Snow water equivalent estimation using blackbox optimization *

Stéphane Alarie † Charles Audet‡ Vincent Garnier‡
Sébastien Le Digabel‡ Louis-Alexandre Leclaire†

February 23, 2011

Abstract: Accurate measurements of snow water equivalent (SWE) is an important factor in managing water resources for hydroelectric power generation. SWE over a catchment area may be estimated via kriging on measures obtained by snow monitoring devices positioned at strategic locations. The question studied in this paper is to find the device locations that minimize the kriging interpolation error of the SWE. This is done first by formulating a simulator blackbox that takes a set of locations as inputs and returns the interpolation error, and then to minimize this error using the mesh adaptive direct search (MADS) algorithm designed for blackbox optimization. The fact that the optimization variables represent planar coordinates is used to devise algorithmic strategies that dynamically groups subsets of variables. The methodology is applied to three water-resource systems in the province of Québec on the blackbox simulator and on a surrogate with various grouping strategies.

Keywords: Snow Water Equivalent, Mesh Adaptive Direct Search, Blackbox optimization, Groups of variables, Kriging.

*Works of the second author is supported by NSERC grant 239436-05, AFOSR FA9550-07-1-0302, and ExxonMobil Upstream Research Company.

†Expertise Réseaux électriques et Mathématiques, Institut de recherche d'Hydro-Québec (IREQ), 1800, boul. Lionel-Boulet, Varennes, Québec, J3X 1S1, Canada, alarie.stephane@ireq.ca, leclaire.louis-alexandre@ireq.ca.

‡GERAD and Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal, Québec, H3C 3A7, Canada, www.gerad.ca/Charles.Audet, Charles.Audet@gerad.ca, Vincent.Garnier@gerad.ca, www.gerad.ca/Sebastien.Le.Digabel, Sebastien.Le.Digabel@gerad.ca.

Résumé: L'équivalent en eau de la neige (ÉEN) est un facteur prépondérant dans la gestion des ressources hydroélectriques en période de fonte. Les valeurs de l'ÉEN sur un bassin versant peuvent être estimées par krigeage des mesures provenant d'appareils positionnés sur le territoire. La question étudiée dans le présent document consiste à identifier les emplacements qui minimisent l'erreur d'interpolation résultant du krigeage des mesures d'ÉEN. Cela se fait d'abord par la création d'un simulateur qui prend un ensemble de positions en amont et en renvoie l'erreur d'interpolation, puis de minimiser cette erreur en utilisant l'algorithme de recherche directe (MADS) conçu pour l'optimisation de boîtes noires. Le fait que les variables d'optimisation représentent des coordonnées planaires est utilisé pour concevoir des stratégies algorithmiques groupant des sous-ensembles de variables de façon dynamique. La méthodologie est appliquée à trois systèmes hydriques du Québec et testée pour différentes stratégies de regroupement sur le simulateur et un substitut.
Mots clefs: Équivalent en eau de la neige, Recherche directe sur treillis adaptifs (MADS), Optimisation de boîtes noires, Groupes de variables, Krigeage.

1 Introduction

In the province of Québec, 98% of electricity is generated by HYDRO-QUÉBEC (HQ) and originates from hydroelectric sources [29]. To supply the 60 hydroelectric generating stations, some 90 catchment areas over a vast region of 500,000 km^2 are used to collect and stock the required water. Approximately 30% of the hydraulic reserves stem from melting snow. This value reaches 40% in the northern regions. Accurate measurements of the *Snow Water Equivalent* (SWE) is therefore of crucial concern during spring so that the inflows from the snowmelt flood are not wasted.

SWE estimates can be obtained by regularly measuring the density of snow samples at specific locations called *snow courses*. Due to the broad territory to cover, a snow course can be physically visited at best every two weeks. Storms and other weather events may temporarily prevent access to a snow course, leading to incomplete sets of measurements. There are several strategies to handle the lack of measurements from snow courses. Aircrafts measuring natural gamma radiation from the soil are used in the United States by the National Weather Service to correct a snow accumulation and ablation model that infers the SWE variation from temperature and precipitation data [12]. In [17], snow course measurements are combined with data from aerial markers read from low-flying aircraft, snow pillows (snow accumulation measured by hydrostatic pressure) and airborne gamma radiation campaigns through a conditional autoregressive model to improve SWE estimates. An artificial neural network has been proposed to evaluate SWE in taiga environments [22]. The neural network trained for rendering kriged measurements from snow courses is supplied with satellite

passive microwave data (SSM/I sensors). A multiple linear regression is used to generate gridded estimates of SWE for mountainous areas in south-central Idaho [26]. The inputs consist of data from snow courses and snow pillows (SNOTEL network in western United States), binary data on snowed areas recorded by satellites (MODIS images) and physiographic variables.

Recently, a monitoring apparatus, named GMON (for *Gamma MONitor*), has been developed to infer SWE in distant wilderness locations [14, 37]. A GMON consists of a gamma ray monitoring device fastened to a post approximately three meters above the snow level and covering an area of ~ 100 m^2. Since the gamma rays from the ground are attenuated by the quantity of water between the ground and the monitor, the snow pack covered by the monitor can be inferred by the gamma ray measures. In addition to being almost maintenance-free, the main advantage of a GMON is to provide daily SWE measurements by satellite communications. One could consider installing a GMON at each snow course to bring an immediate benefit to the actual data collection. However, since hydrologists are mostly interested by gridded estimates of SWE over a catchment area, it is then preferable to install GMONs at locations that minimize the overall estimate error from the underlying kriging interpolation.

Monitoring network design based on kriging is a regular practice in hydrology. Applications include groundwater management where networks are designed to detect contaminants [11, 20, 38], to measure water level [44], and to predict groundwater flow [28]. Other applications consist in the location of rain gauges to estimate the precipitation over a given area [13, 34, 39, 41]. Similar network designs concern acid deposition [45] and surface temperatures in lakes and reservoirs [32].

The present paper studies the question of positioning GMONs in an optimal way. The codes and simulations defining the optimization problem are developed by HQ. These tools are proprietary and cannot be executed outside of HQ. Therefore, a simpler surrogate optimization problem is proposed in the present work to replace this blackbox problem. This surrogate is easier to manipulate, requires less computational resources and shares some similarities with the true problem. The development and testing of new algorithmic strategies are conducted on the surrogate before being applied to the true HQ problem.

The optimization is conducted by the Mesh Adaptive Direct Search algorithm (MADS [6]) for blackbox optimization. MADS has been successfully used to solve real problems in [4, 9, 33, 40] and was previously applied to positioning problems: In [5] to find the optimal locations of tsunami detection buoys maximizing warning time to coastal cities, and in [8, 23] as part of a study to compare derivative-free solvers for groundwater supply and hydraulic capture community problems.

The present work develops and uses new features of the MADS algorithm. The main new feature originates from the observation that the variables defining the problem can be grouped in pairs of

planar coordinates. We propose and study various strategies for grouping and regrouping variables to enhance the practical efficiency while maintaining a satisfactory convergence analysis. Another major difficulty of both the true and the surrogate functions to optimize is that, on a given catchment area, the possible locations for the GMONs form a very fragmented domain. This paper introduces a way to address this issue, exploiting the MADS flexibility.

