Experimental Design for Cloud Seeding

Roberto Benedetti

University of Chieti-Pescara, Italy

Federica Piersimoni

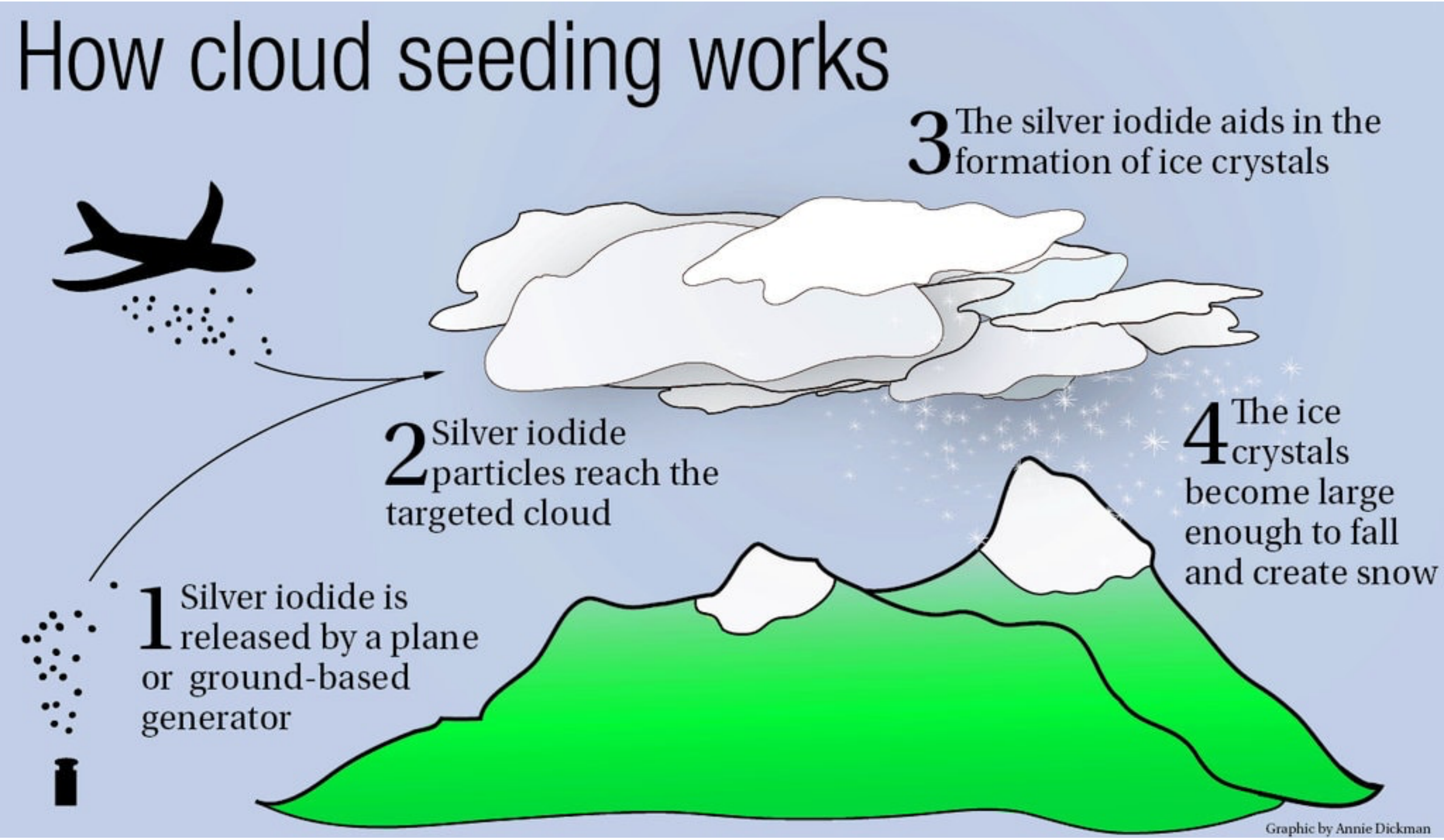
ISTAT, Italian National Statistical Institute, Rome, Italy

