An Hybrid Refinement Methodology for Multivariate Simulation in Home Health Telecare

Florence DUCHÊNE[1], Catherine GARBAY[1], Vincent RIALLE[1,2].

[1] *Laboratory TIMC-IMAG UMR CNRS 5525, Grenoble, France*
[2] *Department of Medical Informatics (SIIM), Michallon hospital, Grenoble, France*

***Abstract* — This paper deals with monitoring the health status of a patient at home so as to detect critical evolutions over a scale of several days or weeks. A multivariate simulation is used as a way to overcome the lack of experimental data required for studying this decision-making issue. The simulation process needs to be driven in a rigorous and robust way. Due to that, and to deal with heterogeneity and complexity, we make use of an hybrid and refinement methodology. It involves the fusion of several types of models and knowledge, as well as the implementation of a cascade structure for the simulation process.**

I. Introduction

Home health telecare concentrates on helping elderly people to remain living independently, and on enhancing their feeling of safety and security, in addition to the family and the nursing cares. Such systems may be particularly suited to elderly people living on their own, and more generally to people exposed to risks of motor (fall, etc.) or cognitive (depression, confusion, senile dementia, etc.) disorders, or needing specific medical care (diabetics, asthmatics, etc.). A remote health care system is built upon a global medical information system made up of (1) a provision of automatic devices and various sensors of different types (physiology, environment, and activity) installed and networked in the person's home, (2) a local intelligence unit (LIU) located at home and devoted to sensor-data processing and management, and responsible for broadcasting messages and alarms; and (3) a remote control center which ensures the response in case of emergency.

Experiments in remote health care systems carried out in the world are scattered and vary in their purposes and concepts. They focus either on implementing a generic architecture for the integrated medical information system, on improving the daily life of patients using various automatic devices, specific equipment, and basic alarms, or on providing health care services to patients suffering from specific diseases like asthma, diabetes, cardiac, pulmonary, or Alzheimer's. Rialle *et al.* have presented in [1] an overview of projects related to home health telecare.

Most of the current works dealing with home health telecare are focused on the architecture issue, dedicated to specific pathologies, or concern basic alarms related to the person's situation at home. Basic alarms are raised by "smart" sensors or low layers of the LIU when a problem occurs at a short temporal scale: either one parameter overpasses a critical value (nocturia, pollakisuria, fall, hypertensive crisis, etc.), or a critical scenario involving the value of possibly more than one parameter is recognized (asthma crisis, etc.). Our focus is on the broadcasting of high level alarms about the person's health status, which concern a larger temporal scale. They correspond to slow changes in the person's behavior – along several days or weeks – that are not easily spotted by a daily visit from caregivers. These changes are observed through global trends in the variation of several complementary parameters representative of the person's health status (change in sleeping time, weight, appearance of high blood pressure, decrease of activity, aging, fatigue, etc.).

Dealing with this decision-making issue and considering the lack of experimental data, we propose to set up a simulation process. The aim is to generate multivariate time-series relevant for the study of the person's behavior. The simulation process is designed to preserve the problem's complexity and requires as input commonsense and academic knowledge, as well as knowledge extracted from a set of experimental data. The methodology is based on cycles of refinement according to the outcomes of the validation stages.

These and related issues will be described in the remainder of the paper. Section II explains the reasons for a simulation process, section III the methodology used. Section IV describes the context of simulation considering home health telecare, while section V details the steps of model building and validation. Section VI presents the results of the experimentation of this model, followed by a discussion of these results, and some conclusions and perspectives.

II. Why a simulation process?

A. *Motivations to set up a simulation process*

The study of any decision-making process requires realistic and accurate data collection. Research projects about home health telecare are as yet only at their first stages of development, and collection of data in realistic environments has just started. Moreover, a full study entails consideration of several profiles of people facing many types of situations. Then, collecting complete and representative sets of data may be a quite hard task, especially to hold data corresponding to rare events. For these reasons, many researchers have turned to simulation as a way to overcome the difficulty of collecting large sets of experimental data.