The paper is divided as follows. Section 2 formulates the question of finding the GMON locations that minimize the overall kriging interpolation error as a blackbox optimization problem. Section 3 summarizes the main features of the MADS algorithm for blackbox optimization and proposes a general framework for grouping and regrouping variables. Some strategies are specific to positioning problems and others are applicable to any optimization problem. Section 4 contains extensive numerical results on the surrogate function and on the more expensive true objective function on the Gatineau, Saint-Maurice and La Grande water-resource systems located in the province of Québec.

2 Blackbox description

This section describes the optimization problem of positioning the GMONs in a way that minimizes the kriging interpolation error. The optimization problem that needs to be solved by HQ is referred to as the *true* optimization problem. However, this problem is only internally accessible at the HQ research center and is not available for performing preliminary tests on grouping strategies. Therefore, we constructed a *surrogate* optimization problem [10] designed to mimic some features of the true objective function, with the advantage of being cheaper to evaluate and to be accessible for testing outside of HQ. This section describes both the true and surrogate functions. The next subsection describes the domain where the GMONs can be placed.

2.1 Fragmented domain

The three water-resource systems studied in this work are depicted in Figure 1. These are the *Gatineau* (GAT), *Saint-Maurice* (STM), and *La Grande* (LG) systems and the potential locations are at the one-kilometer scale. The surface of these areas are 24,755, 43,017 and 90,008 km^2, respectively. Dark pixels correspond to the feasible locations for positioning GMONs.

The GAT domain is highly fragmented with only 3.7% feasible locations. The density of feasible locations is the largest with 33.5% on LG. The density on the STM system is 11.7%. The reason why the feasible domain is fragmented is that there are many criteria to satisfy for getting reliable

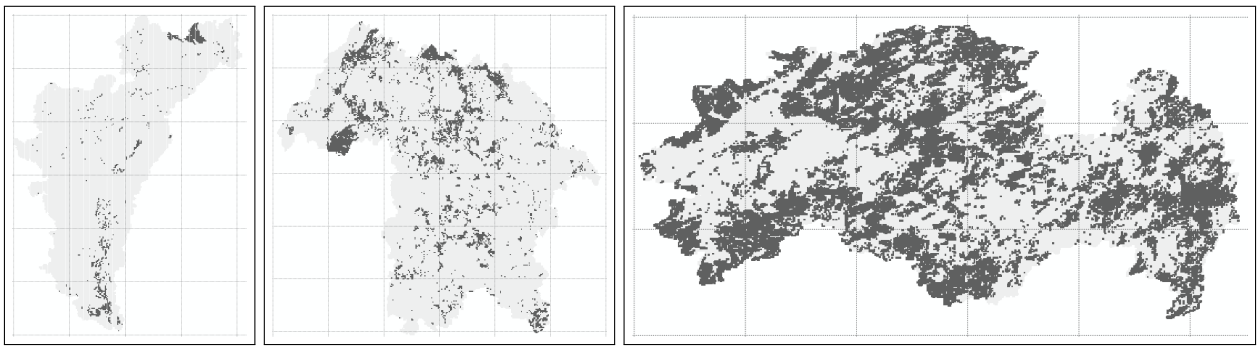

Figure 1: The GAT, STM and LG basins: dark pixels correspond to feasible locations.

SWE data from a GMON. To name a few, a GMON cannot be positioned over water, the slope of the terrain cannot be too great, surrounding vegetation not too dense, and the soil must be sufficiently rich in gamma rays.

The CPU time required to verify if a GMON can or cannot be positioned at a given location is not significant since the position map is explicitly known. Positioning objects in the space $\mathbb{R}^3$ would require more computing resources.

The fragmented nature of the domain can cause numerical difficulties for some optimization methods since attempting to change the location of a GMON will most often lead to infeasible solutions. To address this issue, a preprocessing phase is added so that any location given as input is modified to coincide with a feasible point nearby. This treatment is performed via a deterministic spiral walk on the set of pixels. More precisely, if a given coordinate $x \in \mathbb{N}^2$ is infeasible, then pixels in

$$\{u \in \mathbb{N}^2 : \|u - x\|_\infty = d\}$$

are successively tested for values of d varying from 1 to 15, and by considering the ones closest to x first. The process stops as soon as a feasible pixel is found, and x is moved to it. Otherwise, x is declared infeasible and the evaluation is declared a failure. Again, due to the two-dimensional nature of the problem, the CPU time required by this spiraling strategy is very small.

2.2 The true objective function

Using kriging to generate gridded estimates of SWE over a territory also provides an error estimate at each point of the grid, which is referred to as the *error map*. To evaluate the SWE error from a given set of GMONs, kriging must be performed with the values collected by the GMONs. An important difficulty is that the GMONs are not yet on the territory. We next describe a way to infer

the values returned by the fictive GMONs from any locations based on historical data. Using a date for which numerous SWE measurements are available, we generate gridded SWE estimates labelled as as the reference values for that date. Having such a *reference map*, one can infer the SWE value that a fictive GMON would return from any point on the map.

Therefore, the true objective function is evaluated as follows: Given m GMON locations, the value measured by each GMON is read from the location on the reference map. The gridded SWE estimates is next kriged from those values and a error map is generated. In order to compare various error maps during the optimization process, the values of the error map are summed and returned as the objective function value. This sum gives an appreciation of the quality of the current GMON locations. The smaller the error sum, the better the set of locations.

In the present work, the reference map date is March 15, 1990. Kriging is performed with the ISATIS software package [25]. Ordinary kriging is currently used but kriging with external drift will replace it in future work to consider additional information from secondary variables, such as elevation of measurement locations.

In summary, the computer simulation takes the Cartesian coordinates of the m GMON locations as input, applies the spiralling strategy on each GMON if necessary, and returns an error map or a flag indicating failure. The error map is then converted into a single value by summing the kriging errors at each pixel in the catchment area. This sum is the objective function value to minimize. The whole process requires between 1 and 2 seconds to compute.

2.3 The surrogate objective function

As mentioned above, the true objective function can only be executed from within the HQ installations. The present section describes a surrogate function, developed as a substitute for the true function. Using a surrogate has three important advantages in the present situation. First it allows us to test algorithms and grouping strategies outside of the HQ research center. Second, the CPU time required for the evaluation of the surrogate is very low and testing in a parallel environment is possible. Third, the surrogate is used in Section 4.3 to guide the MADS algorithm for optimizing the true objective function.

Ideally a surrogate shares similarities with the true function while being cheaper to evaluate. Surrogates may be adaptive and dynamically updated based on past evaluations of the true function as in the *surrogate management framework* described in [10] and applied in [36]. However, in the present work, the considered surrogate is non-adaptive meaning that it is simply viewed as another blackbox that implements a simplified model.

The surrogate is based on a potential-well function that decreases with the distance to the GMON. It models the error distribution of the SWE estimate for a set of GMONs. The surrogate is given by Equation (1)

$$f(x) = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \alpha_{ij} e_{ij}(x) \tag{1}$$

where N_i and N_j are the map dimensions and $x \in \mathbb{N}^{2m}$ is a vector composed of the locations of the m GMONs. The boolean value α_{ij} is equal to 1 for all pairs (i, j) belonging to the water-resource system and to 0 otherwise (e.g., for the STM system, $\alpha_{ij} = 0$ if and only if the corresponding pixel from Figure 2 is white).