RL Chambers

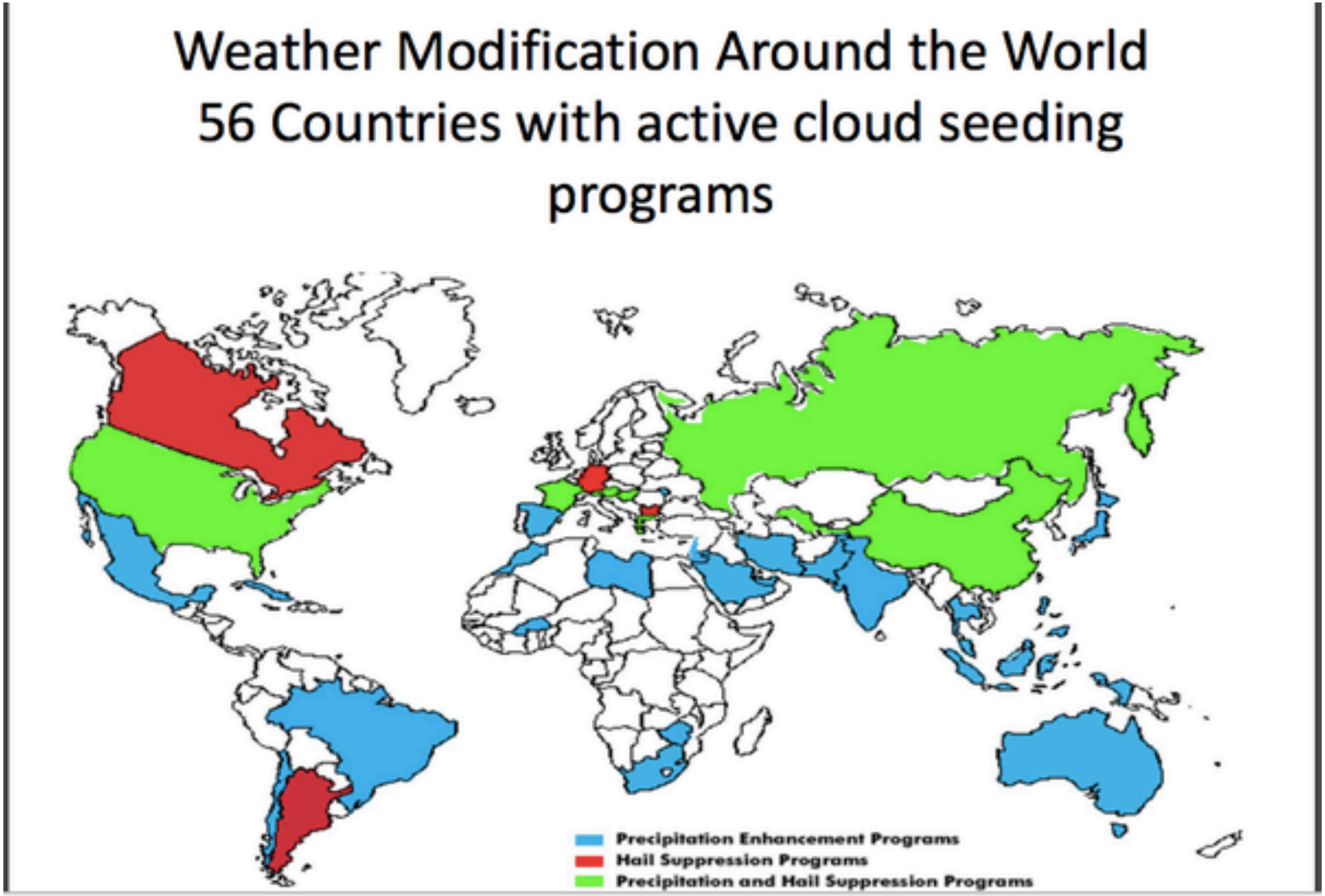
University of Wollongong, Australia

RAIN ENHANCEMENT: traditional solution

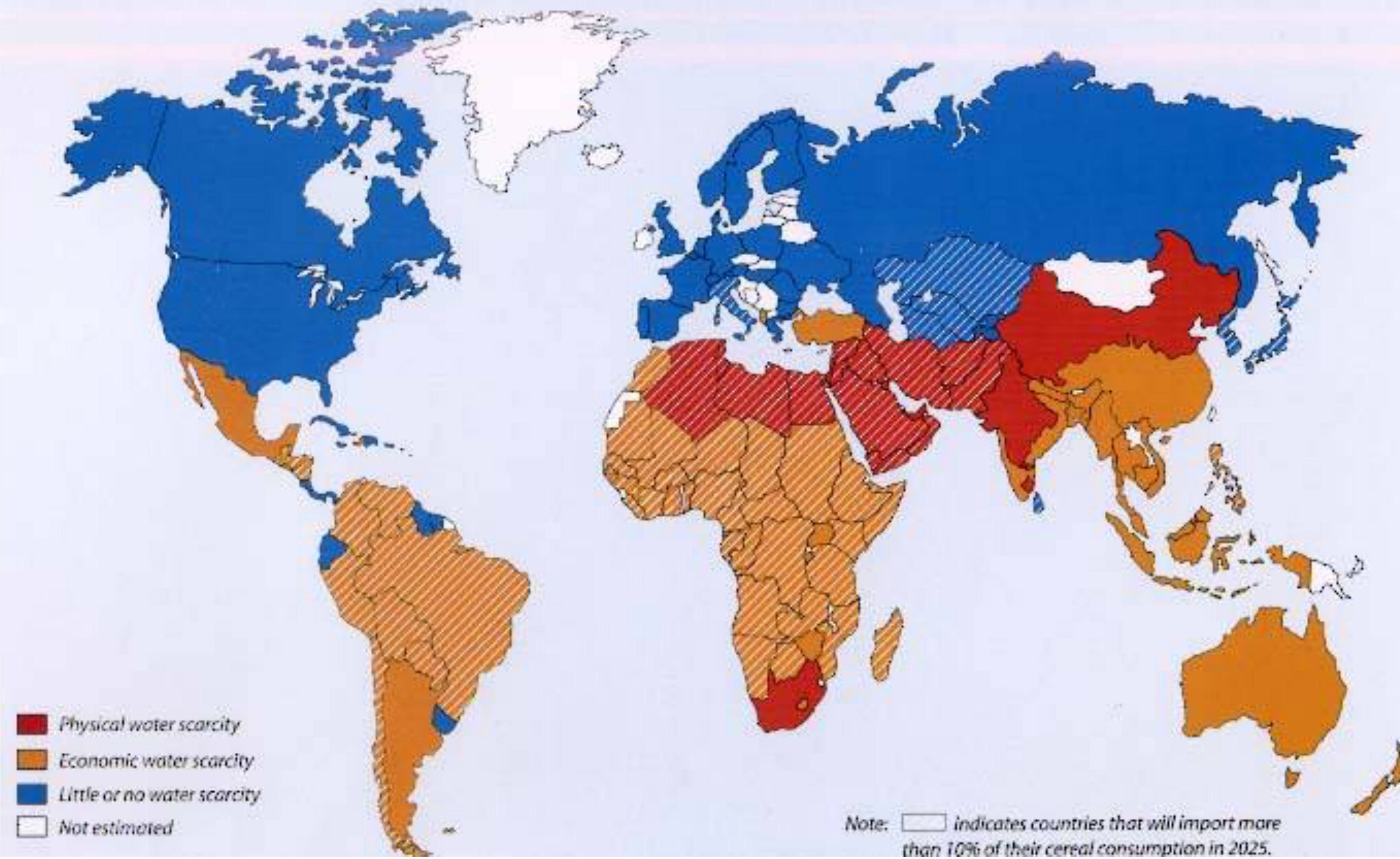
RAIN ENHANCEMENT: revised solution



RAIN ENHANCEMENT: revised solution



Projected Water Scarcity in 2025



Outline

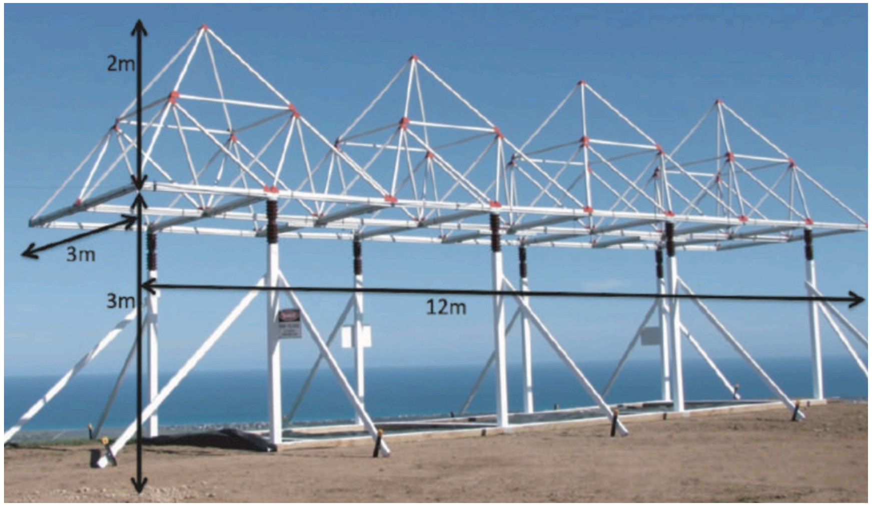
1. The problem: Definition of the study region
2. The data: Setting up the Frame
3. Optimal Selection of New Atlants Positions
4. Random Selection of New Gauge Positions and Gauges to Move
    - Strategies to Coordinate Spatially Balanced Samples
5. Random Selection of the Switching On and Off Daily Sequence of the Atlants
6. Repositioning the Atlants through Response Surface Optimization
7. Concluding Remarks

Main practical problems

3. algorithms that allow us to position any new Atlant in an optimal way;
4. a design to select randomly, but with very clear and binding features, any new Gauge or displacement of existing Gauges;
5. a criterion for the random selection mode of the switching on and off sequence of the Atlants in order to suggest an alternative that should guarantee the randomness of the selection while respecting many constraints such as the non-contiguity in space and time of the switched on or off Atlants and that the frequency of switching on the different Atlants in a given period is the same for all the Atlants.
6. move from an initial phase of the design of the experiment to optimal positioning (operational phase)

Main Tools used

Atlant is a rainfall enhancement system that is low cost and appears to be capable of increasing rainfall. There have been only a limited number of trials of the Atlant system, and there is considerable uncertainty as to how effective the system is, especially in different climatic and geographic systems.



Gauges are simple weather stations to measure the amount of daily rainfall

Sultanate of Oman

Sultanate of Oman

Definition of the study region

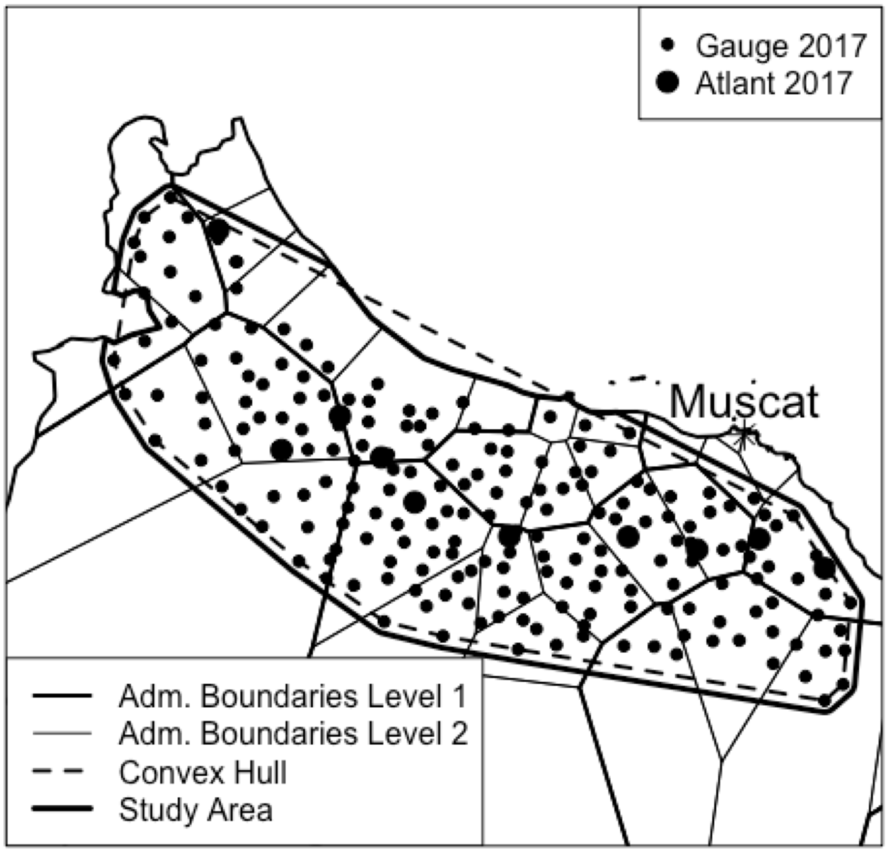
The first problem to be solved is the definition itself of the study region in a geographically unambiguous way, this does not mean that from year to year its form and extension can not be changed but only that it is important to define for which area the results are valid and in which area you can select new instrument locations.