In relation to experimentation, setting up a simulation process enables researchers to have a full and tightly controlled universe of data sets. The advantages are at least five fold: (1) producing large sets of data to experiment with decision-making algorithms, (2) generating data representative of many situations and many peoples' profiles, (3) building a process that is credible and easily understandable by any actor of the system, in contrast with an analytic modeling approach, (4) providing a better *a posteriori* knowledge of the parameters observed at home and the trends of their joint variations, and (5) testing the efficiency and the robustness of detection algorithms by varying the simulation parameters.

B. *Related works*

Simulation can be defined as "the process of designing a model of a real system and conducting experiments with this model for the purpose of understanding the behavior of the system and / or evaluating various strategies for the operation of the system." [2] This entails that simulation includes both the construction of the model and their experimental use.

Simulation is already broadly used in the medical area, but especially for more efficient teaching and training of students and practitioners, without putting patients at risk.

Regarding any research area, Kelton [3] underlines the success experienced by simulation research and practice over the past 25 years. In the context of decision-making, Shannon [2] considers simulation to be one of the most powerful tools available for the design and operation of complex processes and systems. Advantages of simulation over analytical or mathematical models for analyzing systems include: (1) the concept of simulation is easy to comprehend, (2) a simulation model is more credible because its behavior is compared to the real system or it requires fewer assumptions, and (3) it lets us experiment with new and unfamiliar situations.

Even though simulation has many strengths, it is not without drawbacks. According to Shannon [2], the usefulness of a simulation process depends on (1) the quality of the model, (2) the appropriateness and quality of the data, and (3) the accurate specification of the simulation conditions to generate data consistent with the purpose of the study. This includes the need to take particular care in the system definition and conceptual model formulation of the levels of abstraction and simplification in order to neither oversimplify the system, nor carries too much details. In terms of simulation model complexity, Chwif [4] advises to keep a model simple and to add complexity later if it is strictly necessary. The aim is then to determine the best complex level of a given model that still maintains its validity. However, Chwif also highlights a lack of methodologies to lead a modeler to obtain a simpler model.

More generally, Kelton [3] underlines that simulation faces general methodological problems concerning how to model, how to plan a course of simulation experimentation, and how to interpret the results. Simulation methodologies are scattered and vary widely according to the context of their application. According to Kelton, it even seems that simulation research may have wandered off in directions that are not particularly tied to demand derived from realistic applications. In [5], Ören also highlights the diversity of simulation methodologies. He presents a simulation taxonomy and points out many types of simulation whose relevance depends on varying features like the time or the functional relation of descriptive variables. This leads to the definition of almost up to 100 types of simulation. Then, extracting general guidelines to set up a simulation process becomes almost impossible.

However, there is a critical step required in any simulation process: the validation and the verification of the model and their behavior. This issue is tackled in several papers like [2,6]. Verification and validation aim at checking respectively whether (1) the simulation process operates the way the analyst intended and (2) it behaves the way the real system does or will. In [6], Sargent exclusively discusses validation and verification of simulation models. He particularly discusses how model validation and verification relate to the model development process, and presents the main steps of the process and techniques used to complete these tasks.

In our context of research, the simulation process includes several original features related to the use of an incremental and hybrid approach. The methodology is incremental to allow the process refinement at two levels: (1) within the scope of the problem solving scheme and (2) at the stage of the simulation model building and validation considering a given purpose and context. An hybrid approach is required (1) to deal with heterogeneous data, (2) to support the generation of multidimensional and correlated data sets, and (3) to integrate different kinds of knowledge in the process: commonsense and academic knowledge, as well as new knowledge extracted from experimental data sets.

III. METHODOLOGY FOR SIMULATION

A. *Simulation as part of a problem solving scheme*

Setting up a simulation process makes sense only in its context and purpose of use. A simulation process should be considered as part of a problem solving scheme (figure 1). The generation of relevant time-series for the issue studied, as well as the data collection if required, is led by contextual information related to the general purpose and context of the decision issue. This aims at narrowing and specifying the space of information and knowledge to consider by answering questions like: what are the relevant observations to set up ? or which level of detail to consider? Data collected from experiments or generated by a simulation process are then used to test appropriate methods of decision making to solve the problem. The sensitivity and specificity related to these algorithms must match the problem requirements.

