1. Health Management and EMS

2. Problems arising in EMS management

3. Preliminaries
   - Linear Programming
   - Simulation

4. Ambulance Management in Milano
   - The performance of the actual EMS
   - Actions for improving the EMS performance
   - An alternative set of ambulance posts

5. Conclusions
Health Care Management (HCM) is a research field where academics should meet practitioners in order to provide good and implementable solutions to management problems arising in the health care delivery: operating room planning, patient flow, appointment scheduling, clinical pathways management, ...

HCM problems are more challenging than other management problems since they involve people with diseases and many social issues.

Patient-centered is the current trend in HCM: it means that solutions should take into account the patient perspective and not only the point of view of the health care provider.

It is required to deal with fairness and efficiency (not only cost) at the same time.
HCM problems usually require to adopt **unconventional** (combination) of solution methodologies.

Among many others, the problems arising in the management of an **Emergency Medical Service (EMS)** are interesting since they have all the above characteristics plus (very often) a lot (terabytes) of raw data are available.

These data represent a sort of a mine of hidden information which are usually unexplored and, by consequence, not exploited to improve the EMS capability.
An **Emergency Medical Service (EMS)** is in charge of providing pre-hospital (or out-of-hospital) acute care to patients with illnesses and injuries.

EMSs play a fundamental **role** in providing a good quality of health care services to citizens, as they provide the first answer in distressing situations.

Besides, their importance is increasing due to the **ageing of population**: in Italy, the population over 65 will increase up to the 30% (between 2030 and 2040) of the total population (now is 20%).

Moreover, about the 38% of residents in Italy is suffering from one **major chronic diseases** whose incidence increases with age.
Actually, an EMS should be able to meet the population demand in a \textit{fair} manner, concerning both \textit{horizontal} and \textit{vertical} equity, that is providing the same treatments for the same needs, and different treatments for different needs.

From an operational point of view, EMS management is a \textit{series of chronological steps}, each of them requiring precise decisions and actions whose positive outcomes largely depends on the organizational model underlying the EMS system management.

Modern EMS are organized as a \textit{district based system} along the lines of national laws. National regulation usually imposes or suggests a \textit{set of targets} defining the quality of health service provided in terms of the number of ambulances available per capita, the time threshold to get the scene, and so on.
To achieve these targets, remarkable differences can be observed in the EMS organizational model.

Furthermore, even if the EMS workflow is actually the same, such differences (structural organization, type of available resources, accessibility conditions ...) raise when considering different European countries.

Demand for ambulances is known to fluctuate spatially and temporally by day of the week, and time of day. Indeed, daily practice shows that EMS is an extremely dynamic system in which the emergency demand (stochastic in nature) changes during the day and over the district.

Huge amount of data are available due to law requirements.
An analysis, supported by **quantitative methods**, is therefore needed to provide, for instance, an accurate sizing of the resources and their careful utilization in order to guarantee a good efficiency and the fairness of the service.

From this perspective, **Operational Research** has already proven its potentiality playing a fundamental role in the analysis of the EMS organizational model as reported in [DeCEMBRIA, 2008].
The **key factors** in a successful treatment of an injury are:

- early detection,
- early reporting,
- **early response**,  
- good on scene care,
- care in transit,
- transfer to definitive care.

The **early response** is strongly influenced by the performance of ambulances, which are sent to rescue the patient.

The **resources of an EMS**, such as ambulances, are usually **limited** and therefore their management has a considerable **impact** on the overall system performance.
In this talk, we first present some **topics** in EMS management. Those topics show how EMS management could be **challenging**.

Then, we introduce some **preliminary** concepts for the development of linear programming and simulation models.

Finally, we report about the collaboration with the EMS of Milano (118) during the **DeCEMBRIA** research project funded by Regione Lombardia.
The management of emergency vehicles, especially ambulances, determines a class of problems dealing with the deployment of the emergency vehicles, their dynamic re-positioning and their dispatching.

Widening the focus on EMS evaluation, literature lacks in the comparison among different solutions, the trade-off analysis of such solutions, and in the definition of a set of unique evaluation metrics taking into account efficiency and fairness.

Note that strategic, tactical and operational levels are embedded in the previous general problems: as a matter of fact, literature reports a series of mixed applications in which, for instance, the strategic level decisions takes also into account some tactical or operational aspects
DEMAND FORECASTING

It is well known in literature [Channouf et al., 2007, Setzler et al., 2009] that emergency demand is not static, but, rather, fluctuates throughout the week, day of the week, and hour by hour within a given day.

There are many models to predict the temporal emergency demand [Williams et al., 2010, Righini et al., 2011]. On the other side, few models has been developed to predict the demand also spatially [Micheletti et al., 2010].

A promising idea is that to combine such methods with data mining. To the best of our knowledge, there is no literature about the application of data mining techniques to EMS even if there is fruitful field of research regarding the combination of data mining methods with Geographic Information Systems (GIS) in order to perform spatial analysis of geographic data [Zeitouni, 2000] and [Shang et al., 2011].
The literature on location models in general and ambulance location problems in particular, is rich and diverse [Brotcorne et al., 2003] and [Goldberg, 2004].

The first wave of published location models were deterministic in nature.

The uncertainty of ambulance availability was subsequently addressed by probabilistic location models but using simplifying assumptions.

To increase the realism of prescriptive models by reducing/eliminating simplifying assumptions, researchers have begun utilizing the descriptive hypercube model.
Larson’s hypercube model [Larson, 1974, Larson, 1975] represents an important milestone in that it introduces a spatially distributed queuing framework for facility location problems.

An important recent development in location science improves the definition of coverage used.

[Karasakal et al., 2004] introduced the notion of partial coverage which considers demand points outside a given distance threshold.