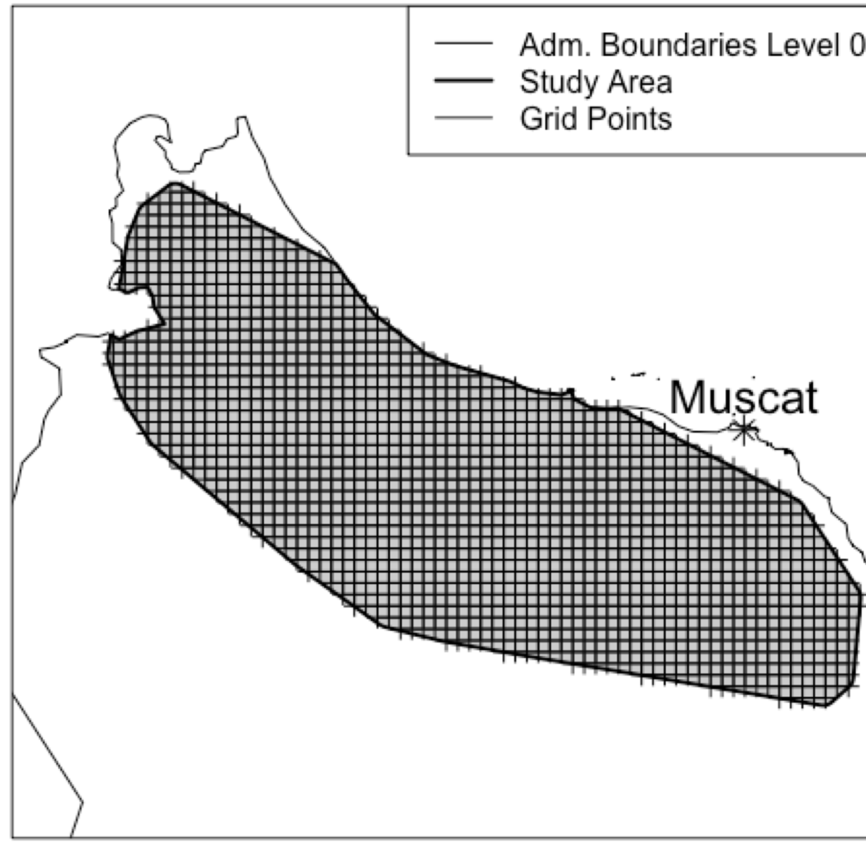
To this purpose, the minimum convex polygon (Convex Hull) was calculated, containing the position of the Gauge and Atlants of 2017, to which a buffer zone of 0.05 degrees of amplitude was added. To avoid that a part of the area of interest fell outside of Oman or even at sea, the obtained polygon was then intersected with the administrative limits of Oman downloaded from GADM, a free spatial database of the location of the world's administrative boundaries for use in GIS (http://www.gadm.org).

Definition of the study region

Finally a grid has been generated starting in the point with coordinates {55.7, 22.4} and ending in the point with coordinates {59.4,24.7} with a resolution 0.05 degrees in Latitude and Longitude and then intersected with the polygon representing the study region retaining only the points that fall inside it. In this way a population of size 1450 of possible candidate positions both for Atlants and Gauges has been obtained.

Definition of the study region





Optimal Selection of New Atlants Positions

A spatial finite population $U = \{1,2,\ldots,i,\ldots,N\}$ of size N=1450 is recorded on a frame together with a set of k auxiliary variables $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_k\}$ and a set of d coordinates $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2,\ldots, \mathbf{c}_d\}$ obtained by the geo-coding of each unit. Obviously the case d = 2, even if it does not represent a constraint, covers most of the applications of sampling in the environmental and meteorological field. From C we can always derive, according to any distance definition, a matrix $\mathbf{D}_U = \{d_{ij}; i=1,\ldots,N, j=1,\ldots,N\}$ which specifies how far are all the pairs of units in the population. To use some covariates we always assume that there is, at least approximate, a certain degree of dependence between the *survey* variable and the set **X** even if not specified in detail.

Optimal Selection of New Atlants Positions

With regard to the use of the set **C**, the widely used distance matrix as a synthesis of the spatial information emphasizes the importance of the spread of the sample over the study region as a feature which can be related, but not necessarily, to this dependence but also to some form of similarity between adjacent units.

The use of the matrix $\mathbf{D}_U$ as a synthesis of the spatial information implies the hypothesis that the dependence does not change with the position of the unit i and the direction, i.e. that the random field $Y(i)$ is homogeneous and isotropic, i.e. its distribution does not change if we shift or rotate the space of the coordinates (Cressie, 1993).

Optimal Selection of New Atlants Positions

$\mathbf{D}_U$ is a very important tool to emphasize the importance to spread the sample over U, a property that can be related, through a variogram (Cressie, 1993), to the spatial dependence of $Y(i)$ but also to some form of similarity between adjacent units as a spatial clustering or a spatial stratification. If the units are points, as in our case, for the definition of $\mathbf{D}_U$, we can simply resort to simple concepts of distance between sets of coordinates.

Setting up the Frame

While for the Gauges the *Y*(*i*) can be undoubtedly defined as the daily rainfall, if we focus on the Atlants, the *Y*(*i*) is more difficult to be defined as it concerns the objective itself of the experiment: the effects of ionisation on the microphysical processes of precipitation formation which is undoubtedly a latent variable whose values are impossible to be collected with a ground survey.

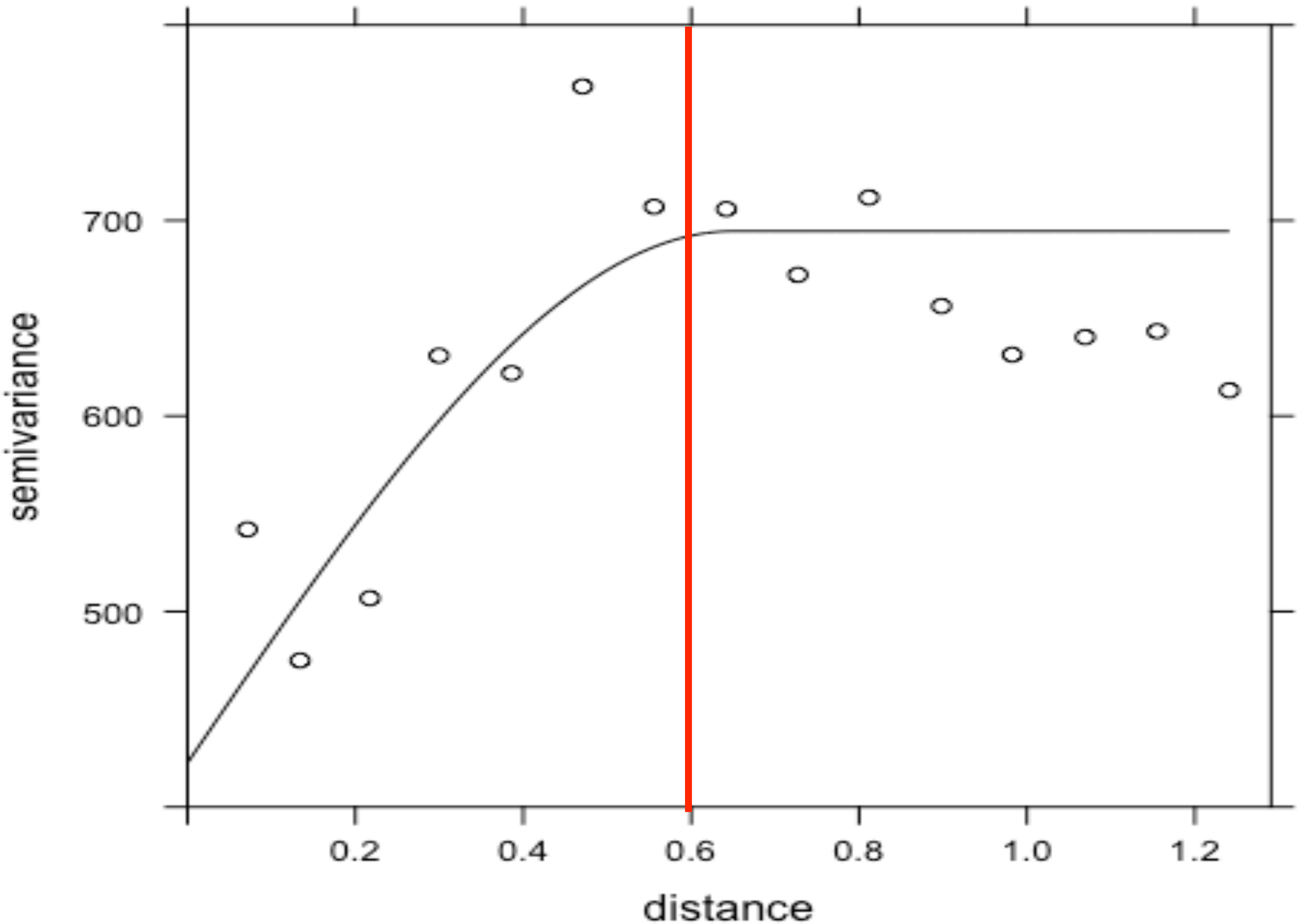
From these considerations it follows that the total amount of rain detected in previous years can be considered a good auxiliary variable for the positioning of the Gauges while for the Atlants we must rely on some rules suggested by common sense and knowledge of the physics of the problem.

Setting up the Frame

With "impacting significantly on precipitation" we refer in particular to the effect that the land-orography has on the choice of an efficient positioning of new Atlants. The elevation of each point is therefore an extremely important parameter that can be used both as it is and through indices of terrain characteristics derived from it as slope and aspect that measure respectively the roughness of the territory surrounding the point and its angle of exposure.