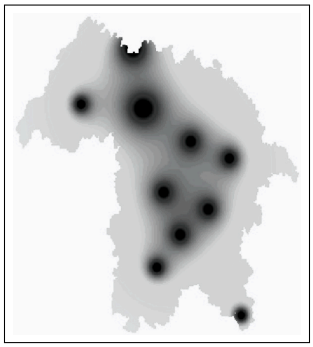

Figure 2: Example of a surrogate error map for the STM water-resource system

The value $e_{ij}(x)$ corresponds to the error at the location (i, j) and depends on the GMON locations. This value equals 1 when there are no errors, and greater than 1 otherwise. It is calculated as follows

$$e_{ij}(x) = \begin{cases} 1 & \text{if } d_{ij}(x) \geq 1, \\ (d_{ij}(x))^{-0.5} & \text{if } 0.030 \leq d_{ij}(x) < 1, \\ (d_{ij}(x))^{-0.6} & \text{if } 0.025 \leq d_{ij}(x) < 0.03, \\ (d_{ij}(x))^{-0.75} & \text{if } 0.020 \leq d_{ij}(x) < 0.025, \\ (d_{ij}(x))^{-1} & \text{if } d_{ij}(x) < 0.020, \end{cases}$$

with

$$d_{ij}(x) = \sum_{q=1}^{m} c_{ij}(x_{2q-1}, x_{2q})$$

and where

$$c_{ij}(x_u, x_v) = \begin{cases} 1 & \text{if } x_u = i \text{ and } x_v = j, \\ \frac{0.8}{\sqrt{(x_u - i)^2 + (x_v - j)^2}} & \text{otherwise.} \end{cases}$$

The error e_{ij} is minimal at the GMON location (i, j). Figure 2 illustrates the values of e_{ij} on the STM system for 12 GMONs. The darker the pixel, the lower is the SWE error. In this example, there are two locations covered by two GMONs, as seen by the larger circles in the top part of the figure.

The constant values appearing in the definition of c_{ij} and of e_{ij} were not arbitrarily chosen. In fact, a large number of values for these constants were considered and submitted to experts at HQ. The corresponding error maps such as the one represented in Figure 2 were compared to true error maps and analyzed, and the experts agreed on the ones that gave satisfactory visual representations of the errors.

Note finally that the surrogate blackbox also uses the preprocessing spiralling strategy described in Section 2.1 to handle the fragmented domain.

3 Mesh Adaptive Direct Search and groups of variables

MADS [6] is a directional direct search algorithm from the family of derivative-free methods reviewed in [16]. MADS is designed to solve constrained blackbox minimization of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ over a domain $\Omega \in \mathbb{R}^n$. The functions defining f and Ω are typically obtained through a computer simulation which is often expensive and may fail to return a result even at some feasible trial points.

A mechanism for handling groups of variables is detailed in [7] in the context of parallelism, where the difficulty of dealing with numerous variables is addressed by working on subsets of variables. That work focusses on the parallel framework based on generic groups and no specific strategy was used to construct these groups. The present study proposes strategies for grouping variables, some are specific to positioning problems while others are generic. Another difference with the present work is the way by which the convergence analysis is ensured. In [7], a specialized parallel process considers all the variables with a reduced number of directions. Here the convergence analysis is derived from the grouping strategy evolution.

3.1 MADS overview

The MADS algorithm evolved from the generalized pattern search (GPS) [43] which itself evolved from the coordinate (or compass) search (CS) [18]. These are all iterative algorithms where every iteration attempts to generate a feasible trial point in $\mathbb{R}^n$ having a better objective function value than the current best solution, denoted $x_k \in \mathbb{R}^n$. If iteration k is a success, then iteration $k + 1$ starts

with a new incumbent x_{k+1}, otherwise it starts with $x_{k+1} = x_k$. Each iteration tests a finite number of trial points and the algorithms differ by the way that these trial points are generated.

There are some conditions that these trial points must satisfy. Some of these points must be generated during what is called the POLL step at each iteration. The convergence analyses of these methods rely on the study of the points generated during the POLL step. In CS, there are exactly $2n$ POLL points at each iteration:

$$\{x_k \pm \Delta_k e_i : i = 1, 2, \ldots, n\},$$

where e_i is the ith coordinate vector in $\mathbb{R}^n$, and Δ_k is a strictly positive real number that depends on the iteration number k. This set of POLL points can also be written as the union of n sets

$$\{x_k \pm \Delta_k e_1\} \cup \{x_k \pm \Delta_k e_2\} \cup \ldots \cup \{x_k \pm \Delta_k e_n\}.$$

In MADS, the POLL points can be generated with more flexibility, as the method ensures that the set of normalized directions

$$\bigcup_{k=0}^{\infty} \left\{ \frac{t_k - x_k}{\|t_k - x_k\|} : t_k \text{ is a POLL point generated at iteration } k \right\}$$

is dense in the unit sphere. The interested reader may consult [2] for details about how such directions may be constructed in practise. For CS, the union of normalized directions is simply the finite set $\{\pm e_1, \pm e_2, \ldots, \pm e_n\}$. For GPS, this union also consists of a finite set, and the convergence results of CS and GPS are limited to these sets. Only the MADS algorithm ensures that the Clarke generalized derivative [15] $f^\circ(\hat{x}; v)$ (evaluated at some accumulation point of the limit $\hat{x}$ of unsuccessful iterates x_k, where f is locally Lipschitz) is nonnegative for every direction v in the hypertangent directions.

3.2 Grouping variables

There are cases where there is some known structure concerning the variables. In [23] for example, the variables are partitioned into groups of three, where two of them define the location of a pump, and the third a pumping rate. If one wants to modify a solution, it seems more natural to explore modifications of the locations and pumping rate of a single pump rather than simultaneously changing the x coordinate of a first pump, the y coordinate of a second pump, and the pumping rate of a third one.

More generally, let $N = \{1, 2, \ldots, n\}$ be the set of indices of all the optimization variables. Groups of variables are obtained by considering a finite number of subsets $N_q \subseteq N$ with $q \in Q$. The number of variables in the group N_q is denoted by $n_q = |N_q|$.

At iteration k, the MADS algorithm generates trial points on the mesh

$$M_k = \left\{ x_k + \Delta_k^m Dz : z \in \mathbb{Z}_+^{|D|} \right\}$$

where $\mathbb{Z}_+$ is the set of the nonnegative integers, Δ_k^m is a strictly positive real number called the mesh size parameter, and D is a constant positive spanning matrix [3, 19]. The MADS algorithm with groups of variables requires more structure on the matrix D: for each group N_q, there exists a subset D_q of the columns of D that positively spans the subspace defined by the variables of N_q. Notice that this supplementary condition is trivially satisfied with $D = [I; -I]$ used in the two existing MADS implementations LTMADS [6] and ORTHOMADS [2] as well as in CS.