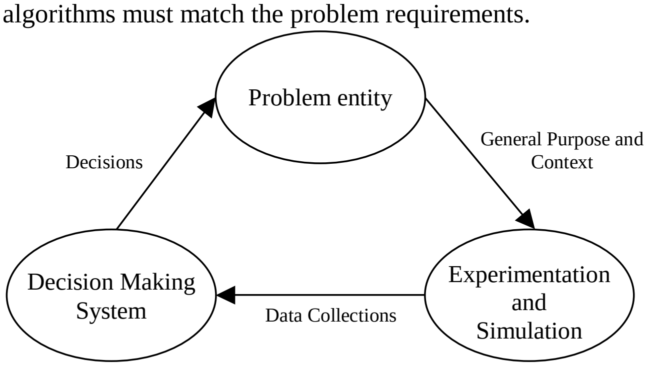

Figure 1. A simulation process as part of a problem solving scheme

The decision issue about the health status of a person at home involves some complex phenomena (such as the observation of more or less correlated parameters that may be sensitive to a deterioration in health) that needs to be saved in a simulation process. However, a difficulty lies in the lack of *a priori* knowledge about the joint variations of these parameters, and that requires a reliance on the diversity of informational sources, which are *a priori* knowledge – that is commonsense and academic knowledge – and experimental data sets. Some new useful knowledge – so called "extracted knowledge" – may be extracted from the experimental data.

The next critical issue related to simulation concerns the level of detail required. The relevance and the level of detail of the knowledge involved in the simulation process must be suited to the decision's purpose, so that the generated time-series are also appropriate to the decision making. The study of slow changes in a person's behaviour at home rests on the observation of mean- or long-term trends, which is a "high level" issue. Moreover, the aim is not to interpret precisely the problem that occurred, but to set up the context of occurrence of the changes. Then, while remaining realistic, the sequences of data used for the decision making may not require a high degree of detail and accuracy. It is the joint variation of these parameters that is more crucial. This highlights a compromise to be found between the necessity to save the complexity of phenomena and the restriction to a level of detail that meets the decision's purpose, as will be discussed at the end of this paper.

Once a simulation process has been implemented and experimented, the results of matching between the outputs of the decision making process and the problem requirements may entail a refinement of the general purpose and context of the experimentation and simulation in order to get better sensitivity and specificity. In addition to the refinement of the decision making system, the set of the parameters or the level of detail in their observation might be changed. For instance, more precision in the values of these parameters may be required in case of low sensitivity, or the contrary, in case of low specificity.

B. The simulation process : a cycle of refinement

On its own side, the simulation process is also a cycle of refinement to better match the requirements of the problem entity [6]. The simulation process includes sequentially: (1) the conceptual model building according to the analysis of the general purpose and context of the problem; (2) the implementation of this model in a computerized model; and (3) experimentation to generate large sets of data.

The critical points of verification and validation are relevant at all these steps of simulation, so that they are conducted as part of the simulation development process (figure 2). Their aim is: (1) to validate the theories and assumptions underlying the conceptual model and to check the model's structure and logic according to the intended purpose; (2) to ensure their right implementation, that is the computer programming is correct; and (3) to determine whether the model's output behavior has the accuracy required for their intended purpose in reference to experimental sequences of data (operational validation).