Annual average rainfall has been predicted on the population U by using a kriging model (Cressie, 1993) with a spherical semivariogram from the annual average rainfall collected in 2017 in each Gauge.

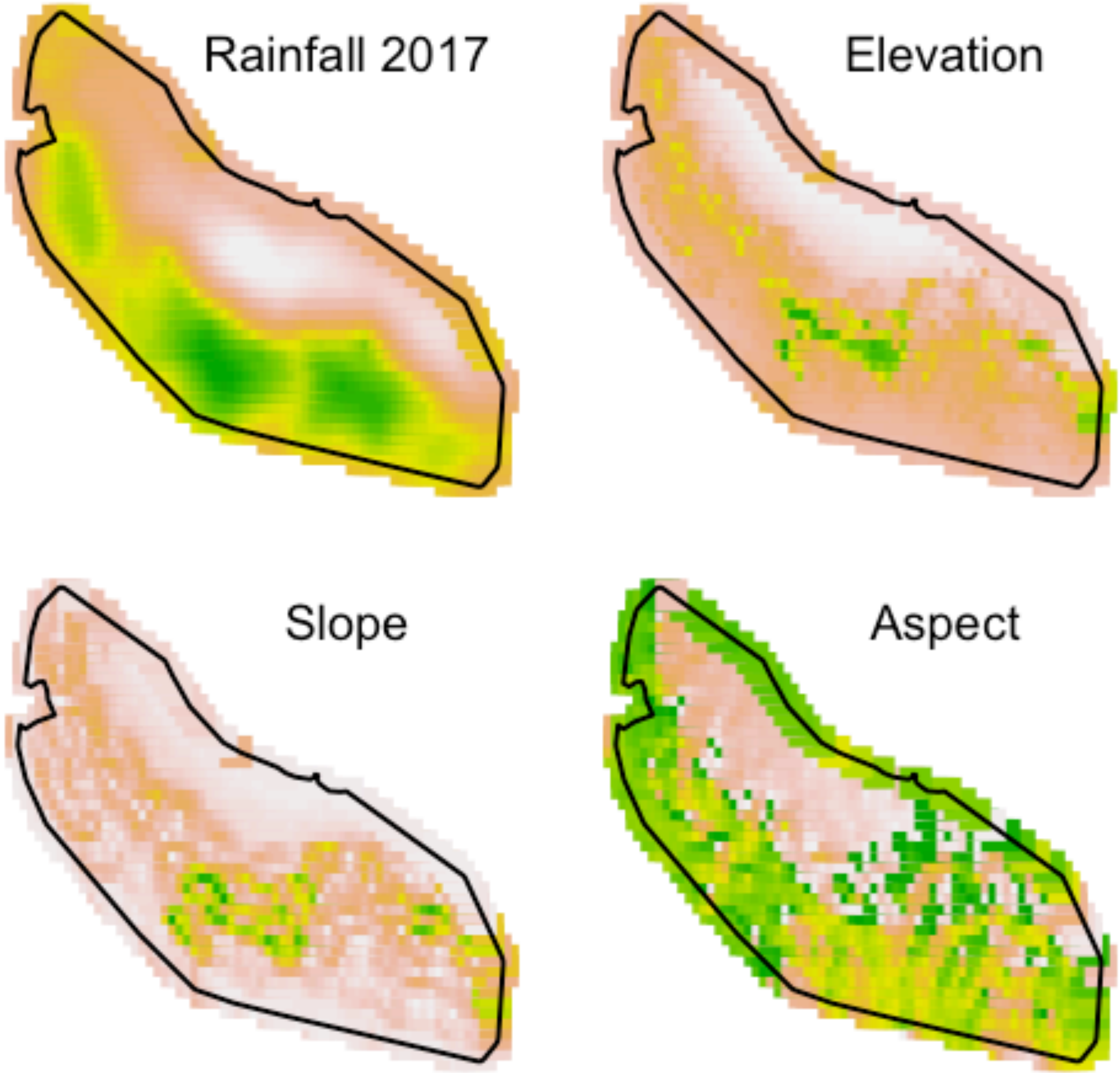
Setting up the Frame: rainfall variogram



Setting up the Frame

The elevations for each point of the grid *U* have been downloaded from Mapzen (https://mapzen.com) that unfortunately has shut down its services in January 2018. However there are several Digital Elevation Model (DEM) data sources, with a very high detailed resolution (up to 30 meters), that are freely available and that can be used in the future for our purposes, such as: http://www.webgis.com/terraindata.html.

