URSI GASS 2020
Rome
29 Aug-5 Sept, 2020

National Research
Council of Italy

Stochastic Dosimetry and Machine Learning: Innovative Approaches for Facing Challenges in Exposure Assessment in Realistic Scenarios

Emma Chiaramello, Gabriella Tognola, Marta Bonato, Silvia Gallucci, Isabelle Magne†, Martine Souques†, Serena Fiocchi, Marta Parazzini, Paolo Ravazzani

CNR-IEIIT, Milan, Italy
gabriella.tognola@ieiit.cnr.it

†EDF Electricité de France,
Paris, France

Institute of
Electronics
Computer and
Telecommunication
Engineering

Objectives

- Assess and identify most relevant features that influence the exposure scenarios and characterize the uncertainty and variability of the real exposure scenarios with advanced statistical approaches.

- In particular:

  - PART ONE: application of **stochastic dosimetry** to estimate indoor children exposure to RF generated by WLAN;

  - PART TWO: application of ***K*-means clustering** to identify recurrent patterns of indoor residential exposure to extremely low-frequency (ELF) field in children.

Part 1 - An application of Stochastic Dosimetry to estimate indoor children exposure to WLAN source

Stochastic dosimetry – rationale

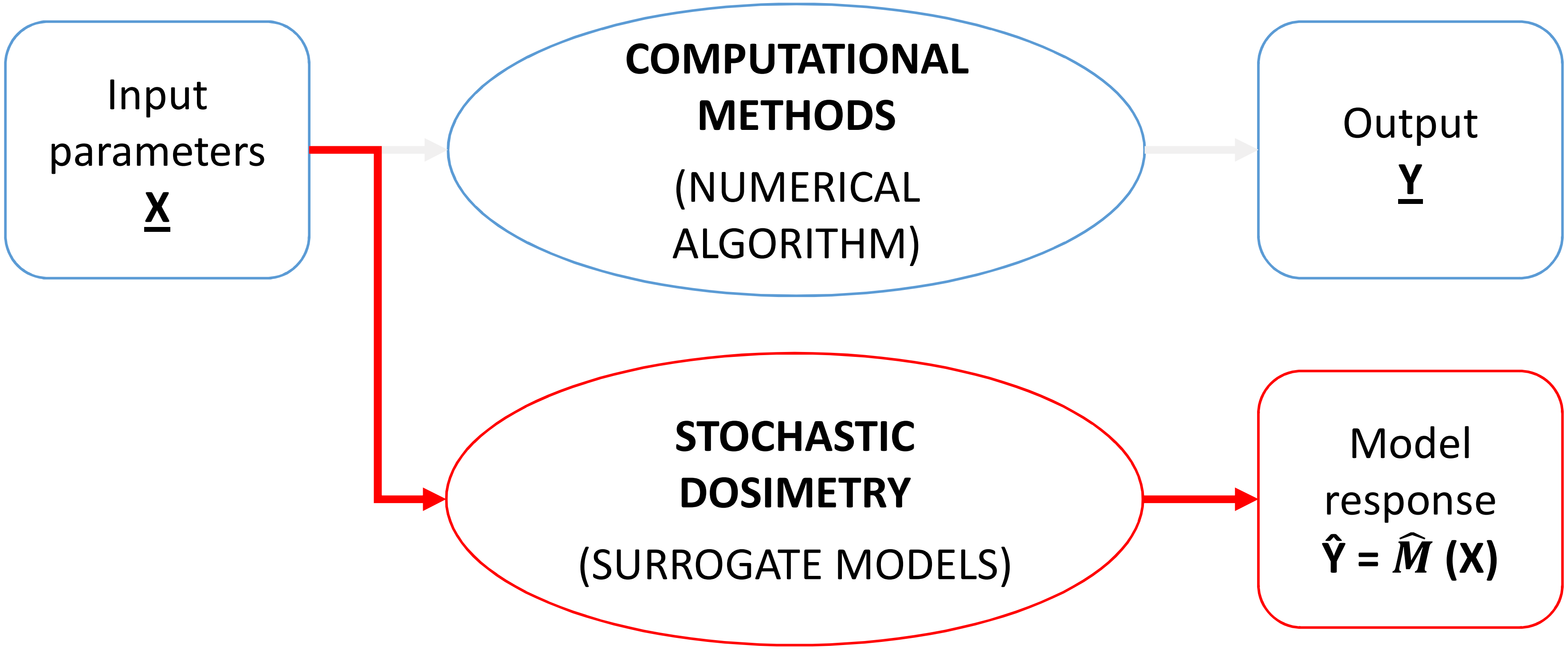
Input parameters **X** → **COMPUTATIONAL METHODS** (NUMERICAL ALGORITHM) → Output **Y**

EMF human exposure in real conditions depends on a great degree on several of parameters; each of these parameters is intrinsically affected by variability and uncertainty.