Figure 3 illustrates a grouping strategy on a problem with six variables representing the locations in $\mathbb{R}^2$ of three GMONs. The initial location vector $x_k \in \mathbb{R}^6$ is represented in the three subfigures by the $\bullet$ symbols. Trial points generated by CS will only differ from x_k by a single variable, as represented by the $\otimes$ symbol in the leftmost illustration, in which a single GMON moved vertically. With MADS, the three GMONs are allowed to move simultaneously on a fine mesh as shown in the central illustration. If three groups are formed with the new strategy, then a single GMON can be moved on the fine mesh, as depicted on the third subfigure.

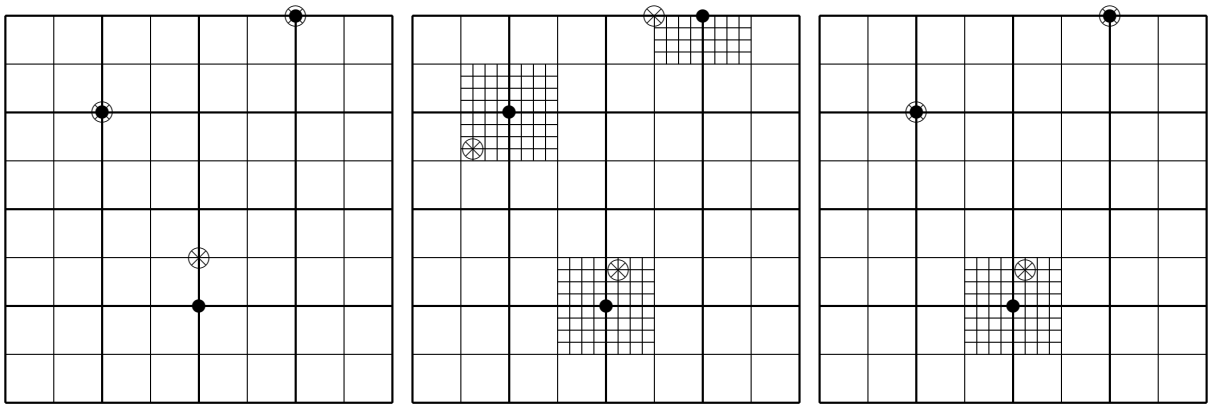

Figure 3: The $\bullet$ symbols represent the initial locations of three GMONs and the $\otimes$ symbols the locations provided by a local POLL.

With this strategy, CS can be seen as a specific grouping instance in which there are n groups, each with a single variable, and MADS can be seen as an instance with a single group containing all n variables.

3.3 Clarke stationarity in subspaces

The MADS algorithm with groups of variables belongs to the class of MADS algorithm, and therefore it inherits its convergence analysis. The fundamental result is Theorem 3.12 of [6] that relies on the definition of a refining point $\hat{x}$. A refining point is the limit of a sequence of convergent feasible unsuccessful iterates $\{x_k\}_{k\in K}$. The results ensures that if the objective function f is Lipschitz near the refining point $\hat{x}$, then the Clarke directional derivative satisfies $f^\circ(\hat{x};d) \geq 0$ for every refining direction d. The set of refining directions is:

$$\bigcup_{k\in K} \left\{ \frac{t_k - x_k}{\|t_k - x_k\|} : t_k \text{ is a trial point generated at the unsuccessful iteration } k \right\} .$$

With the standard MADS implementaions [2, 6], the set of refining directions is dense in the unit sphere. With CS, the set of refining directions is simply composed of the $2n$ positive and negative standard orthogonal directions.

Theorem 3.1 *Let F_q be the linear subset of $\mathbb{R}^n$ spanned by the variables in the group N_q, for every $q \in Q$. If the directions used to generate the* POLL *points are using* MADS *in each n_q dimensional subspace F_q, then the set of refining directions will be dense in the union of the subspaces $\bigcup_{q\in Q} F_q$.*

Proof. The additional restriction on the set of directions D used in conjunction with the groups of variables makes it such that the directions used by the algorithm to generate the POLL points will necessarily be contained in one or more subspaces F_q for $q \in Q$. Using the MADS directions in the subspaces implies that positive bases will be used in each subspace, leading to asymptotically dense set of directions in each F_q. Therefore, the set of refining directions is dense in $\bigcup_{q\in Q} F_q$. □

Consider for example a problem in $\mathbb{R}^5$ with three groups $N_1 = \{1, 2, 3\}$, $N_2 = \{3, 4\}$ and $N_3 = \{5\}$. This would mean that the convergence result could only guarantee that the Clarke derivatives are nonnegative in the 3-dimensional subspace $\{x \in \mathbb{R}^5 : x_4 = x_5 = 0\}$, the 2-dimensional subspace $\{x \in \mathbb{R}^5 : x_1 = x_2 = x_5 = 0\}$ and the 1-dimensional subspace $\{x \in \mathbb{R}^5 : x_1 = x_2 = x_3 = x_4 = 0\}$. It would be possible to have a problem where $f'(\hat{x};d) < 0$ at a refining point in the direction $d = (0, 0, 0, 1, 1)$ for example.

In summary it appears natural to explore subspaces associated to groups of variables when the problem possesses such a structure. However restricting the exploration to these groups weakens the theoretical convergence analysis and may lead to practical difficulties by making it impossible to escape from an undesirable region by varying only a subset of variables.

A trivial way to ensure that the set of refining directions be dense in the unit sphere is to make sure that one of the group contains all the variables: $N_q = N$ for one q in Q. This strategy is easy

to implement but requires a large number of function evaluations. A more sophisticated strategy consists in periodically reforming the groups in such a way that after finitely many reconfigurations a single group contains all the variables. The easiest way to implement this strategy is to simply set $Q = \{1\}$ and $N_1 = N$ when a stopping criteria is met for the original grouping. Another strategy consists in progressively merging groups together as proposed in the next subsection.

3.4 Grouping and dynamic regrouping strategies

Grouping and regrouping the n variables requires three elements: an *initial grouping strategy*, a *stopping criteria for regrouping* and a *regrouping strategy*. In this section, we propose strategies for these three elements. Every combinations of these strategies are tested in Section 4.

Initial grouping strategy

In the context of GMON positioning, n is an even number and (x_{2q-1}, x_{2q}) represents the Cartesian locations of the q-th GMON. Let m denote the number of GMONs, and therefore there are $n = 2m$ variables. The *initial grouping* strategies studied in this paper are chosen from the following:

INDIV : Individual variable: $N_q = \{q\}, q = 1, 2, \ldots, n$.
PAIRS : Pairs by GMON in Cartesian coordinates: $N_q = \{2q-1, 2q\}$, $q = 1, 2, ..., m$.
ALL : All n variables belong to a single group: $N_1 = N$.

Stopping criteria for reconfiguration

The reconfiguration should be frequent enough to be useful, but should not impede the progress of the algorithm. To limit the number of tests and generate useful results for the true location problem from HQ, we opted for the following strategy:

Step 1 Set $n_r = 1$.
Step 2 Launch MADS. Terminate after n_r consecutive failed MADS iterations.
Step 3 Set $n_r \leftarrow n_r + 1$ and go to Step 2.