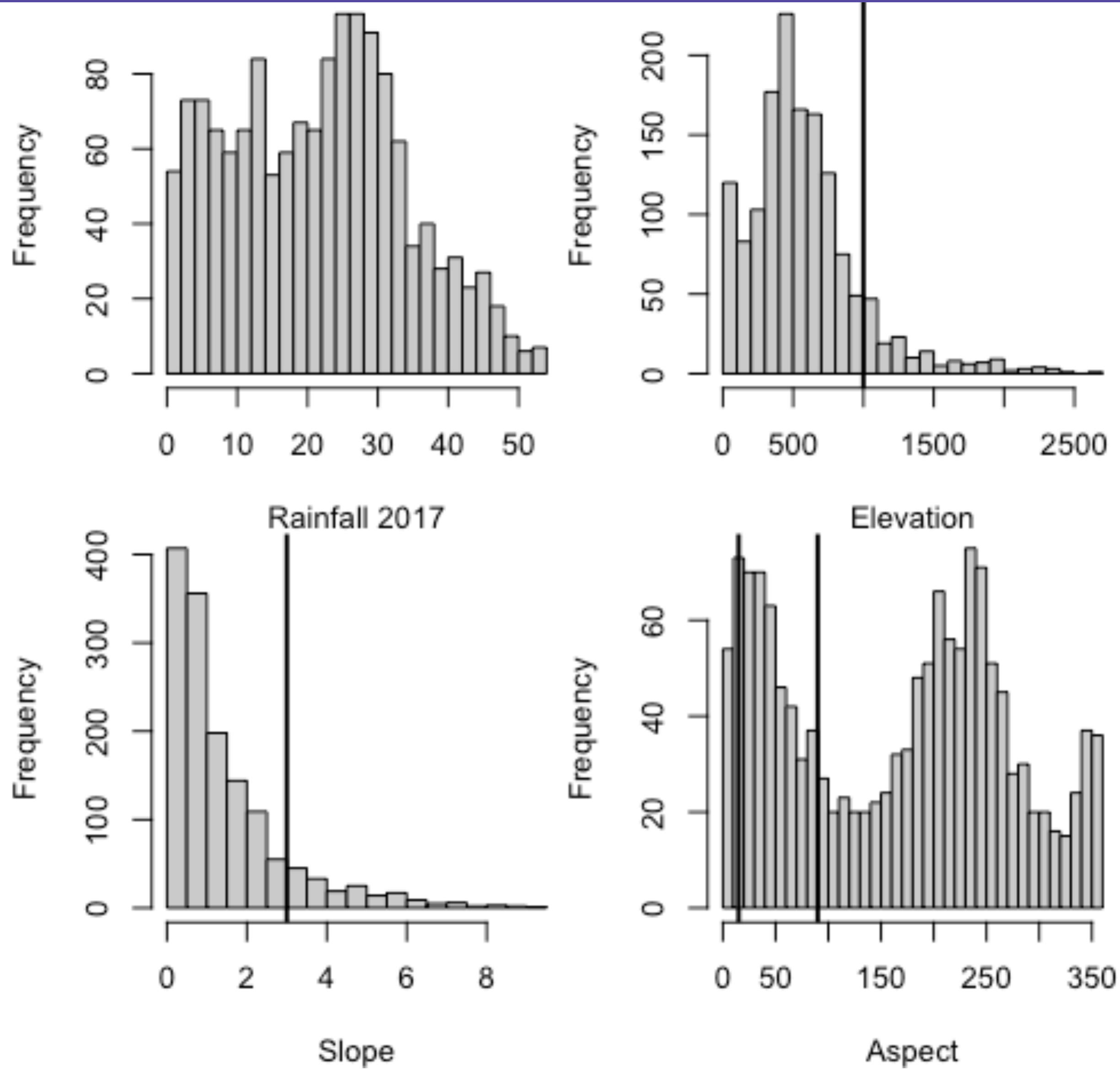
Setting up the Frame



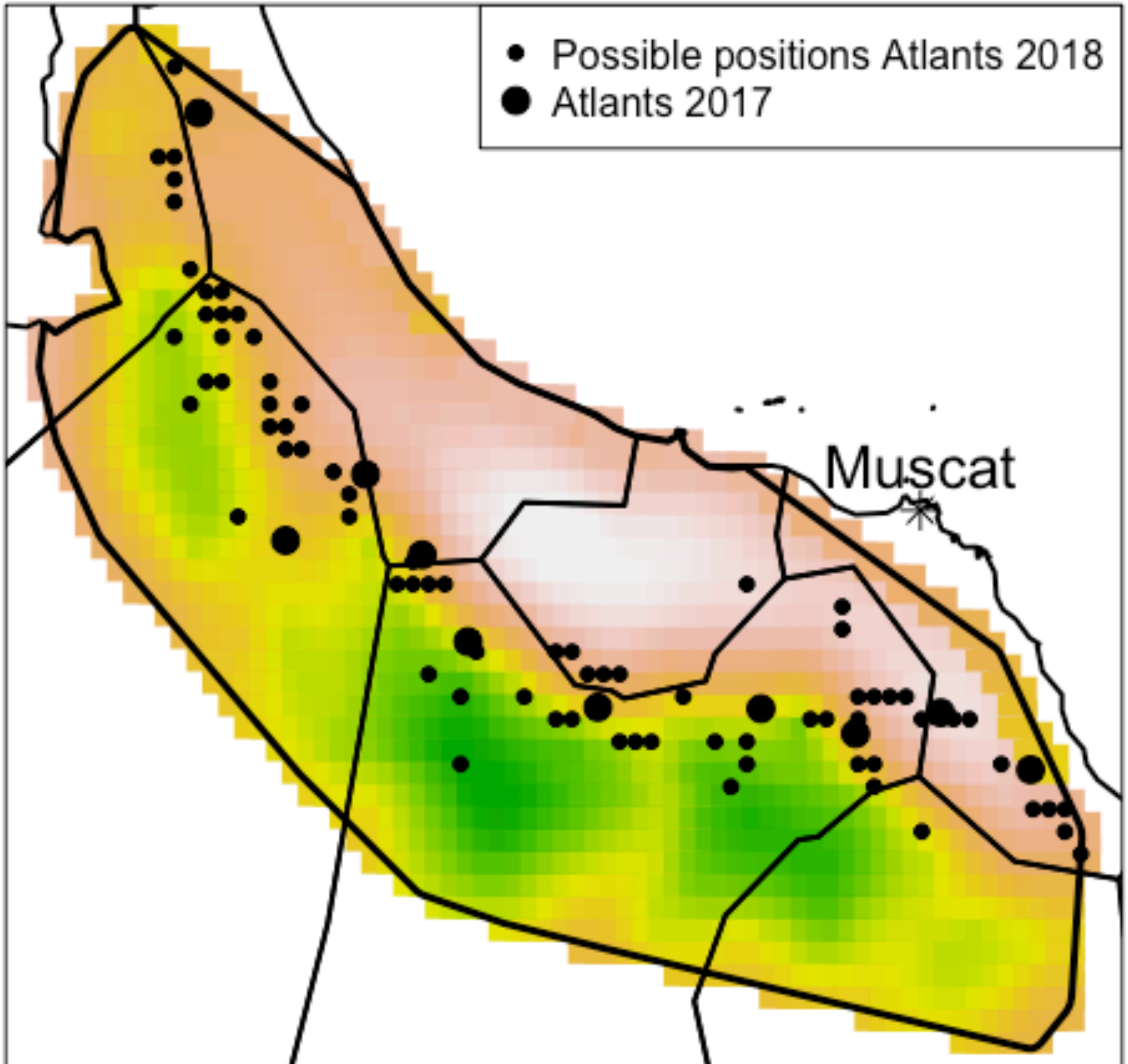
Setting up the Frame

On the basis of this ancillary information, an U_P subpopulation of possible positions for the new Atlants can be selected from U. In order to be more effective, for the effect of ionization, it is very important that the Atlants are positioned at elevated heights and with a good exposure towards the North-East so that they can be reached by the wind coming from the sea. Furthermore, for purely technical reasons, an Atlant can not be placed in areas that are too steep. It follows that some thresholds have been imposed on the variables deriving from the DEM. In particular we considered as valid points for the positioning of the new Atlants only those points of U that: have an elevation of more than 1000 meters, an aspect smaller than 3 and an angle of exposure between 15 and 90 degrees. These restrictions imposed to U resulted in a list of 77 possible sites.

Setting up the Frame: thresholding



Setting up the Frame: Atlants possible positions



Optimal Sampling Design

Considering only without replacement samples, a convenient way to state the sample selection process is to assume that, for each unit *i* of the frame, a vector $S=\{s_i;\ i=1,\ldots,N\}$ is generated that is equal to 1 if a unit is selected in the sample and 0 otherwise.

To select *S* the logic of the optimal design is conceived so that preferential sampling (Diggle and Menezes, 2010) can be used allowing explicitly for the minimization of a criterion *F* linked with some summary statistics *Q*, usually an utility function, arising from model x. More formally the set of *n* units that constitute the sample *S* is the result of an optimization problem such as

$$\max_S [Q(\xi)]$$

(Rogerson and Delmelle, 2004; Müller and Stehlìk, 2010). This is a combinatorial optimization problem that can be solved through some heuristics or by using the well known *Simulated Annealing* (SA) algorithm that has shown to provide promising results (Benedetti and Palma, 1995).

Simulated Annealing

The basic idea behind this algorithm was originally introduced in the statistical mechanics by Metropolis *et al*. (1953). Kirkpatrick *et al*. (1983) and Cerny (1985) incorporated the Metropolis scheme in a procedure analogous to chemical annealing to define a search algorithm for solving combinatorial optimization problems as the travelling salesman problem. The convergence behavior of SA algorithm to the global optimum was analyzed extensively (van Laarhoven and Arts, 1987). For further details, see also Geman and Geman (1984), where is shown that a necessary and sufficient condition to reach a global optimum is that the temperature decreases logarithmically with time. In the approach of Geman *et al*. (1990), a spatial combinatorial optimization problem can be viewed as a *Markov Random Field* (MRF). The probability measure of a MRF adopting the Gibbs representation is:

Simulated Annealing

$$P(\mathbf{D}_U, \mathbf{C}, S) = \frac{e^{-\frac{Q(\mathbf{D}_U, \mathbf{C}, s)}{T}}}{\sum_{s \in S} e^{-\frac{Q(\mathbf{D}_U, \mathbf{C}, s)}{T}}}$$

where T is a control parameter called temperature. When we are dealing with constrained optimization problems, the function $Q(\mathbf{D}_U, \mathbf{C}, s)$ can be expressed as the sum of two terms: an interaction term $M(\mathbf{D}_U, \mathbf{C}, s)$ which depends on the known distance matrix and the labels of the sample, and a penalty term V(s), that depends only on s. At the *k-th* iteration, the energy function is defined as:

$$Q_k(\mathbf{D}_U, \mathbf{C}, s) = -(M_k(\mathbf{D}_U, \mathbf{C}, s) - \lambda\, V_k(s)),$$

where λ is a tuning parameter. The interaction term is given by:

Simulated Annealing

$$M_k(\mathbf{D}_U, \mathbf{C}, s) =$$
$$(\alpha)\left(\sum_{i\in S_A}\sum_{j\in S_E}\bar{d}_{i,j} + \frac{\sum_{i\in S_A}\sum_{j\in S_A}\bar{d}_{i,j}}{2}\right) + (1-\alpha)\left(\frac{\sum_{i\in S_A}\mathbf{c}_{i,DEM}-1000}{1000}\right),$$

where α is the weight of the convex linear combination of the two components that we want to maximize, the distance between Atlants and their average elevations, S_A and S_E are respectively the new and the existing Atlant positions and $\mathbf{c}_{DEM}$ is the auxiliary of the elevations for each point of the acceptable positions of a new Atlant. Notice that $\bar{d}_{i,j}$ is a truncated version of the distance as its value after a given threshold, i.e. it is equal to:

$$\bar{d}_{ij} = \begin{cases} d_{ij} & \text{if } d_{ij} < d_{th} \\ d_{th} & \text{otherwise} \end{cases}.$$

Simulated Annealing

The adjacency constraint are formalized through the penalty function:

$$V_k(s) = \left(\sum_{i \in S_A} \sum_{j \in S_E} I_{d_{ij} < d_{min}} + \frac{\sum_{i \in S_A} \sum_{i \neq j, j \in S_A} I_{d_{ij} < d_{min}}}{2}\right),$$

where d_{min} is a threshold value representing the minimum distance allowed for the Atlants and $I_{d_{ij} < d_{min}}$ is an indicator function that is equal to 1 if the inequality is true and 0 otherwise. This model discourages configurations with low distances and low average altitude avoiding that even one Atlant has a distance lower than d_{min} to any other Atlants.