Thus to assess realistic EMF exposure it would be requested to run a high number of time-consuming deterministic simulations to take into account the variability of the parameters.

High Input Variability → **Very high number of time-consuming simulations**

Stochastic dosimetry – rationale



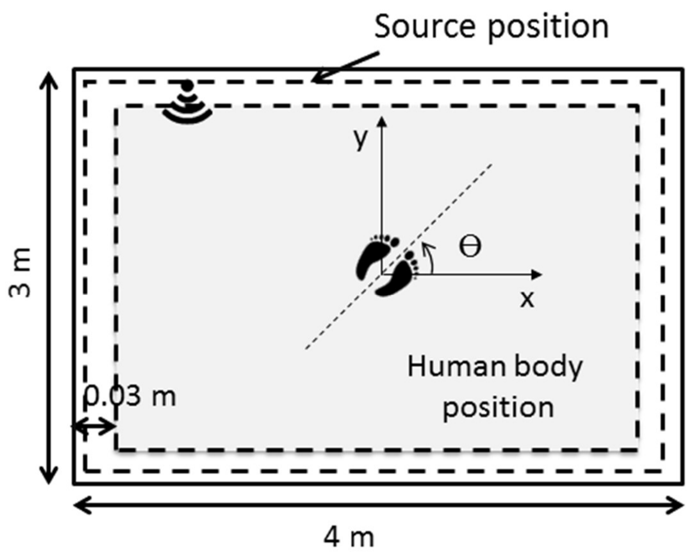
Stochastic dosimetry uses advanced statistics to build surrogate models able to estimate the distribution of the EMF exposure quantities of interest, considering the variability of the exposure scenario.

A surrogate model $\widehat{M}$ is an approximation of the original model M. The procedure follows 3 steps:

- ✓ A probabilistic model of input parameters has to be defined, i.e. the input parameters X are modelled by a random vector;
- ✓ A limited set of runs of the original model M called experimental design has to be estimated by computational methods;
- ✓ A surrogate model has to be designed using a proper statistical method.

Stochastic dosimetry – An application of Low Rank Tensor Approximation (LRA) to estimate indoor children exposure to WLAN source

Whole-Body SAR (WB SAR) induced by a WLAN source (operating at 2.4 GHz) has been assessed in child tissues when varying the position of the source on the wall and the position of the child in a 3x4 m^2 room by using surrogate models based on Low Rank Tensor Approximation (LRA).



Input parameters:

- Horizontal location L of the source
- Source height Z
- Human Body coordinates x and y
- Human body rotation Θ

	Range of variations
L	[0, 13.59] m
Z	[0.25, 2] m
X	[0.3, 3.7] m
Y	[0.3, 2.7] m
θ	[0, 359] °

Low Rank Tensor Approximation (LRA) to estimate indoor children exposure to WLAN source

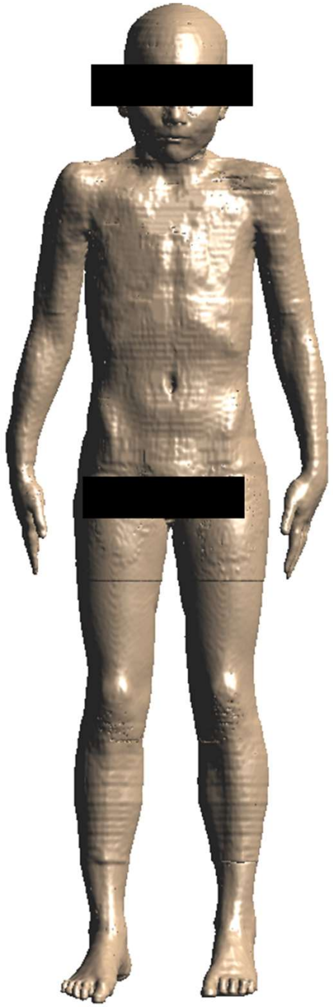
Design of the experiment

- Source: WLAN 2.4 GHz
- Anatomical model: Eartha, 8 years old girl
- Latin Hypercube Sampling
- N = 350 simulations

Analysis of the exposure

- Whole-Body Specific Absorption Rate (WB SAR),

The SAR values were computed by combination of Spherical Wave Expansion (SWE) and FDTD *



Eartha
8 years

* Y. Pinto, and J. Wiart. "Statistical analysis and surrogate modeling of indoor exposure induced from a WLAN source." Antennas and Propagation (EUCAP), 2017 11th European Conference on. IEEE, 2017.

