to estimate. That is why rather than reducing directly the database by a PCA variant, we will operate at the level of the design cost function, and deduce the way the database should be reduced to retain most of the objective criterion. Saunier *et al.* (2009) have defined two approaches named PCA-1 and PCA-2 however, we introduce only the PCA-1 technique which is the one used for all computations. We consider H as the Jacobian matrix in IR^{nxq}. It has components $[H]_{k,i} = c_k$ and represents the collection of all accidents. The aim of this technique is to carry out component analysis to get a truncated cost function which would then be used as a substitute for the total cost function to be used in the network design optimisation.

The Jacobian matrix H has a singular value decomposition:

$$H = \sum_{i=1}^{q} \lambda_i v_i u_i^T$$

where $u_{i=1,\ldots,q}$ is an orthonormal basis of IR^q, $v_{i=1,\ldots,q}$ is an orthonormal vector set of IR, and the $\lambda_{i=1,\ldots,q}$ are the singular values of H. It is assumed that $q \le n$. We can prove for $\alpha = 2$ that the reference cost function (1) is equivalent to:

$$J_2 = \left(\frac{1}{N} \sum_{i=1}^{q} \sum_{k=1}^{n} \lambda_i^2 \left| \overline{v_k^i} - v_k^i \right|^2 \right)^{\frac{1}{2}}$$

where $\overline{v_i}$ is the interpolation of v_i from the measurements on a given network.

A truncation of the expansion is then obtained by retaining the $r \le q$ first eigenvectors and:

$$J_2^{PCA-1} = \left(\frac{1}{N} \sum_{i=1}^{r} \sum_{k=1}^{n} \lambda_i^2 \left| \overline{v_k^i} - v_k^i \right|^2 \right)^{\frac{1}{2}} \quad (2)$$

Theoretically, the formula (2) is not extensible for $\alpha \ne 2$. However, several computations for $\alpha = 1$ have shown that optimal network obtained with reference cost function (1) and its truncation (3) were similar. Consequently, the cost function that we shall use is the truncated version of (1) for $\alpha = 1$:

$$J_1^{PCA-1} = \frac{1}{N} \sum_{i=1}^{r} \sum_{k=1}^{n} \lambda_i^2 \left| \overline{v_k^i} - v_k^i \right| \quad (3)$$

where $N = r \times n$.

In practice, one would have to:

1. Compute $H^t H$ of lower dimension $\mathbf{IR}^{qxq}$, and diagonalise it into the eigensystem $\{u_i, \lambda_i^2\}_{i=1,\ldots,q}$,

2. Depending on the inertia $\sum_{i=1}^{r} \lambda_i^2$ choose a truncation order $r \le q$,

3. Then the corresponding singular vectors in state space are given by $v_i = \frac{1}{\lambda i} H u_i$, with $1 \le i \le r$,

4. Use this truncated vector set to perform network design optimisation on the basis of the cost function (3).

Practically, the first modes are obtained via the iterative Lanczos algorithm (Paige, 1970). This method called PCA-1 will be used for all computations.

Optimisation of the network

As in Abida *et al.* (2008), we resort to simulated annealing to minimise the cost function. Simulated annealing is a stochastic optimisation technique inspired from statistical physics (Metropolis *et al.*, 1953; Kirkpatrick *et al.*, 1983). The cost function value $J(\xi)$ is interpreted as the internal energy of the current state ξ of the system (a network here) at some fictitious temperature T. The minimisation is iterative. At each step a new state ξ' is selected randomly. The candidate state ξ' is accepted as the new state of the system if its energy is lower than the former state J (ξ) <J (ξ ').

Otherwise it is accepted with a probability exp (-(J (ξ')-J (ξ))/T), or it is rejected. The temperature of the system is initially high so that the candidate states are often accepted and they can explore the phase state space. Then the temperature is progressively cooled down to the point where the system is frozen and only accepts lower energy states. This algorithm reduces the risk of being trapped into local minima, since it can still escape local minima at finite temperature.

APPLICATIONS

First, a rigorous validation of the PCA-1 approach has been realised by Saunier *et al.* (2009). Consequently, we present directly the principal results of the study performed with PCA-1 approach. First, we define an optimal strategy of deployment for the future renovated network then computations including accidents from foreign power plant in Europe and the French density population are performed.

Sequential deployment

The objective of this section is to determine an optimal strategy for the deployment of the renovated network. As a matter of fact, in practice it will not be possible to deploy the full renovated network in one round, because of the manpower required, and of the necessary calibrations and tests which can be lengthy. A working hypothesis here is to consider a network of 100 stations that will be deployed in 10 rounds of 10 stations. Five scenarios have been contemplated and studied:

- Strategy 1a: The network is deployed by batch of 10 stations. Each batch is optimised on the full domain, given the stations already deployed. Theoretically, this strategy could lead to a sub-optimal final network, as compared to the reference optimal design.
- Strategy 1b: The same as 1a except that the first 20 stations are placed at the nuclear site locations. This strategy could also lead to a sub-optimal final network. Note that the very first 10 stations are optimally chosen within the group of 20 fixed ones.
- Strategy 2a: We consider a pre-computed optimal network of 100 stations, without constraints. The network is deployed by a batch of 10 whose locations are optimally chosen among the remaining sites of the pre-computed optimal network. This way, we guarantee that the final network is optimal.
- Strategy 2b: The same as 2a, except that the pre-computed optimal network of 100 stations has 20 fixed stations (figure 1), one for each nuclear site.
- Strategy 2c: The same as 2b, except that the first 20 stations to be deployed are positioned at the nuclear sites. As for case 1b, the very first 10 stations are optimally chosen within the group of 20 fixed ones.