Simulated Annealing

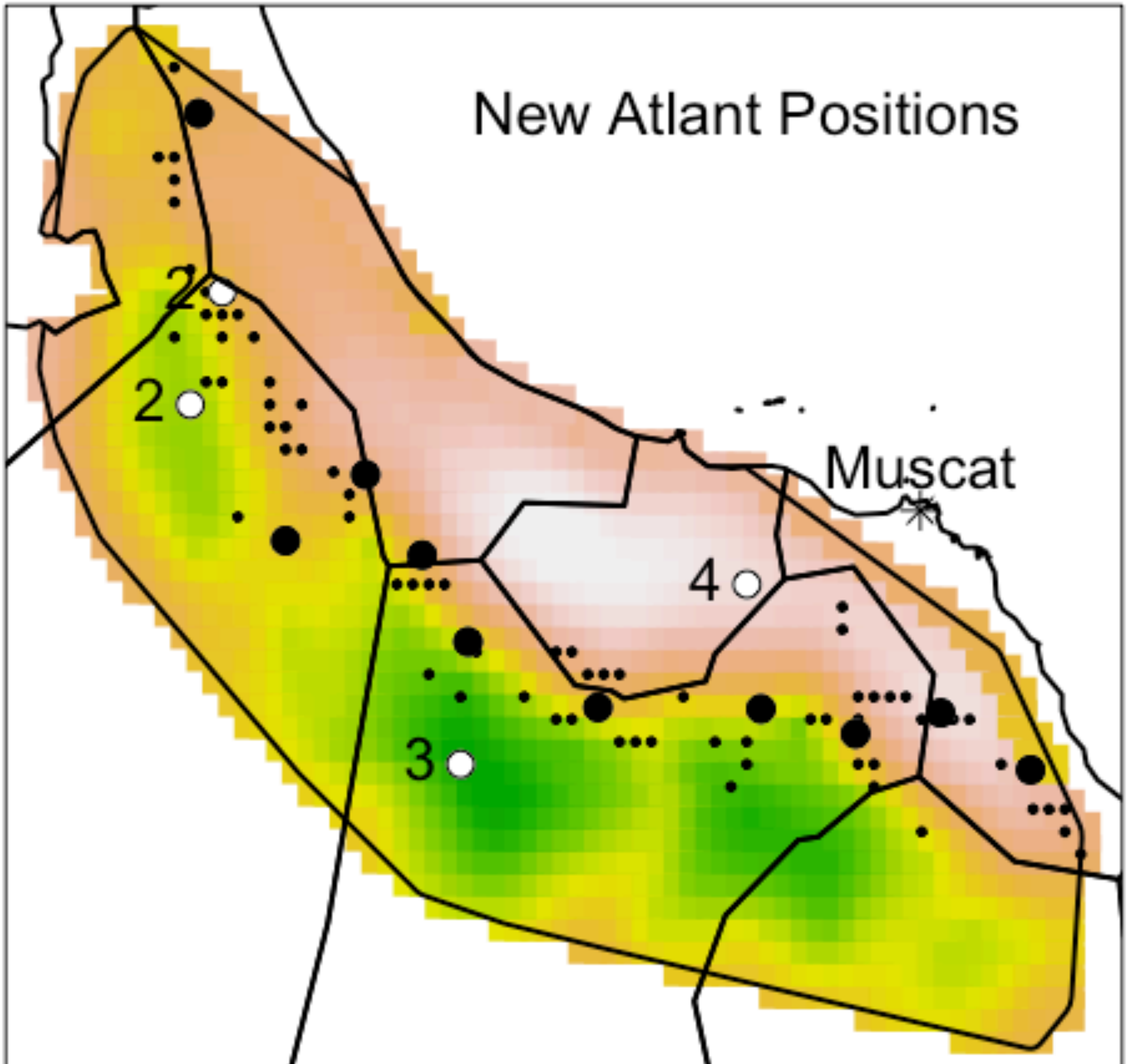
Given a starting configuration s_0 selected through simple random sampling without replacement, the algorithm replaces the solution obtained at the *j-th* iteration s_j with a new solution s_{j+1} chosen at random according to the visiting schedule with probability:

$$p = \begin{cases} 1 & \text{if } Q(s_{j+1}) < Q(s_j) \\ e^{\left(-\frac{Q(s_{j+1}) - Q(s_j)}{T}\right)} & \text{otherwise} \end{cases}$$

This means that a move from a configuration s_j to a worse configuration s_{j+1} is allowed with a probability which depends on the value of the parameter T_j, where T_j indicates the temperature at the *j-th* step of the procedure.

The tuning parameters used are $\alpha = 0.75$, $\lambda=10$, $d_{min}=0.25$ and $d_{th}=0.60$.

Positions of new Atlants



Random Selection of New Gauges Positions and Gauges to Move

Regarding the random positioning of the Gauges we must always bear in mind that geographically distributed observations present particularities that should be appropriately considered when designing an experiment. Traditional sampling designs may be inappropriate when investigating geocoded data, because they might not capture the spatial information present in the units to be sampled. This spatial effect represents valuable information that can lead to considerable improvement in the efficiency of estimates. For these reasons, during the last decades, many contributions have been introduced in the literature (Grafström 2012, Grafström and Schelin 2014, Benedetti et al. 2017, Benedetti and Piersimoni, 2017). Spatially balanced samples have the property to be well-spread over the spatial population of interest.

Strategies to Coordinate Spatially Balanced Samples

To deal with sample selection in space, time and between different populations we must extend the notations used in Section 2.

Let $U_{v,t}$={1,2,…, $i_{v,t}$,…, $N_{v,t}$} be a finite population of a set *v*={*1,2,…,V*} of statistical units, recorded in time *t*={*1,2,…,M*} on a spatial frame together with a set of *k* auxiliary variables $\mathbf{X}_{v,t}=\{\mathbf{x}_{v,t,1},\mathbf{x}_{v,t,2},\ldots,\mathbf{x}_{v,t,j},\ldots,\mathbf{x}_{v,t,k}\}$ and a set of *h* coordinates $\mathbf{C}_{v,t}=\{\mathbf{c}_{v,t,1},\mathbf{c}_{v,t,2},\ldots,\mathbf{c}_{v,t,j},\ldots,\mathbf{c}_{v,t,h}\}$ obtained by the geo-coding of each set *v* of units. From **C** we can always derive a matrix $\mathbf{D}_{Uv,t}=\{d_{ij,vt};\ i=1,\ldots,N_{v,t},\ j=1,\ldots,N_{v,t}\}$. We restrict our attention to populations that do not change over time, i.e. the population size is always $N_{v,t1} = N_{v,t2}$ for each *v*, and the positions of the spatial units are fixed $\mathbf{C}_{v,t1}= \mathbf{C}_{v,t2}$ for each *v*.

Strategies to Coordinate Spatially Balanced Samples

In selecting the sample of the new Gauges, it was preferred to use the stratified version of the LPM algorithm, essentially because it allows for a greater control of the first order inclusion probabilities that are assumed to be proportional to the interpolated total rainfall observed in 2017.

Formally a population consisting of three strata is considered: Atlants of size N_A, Gauges 2017 of size N_G and the regular grid of the new possible positions of size N. The first stratum is censused, i.e. the sample size will be $n_A=N_A$, from the second $n_G= N_G-n_R$ units will picked up, where n_R is the number of units that we want to move and, finally, $n_B= n_R+n_N$ will be selected from the third stratum, where n_N is the number of new Gauges to be positioned.

Strategies to Coordinate Spatially Balanced Samples

In the next Table the parameters used to randomly select the sample of Gauges for 2018 are reported. Notice that two tuning parameters have been added to the definition of the distance matrix between the three strata: β_1 and β_2. The first one is used to handle the spreading of the sample between selected Gauges while the second to drive the selected sample to be closer to existing Atlants.

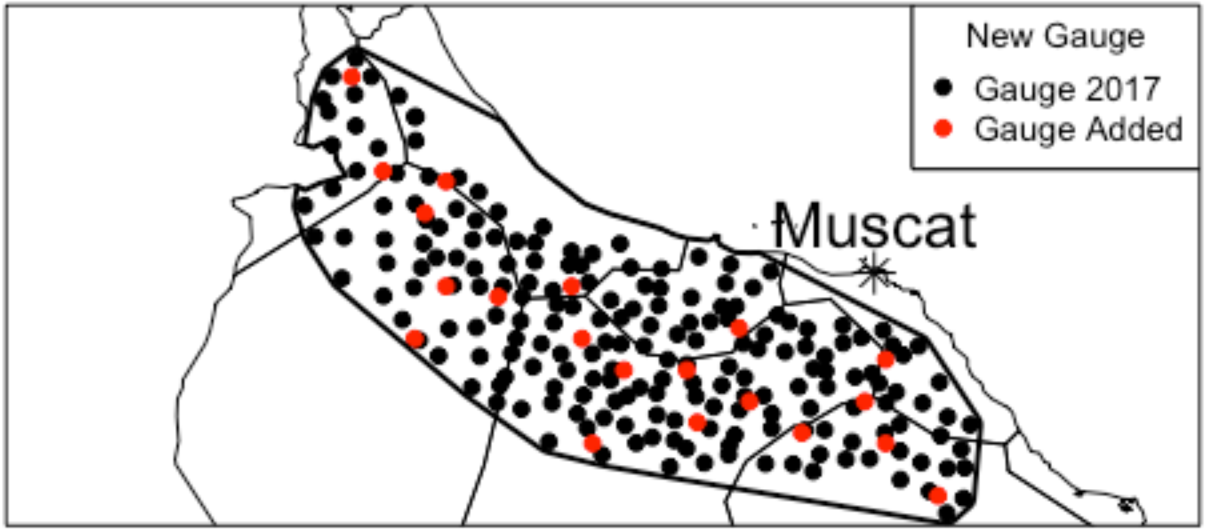
Table 1: Parameters used to Coordinate the Spatially Balanced Random Positions for Gauges