[Eiselt et al., 2009] extended classic location set covering models for softer, gradual coverage functions.
**Problems arising in EMS management**

**AMBULANCE RE-DEPLOYMENT**

Redeployment models, on the other hand, consider operational-level decisions that managers make on a daily, or hour-by-hour, basis, in an attempt to relocate ambulances in response to demand fluctuations over both time and space.

**Real Time Redeployment Models**: typically relocate ambulances every time one is dispatched, or becomes available for dispatch, with the goal of providing maximum coverage at all times [Gendreau et al., 2002].

**Multi-period Redeployment Models**: relocate ambulances w.r.t multiple time intervals [Rajagopalan et al., 2008].
The dispatching decision for EMS calls is generally quite simple: the closest idle vehicle is usually dispatched to the call. There can be complications however and sometimes it could be better to send the second closest vehicle.

Nearest neighbour policy [Harries et al., 1988] has been proven to perform, on the average, uniformly better than the other dispatching rules studied in [Larsen et al., 2002].

The problem here is to select the ambulance which minimizes the coverage loss among the closest ambulances. Exploiting real-time traffic information, on line routing algorithm can provide useful information to drive this optimization process.
INTERPLAY BETWEEN EMS AND EDs

At least in Italy, there is a weak coordination among Emergency Departments (EDs) and EMS. On the contrary, a strong coordination can help to improve the performance of the overall system.

To deal with the interplay between EMS and EDs, it could be interesting to explore the problem of assigning patients to EDs in such a way to equally distribute the workload.

This problem involves two decisions: which ambulance should be selected to serve the emergency request minimizing the coverage loss (dispatching), and which ED should receive the patient to provide him/her the best treatment while distributing the workload among EDs.
No one may say if a EMS is better than the others: organizational models can largely differ.

At European level, two projects were funded in last years to begin identifying best practices and what quality requirement should EMS guarantee: EED, European Emergency Data Project and Hesculaep (ERA-NET). Though these projects have been effective in rising interest amongst multiple European partners and systems), no indication on the best solution is given and only qualitative indications are given.

On the contrary, academic literature provides interesting starting point of analysis, such as those directly concerning EMS [Headrick, 2005] and those more general [Wagstaff et al., 2000, Sen, 2002].

Another open problem is that of developing a reliable procedure for determining the EMS standard cost.
The core of an EMS system is the Operation Centre (OC).

The operators at the OC are in charge of answering the calls and assigning a color code based on the severity of injury, through a phase called triage. After the triage phase the operator usually dispatches an ambulance. Then, he/she is ready to serve another call.

Here we refer to an urgent call as a patient with very severe injury to whom a red or yellow code is assigned: the Italian law states that the response to urgent calls has to be performed within a mandatory time of 8 minutes in the urban areas. From now on, we refer to this mandatory time as LAW time.

EMS should deliver its services in a fair and efficient way.
The staff operating at OC is organized in a set of teams having a fixed number of operators; operators answer to calls and manage the emergency requests.

The **efficiency** is directly connected to the capability of the system to guarantee a fast response at the Operation Centre to each call associated to an urgent request.

The **equity** in terms of access is related to the fact that the *same efficiency* should be always guaranteed independently on the time of the day. To accomplish it, all teams should have good and heterogeneous skills.

To this problem, and to its generalization, is devoted the tomorrow talk.


Many practical problems can be described and solved as linear programming models.
A model (or a program) is the description of a problem which require to minimize (or maximize) a cost or profit function over a given domain.
It can be represented as:

\[
\max \ z = \sum_{j=1}^{n} c_j x_j
\]

soggetto a

\[
\sum_{j=1}^{n} a_{ij} x_j \begin{cases} 
\leq b_i \\
= b_i \\
\geq b_i \end{cases} \quad i = 1, \ldots, m,
\]

\[
x = (x_1, \ldots, x_n) \in X \subseteq \mathbb{R}^n.
\]
The **components** of a linear programming model are:

- a set of **decision variables** that are used to build the model; usually they (partially) correspond to quantities on which solutions are then implemented;

- a (linear) **objective function** $f(x)$ determining the cost (or the profit) of the solution;

- a set of (linear) **constraints** representing the requirements to make a solution feasible, that is, that could be actually implemented.
**0-1 KNAPSACK: DESCRIPTION**

We are given \( n \) projects. For each project, the cost \( a_j \) and the profit \( c_j \) are known.

One can invest on a project or not, but it is not allowed to split a project; further, a budget \( B \) is available.

The problem consists in the selection of a subset of projects in such a way to maximize the expected profit without exceeding the budget constraint.

This problem is known in literature as **0-1 knapsack problem** and it is one of the most studied over the time.
0-1 KNAPSACK: MODEL

- **decision variables**: $x_j = 1$ if the project $j$ is selected, 0 otherwise
- **constraints**: we can not exceed the budget available

$$\sum_{j=1}^{n} a_j x_j \leq b$$

and binary variables

$$x_j \in \{0, 1\} \quad \forall j = 1, \ldots, n$$

- **objective function**: to maximize the expected profit

$$\max \sum_{j=1}^{n} c_j x_j$$
We are given $n$ objects and $m$ bin ($m > n$).

Let $w_j$ be the weight of the object $j$, and $C$ the capacity of each bin (the same for all bins). Without any loss of generality, we suppose that $w_j$ and $C$ are positive integers such that $w_j \leq C$, per $j = 1, \ldots, n$.

The problem consists in the assignment of the $n$ objects to a unique bin in order to minimize the number of bin using and without exceeding the bin capacity.