Low Rank Tensor Approximation (LRA) to estimate indoor children exposure to WLAN source

A **Low Rank Tensor canonical Approximation** of *Y =M(X)* has the form

$$Y = \sum_{l=1}^{R} b_l \left(\prod_{i=1}^{M} v_l^{(i)}(X_i) \right) = \sum_{l=1}^{R} b_l \left(\prod_{i=1}^{M} \left(\sum_{k=0}^{p_i} z_{k,l}^{(i)} P_k^{(i)}(X_i) \right) \right)$$

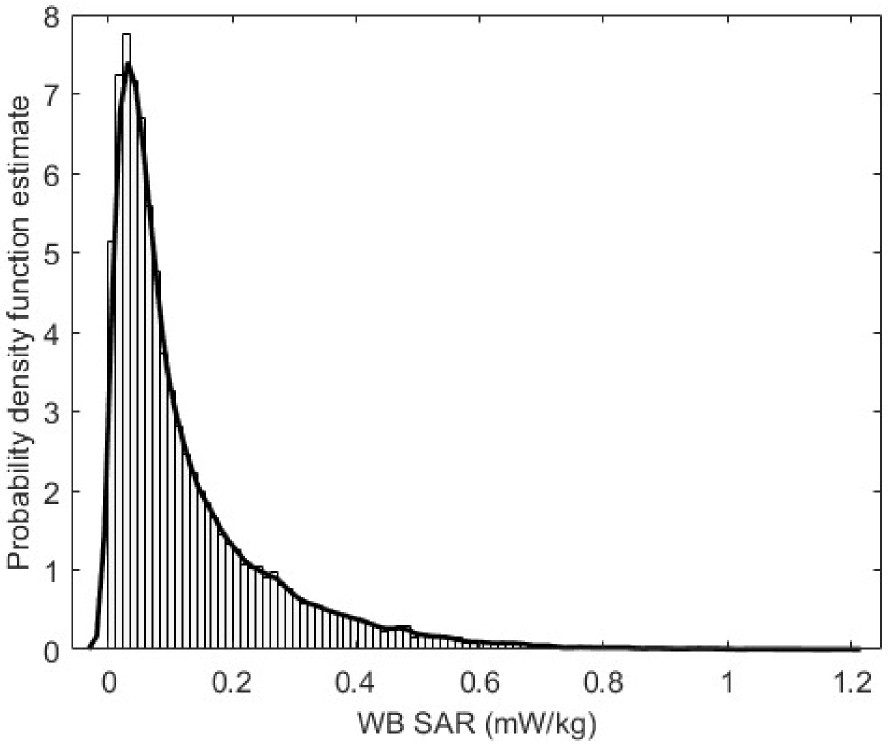
LRA can be built with **greedy approach** based on alternative least-square minimization, that involves sequentially updating of the polynomial coefficients along separate dimensions, and progressive increase of the rank by successively adding rank-one components.

The greedy approach consists of 2 steps:

- **correction step:** where a rank-1 tensor is built
- **updating step:** the normalizing coefficients are determined/updated.

Low Rank Tensor Approximation (LRA) to estimate indoor children exposure to WLAN source - Results

Histogram and probability density function of the WB SAR values obtained in 100,000 random positions of the WLAN source and the child by the LRA model.