Regrouping strategy

A coherent set of rules that manages the merger, division and conservation of groups of variables is called *regrouping strategy*. The regrouping algorithms used are:

STATIC	:	Keep the same groups.
DIST	:	Merge by smallest distance.
CLUSTER	:	Merge by distance with a *K-means* method.
MVT	:	Group all variables that did not change in the previous run.
REGRES	:	Use linear regression to identify important variables.

Below is a more detailed description of these regrouping strategies.

STATIC**:** This strategy does not modify the groups. It is presented for comparison purposes as it corresponds to existing optimization algorithms. Combining the STATIC regrouping strategy with INDIV initial grouping is equivalent to the CS algorithm. Combining it with the ALL initial grouping is equivalent to the MADS algorithm.

DIST**:** This regrouping strategy is specific to positioning problems. The first time that it is invoked, it groups the four variables associated to the two GMONs that are the closest to each other. The remaining variables are grouped in pairs as in the initial grouping PAIRS. Then, every time that the strategy is called, it moves to the group the pair of variables corresponding to the closest GMON. Ties are broken arbitrarily. This regrouping strategy ensures that after m calls, there will be a single group containing all variables.

CLUSTER**:** This regrouping strategy is specific to positioning problems. The first time that the strategy is called, the variables are grouped into m pairs, as in the initial grouping PAIRS. Then, each time that the strategy is invoked, the pairs are clustered using a K-means algorithm [27]. In the numerical tests, we use the `kmeans()` function from the `R` software [42]. The number of clusters, initialized to m, decreases incrementally with each new call to this strategy, resulting in a single group containing all variables after m calls.

MVT**:** This strategy compares the initial solution x_0 supplied to the previous run of the MADS algorithm (from Step 2) with the final solution x_k produced by that same run. All variables associated to the GMONs that did not move are grouped together. This strategy allows stationary GMONs to be moved together in order to unlock suboptimal configurations. As a local optimal configuration is approached, the movement of the GMONs become less frequent, and will necessarily entail a single group containing all variables.

REGRES: In optimization, the objective function value is often more sensitive to some variables than others. For example, analyses of variance are performed in [10] to reduce the size of an helicopter rotor design problem. The REGRES strategy consists of a linear regression with simple rank transformation [30, 31] using trial points evaluated during the previous step. The `lm()` function from the `R` software is used to rank the variables in decreasing order of regression coefficient magnitude. The first variables have priority and are grouped individually (one group per variable), while the remaining ones are placed in a single group. Each time this strategy is invoked, the set of priority variables is reduced by two elements, leading to a single group after m calls.

All regrouping strategies presented above (except STATIC) ensure that after finitely many reconfigurations, there will be a single group containing all the variables. This implies, at that point, that the algorithm corresponds to the MADS algorithm described in [6] and inherits from its convergence analysis.

4 Numerical experiments

Numerical experiments are conducted in two phases. The first one studies combinations of preprocessing, grouping and regrouping strategies on the surrogate objective function. The aim of this first phase is to identify a limited number of strategies that stand out as being preferable to others. These specific strategies are then tested on the true problem in the second phase.

4.1 Testing methodology

For each of the three water-resource systems, we study the situations in which the number m of GMONs varies from 5 to 10. This leads to a total of 18 instances. For the tests conducted on the surrogate function, realist initial GMON locations were generated by visually spreading the m GMONs over the domain for each of the instances. For the tests on the true function, the starting point for a given instance consists of the best solution produced by all runs made while optimizing the surrogate problem.

The NOMAD blackbox optimization software [1, 35] is used for the numerical tests of this paper, with all default settings. In particular, these settings ensure that the components of the initial mesh size parameter Δ_0^m associated to the variables with even (odd) index are set to one tenth of the maximum length (width) of the considered map. The settings also ensure an opportunistic approach, i.e., an iteration of the MADS algorithm terminates as soon as a new improvement is made.

The regrouping strategies consist in launching a series of NOMAD runs with different sets of parameters and stopping criteria. More precisely, every time a regrouping strategy is performed, NOMAD is launched from the best solution from the previous run, with the same mesh size parameter from the end of the previous run.

For comparison purposes, we set the overall stopping criteria to be a limit of 1000 on the total number of blackbox evaluations, regardless of the problem. This value is reasonable in terms of computing time, and reveals the evolution of different algorithms, both on the problem of HQ and on the surrogate.

Tables are presented below to summarize the results of the numerical experiments. Each row corresponds to a specific combination of initial grouping and regrouping strategies. Columns give the average relative improvement (%) in objective function value after a total evaluation budgets (TEB) chosen in the set $\{250, 500, 1000\}$. The average is taken over the 18 instances, and the relative improvement is computed with respect to the objective function value of the initial solution. The rows are presented by decreasing order of the mean values.

The numerical experiments are conducted in two phases. Section 4.2 establishes a basis for comparison by launching the MADS algorithm without any regrouping strategy, and analyses the effect of the spiralling strategy proposed in Section 2.1. Then, computational results are reported for combinations of grouping and regrouping strategies on the surrogate problem. The most promising strategies are then tested in Section 4.3 on the true function from HQ.

4.2 Testing strategies on the surrogate optimization problem

The first three rows of Table 1 summarize the performance of three static strategies, without the spiralling strategy designed to produce feasible locations. The INDIV strategy corresponds to coordinate search, in which the algorithm only changes a single variable at a time. The ALL strategy corresponds to the MADS algorithm, and allows simultaneous variations of all variables. The PAIRS strategy is intermediate, and attempts to move the two variables defining the location of a single GMON. The three bottom rows of the table compare the same strategies but uses the spiralling treatment to move the trial points to nearby feasible locations. Without any surprise, the spiralling treatment significantly allow all strategies to generate better solutions. Therefore, all further tests are performed with the spiralling treatment.

The next series of tests were performed to exploit the fact that the MADS algorithm uses an opportunistic strategy when processing a list of trial points. This means that the evaluation of the trial points in the list can be interrupted as soon as a new incumbent solution is generated. Therefore, the

Table 1: Effect of the spiralling strategy for feasibility for static strategies. Numbers correspond to the average relative improvement compared to the starting solution.

Without spiralling

Initial	Regroup	TEB = 250	TEB = 500	TEB = 1000	Mean
PAIRS	STATIC	0.89	0.92	0.92	0.91
ALL	STATIC	0.77	0.77	0.77	0.77
INDIV	STATIC	0.42	0.42	0.42	0.42

With spiralling

Initial	Regroup	TEB = 250	TEB = 500	TEB = 1000	Mean
PAIRS	STATIC	1.31	1.60	1.71	1.54
ALL	STATIC	1.23	1.55	1.71	1.50
INDIV	STATIC	1.25	1.53	1.70	1.49

order in which the trial points are evaluated can affect the number of function evaluations required by the algorithm. Table 2 shows the performance of the same three strategies when ordering the trial points following the values on the error map associated to the GMON locations. The rationale behind this ordering is that positioning GMONs where the error is large is probably more efficient than positioning them in places where the estimation has a low error. The ordering strategy helped the algorithm on almost all instances tested, and therefore it will be applied in the remaining tests on the surrogate function.

Table 2: Effect of ordering trial points for static strategies.