This problem is called bin packing problem.
**Bin-Packing: Model /1**

- **decision variables:** $y_i = 1$ if bin $i$ is used, 0 otherwise ($i = 1, \ldots, m$)
  
  $x_{ij} = 1$ if object $j$ is assigned to bin $i$, 0 otherwise ($i = 1, \ldots, m$ and $j = 1, \ldots, n$).

- **constraints:** an object must be assigned to only one bin

  $$\sum_{i=1}^{m} x_{ij} = 1, \quad \forall j = 1, \ldots, n$$

  the assigned objects cannot exceed the bin capacity

  $$\sum_{j=1}^{n} w_j x_{ij} \leq C y_i, \quad \forall i = 1, \ldots, m$$
Bin-Packing: Model /2

- variables $x_{ij}$ and $y_i$ are binary

\[ x_{ij} \in \{0, 1\} \quad \forall i = 1, \ldots, m, j = 1, \ldots, n \]

- objective function: we would minimize the number of bin used

\[ \min \sum_{i=1}^{m} y_i \]
When analytical models, such as linear programming, are not easy to solve, simulation could be an alternative approach to find a solution for a problem.

The term simulation refers to the activity of the replicate, by means of suitable models, an actual system or a system to be designed in order to study, in the first case, the effects of possible actions or events somehow predictable, or, in second, to evaluate various alternative design choices.

Three different simulation methodologies are usually employed: Discrete Event Simulation (DES), Agent Based Simulation (ABS) and System Dynamics (SD).
**DISCRETE EVENT SIMULATION**

DES models the operation of a system as a *discrete sequence of events in time* [Pidd, 1998].

DES approach is used to model the work-flow underlying a system: this allows to identify system bottlenecks, to evaluate the resource dimensioning and response time.

**Example:**

![Diagram of Ambulances Area](image-url)
AGENT BASED SIMULATION

An **agent based model** allows to track the behaviour of each individual acting in the simulated environment and the interaction among them with a view to assessing their effects on the system as a whole [Gilbert, 2008].

A **set of rules** describes the agent behaviour and its interaction with the environment; as a consequence, the state of each agent is determined [Gilbert et al., 2000].

A classical example is the agent based version of the **Bass model** of product or innovation diffusion:

- people are globally influenced by **advertising** and also can **contact** and influence each other
- people have geographical locations and can only contact their neighbours within a certain range
**SYSTEM DYNAMICS**

SD is an approach to **policy** analysis and design [Sterman, 2000].

It applies to **dynamic** problems arising in complex social, managerial, economic, or ecological systems, that is any dynamic systems characterized by **interdependence**, **mutual interaction**, **information feedback**, and **circular causality**.

Conceptually, the **feedback** concept is at the heart of the system dynamics approach.

A **feedback loop** exists when information resulting from some action travels through a system and eventually returns in some form to its point of origin, potentially influencing future action(s).
Afghanistan Stability / COIN Dynamics
REFERENCES


In the following we report about the collaboration with the EMS of Milano (118) during the DeCEMBRIA research project funded by Regione Lombardia.

Starting from the huge amount of data collected by 118, the question arose if such huge amount of data could be exploited and decision making tools could be applied so as to provide suggestions for decision makers.

Three steps were developed: first the current EMS system performance was evaluated through statistical analysis based on the collected data; then a simulation model was developed in order to study operational policies which can improve the system performance; and finally an optimization model was studied with the purpose of defining alternative sets of posts.
Joint works with:

- **Giuliana Carello**, Politecnico di Milano
- **Daniela Morale**, Università degli Studi di Milano
As mentioned, the **EMS of Milano** collects, via the **Operations Centre (OC)**, a huge amount of data describing the ambulance services or **missions**, from the instant in which a call is received by the operator to the instant in which the ambulance leaves the hospital and comes back to an ambulance post.

The Milano EMS uses **two types of ambulances**, which differs in the kind of applied contract. The **first** set composed of 29 ambulances is always available and represents a fixed cost which does not depend upon the number of performed missions; the **second** set can be summoned if needed and is paid for each performed mission.

Ambulances of the first set are located in the **ambulance posts**, while ambulances of the second one wait in the headquarters of the volunteering organizations which own them.

We denote the ambulances in the first set as **prepaid ambulances**.
# Frequencies of the Ambulance Requests

<table>
<thead>
<tr>
<th></th>
<th>Ambulance request</th>
<th>Prepaid ambulances</th>
<th>7 a.m. – 11 p.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>urgent</td>
<td>51413</td>
<td>41647</td>
<td>34663</td>
</tr>
<tr>
<td>non urgent</td>
<td>44681</td>
<td>36368</td>
<td>29808</td>
</tr>
<tr>
<td></td>
<td>96094</td>
<td>78015</td>
<td>64471</td>
</tr>
</tbody>
</table>

**Table:** Frequencies of the ambulance requests. The first column lists the total requests of ambulances, which may or may not be served by ambulances already in agreement with 118 (prepaid); the second the services covered by prepaid ambulance during the whole day; the last column the number of service requests covered by prepaid ambulances in the time period 7 a.m. – 11 p.m.
The EMS needed to evaluate, from a quantitative point of view, its capability of satisfying the emergency demand arising in different points of the urban area.

This means that a statistical analysis of the available data with the goal of gathering together system performance and spatial information was needed.
We focus on the performance of the 29 prepaid ambulances, located in the 29 ambulance posts. The rationale is to guarantee that the EMS serves the largest number of requests with prepaid ambulances.