Worth noting:

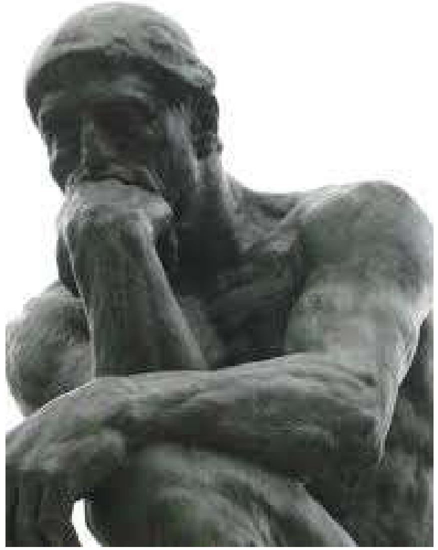
- positive-skewed shape;
- WB SAR values showed mean, median and max values equal to 0.13 mW/kg, 0.07 mW/kg, and 1.40 mW/kg;
- Quantile Coefficient of Dispersion (QCD) equal to 65% → high variability in WB SAR as a function of the relative positions between the WLAN source and the child;
- The probability density function could be approximated by a Gamma distribution with parameters a = 1.04, and b = 0.12, (with R^2 = 0.97).

Part 2 - An application of Cluster Analysis to assess indoor residential exposure to extremely low-frequency (ELF) field in children

Cluster analysis - rationale

- Cluster analysis aims to:
  - **Group similar data objects** into **clusters** based on similarities between the data
- It is an **unsupervised learning** approach:
  - no predefined clusters (i.e., no hints on what could be the clusters because learning is achieved by observation vs. learning by examples: supervised)
- Typical **applications**:
  - Exploratory analysis, as a stand-alone tool to infer new/deeper knowledge on the data;
  - Pre-processing step for variable reduction.

Cluster analysis - rationale



Starting from the observation of the measurements data

- Is it possible to describe how the data are organized or clustered, that is, to find **recurrent patterns** in the original data sample?

Measurement data sample → Cluster analysis → Recurrent patterns → Infer new/deeper knowlegde

Cluster analysis – *K*-means

- *k*-means is an unsupervised Machine Learning problem.
- The goal is to **partition the observations** into groups ("clusters") so that the **distance from the center of the cluster** ("centroid") within those assigned to the same cluster tend to be **smaller** than those in different clusters.

Observations: x_{ij} for $i = 1, 2, \ldots, N,$

Variables: $j = 1, 2, \ldots, p$ (also called *attributes*)

Distance between values of the *j*-th variable: $$d_j(x_{ij}, x_{i'j}) = (x_{ij} - x_{i'j})^2.$$

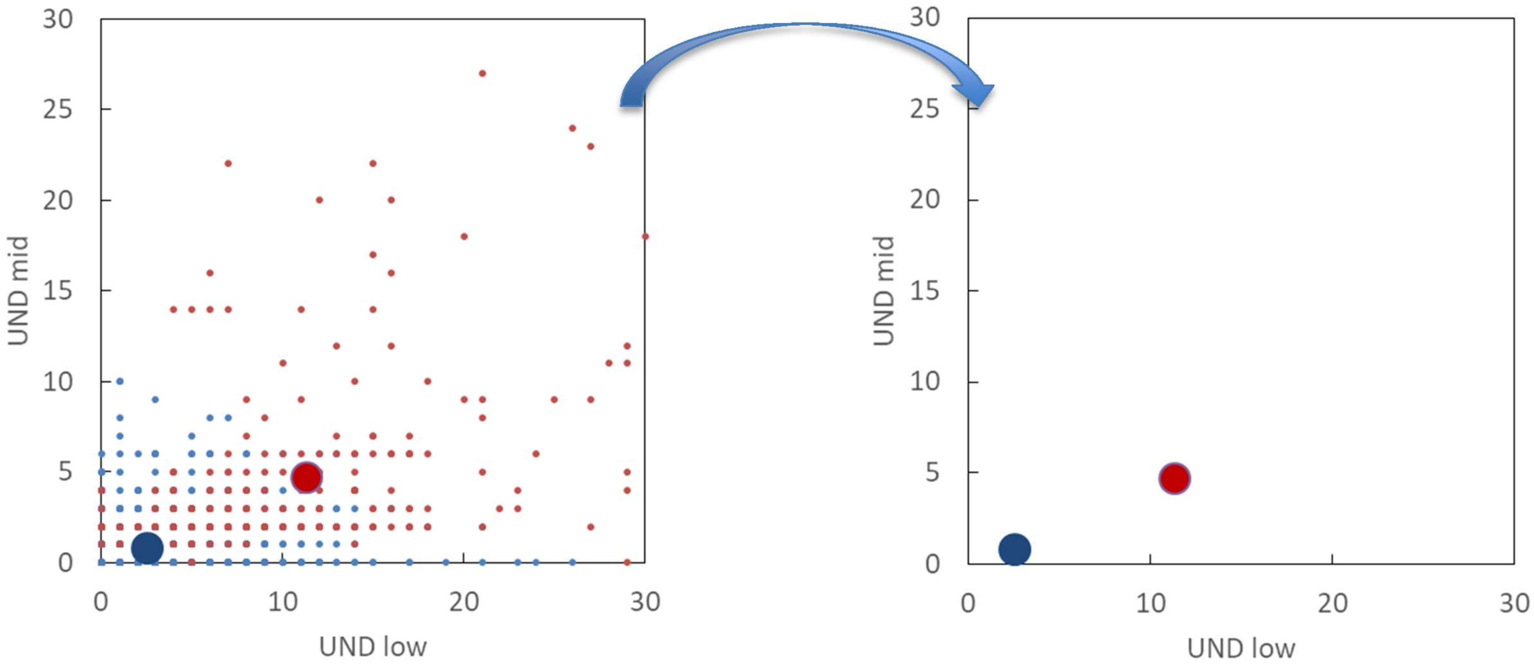
Overall distance between two observations *i* and *i'*: $$D(x_i, x_{i'}) = \sum_{j=1}^{p} d_j(x_{ij}, x_{i'j})$$