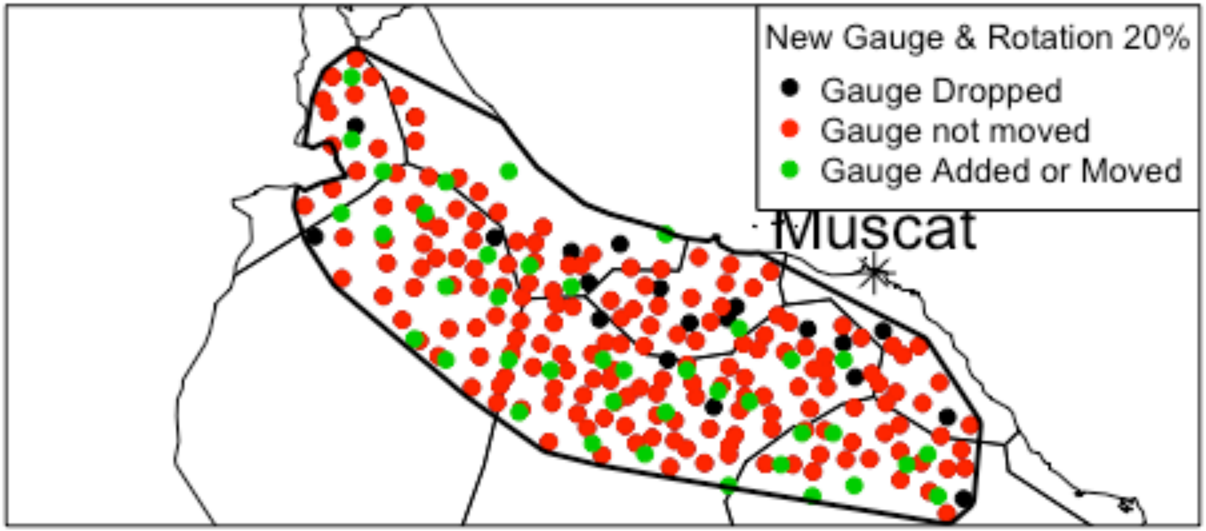
Definition	Strata Code	Pop. Size	Sample Size	π_i	$\mathbf{D}_U$
Atlants	1	N_A	N_A	1	$(d_{ij})^{\beta_1}$ $(d_{ij})^{\beta_2}$ $(d_{ij})^{\beta_2}$
Gauges 2017	2	N_G	$n_G = N_G - n_R$	$\frac{\mathbf{c}_{i,P}}{\sum_{i \in G} \mathbf{c}_{i,P}} n_G$	$(d_{ij})^{\beta_2}$ $(d_{ij})^{\beta_1}$ $(d_{ij})^{\beta_1}$
New Positions	3	N	$n_B = n_R + n_N$	$\frac{\mathbf{c}_{i,P}}{\sum_{i \in U} \mathbf{c}_{i,P}} n_B$	$(d_{ij})^{\beta_2}$ $(d_{ij})^{\beta_1}$ $(d_{ij})^{\beta_1}$

Strategies to Coordinate Spatially Balanced Samples

The proposal presented is to select new Gauges based on two hypotheses, the first to add 20 new sites and the second to add the same number of sites and to move 10% of existing sites. The parameters β_1 and β_2 were respectively set equal to 5 and -15.

Strategies to Coordinate Spatially Balanced Samples





Random Selection of the Switching On and Off Daily Sequence of the Atlants

The 2018 trial will employed a balanced randomized design constrained as much as possible to avoid configurations that have a too high contiguity pattern in space and time. Assuming that there will be N_A=12 operational Atlants, and that a 30km corridor footprint model used for the analysis essentially isolates the Atlants from one another, there are no concrete reasons to 'pair' Atlants (operational versus control), particularly since there does not seem to be any reasonable way of spatially matching the Atlants locations.

The design aim is to have n_A=6 (equal to N_A/2, this is not a constraint for the algorithm but only a practical example) Atlants operational every day (ensuring within day effects over the entire trial area were estimated with maximum precision), implying 924 different combinations of operating states on a day.

Random Selection of the Switching On and Off Daily Sequence of the Atlants

A nominal trial period of *ND* days was therefore used (in the example *ND*=200). This period can also be further segmented into equal-sized blocks, with these 924 combinations randomly allocated to days within each block or within the whole period. Each Atlant should be on for $ND(n_A/N_A)$ days and off for $ND(1-n_A/N_A)$ days. And within each day, n_A Atlants are on and (N_A-n_A) are off.

Assuming that ordering the Atlants as 1 to 12 reflects spatial ordering, then no more than two contiguous Atlants can be operated on a day. In general we can introduce a contiguity matrix **W** whose generic value is 1 if the two Atlants are contiguous and 0 otherwise.

Random Selection of the Switching On and Off Daily Sequence of the Atlants

Given these constraints, the basic approach is to randomly choose a design that meets all three constraints, and avoid that an Atlant is operated more than two days in a row (in general a threshold value th_t). Let define a possible Atlants configuration as a vector $\mathbf{h}_i$ of length N_A whose generic value is 1 if the Atlant is on and 0 if the Atlant is off and whose index ranges in the set $V_c = \{1,\ldots,i,\ldots, N_v\}$.

When we remove all combinations where the number of contiguous Atlants that are switched on are more than a given threshold th_s (in our example it is set equal to 2), i.e. $S_c = \{i: \mathbf{h}_i^T \mathbf{W} \mathbf{h}_i < th_s\}$, the cardinality of the set $S_c = \{1,\ldots, N_S\}$ of possible combinations reduces to N_S=112 possible samples for each day.

Random Selection of the Switching On and Off Daily Sequence of the Atlants

From S_c we can always derive a matrix $\mathbf{D}_S = \{d_{ij};\ i=1,\ldots,N_S,\ j=1,\ldots,N_S\}$ whose generic value d_{ij} is equal to the number of Atlants that have a different status in configurations *i* and *j* and a transformed distance matrix $\mathbf{K}_S$ whose generic value k_{ij} is equal to 1 if $d_{ij}>2th_t$ and 0 otherwise.

In the first day, u=1, a configuration $\mathbf{h}_1$ is selected at random from S_c and in subsequent days $\mathbf{h}_u$ is selected at random from the set of configurations whose distance $k_{u,u-1}$ from $\mathbf{h}_{u-1}$ is equal to 1.

Following this simple and quick procedure we guarantee the randomness and the respect of the spatial and time contiguity constraint but we cannot avoid that the Atlants are not switched on with the same frequency in the period.

Random Selection of the Switching On and Off Daily Sequence of the Atlants

To balance this sequence of configurations an iterative algorithm has been used in which every step consists in running a non-homogeneous Markov-Chain, with the temperature T being reduced at each step. Specifically, let *Fs* be a vector of the observed totals (or frequencies) for each Atlants and let *F* be the vector of the requested totals (or frequencies). The energy function can be defined as *fs=gFs-F* where *g* · is a distance function (usually Euclidean) between the observed totals and the requested totals and it can be evaluated for any possible configuration. Of course, $f(s)=0$ when all the requested totals are respected. The algorithm uses a logic in which given a configuration s^j at the *j*-th iteration, another configuration, say s^{j+1}, is chosen according to a visiting schedule. The status, intuitively, is exchanged if $f(s^{j+1})<f(s^j)$.

Random Selection of the Switching On and Off Daily Sequence of the Atlants

For any suitable choice of the stopping criterion, the final configuration is therefore obtained. Generally, this result implies that a local minimum and not necessarily a global one is reached; however this is not a problem if it is equal to the lower bound 0. More formally, the algorithm proposed can be summarized as follows. The procedure starts at time j=0, with an initial configuration generated according to the rules used to avoid contiguous Atlants in space and time, randomly selected one day of the configuration, say m, and an alternative configuration l of the Atlants from the set S_c such that it does not violate the contiguity constraint in time, i.e. from the set S_v ={i: $k_{i,m-1}\times k_{i,m+1}$=1}. The two daily configurations are exchanged if $f(s^{j+1})<f(s^j)$. The algorithm is run in steps (index by j=0,1,2,…). Each step consists in *ND* iterations.

Random Selection of the Switching On and Off Daily Sequence of the Atlants

Our computational experience indicates that when the constraints can be easily satisfied then the algorithm is quite efficient and a global minimum is easily reached in 5 or 6 iterations. Clearly, when we are dealing with a huge number of constraints arising from several equal-sized blocks a solution could not exist and it could be more difficult to find a configuration that respects all the contiguity and frequency requirements.