<table>
<thead>
<tr>
<th>post</th>
<th># call</th>
<th>post</th>
<th># call</th>
<th>post</th>
<th># call</th>
<th>post</th>
<th># call</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>894</td>
<td>9</td>
<td>1055</td>
<td>17</td>
<td>1807</td>
<td>25</td>
<td>616</td>
</tr>
<tr>
<td>2</td>
<td>884</td>
<td>10</td>
<td>1847</td>
<td>18</td>
<td>1241</td>
<td>26</td>
<td>190</td>
</tr>
<tr>
<td>3</td>
<td>1546</td>
<td>11</td>
<td>492</td>
<td>19</td>
<td>1909</td>
<td>27</td>
<td>912</td>
</tr>
<tr>
<td>4</td>
<td>1666</td>
<td>12</td>
<td>2096</td>
<td>20</td>
<td>1659</td>
<td>28</td>
<td>534</td>
</tr>
<tr>
<td>5</td>
<td>2019</td>
<td>13</td>
<td>681</td>
<td>21</td>
<td>1444</td>
<td>29</td>
<td>858</td>
</tr>
<tr>
<td>6</td>
<td>1284</td>
<td>14</td>
<td>1389</td>
<td>22</td>
<td>260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>912</td>
<td>15</td>
<td>2129</td>
<td>23</td>
<td>1128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1493</td>
<td>16</td>
<td>903</td>
<td>24</td>
<td>815</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table: Frequencies per ambulance post, time period 7 a.m – 11 p.m.
In order to estimate the area covered by the ambulances, we consider a random variable which describes the Euclidean distance between the post and the scenes travelled within the allowed travel time, according to the collected data detailed in Table 2.

The use of Euclidean distance is due to the fact that there is a lack of information on the trajectory of each ambulances: actually, we do not have any information about the routes followed by the ambulance drivers and, by consequence, GIS distances can not be measured.

To verify if it is possible to consider the Euclidean distance instead of the real one, we have performed a regression among the two distances. It comes out that it is statistically significant to consider the linear relation $d_{GIS} = 1.4 \times d_E$ ($p$-value < 0.05), where $d_{GIS}$ is the GIS distance while $d_E$ is the Euclidean distance.
Figure: Box-Whisker plot for the Euclidean distance (in meter) covered by the ambulances within the LAW time, starting from each of the 29 posts.
**ACTUAL POST COVERING ANALYSIS /3**

<table>
<thead>
<tr>
<th>post</th>
<th>%</th>
<th>post</th>
<th>%</th>
<th>post</th>
<th>%</th>
<th>post</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62.1%</td>
<td>9</td>
<td>61.1%</td>
<td>17</td>
<td>58.4%</td>
<td>25</td>
<td>66.8%</td>
</tr>
<tr>
<td>2</td>
<td>59.1%</td>
<td>10</td>
<td>54.0%</td>
<td>18</td>
<td>66.2%</td>
<td>26</td>
<td>57.4%</td>
</tr>
<tr>
<td>3</td>
<td>61.1%</td>
<td>11</td>
<td>64.9%</td>
<td>19</td>
<td>53.5%</td>
<td>27</td>
<td>46.4%</td>
</tr>
<tr>
<td>4</td>
<td>63.0%</td>
<td>12</td>
<td>63.4%</td>
<td>20</td>
<td>61.4%</td>
<td>28</td>
<td>57.6%</td>
</tr>
<tr>
<td>5</td>
<td>60.1%</td>
<td>13</td>
<td>66.3%</td>
<td>21</td>
<td>61.7%</td>
<td>29</td>
<td>49.4%</td>
</tr>
<tr>
<td>6</td>
<td>61.4%</td>
<td>14</td>
<td>59.0%</td>
<td>22</td>
<td>51.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>65.2%</td>
<td>15</td>
<td>57.5%</td>
<td>23</td>
<td>71.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>65.2%</td>
<td>16</td>
<td>69.3%</td>
<td>24</td>
<td>48.6%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Percentage of demands served within the LAW time for each post.
The urban area of Milano is not **completely covered** by the posts within the LAW time. The average estimated percentage of the demand served within the LAW time over all posts is **60.1%** with a **95%** confidence interval given by **[56.13%, 64.06%]**.

A wider analysis for estimating and forecasting the demand of ambulance service in the area of Milano is reported in [Micheletti et al., 2010] and [Righini et al., 2011].

The ambulance post covering analysis have shown that there is **room for improving** the performance of the EMS system.

To achieve this goal, two different **actions** were often taken into account, that is to increase the **average ambulance speed** and to **add a new ambulance**. A further action, suggested by the preliminary analysis, was meant to increase the **time availability** of an ambulance. These actions, especially the first one, can require a huge investment without any guarantee of return in terms of improved performance.

To **overcome** this limitation, a **simulation model** were developed in order to evaluate the behaviour of the EMS system when a critical parameter, such as speed or number of ambulances, changes.
INTRODUCTION

In the development of a simulation model for an EMS, one of the most critical issues to be addressed is how to model the movement of an ambulance in the system.

The simulation models already proposed in literature are usually based on a discrete event simulation (DES) approach.

In DES, the actual movement of the ambulance is not an active part of the simulation model.
On the contrary, in our preliminary work [ORAHS 2008] and in its extension [WHCM, 2010], we proposed an agent based model (ABS-EMS) in which the ambulance movement is a **crucial** part of the model.

As a matter of fact, the agent modelling the ambulance **replicates** its movement on the euclidean space or on the GIS map.

This characteristic makes the model more **flexible** when testing different ambulance **management policies**: for instance, it naturally allows to reroute an ambulance while it is moving if a more serious emergency request occurs nearby.
REFERENCES


ABS-EMS is intended to evaluate the EMS performance starting from a set of posts.

Due to difficulty of having a reliable emergency demand generator in term of both spatial and temporal distribution, each emergency request is generated by using the real data of a given day.

This choice is also motivated by the need of the EMS managers to evaluate the system performance during some selected critical days, which are representative of typical emergency scenario.

It is composed of two types of agents, that is “Operation Centre” and “Ambulance”.