Obviously this study requires a lot of computations, and we heavily rely on the global PCA-1 approach. The mean error for each one of these strategies is reported in figure 2. First of all, fixing the first 20 stations close to the nuclear sites leads to poor intermediary networks (strategies 1b and 2c). Indeed these stations measure very high activity concentrations that are reported by interpolation on large nearby areas because the network is not dense enough yet. Strategies 1b and 2c become competitive at the end of the deployment (80–90 stations), though still behind. Avoiding to fix the first 20 stations (strategies 1a, 2a and 2b) leads to satisfying intermediary networks. Compared to strategy 1a, strategy 2a leads to a better final network, as expected. However it could have been expected that strategy 1a leads to better intermediary networks than strategy 2a, because of the larger degrees of freedom in the optimisation. It turns out that except for the first batch of 10 stations, this is not the case. This favours a deployment on a pre-computed optimal network of 100 stations (type 2 strategies). Since strategy 2c has been shown to be weaker, strategies 2a and 2b are the preferred ones. Fixing 20 stations near the nuclear sites prevents the design from ignoring the importance of one particular nuclear site that would not be in the central part of the domain. That is why, ultimately, strategy 2b is the one we would opt for. The initial, intermediary and final networks for strategy 2b are displayed on figure 2.

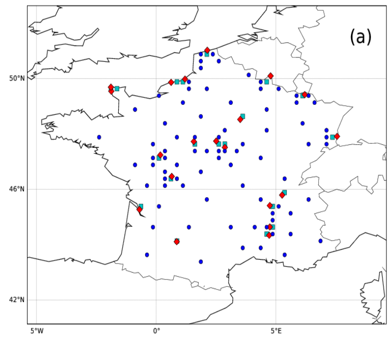

Figure 1. Reference optimal network for 100 stations. The red rhombuses are the nuclear sites. The green squares stand for the fixed stations. The blue disks are the optimised monitoring stations.

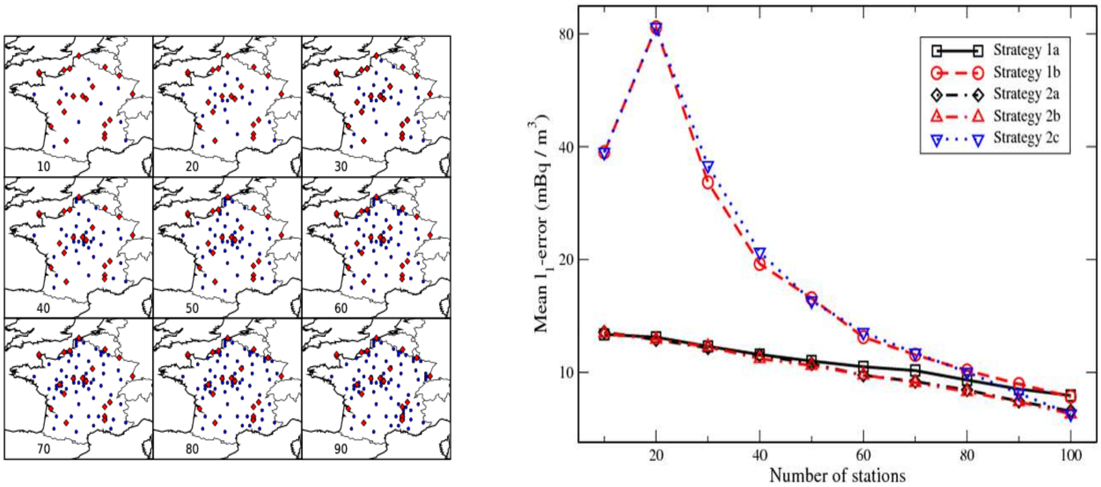

Figure 2. Sequential deployment of the network following strategy 2b (left). The red rhombuses are the nuclear sites. The blue disks are the optimised monitoring stations. Mean error for deployment strategies (right)

Taking into account foreign sources

In this section, the impact of potential major accidents on foreign (in this context: non-French) nuclear power plants is taken into account. All European countries nuclear power plants within a radius of 700 km from one arbitrary but central location (2°25'E, 48°37'N) were retained. There are 34 such power plants, for a total of 54 nuclear sites. An augmented database is built with accident simulations every 21 h over year 2004, and for the 34 additional sites. Again one seeks an optimal network of 100 stations to be deployed on the French territory.