Repositioning the Atlants through Response Surface Optimization

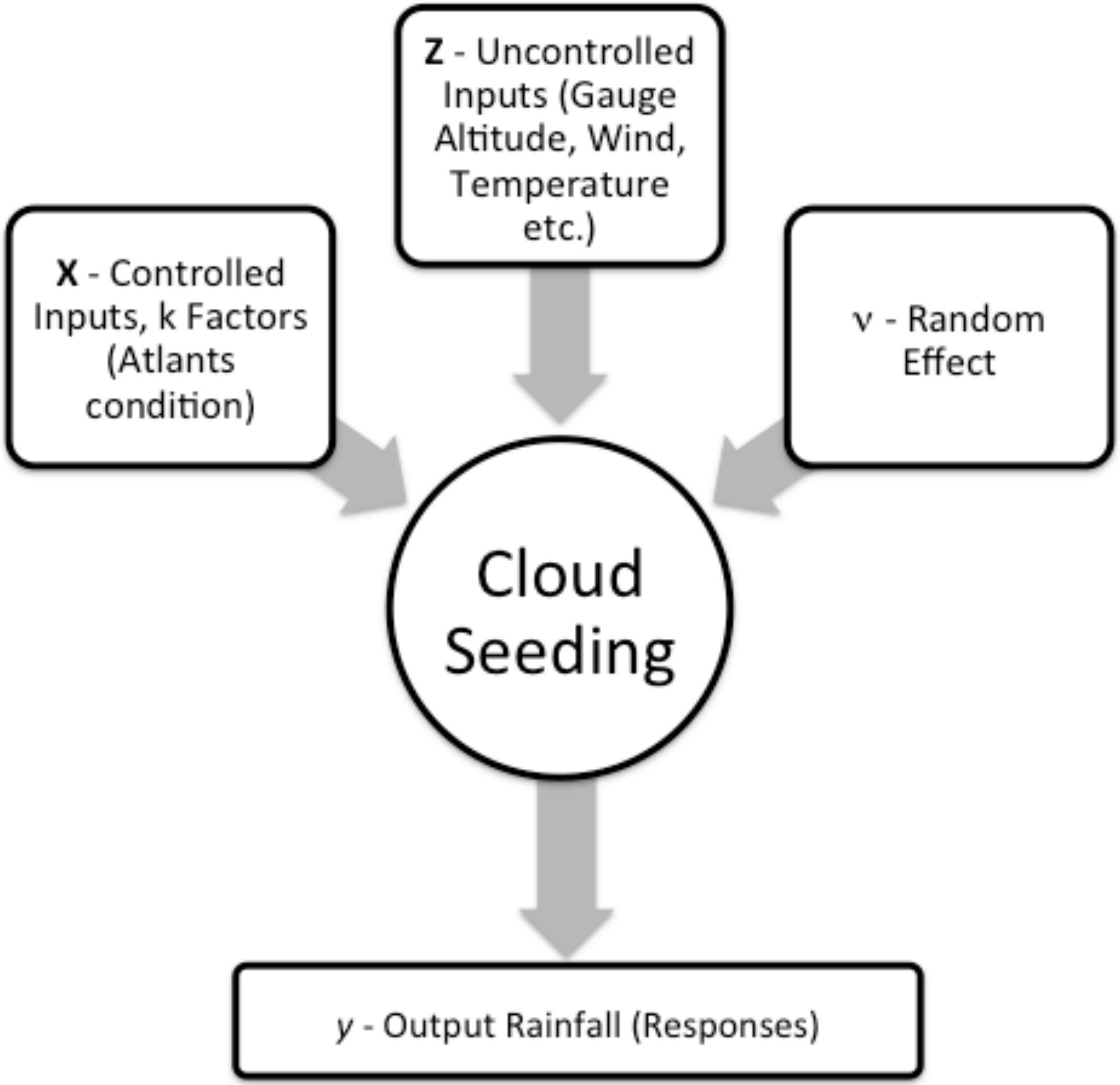
It is reasonable to assume that the daily rainfall is the outcome, say *y*, of our experiment and that it is dependent on the experimental conditions.

A useful model is the mixed effect that contains both fixed and random effects. It is particularly useful in settings where repeated measurements (of rainfall) are made on the same statistical units (the gauges in our longitudinal study). Such models are often preferred over some traditional approaches such as repeated measures ANOVA.

According to the mixed effect logic the outcome can be described as a function of the experimental variables (see Figure):

$$y = f(\mathbf{X}, \mathbf{Z}) + \mathbf{V}\boldsymbol{\nu} + \epsilon,$$

Repositioning the Atlants through Response Surface Optimization



Repositioning the Atlants through Response Surface Optimization

where **X** is a known design matrix of *k* factors, ie experimental variables whose values are controlled, **Z** is a known design matrix of *h* auxiliary variables defining the experimental conditions, **V** is a known design matrix relating *y* with n and n is an unknown vector of random effects, with mean E (n) =0 and variance-covariance matrix var(n) = G. Notice that the columns of **X** are binary variables (1=switched on and 0=switched off) indicating the condition of each Atlant for each day defined. The function *f*(**X**,**Z**) is usually approximated by a polynomial function and should represent a good description of the relationship between the experimental variables and the responses within the experimental domain defined by the study region. The simplest polynomial model contains only linear terms and describes only the linear relationship between the experimental variables and the responses.

Repositioning the Atlants through Response Surface Optimization

This simple solution, however, would be conditioned by the estimates made in the mixed-effect model, while the information on the positioning of the Atlants could be introduced directly in the model through a re-parameterization. It could be reasonable to assume that the vector $\boldsymbol{\beta}$ is a function of the geographical coordinates $\{\mathbf{c}_1,\mathbf{c}_2\}$ of the Atlants and of a set $\mathbf{A}$ of covariates known for each Atlant:

$$\beta(\mathbf{c}_1,\mathbf{c}_2,\mathbf{A}) = \boldsymbol{\delta}P(\mathbf{c}_1,\mathbf{c}_2,p) + \boldsymbol{\phi}\mathbf{A} + \boldsymbol{\zeta},$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\phi}$ are vector of parameters, P is a polynomial of fixed order p in the geographic coordinates and ζ is an error term with $E(\zeta)=0$. By replacing the expected value of this function in the initial model we could then re-parameterize it and estimate the effects of the Atlants through the parameters $\boldsymbol{\delta}$ and $\boldsymbol{\phi}$ that could be subsequently used to extend these estimates to any position whose coordinates $\{\mathbf{c}_1,\mathbf{c}_2\}$ and covariates $\mathbf{A}$ are known.

Repositioning the Atlants through Response Surface Optimization

For example, suppose that it is possible to assume a linear trend for vector $\boldsymbol{\beta}$:

$$\beta(\mathbf{c}_1, \mathbf{c}_2, \mathbf{A}) = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 \mathbf{c}_1 + \boldsymbol{\delta}_2 \mathbf{c}_2 + \boldsymbol{\zeta},$$

in this case the mixed effects model to estimate would become:

$$y = \boldsymbol{\delta}_0 \sum_{j=1}^{k} \mathbf{x}_j + \boldsymbol{\delta}_1 \sum_{j=1}^{k} \mathbf{c}_{1j}\mathbf{x}_j + \boldsymbol{\delta}_2 \sum_{j=1}^{k} \mathbf{c}_{2j}\mathbf{x}_j + \boldsymbol{\delta}\mathbf{Z} + \mathbf{V}\boldsymbol{\nu} + \boldsymbol{\epsilon}.$$

Notice that the interaction terms between the controlled factors $\mathbf{x}_\mathbf{j}$ of the status of each Atlants and its covariates (including its geographic position) plays an essential role in the estimate of the parameter vector $\boldsymbol{\delta}$ that will be used to estimate the potential effect of an Atlant in any position of the study region.

Once the parameter estimates $\hat{\boldsymbol{\delta}}$ have been obtained, we could therefore easily predict the vector $\hat{\boldsymbol{\beta}}$ for each possible position for the Atlants.

Recalling the optimization algorithm used in Section 2.2 we could then proceed to define the optimal position of a number Q of Atlants that maximize their effects on rainfall in the study area.

Using the same adjacency constraint, at the *k-th* iteration, the interaction term of the energy function could be defined as:

$$M_k\left(\mathbf{D}_U, \hat{\boldsymbol{\beta}}, s\right) = (\alpha)\left(\sum_{i \in S_A} \sum_{j \in S_E} \bar{d}_{i,j} + \frac{\sum_{i \in S_A} \sum_{j \in S_A} \bar{d}_{i,j}}{2}\right) + (1-\alpha)\left(\sum_{i \in S_A} \hat{\boldsymbol{\beta}}(i)\right),$$

where again α is the weight of the convex linear combination of the two components that we want to maximize, the distance between Atlants and their effect on rainfall, S_A and S_E are respectively the new and the existing Atlant positions and $\hat{\boldsymbol{\beta}}(i)$ are the predicted values of the effects for each point of the acceptable positions of a new Atlant.

Concluding remarks

- processing data from 2013 to 2017, it was found that cloud seeding entails more than 20% increase in rainfall.
- during our professional lives we have made some sample designs but this is by far the most complex ever treated.
- from a point of view of the publications we do not know where to start. Many topics are covered, should we put them all in one paper? Partitioning the project in many sub-aspects? On which journal, applied or environmental statistics?
- we have no competitors available in any of the proposed solutions, how can we evaluate the efficiency and effectiveness of the proposed solutions?
- the final phase of the implementation is still lacking, when the positioning will not be directed to an experimental design but to an optimization of the effects of cloud seeding

Thank you for your attention!