Initial	Regroup	TEB = 250	TEB = 500	TEB = 1000	Mean
INDIV	STATIC	1.40	1.65	1.74	1.60
PAIRS	STATIC	1.36	1.61	1.70	1.56
ALL	STATIC	1.30	1.54	1.64	1.50

These numerical results suggest that moving individual variables is preferable to moving pairs of variables, and that both these strategies are preferable to changing simultaneously the values of all the variables. These conclusions are unusual since the theoretical convergence results are stronger

with the ALL strategy. This behavior can be explained by the use of the error map for ordering the trial points. Indeed, the error map does not reflect the consequences of simultaneously varying several variables, as it only quantifies potential locations for GMONs.

Table 3 shows the results of applying all three initial groupings with the four dynamic regrouping strategies for a total of 12 combinations.

Table 3: Grouping and regrouping strategies on the surrogate.

Initial	Regroup	TEB = 250	TEB = 500	TEB = 1000	Mean
INDIV	DIST	1.38	1.59	1.72	1.56
PAIRS	DIST	1.33	1.60	1.71	1.54
ALL	DIST	1.30	1.59	1.73	1.54
PAIRS	REGRES	1.37	1.56	1.74	1.56
ALL	REGRES	1.33	1.57	1.75	1.55
INDIV	REGRES	1.35	1.59	1.69	1.54
PAIRS	CLUSTER	1.34	1.60	1.70	1.55
INDIV	CLUSTER	1.33	1.59	1.70	1.54
ALL	CLUSTER	1.29	1.56	1.75	1.54
INDIV	MVT	1.35	1.57	1.68	1.53
ALL	MVT	1.34	1.56	1.69	1.53
PAIRS	MVT	1.30	1.55	1.72	1.53

HQ provided computer resources to apply six algorithmic strategies on the original problem. For each of the five regrouping strategies, we selected the initial grouping that produced the best mean improvement value. Table 4 lists the selected strategies. The reason why the mean value criterion is used instead of the best solution after 1000 evaluations in to favor robust strategies, i.e., strategies whose performance is less likely to fluctuate. For example, Table 3 reveals that the ALL REGRES strategy found the overall best solution, but its performance after 250 evaluations is inferior to all the selected strategies. For comparison purposes, and even though it did not produce as good results, we added the ALL STATIC strategy which corresponds to applying the MADS algorithm directly without any grouping strategies.

Table 4: Selected initial and regrouping strategies on the surrogate.

Initial	Regroup	TEB = 250	TEB = 500	TEB = 1000	Mean
		Dynamic			
INDIV	DIST	1.38	1.59	1.72	1.56
PAIRS	REGRES	1.37	1.56	1.74	1.56
PAIRS	CLUSTER	1.34	1.60	1.70	1.55
INDIV	MVT	1.35	1.57	1.68	1.53
		Static			
INDIV	STATIC	1.40	1.65	1.74	1.60
ALL	STATIC	1.30	1.54	1.64	1.50

4.3 Numerical results on the true objective function

This section compares the performance of the six selected grouping and regrouping strategies on the true objective function. In all tests, the spiralling treatment for feasibility is applied since the three domains representing the feasible locations are identical to those studied with the surrogate. As mentioned previously, for each instance, the starting point consists of the best solution produced by all runs made with the surrogate.

The surrogate function on which the numerical experiments of the previous section were conducted is still used to guide the optimization algorithm. At each iteration, before invoking the evaluation of the function with ISATIS, the trial points are evaluated and ordered by their surrogate objective value. This strategy aims at evaluating the true objective function value at the most promising trial points first. Preliminary tests showed that this ordering is preferable to using the error map as in Section 4.2.

Unlike the work done in the previous section, the objective is not to identify promising strategies, but instead we want to compare which strategy is the most efficient on the 18 instances. Therefore, we only consider the final relative improvement after 1000 function calls for each instance, and do not consider the evaluations budgets of 250 and 500. Table 5 shows the detailed performance of the six selected combinations of grouping and regrouping strategies for each of the three water-resource systems and for the six possible number of GMONs. The rows in the table are not sorted according to the mean value but are listed with the same order as in Table 4 . The shaded entries indicate which strategy gave the best result for a given column.

Table 5: Selected strategies on the three catchment areas with the true objective function.

Saint-Maurice (STM)

		$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	Mean
INDIV	DIST	3.02	5.30	6.19	7.86	7.37	7.66	6.23
PAIRS	REGRES	3.07	5.31	6.76	7.43	6.55	7.77	6.15
PAIRS	CLUSTER	3.04	5.31	6.75	7.61	6.48	7.44	6.10
INDIV	MVT	3.07	5.28	6.72	7.75	7.33	7.51	6.28
INDIV	STATIC	2.97	5.29	6.58	7.41	6.47	7.74	6.08
ALL	STATIC	3.06	5.29	6.80	7.73	7.06	7.62	6.26

Gatineau (GAT)

		$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	Mean
INDIV	DIST	1.77	4.28	5.74	6.61	5.28	4.08	4.63
PAIRS	REGRES	1.80	4.67	5.65	5.74	5.31	4.27	4.58
PAIRS	CLUSTER	1.80	4.70	5.74	6.60	5.13	4.23	4.70
INDIV	MVT	1.80	4.30	5.74	6.56	5.96	4.24	4.77
INDIV	STATIC	1.77	4.28	5.57	6.44	5.55	3.90	4.58
ALL	STATIC	1.77	4.20	5.69	6.62	4.50	4.21	4.50

La Grande (LG)

		$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	Mean
INDIV	DIST	1.05	2.37	4.14	5.10	6.54	6.28	4.25
PAIRS	REGRES	1.07	2.51	4.22	4.82	6.88	6.40	4.32
PAIRS	CLUSTER	1.07	2.29	4.20	5.09	6.90	6.15	4.28
INDIV	MVT	1.09	2.47	4.24	5.09	6.82	6.31	4.34
INDIV	STATIC	1.05	2.45	4.21	4.92	6.83	6.25	4.28
ALL	STATIC	1.06	2.43	4.23	5.07	6.79	6.04	4.27

The main difference with these results on the true problem with those on the surrogate is that the INDIV STATIC strategy did not perform well. On each of the 18 instances considered, it never found the best solution, and on 7 instances it found the worst solution. The opposite behavior was observed on the surrogate problem.

Another observation is that the ALL STATIC strategy failed to approach the best solution on a large number of test cases. Theses observations suggest that the surrogate function was possibly too simple and did not oppose a challenge to the simple static strategies. The true objective function however is more complex and introduced more nonlinearities and possibly discontinuities which caused difficulties to the static strategies.

Each of the dynamic strategies found the best solution on at least four instances. The PAIRS REGRES strategy stands out by being the only combination that found the best strategy in 7 out of the 18 instances. The INDIV MVT strategy produced the worst solution on only one instance (STM with $m = 6$) and this instance has the smallest gap between the best and worst solutions.