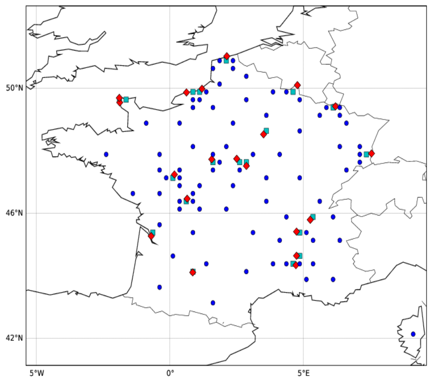

Figure 3. Optimal network of 100 stations in taking account 34 European nuclear power plants outside French borders.

The objective function runs only on the grid cells in France which means that one seeks an optimal mapping only in France. The optimised network with PCA-1 method is given on Figure 2. In comparison with reference network (figure 1), the new network is densified at the borders, at the expense of the centre of France and the South-East part. Moreover, the network is optimised so as to monitor potential accidents from Great-Britain, Switzerland and possibly Germany. The densification is less obvious in the North-East part of France, contrary to our expectations. This may be explained by the West–East dominant wind climatology: virtual accidents from Britain have (on average) higher activity concentration fields over France than those originating from Germany.

Taking into account the French population density

In this section, the population density W criterion is taken into account in the optimisation method. It leads to a new definition of the cost function which can be written for $\alpha = 1$:

$$J_1 = \frac{1}{N}\sum_{i=1}^{q}\sum_{k=1}^{n} W_k \left|\overline{c_k^i} - c_k^i\right| \quad (4)$$

For this computation, only French power plants are considered. PCA-1 approach has not been extended to the cost function (4) with density population; therefore it can't be used here. The network optimised from complete accidental database is displayed in figure 4. We notice that the distribution of stations is significantly different from reference network without constraints (figure 1). As a matter of fact, the optimal network with population is marked by many stations in Rhône valley. On the other hand, we notice that the optimal network with population is less concentrated in the vicinity of western central power plants.

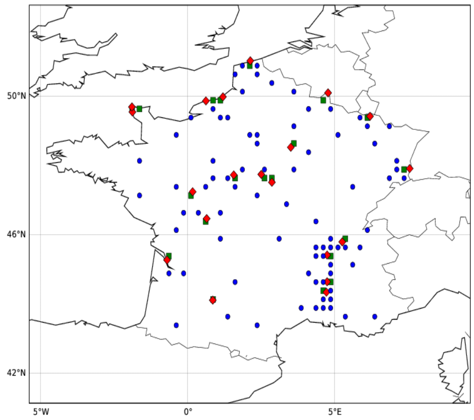

Figure 4. Optimal network of 100 stations in taking account French density population

CONCLUSION

A reduction method has been developed to speed-up the optimisation of an atmospheric monitoring network. The performance of a given network is evaluated on a cost function that measures the discrepancy between an accidental scenario and the related plume reconstruction thanks to the network observations. This ability is evaluated on a large database of accidents that typically represents one year of meteorology, many potential sources and release dates of accidents. The stochastic optimisation of the design requires a large number of evaluations of the cost function. Even with appropriate parallelisation of the cost function evaluation, the optimisation is often still demanding, and the number of accidents/configurations should be reduced. We have shown that this can be done using methods based on principal component analysis of the accidents database.

This PCA-1 method is very flexible and consequently it allows to deal with essential questions. First, given that the network will be deployed sequentially over several years, it is important to define an optimal strategy. The most appealing one is based on a pre-computed 100-station optimal network, with 20 of the stations attached to the nuclear potential sources. At each step, 10 stations are deployed with locations optimally chosen among the pre-computed network station locations. The intermediary networks show a good performance and the final one is optimal by definition. Secondly, the method was applied to an extended optimisation which included additional nuclear sites outside the borders of France. It was shown that the stations in the centre of France in the original design were depleted to the profit of the North-West and East of France, so as to monitor the British, German and Swiss nuclear plants. Finally, the density population has been included in the method and computations have demonstrated that the stations in the western central part of France in reference network were relocated along the Rhône valley.

REFERENCES

Abida, R., Bocquet, M., Vercauteren, N., and Isnard, O., 2008: Design of a monitoring network over France in case of a radiological accidental release. *Atmospheric Environment*, **42,** 5205–521

Anthes, R.A., Warner, T.T., 1978: Development of hydrodynamic models suitable for air pollution and mesometeorological studies. *Mon.Weath.* Rev.106, 1045–1078.

Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P., 1983: Optimization by simulated annealing. *Science* **220,** 671–680.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953: Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21,** 1087–1092.

Paige, C.C., 1970: Practical use of the symmetric lanczos process with re-orthogonalization. *BIT Numer. Math*. **10,** 183–195.

Quélo, D., M. Krysta, M. Bocquet, O. Isnard, Y. Minier and B. Sportisse, 2007: Validation of the Polyphemus platform on the ETEX, Chernobyl and Algeciras cases. *Atmospheric Environment,* **41,** 5300–5315.

Saunier, O., Bocquet, M., Mathieu, A. and Isnard, O., 2009: Model reduction via principal component truncation for the optimal design of atmospheric monitoring networks. *Atmospheric Environment,* **43**, 4940–4950.