**ABS-EMS /1: DESCRIPTION**
Figure: Behaviour of the agent “Operation Centre”.
ABS-EMS /3: STANDARD AMBULANCE

Figure: Behaviour of the agent “standard ambulance”.
ABS-EMS /4: SMART AMBULANCE

Figure: Behaviour of the agent “smart ambulance” (detail).
SCENARIOS

7 different scenarios: representative of different and critical levels of emergency load and different compositions of the emergency demand.

<table>
<thead>
<tr>
<th>day</th>
<th>total</th>
<th>hospitalisation</th>
<th>number of missions</th>
<th>urgent</th>
<th>non urgent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>not hospitalisation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 25</td>
<td>256</td>
<td>221</td>
<td>35</td>
<td>153</td>
<td>103</td>
</tr>
<tr>
<td>Feb 02</td>
<td>298</td>
<td>251</td>
<td>47</td>
<td>172</td>
<td>126</td>
</tr>
<tr>
<td>Mar 08</td>
<td>270</td>
<td>224</td>
<td>46</td>
<td>170</td>
<td>100</td>
</tr>
<tr>
<td>Apr 20</td>
<td>268</td>
<td>236</td>
<td>32</td>
<td>146</td>
<td>122</td>
</tr>
<tr>
<td>May 20</td>
<td>303</td>
<td>251</td>
<td>52</td>
<td>204</td>
<td>99</td>
</tr>
<tr>
<td>Jun 07</td>
<td>284</td>
<td>235</td>
<td>49</td>
<td>156</td>
<td>128</td>
</tr>
<tr>
<td>Sep 05</td>
<td>250</td>
<td>219</td>
<td>31</td>
<td>142</td>
<td>108</td>
</tr>
</tbody>
</table>

Experiment: execution of ABS-EMS on all of the 7 scenarios. For each experiment, we report the percentage of urgent (U) and non-urgent (nU) calls not served within LAW time.
**ABS-EMS /5: MODEL VALIDATION**

Table reports the percentage of calls not served within the LAW time running the ABS-EMS over the 7 test scenarios starting from the actual post location using 29 ambulances.

<table>
<thead>
<tr>
<th></th>
<th>Jan 25</th>
<th>Feb 02</th>
<th>Mar 08</th>
<th>Apr 20</th>
<th>May 20</th>
<th>Jun 07</th>
<th>Sep 05</th>
<th>avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>29.41%</td>
<td>48.26%</td>
<td>32.35%</td>
<td>29.45%</td>
<td>45.59%</td>
<td>43.59%</td>
<td>23.94%</td>
<td>36.08%</td>
</tr>
<tr>
<td>nU</td>
<td>38.83%</td>
<td>64.29%</td>
<td>41.00%</td>
<td>41.80%</td>
<td>39.39%</td>
<td>50.00%</td>
<td>28.70%</td>
<td>43.43%</td>
</tr>
</tbody>
</table>

Table: ABS-EMS validation: (last column report the mean percentage).

As reported, the average value obtained over all the posts is 60.1% with a 95% confidence interval [56.13%, 64.06%].

Since the average value 63.92% belongs to the estimated confidence interval, we can consider the simulation outcomes enough representative of the EMS behaviour.
Concerning the **ambulance speed**, our study showed that in the time period 7 a.m – 11 p.m. it is usually very close to its average value of 25.8 km/h. Note that this behaviour is the same **empirically** observed by ambulance drivers in their experience. By consequence, in agreement with EMS managers, we set the speed of the ambulance equal to its average value. Ambulances are modelled as standard ambulances.

By decreasing the speed to 20.8 km/h, we would represent the case in which **traffic jam increases** in the urban area.

On the contrary, by increasing the speed to 30.8 km/h, we would to represent the case in which the **municipality operates against traffic jam**, that is, for instance, by arranging reserved lanes and green wave on the main streets for ambulances.
**ACTION 1: AVERAGE SPEED ANALYSIS /2**

<table>
<thead>
<tr>
<th></th>
<th>Jan 25</th>
<th>Feb 02</th>
<th>Mar 08</th>
<th>Apr 20</th>
<th>May 20</th>
<th>Jun 07</th>
<th>Sep 05</th>
<th>avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.8</td>
<td>U</td>
<td>54.25%</td>
<td>64.53%</td>
<td>54.71%</td>
<td>48.63%</td>
<td>75.98%</td>
<td>66.03%</td>
<td>45.77%</td>
</tr>
<tr>
<td></td>
<td>nU</td>
<td>61.17%</td>
<td>88.10%</td>
<td>66.00%</td>
<td>56.56%</td>
<td>72.73%</td>
<td>72.66%</td>
<td>50.00%</td>
</tr>
<tr>
<td>30.8</td>
<td>U</td>
<td>16.99%</td>
<td>28.49%</td>
<td>21.18%</td>
<td>17.81%</td>
<td>38.24%</td>
<td>26.28%</td>
<td>14.79%</td>
</tr>
<tr>
<td></td>
<td>nU</td>
<td>28.16%</td>
<td>47.62%</td>
<td>26.00%</td>
<td>31.15%</td>
<td>28.28%</td>
<td>29.69%</td>
<td>17.59%</td>
</tr>
</tbody>
</table>

**Table:** Action 1: percentage of calls not served within LAW time for each scenario and for each average ambulance speed tested (last column report the mean percentage).

The **worsening** and the **improvement** of the EMS performance are not proportional to the decrement and to the increment of the average speed.

Because of **traffic congestion**, note that in Milano the average speed is slightly decreasing along the years.
**ACTION 2: ADDING A NEW AMBULANCE /1**

The simulation experiment consists in the evaluation of the impact of adding one ambulance but keeping the same number of posts (29) and their corresponding location. This scenario models the solution that would be implemented temporarily by EMS management until a new post is activated.