Table 6 orders the six strategies by taking the average over all instances. The shaded cells indicate again the best values for each column. Although the numerical values share the same magnitude, three conclusions can be drawn for this table.

i) For each of the STM, GAT and LG systems, the INDIV MVT strategy dominates all others. Each of the other strategies produced significantly worse averages on at least one of the three systems.

ii) Both static strategies did worse on average than each of the dynamic ones. This observation suggests that dynamically regrouping the variables is preferable than either moving individual variables, or moving all variables simultaneously.

iii) Even if the variables represent locations in $\mathbb{R}^2$, pairing them at the initial iteration does not appear to be beneficial. Both strategies with initial individual variables are preferable on average to the ones where the variables are paired.

Another way to compare different algorithmic strategies, consist in analyzing their performance profile [21]. Figure 4 represents the performance of the strategies versus a scalar α. It shows the proportion out of the 18 test instances for which the strategy produced a solution with an objective function value within $\ln(\alpha)\%$ of the best known value. The higher the curve is, the better is the strategy.

The figure clearly shows that the INDIV MVT strategy dominates all other strategies for most values of α. The INDIV STATIC strategy is dominated by all other strategies on the left part of the

Table 6: Summary of selected strategies on the three catchment areas with the true objective.

		STM	GAT	LG	Mean
INDIV	MVT	6.28	4.77	4.34	5.13
INDIV	DIST	6.23	4.63	4.25	5.04
PAIRS	CLUSTER	6.10	4.70	4.28	5.03
PAIRS	REGRES	6.15	4.58	4.32	5.02
ALL	STATIC	6.26	4.50	4.27	5.01
INDIV	STATIC	6.08	4.58	4.28	4.98

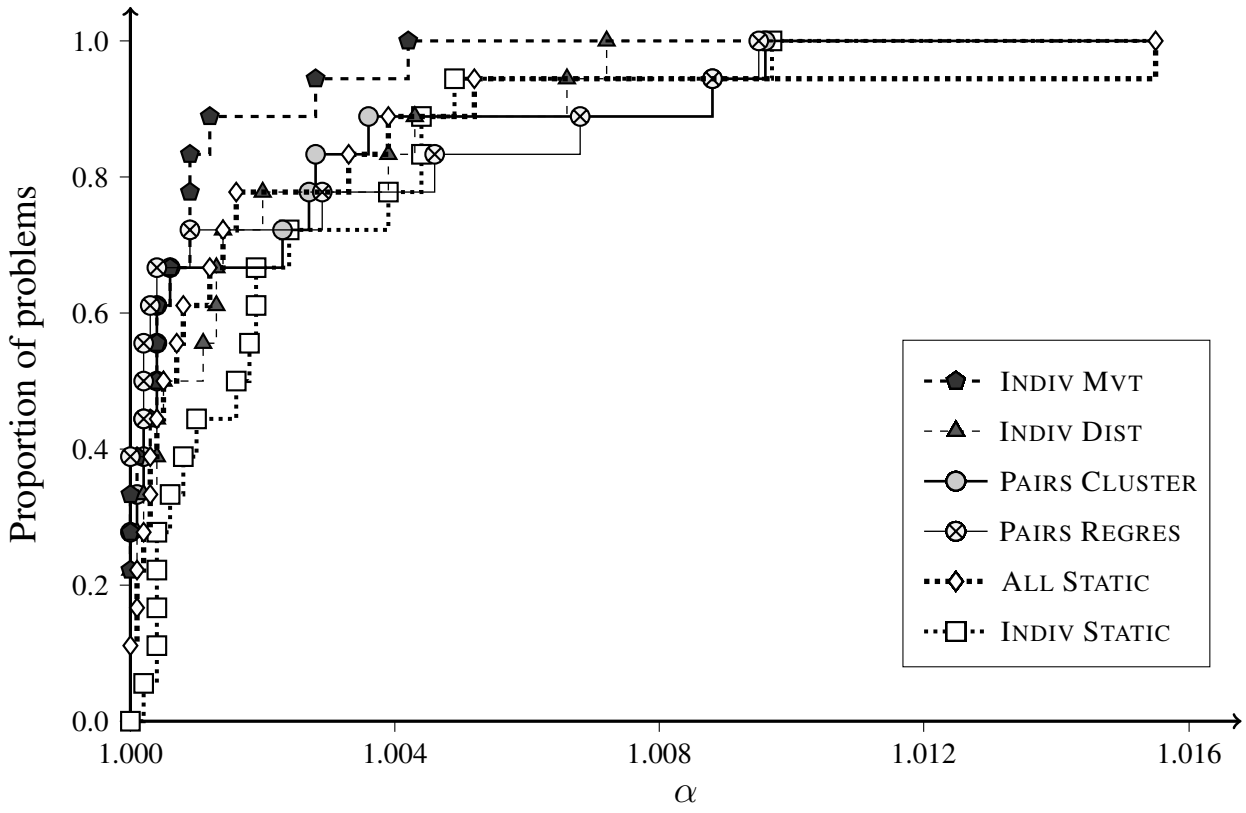

Figure 4: Performance profile [21] for the six selected algorithms applied to the 18 instances using the true objective

figure. This is a consequence of the fact that it never found the best solution on any of the instances. The four other strategies are comparable on the left part of the graph, but when α belongs to the interval $[1.01, 1.015]$, the ALL STATIC strategy is the only one that is does not reach 100%. This indicates that this strategy performed very poorly on some of the test problems. Indeed, analyzing further Table 5 reveals that it did not produce a good solution on GAT with $m = 9$ and on LG with $m = 10$.

5 Discussion

The motivation for this work was to find the GMON locations that minimize the snow water equivalent estimation error. Our prior work in tsunami buoy positioning [5] and in hydraulic capture problem [8, 23] encouraged us to further study grouping strategies for the variables. The present paper develops these ideas by formally defining initial grouping strategies for positioning problems and by introducing regrouping strategies, some specific to positioning, and others generic. The proposed regrouping strategies have the property that after a finite number of iterations all variables belong to the same group and therefore the entire convergence analysis of MADS applies.

The true objective function was not available at the time that this research project was initiated and this lead us to perform exhaustive testing on the surrogate function. Additional initial and regrouping strategies are found in [24] but only the most efficient and intuitive ones are reported in the present paper.

The usefulness of the regrouping strategies is not apparent from the numerical tests on the surrogate function. This is probably because the surrogate function did not capture enough of the complexity of the true objective. However when considering the true function, one strategy stood out as being preferable to the others. This strategy is not specific to positioning problems as it forms groups with idle variables and could be used in a more general context. Furthermore the numerical experiments suggest that dynamically regrouping the variables improves the quality of the final solution compared to the standard MADS and CS methods.

References

[1] M.A. Abramson, C. Audet, G. Couture, J.E. Dennis, Jr., and S. Le Digabel. The NOMAD project. Software available at http://www.gerad.ca/nomad.

[2] M.A. Abramson, C. Audet, J.E. Dennis, Jr., and S. Le Digabel. OrthoMADS: A deterministic MADS instance with orthogonal directions. *SIAM Journal on Optimization*, 20(2):948–966, 2009.

[3] C. Audet. A short proof on the cardinality of maximal positive bases. *Optimization Letters*, 5(1):191–194, 2011.

[4] C. Audet, V. Béchard, and J. Chaouki. Spent potliner treatment process optimization using a MADS algorithm. *Optimization and Engineering*, 9(2):143–160, 2007.