[Tognola G. et al., 'Use of Machine Learning in the Analysis of Indoor ELF MF Exposure in Children', *Int. J. Environ. Res. Public Health,* **2019**]

Cluster analysis – *K*-means

Advantages of clustering:

- The properties of data in the cluster are described by using the coordinates of the centroid instead of the values of the original data;
- **The centroid is the *representative of* the features of the data in the cluster.**



Measured data sample

Data come from the EXPERS study database, subsidized by the French Ministry of Health, EDF and RTE, and carried out by Supélec, EDF and RTE.

[Magne et al., JESEE, 2017]

Subjects: 977 children (0-14 yrs, 8.4±4.2 yrs)

Where: in France, during cold season

When: 24h a day

With: personal dosimeters

Measured variables:

More than 50 items, including:

- B field
- characteristics of power transmission lines near measurement point
- starting time, duration and type of activities performed during exposure measurement
- house characteristics (size and age)
- family size
- type of heating appliances used at home

Measured data sample (cont.d)

Variables considered in cluster analysis:

B field @50Hz: geometric mean of magnetic field (μT), measured indoor, during the day, 50 Hz component

UND: **underground cables (number)**
- *low*: 400 V (within 40 m)
- *mid*: 20 kV (within 40 m)
- *high*: 63 to 150 kV (within 20 m)
- *extra high*: 225 kV (within 20 m)

OVHD: overhead power lines (number)
- *low*: 400 V (within 40 m)
- *mid*: 20 kV (within 40 m)
- *high*: 63 to 150 kV (within 70-100 m)
- *extra high*: 225 kV (within 125 m)
- *ultra high*: 400 kV (within 200 m)

Substations: number of MV/LV (20 kV/400V) substations (within 40 m)

Further variables considered in an association effect analysis on the clusters:

- House heating;
- Residence age;
- Residence type;
- Family size (number of household residents).

K-means results – Centroid analysis

Variable	Centroid Coordinate (i.e., mean value of samples in the cluster)		
	Cluster 1	Cluster 2	Cluster 3
UNDlow (n. of lines)	1.6	13.9	3.0
UNDmid (n. of lines)	0.3	5.8	0.7
UNDhigh (n. of lines)	0.0	0.0	0.0
UNDextra (n. of lines)	0.0	0.0	0.0
OVHDlow (n. of lines)	1.6	1.0	1.4
OVHDmid (n. of lines)	0.2	0.0	0.1
OVHDhigh (n. of lines)	0.5	0.0	0.0
OVHDextra (n. of lines)	0.5	0.0	0.0
OVDHultra (n. of lines)	0.4	0.0	0.0
Substation (n.)	0.2	0.9	0.0
B (µT) **Q1** **Q3**	0.126 0.045 0.225	0.036 0.010 0.050	0.025 0 0.020

[Tognola G. et al., 'Cluster Analysis of Residential Personal Exposure to ELF Magnetic Field in Children: Effect of Environmental Variables', *Int. J. Environ. Res. Public Health,* **2019**]

Worth noting:

- Three clusters are identified: the **highest exposures** are associated to children living close to overhead lines of high (63−150 kV), extra-high (225 kV) and ultra-high voltage (400 kV).