<table>
<thead>
<tr>
<th></th>
<th>Jan 25</th>
<th>Feb 02</th>
<th>Mar 08</th>
<th>Apr 20</th>
<th>May 20</th>
<th>Jun 07</th>
<th>Sep 05</th>
<th>avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nU</td>
<td>33.01%</td>
<td>57.94%</td>
<td>38.00%</td>
<td>35.25%</td>
<td>37.37%</td>
<td>47.66%</td>
<td>28.70%</td>
<td>39.70%</td>
</tr>
<tr>
<td>nU</td>
<td>32.03%</td>
<td>44.77%</td>
<td>38.24%</td>
<td>31.51%</td>
<td>48.53%</td>
<td>48.72%</td>
<td>28.87%</td>
<td>38.95%</td>
</tr>
</tbody>
</table>

**Table:** Action 2: percentage of calls not served within LAW time for each scenario (last column report the mean percentage).
**ACTION 2: ADDING A NEW AMBULANCE /2**

We observe that the average performance of the system slightly decreases w.r.t the actual one.

Adding a new ambulances maintaining the same number of posts practically means that each post should host, if needed, two ambulances at the same time, instead of only one.

**Facts**: EMS implements a nearest neighbour policy and emergency demand is not static!

This implies that, during the simulation, it could happen that some posts are uncovered whilst some other posts are covered by two ambulances as ambulances tend to follow the emergency demand.

Basically, ambulances tend to be gathered in few posts determining an unbalanced global coverage.
ACTION 3: AMBULANCE TIME AVAILABILITY /1

The current system does not allow the assignment of ambulances when they are travelling to an ambulance post, as their position is unknown since the prepaid ambulances are not equipped with a Global Positioning System (GPS).

Despite the clear advantages determined by such a system, the past years were characterised by a lack of political will to support the new organisation regarding especially the need of a secure communication link between ambulances and OC.

In the current experiment, we consider smart ambulances instead of the standard one.
**ACTION 3: AMBULANCE TIME AVAILABILITY /2**

<table>
<thead>
<tr>
<th></th>
<th>Jan 25</th>
<th>Feb 02</th>
<th>Mar 08</th>
<th>Apr 20</th>
<th>May 20</th>
<th>Jun 07</th>
<th>Sep 05</th>
<th>avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>17.65%</td>
<td>28.49%</td>
<td>21.18%</td>
<td>26.71%</td>
<td>25.49%</td>
<td>26.28%</td>
<td>15.49%</td>
<td>23.04%</td>
</tr>
<tr>
<td>$n_u$</td>
<td>30.10%</td>
<td>50.00%</td>
<td>30.00%</td>
<td>23.77%</td>
<td>19.19%</td>
<td>35.94%</td>
<td>22.22%</td>
<td>30.17%</td>
</tr>
</tbody>
</table>

Table: Action 3: percentage of calls not served within LAW time for each scenario (last column report the mean percentage).

We observe a performance **improvement** with respect to the same solution without smart ambulances.

Increasing the time availability of an ambulance, catching up the time spent when they are en route to a location while they are not serving a call, can significantly increase the EMS performance.
**ACTION 3: AMBULANCE TIME AVAILABILITY /3**

Furthermore, if we consider the scenario “*May 20*”, we can observe a large percentage reduction (about the 20%) of urgent calls not served within LAW time. This scenario represents the EMS heavy day, i.e., day with a large number of total missions and, among them, a large number of urgent calls.

The same trend is confirmed by the results for scenarios “Feb 02” and “Jun 07”.

We also observe that the improvements obtained adopting smart ambulances is equivalent, on average, to that obtained by increasing the average speed.

The investment for smart ambulances seems to be more trustworthy than the one needed to increase the average speed.
ABS-EMS allows to evaluate different actions whose target is to improve the operational efficiency of the current EMS system.

Clearly, the improvement of such actions strongly depends on the current set of ambulance posts: the strategic decision on the ambulance posts location has an impact on the operational efficiency.

But what happens if a different set of ambulance posts is used? Is the current set of posts really efficient?

The following study is devoted to the problem of finding an efficient set of ambulance posts in the urban area of Milano.
**OPTIMIZATION MODELS**

We introduce some standard optimization models from literature (see, e.g., [Brotcorne et al., 2003, Goldberg, 2004] and one ad-hoc model.

Models discussed are **static** and **deterministic**.

They do not take into account the **temporal** aspect of the problem, e.g., the different latent time that an ambulance operator spends in the hospital. Moreover, they do not consider the fact that when an ambulance is available, it may start a new service, on the way back to a post.

Models take into account only the **coverage** and **capacity** aspects of the problem.
REFERENCES


**Basic Coverage Model**

- Only coverage constraints
- Minimizing number of posts
- Toregas et al, 1971

**Data**

\( \mathcal{V} \): set of demand points

\( \mathcal{W} \): set of candidate posts

\( \mathcal{W}_i \): set of candidate posts to cover \( i \in \mathcal{V} \)

**Variables**

\[ x_j = \begin{cases} 
1 & \text{if an ambulance post is set in location } j, \\
0 & \text{otherwise}. 
\end{cases} \]
BASIC COVERAGE MODEL

LSCM

\[
\begin{align*}
\min & \quad \sum_{j \in \mathcal{W}} x_j \\
\text{s.t.} & \quad \sum_{j \in \mathcal{W}_i} x_j \geq 1 \quad \forall i \in \mathcal{V} \\
& \quad x_j \in \{0, 1\} \quad \forall j \in \mathcal{W}
\end{align*}
\]

Results:

- Gives minimum number of posts/ambulances (17-20)
- Assumes ambulances are always available
- Different solutions for different time intervals
- Reasonable CPU time - less than 3 sec.
BACOP1 [Hogan et al., 1986] models the double coverage of at least a fraction of the demand.