[5] C. Audet, G. Couture, J.E. Dennis, Jr., M.A. Abramson, F. Gonzalez, V. Titov, and M. Spillane. Optimal placement of tsunami warning buoys using mesh adaptive direct searches. In *Optimization days*, Montreal, May 2005. Slides available at http://www.gerad.ca/Charles.Audet/Slides/Tsunami_JOPT.pdf.

[6] C. Audet and J.E. Dennis, Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006.

[7] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. Parallel space decomposition of the mesh adaptive direct search algorithm. *SIAM Journal on Optimization*, 19(3):1150–1170, 2008.

[8] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. Globalization strategies for mesh adaptive direct search. *Computational Optimization and Applications*, 46(2):193–215, 2010.

[9] C. Audet and D. Orban. Finding optimal algorithmic parameters using derivative-free optimization. *SIAM Journal on Optimization*, 17(3):642–664, 2006.

[10] A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.B. Serafini, V. Torczon, and M.W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization*, 17(1):1–13, 1999.

[11] J. Carrera, E. Usunoff, and F. Szidarovszky. A method for optimal observation network design for groundwater management. *Journal of Hydrology*, 73(1–2):147–163, 1984.

[12] S.S. Carroll. Estimating the uncertainty in spatial estimates of areal snow water equivalent. *Nordic Hydrology*, 27(5):295–312, 1995.

[13] Y.-C. Chen, C. Wei, and H.-C. Yeh. Rainfall network design using kriging and entropy. *Hydrological Processes*, 22(3):340–346, 2008.

[14] Y. Choquette, P. Lavigne, P. Ducharme, A. Houdayer, and J.-P. Martin. Apparatus and Method for Monitoring Snow Water Equivalent and Soil Moisture Content Using Natural Gamma Radiation, September 2010. US Patent No. 7800051 B2.

[15] F.H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley, New York, 1983. Reissued in 1990 by SIAM Publications, Philadelphia, as Vol. 5 in the series Classics in Applied Mathematics.

[16] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to Derivative-Free Optimization*. MPS/SIAM Book Series on Optimization. SIAM, Philadelphia, 2009.

[17] M.K. Cowles, D.L. Zimmerman, A. Christ, and D.L. McGinnis. Combining snow water equivalent data from multiple sources to estimate spatio-temporal trends and compare measurement systems. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(4):536–557, 2002.

[18] W.C. Davidon. Variable metric method for minimization. *SIAM Journal on Optimization*, 1(1):1–17, 1991.

[19] C. Davis. Theory of positive linear dependence. *American Journal of Mathematics*, 76:733–746, 1954.

[20] A. Dhar and B. Datta. Global optimal design of ground water monitoring network using embedded kriging. *Ground Water*, 47(6):806–815, 2009.

[21] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

[22] N.D. Évora, D. Tapsoba, and D. De Sève. Combining artificial neural network models, geostatistics, and passive microwave data for snow water equivalent retrieval and mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):1925–1939, 2008.

[23] K.R. Fowler, J.P. Reese, C.E. Kees, J.E. Dennis Jr., C.T. Kelley, C.T. Miller, C. Audet, A.J. Booker, G. Couture, R.W. Darwin, M.W. Farthing, D.E. Finkel, J.M. Gablonsky, G. Gray, and T.G. Kolda. Comparison of derivative-free optimization methods for groundwater supply and hydraulic capture community problems. *Advances in Water Resources*, 31(5):743–757, 2008.

[24] V. Garnier. La gestion des groupes de variables en recherche directe. Master's thesis, École Polytechnique de Montréal, June 2010.

[25] Geovariances. Isatis version 10.03. http://www.geovariances.com.

[26] B.J. Harshburger, K.S. Humes, V.P. Walden, T.R. Blandford, B.C. Moore, and R.J. Dezzani. Spatial interpolation of snow water equivalency using surface observations and remotely sensed images of snow-covered area. *Hydrological Processes*, 24(10):1285–1295, 2010.

[27] J.A. Hartigan and M.A. Wong. A K-means clustering algorithm. *Applied Statistics*, 28:100–108, 1979.

[28] H.J.W.M. Hendricks Franssen and F. Stauffer. Optimal design of measurement networks for groundwater flow predictions. *Advances in Water Resources*, 28(5):451–465, 2005.

[29] Hydro-Québec. Shaping the future, annual report 2009. Available at http://www.hydroquebec.com/publications/en/annual_report.

[30] R.L. Iman and W.J. Conover. The use of the rank transformation in regression. *Technometrics*, 21(4):499–509, 1979.

[31] J. Jacques. *Contributions à l'analyse de sensibilité et à l'analyse discriminante généralisée*. PhD thesis, Université Joseph Fourier - Grenoble I, December 2005.

[32] N. Jimênez, F.M. Toro, J.I. Vélez, and N. Aguirre. A methodology for the design of quasi-optimal monitoring networks for lakes and reservoirs. *Journal of Hydroinformatics*, 7(2):105–116, 2005.

[33] L.A. Sweatlock K. Diest and D.E. Marthaler. Metamaterials design using gradient-free numerical optimization. *Journal of Applied Physics*, 108(8):1–5, 2010.

[34] A.H.M. Kassim and N.T. Kottegoda. Rainfall network design through comparative kriging methods / La planification des réseaux de pluviomètres par les méthodes comparatives de krigeage. *Hydrological Sciences Journal*, 36(3):223–240, 1991.

[35] S. Le Digabel. Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm, 2010. To appear in ACM Transactions on Mathematical Software.

[36] A.L. Marsden, M. Wang, J.E. Dennis, Jr., and P. Moin. Optimal aeroacoustic shape design using the surrogate management framework. *Optimization and Engineering*, 5(2):235–262, 2004.

[37] J.-P. Martin, A. Houdayer, C. Lebel, Y. Choquette, P. Lavigne, and P. Ducharme. An unattended gamma monitor for the determination of snow water equivalent (SWE) using the natural ground gamma radiation. In *Nuclear Science Symposium Conference Record*, pages 983–988. IEEE, 2008.

[38] L.M. Nunes, M. Conceição Cunha, and L. Ribeiro. Monitoring network optimisation using both spatial and temporal information. In F. Khosrowshahi, editor, *Proceedings of the 3rd International Conference on Decision Making in Urban & Civil Engineering*, London, November 2002.

[39] E. Pardo-Igúzquiza. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology*, 210(1–4):206–220, 1998.

[40] S. Sankaran, C. Audet, and A.L. Marsden. A method for stochastic constrained optimization using derivative-free surrogate pattern search and collocation. *Journal of Computational Physics*, 229(12):4664–4682, 2010.

[41] U.M. Shamsi, R.G. Quimpo, and G.N. Yoganarasimhan. An application of kriging to rainfall network design. *Nordic Hydrology*, 19(3):137–152, 1988.

[42] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. http://www.R-project.org.

[43] V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1):1–25, 1997.

[44] F.-G. Yang, S.-Y. Cao, X.-N. Liu, and K.-J. Yang. Design of groundwater level monitoring network with ordinary kriging. *Journal of Hydrodynamics*, 20(3):339–346, 2008.

[45] D.L. Zimmerman. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 17(6):635–652, 2006.