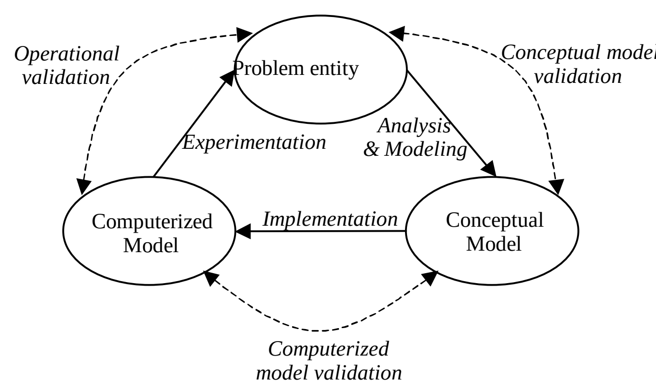

Figure 2. Simulation process

C. An hybrid simulation scheme

Fusion is commonly used to deal with complexity and heterogeneity. In setting up the simulation process, we considered it essential to integrate several kinds of knowledge at each stage of its conception and validation (figure 3), that is: (1) <u>commonsense knowledge</u>, (2) <u>academic knowledge</u>, and (3) <u>knowledge extracted from experimental data sets</u>. Commonsense knowledge is exclusively qualitative, whereas academic and extracted knowledge may be either qualitative or quantitative. Qualitative knowledge is interesting at two levels: (1) it gives an idea of fundamental concepts underlying the simulation model, and (2) commonsense knowledge from experts may be used for the validation of both the conceptual model and the data produced by the simulation when facing a lack of quantitative knowledge and / or experimental data. Quantitative knowledge allows validation and / or quantification of the concepts in order to get a model ready for implementation. It may be extracted from one part of the experimental data set (modeling data set). The other part is used for the operational validation (validation data set).

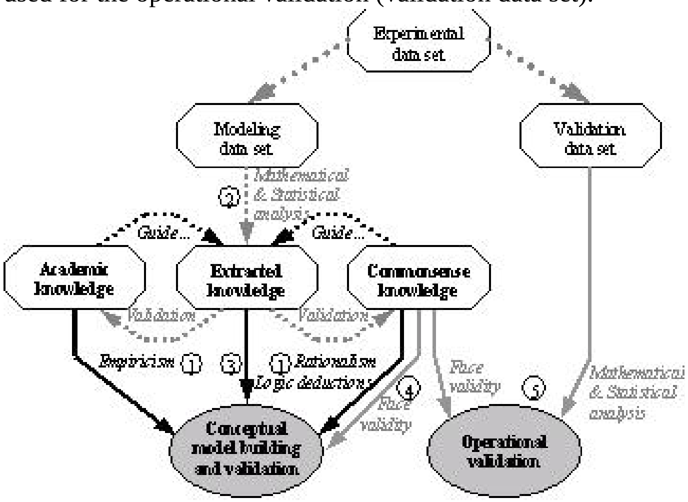

Figure 3. Integration of knowledge in the simulation process
1. Definition of the underlying conceptual model
2. Validation of some concepts using knowledge from experimental data
3. Integration of extracted knowledge in the model
4. Validation of other concepts of modeling by experts
5. Validation of data sets produced using the model implementation

The kinds of knowledge involved at the different stages of the simulation process determine the techniques of modeling and validation [6]: (1) rationalism and logic deductions integrate commonsense knowledge and assume that everyone knows whether the underlying assumptions are true; (2) empiricism involves academic knowledge and then requires every assumptions to be empirically validated; (3) mathematical analysis and statistical methods are used to test theories and assumptions underlying the conceptual model using experimental data set, and produce new so called extracted knowledge; and (4) face validity involves asking people knowledgeable about the system whether the model looks reasonable, calling on their commonsense knowledge. Concerning the operational validation, the major attribute affecting the selection of a relevant technique is whether the system is observable. If data about the real system are available, appropriate techniques consist of comparing the model and system input-output behaviors, for instance by the means of graphical displays or using statistical tests and procedures. Otherwise, the data sets produced by the model are validated subjectively by experts, or by comparisons to the results of other models.

IV. THE CONTEXT OF HOME HEALTH TELECARE

A. Purpose and requirements

In the context of detecting bad trends in health status, we aim to learn the person's lifestyle to build a sort of profile, which is sensitive to any critical deviation, and then to detect any unusual behavior in comparison with this profile. This approach towards the decision-making issue is required because it is inconceivable to describe all possible critical situations of any nature and level, just as we do not yet have any way of learning the occurrence of such situations (monitoring of persons getting to critical situations and collecting the corresponding data). In that context, the purpose of the simulation is, first, to generate sequences of data representative of usual conditions of life. Later, the simulation of the disruption of these sequences of data, in a more or less realistic manner, will allow us to test the efficiency and robustness of decision algorithms, expected to detect any unusual situation.