An additional binary variable is introduced

\[ u_i = \begin{cases} 
1 & \text{if } i \text{ is covered twice,} \\
0 & \text{otherwise.} 
\end{cases} \]

The importance of double coverage is represented by the objective function, aiming at \textbf{maximizing} the amount of demand covered by at least two posts, while the covering by one post is guaranteed.
**BACOP1**

\[
\begin{align*}
\text{max} & \quad \sum_{i \in \mathcal{V}} d_i u_i \\
\text{s.t.} & \quad \sum_{j \in \mathcal{W}_i} x_j \geq 1 + u_i \quad \forall i \in \mathcal{V} \\
& \quad \sum_{j \in \mathcal{W}_i} x_j = p \\
& \quad x_j \in \{0, 1\} \quad \forall j \in \mathcal{W}, \quad u_i \in \{0, 1\} \quad \forall i \in \mathcal{V}
\end{align*}
\]

**Parameters:**

- \( d_i \): number of emergency calls in grid square \( i \)
- \( p \): number of ambulance posts to be located.
BACOP2 [Hogan et al., 1986] differs from BACOP1 with respect to its objective function.

Aims at **maximizing** a combination, using parameter $\theta$, of the demands covered once and twice. The coverage of all urban area is not required in this case.

A new binary variable $y_i$ is introduced, for each subarea $i \in \mathcal{V}$, such that

$$y_i = \begin{cases} 1, & \text{if } i \text{ is covered,} \\ 0, & \text{otherwise.} \end{cases}$$
BACOP2

\[
\begin{align*}
\text{max} & \quad \theta \sum_{i \in \mathcal{V}} d_i y_i + (1 - \theta) \sum_{i \in \mathcal{V}} d_i u_i \\
\text{s.t.} & \quad \sum_{j \in \mathcal{W}_i} x_j - y_i - u_i \geq 0 \quad \forall i \in \mathcal{V} \\
& \quad u_i - y_i \leq 0 \quad \forall i \in \mathcal{V} \\
& \quad x_j \in \{0, 1\} \forall j \in \mathcal{W}, \quad u_i, y_i \in \{0, 1\} \forall i \in \mathcal{V}
\end{align*}
\]
BACKUP MODELS: RESULTS

BACOP1

- Current number of posts ⇒ more than 90% covered twice
- Minimum number (provided by LSCM) ⇒ about 50%
- Minimum number plus one ⇒ about 60%
- Reasonable CPU time (maximum 20 minutes)

BACOP2

- With 29 posts about 99% covered once and about 98% covered twice
- Reasonable CPU time (about 100 sec on the average)
BACKUP MODELS: RESULTS
AD-HOC MODEL: DESCRIPTION

As some of the features of the considered real life case are not present in the previous models, we developed a new model tailored on the Milano case.

An ambulance can perform a limited number of missions during a given time interval. We denote this number as **ambulance capacity**.

The main **advantage** is to take into account the ambulance availability in a static deterministic optimisation model.

The problem represented by the model is to assign the urgent demand of each subarea to the posts, respecting the constraint on the time limit, while the non urgent demand can be assigned to any post in such a way they are served within a reasonable time limit.

The model aims at **minimizing** the number of needed ambulances.
**AD-HOC MODEL: DATA**

**data**

- $d^r_i$, $d^y_i$ and $d^g_i$ : number of red, yellow or green calls from subarea $i$
- $\mathcal{W}_i^1 \subseteq \mathcal{W}$ : set of candidate posts close to demand point $i$
- $\mathcal{W}_i^2 \subseteq \mathcal{W}$ : set of candidate posts "not so far from" demand point $i$
- $k_j$ : number of missions of posts $j$ in the considered time interval

**variable**

- $z_{ij}$ : fraction of red/yellow demand of point $i$ assigned to post $j$
- $w_{ij}$ : fraction of green demand of point $i$ assigned to post $j$
- $x_j$ : number of ambulances in post $j$
**AD-HOC MODEL: FORMULATION**

**LPCC (Lower-Priority Calls Coverage)**

(LPCC)

\[
\begin{align*}
\text{(7)} & \quad \min \sum_{j \in \mathcal{W}} x_j, \\
\text{(8)} & \quad \text{s.t.} \quad \sum_{j \in \mathcal{W}_{i}^1} x_j \geq 1, \quad \forall i \in \mathcal{V}; \\
\text{(9)} & \quad \sum_{j \in \mathcal{W}_{i}^1} z_{ij} = 1, \quad \forall i \in \mathcal{V}; \\
\text{(10)} & \quad \sum_{j \in \mathcal{W}} w_{ij} = 1, \quad \forall i \in \mathcal{V}; \\
\text{(11)} & \quad \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{W}_{i}^2} d_i^g w_{ij} \geq q \sum_{i \in \mathcal{V}} d_i^g, \\
\text{(12)} & \quad \sum_{i \in \mathcal{V}} (d_i^r + d_i^y) z_{ij} + d_i^g w_{ij} \leq k_j x_j, \quad \forall j \in \mathcal{W}; \\
\text{(13)} & \quad x_j \in \mathbb{Z}_+, \quad \forall j \in \mathcal{W}; \\
\text{(14)} & \quad w_{ij} \in [0, 1], \quad \forall i \in \mathcal{V}, j \in \mathcal{W}; \\
\text{(15)} & \quad z_{ij} \in [0, 1], \quad \forall i \in \mathcal{V}, j \in \mathcal{W};
\end{align*}
\]
AD-HOC MODEL: CONSTRAINTS