K-means results – Centroid analysis

Variable	Centroid Coordinate (i.e., mean value of samples in the cluster)		
	Cluster 1	Cluster 2	Cluster 3
UNDlow	1.6	13.9	3.0
UNDmid	0.3	5.8	0.7
UNDhigh	0.0	0.0	0.0
UNDextra	0.0	0.0	0.0
OVHDlow	1.6	1.0	1.4
OVHDmid	0.2	0.0	0.1
OVHDhigh	0.5	0.0	0.0
OVHDextra	0.5	0.0	0.0
OVDHultra	0.4	0.0	0.0
Substation	0.2	0.9	0.0
B (μT) Q1 Q3	0.126 0.045 0.225	0.036 0.010 0.050	0.025 0 0.020

[Tognola G. et al., 'Cluster Analysis of Residential Personal Exposure to ELF Magnetic Field in Children: Effect of Environmental Variables', *Int. J. Environ. Res. Public Health,* **2019**]

Worth noting:

- mid exposure levels associated to children living close to underground networks of low (400 V) and mid voltage (20 kV) and substations (20 kV/400 V);
- the lowest exposures being associated to children living more distant from electric networks.

Association effect analysis – Effect of Heating type

Percentage of children across Clusters and heating type

Heating Type	Cluster 1	Cluster 2	Cluster 3	Heating Marginal
Non-electric	0.69 {6} (0.85) {7}	9.94 {86} (9.37) {81}	50.75 {439} (51.17) {443}	61.38 {531}
Mixed	0.23 {2} (0.24) {2}	0.92 {8} (2.66) {23}	16.30 {141} (14.55) {126}	17.45 {151}
Electric	0.47 {4} (0.30) {3}	4.40 {38} (3.23) {28}	16.30 {141} (17.63) {152}	21.17 {183}
Cluster marginal	1.39 {12}	15.26 {132}	83.35 {721}	100.00 {865}

(): percentage expected from uniform distribution (i.e., no association with the cluster)
{}: number of cases

Worth noting:

- electric heating: mostly associated to Cluster 2 (mid residential exposures);
- mixed heating: mostly associated to Cluster 3 (lower residential exposures).

[Tognola G. et al., 'Cluster Analysis of Residential Personal Exposure to ELF Magnetic Field in Children: Effect of Environmental Variables', *Int. J. Environ. Res. Public Health,* **2019**]

Association effect analysis – Effect of Residence type

Percentage of children across Clusters and residence type

Residence Type	Cluster 1	Cluster 2	Cluster 3	Residence Marginal
Individual	1.13 {10} (0.68) [a] {6}	3.28 {29} (8.01) {71}	45.81 {405} (41.53) {367}	50.22 {444}
Terraced	0.12 {1} (0.26) {2}	1.58 {14} (3.07) {27}	17.53 {155} (15.90) {141}	19.23 {170}
Apartment in small building	0 {0} (0.14) {1}	2.49 {22} (1.62) {14}	7.69 {68} (8.42) {75}	10.18 {90}
Apartment in big building	0.11 {1} (0.28) {3}	8.60 {76} (3.25) {29}	11.66 {103} (16.84) {148}	20.37 {180}
Cluster marginal	1.36 {12}	15.95 {141}	82.69 {731}	100.00 {884}

(): percentage expected from uniform distribution (i.e., no association with the cluster)
{}: number of cases

Worth noting:

- individual and terraced homes: mostly associated to Cluster 3 (lower residential exposures);
- big building: mostly associated to Cluster 2 (mid residential exposures).

[Tognola G. et al., 'Cluster Analysis of Residential Personal Exposure to ELF Magnetic Field in Children: Effect of Environmental Variables', *Int. J. Environ. Res. Public Health*, **2019**]

Association effect analysis – Effect of Family size

- Family size was found to differ with the clusters ($p < 0.02$; $\chi^2 = 8.63$);

- Family of greater size for children assigned to Cluster 2 (mid exposure level) than in Cluster 3 (lower exposure level).

Residence age has no statistical significant association with the clusters.

Conclusions

- Stochastic dosimetry based on LRA gave a complete description of the exposure scenario, without limiting the assessment to few “worst case” conditions.
- Although not eliminating the need of simulations, the LRA approach used about 0.35% of the time that would be needed using only computational methods.
- As to cluster analysis, it was useful to perform exploratory analysis on exposure data and infer new and deeper knowledge on the variables that are more important on the exposure patterns of ELF magnetic field in children.