(8): must be at least one ambulance close enough to each subarea, guaranteeing the coverage of the whole city

(9): all the emergency calls are served within the given time limit

(10): all the low priority calls are served from any post

(11): forces at least a given percentage $q$ of the low priority demand to be served within a second, less tight, time limit

(12): the number of missions assigned to a post must not exceed the number of missions that the post can afford given by the number of ambulances assigned to the post ($y_j$) multiplied by the capacity of each ambulance
## AD-HOC MODEL: RESULTS

<table>
<thead>
<tr>
<th>Time interval</th>
<th>#ambulances used</th>
<th>CPU time [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>7a.m.-9a.m.</td>
<td>21</td>
<td>7828.87</td>
</tr>
<tr>
<td>9a.m.-2p.m.</td>
<td>24</td>
<td>13490.10</td>
</tr>
<tr>
<td>2p.m.-4p.m.</td>
<td>21</td>
<td>1842.96</td>
</tr>
<tr>
<td>4p.m.-7p.m.</td>
<td>21</td>
<td>20911.60</td>
</tr>
<tr>
<td>7p.m.-9p.m.</td>
<td>21</td>
<td>33072.40</td>
</tr>
<tr>
<td>9p.m.-11p.m.</td>
<td>20</td>
<td>2247.50</td>
</tr>
<tr>
<td>whole day</td>
<td>25</td>
<td>2545.16</td>
</tr>
</tbody>
</table>
AD-HOC MODEL: RESULTS

Optimal solution of all models provides different posts w.r.t. current ones: Moves posts from city center to city border
Combining optimisation and simulation is a promising methodology as discussed in [Fu, 2002, Glover et al., 2005].

Here we propose an iterative greedy procedure to compute an alternative set of posts for the urban area of Milano.


The procedure starts from the **optimal solution** computed by LPCC: **lower bound** on the number of the ambulances (25).

Such a solution is then **evaluated** via ABS-EMS determining a **ranking** of the ambulance posts with respect to their **utilisation**.

If the number of posts is less than the number of available ambulances, the procedure **adds** a new post in such a way to decrease the highest utilisation value.

The procedure **iterates** from the ABS-EMS evaluation until a post is located for each available ambulance.

Note that the iterative procedure can be used to design a set of posts to **guarantee a given overall system performance** by adding ambulances until the performance reaches a given threshold.
Combining LPCC and ABS-EMS

<table>
<thead>
<tr>
<th></th>
<th>Jan 25</th>
<th>Feb 02</th>
<th>Mar 08</th>
<th>Apr 20</th>
<th>May 20</th>
<th>Jun 07</th>
<th>Sep 05</th>
<th>avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPCC</td>
<td>U</td>
<td>73.86%</td>
<td>76.16%</td>
<td>74.71%</td>
<td>67.81%</td>
<td>82.35%</td>
<td>77.56%</td>
<td>68.31%</td>
</tr>
<tr>
<td></td>
<td>nU</td>
<td>96.12%</td>
<td>97.62%</td>
<td>97.00%</td>
<td>90.98%</td>
<td>80.81%</td>
<td>92.19%</td>
<td>73.15%</td>
</tr>
<tr>
<td>LPCC</td>
<td>U</td>
<td>26.80%</td>
<td>46.51%</td>
<td>37.06%</td>
<td>31.51%</td>
<td>37.25%</td>
<td>34.62%</td>
<td>35.21%</td>
</tr>
<tr>
<td>ABS-EMS</td>
<td>nU</td>
<td>42.72%</td>
<td>60.32%</td>
<td>45.00%</td>
<td>34.43%</td>
<td>49.49%</td>
<td>47.66%</td>
<td>35.19%</td>
</tr>
</tbody>
</table>

Table: Alternative ambulance post evaluation: percentage of calls not served within LAW time for each scenario (last column report the mean percentage).

Evident **improvement** gained by the iterative procedure with respect to the initial solution.

Such a solution is **comparable**, in terms of system performance, with the current location posts. The same remark holds when evaluating actions 1–3 starting from the new set of posts.
The studies reported have shown that there is room for improving the performance of the system following one of the three actions proposed and evaluated.

Furthermore, they support the idea that the inefficiencies were due to the demand peaks during the day while less impact has the current post location.

This claim is supported by the following remarks.
**FINAL REMARKS /2**

The **first one** is the difficulty to find an alternative set of posts: the solution reported in this paper is the best one among 7 different solutions tested in [ORAHS 2008] and obtained applying different strategic optimisation models.

**Then**, we observe that the LPCC solution determines a lower bound of the number of ambulances which is 25 against the 29 currently in service. Note that such a lower bound is the solution value of a deterministic optimization model in which the capacity parameter models the ambulance capacity imposing that emergency demands occurring in the same post **are not simultaneous** which is not true especially when an emergency demand **peak** arises.
To improve the EMS performance, the more promising action to be taken seemed that of increasing the ambulance time availability.

Furthermore, this action seemed to be more reliable in terms of final improvement.

Therefore, the EMS management decided to introduce the concept of smart ambulance within the EMS organisational model.

The introduction of smart ambulances requires an innovation in terms of both technological and human factors determining a change in the EMS organisational model \(\rightarrow\) logistic operators.
Despite the availability of many Information Technology and mathematical decision support tools, in Italy EMSs usually locate and manage their ambulances based on operators’ experience rather than on quantitative tools.

We report a study which proves that statistical modelling, simulation and mathematical programming can be successfully applied to an EMS, in order to evaluate its current performance and to provide suggestions to improve it.

The study shows that the Milano EMS provides a high level of performance, and yet it can benefit from the policy analytics techniques to improve the quality of the delivered service and to better exploit limited and expensive resources.
**Conclusions**

**RELATED PAPERS**