B. Observable parameters

The decision-making system is based on a set of data, recorded at home and in real-time, that may be collected from different classes of sensors: (1) activity (location, position, motion, fall, etc.), (2) environment (temperature, use of doors, window, lighting, etc.), and (3) physiology (blood pressures, weight, etc.). In the definition of these observable parameters, a compromise needs to be found between (a) being easily observable and non invasive, by focusing on the observation of a small set of parameters, and (b) gaining a full appreciation of the person's condition, sensitive to any change in the health status.

A deterioration of a person's health status usually entails behavioral disorders whose observable symptoms range from an increase in the risk of falls, slowness in executing simple actions, forgetfulness in daily activities, to a global decrease in the person's ability to perform activities of daily living (ADL). Clinical practice has already widely exploited this correlation by estimating a patient's health status in terms of their ability to perform ADL such as getting washed, dressing, or feeding themselves. The usefulness of monitoring some parameters related to the activity of a person is often underlined as being an essential part of any health evaluation [7,8], and several projects in home health telecare [9-11] have already integrated in their concept the assessment of the ADL. Representative of both the activity and the health status, the heart rate is another important and easily observable physiological measure [12].

Thus, we decided to consider in a first step of simulation four parameters that can be defined from a provision of sensors and that are representative of both the heart rate and activity of a person at home: (1) the person's moves, (2) their postures, (3) the activity level, and (4) the mean heart rate.

C. Simulation inputs: data and knowledge

1) Experimental data

Experimental data are required to set up a simulation process. In our context, records of data from sensors installed in a person's home are not available. However, we have got data corresponding to the monitoring during two nonconsecutive periods of 24 hours of twelve young, normotensive, and healthy subjects in their everyday life, male or female, between 20 and 30 years old. Data recorded include: (1) the date and time; (2) the annotation of activity types (every 15 minutes), ranked in 14 increasing levels: sleeping, lying down, sitting still, sitting and speaking, sitting and working, eating, standing, standing and working, ridding bicycle, walking slowly, walking quickly, running, climbing stairs, and going down stairs; (3) a kind of activity level, measured on an arbitrary scale, and corresponding to the norm of the acceleration along the anterior-posterior axis averaged every minute; and (4) the mean heart rate (every minute), corresponding to the average of data recorded by an ECG portable device.

This experimental set of data is restricted so that it is relevant in the context of home health telecare: (1) the activity types correspond to activity levels from low to moderate, and (2) the person is supposed to live in a ground floor house. Then we take account of data corresponding to the following activity types: sleeping, lying down, sitting still, sitting and speaking, sitting and working, eating, standing, standing and working, walking slowly. Moreover, we remove data for the four minutes following high level activities, to be free from the effect on heart rate of the time needed to recover after intense activities.

2) A priori knowledge

A priori knowledge corresponds to well-recognized and / or validated knowledge, that is either commonsense or academic knowledge.

In our context, useful commonsense knowledge concerns the activities of daily living (ADL) of a person at home: (1) the distribution of activities within a day, (2) the expected moves and postures during the different kinds of activities, and (3) the mean activity level according to the activity types.

Academic knowledge of interest (taken from [12]) are related to some features of the mean heart rate and its variability in relation to activity types. The mean heart rate is generally computed from ECG records by an average every 30 seconds to 1 minute. Monod and Pottier [12] note the influence of the posture on the mean heart rate: (a) when sitting, the mean heart rate is 10% higher than when lying down, and (b) when standing, it is from 20 to 30% higher than when lying down. The effect of a low activity on the mean heart rate is also described, as follows: (1) quasi-linear increase of the heart rate with the activity level (a saturation effect is observed for the values corresponding to high activity levels), (2) rapid stabilization of the values, and (3) need for 1 to 3 minutes to recover. Moreover, mean heart rate values are characterized by a large variability.

V. CONCEPTUAL MODEL BUILDING AND VALIDATION

A. Principle of model building

The conceptual model building is guided by *a priori* knowledge about the dependence between parameters, which gives an indication of the factors and relative influences that are relevant to consider in the modeling. One requirement of the simulation process is indeed to preserve the problem's complexity, that is especially the joint variations of the parameters. Considering a person moving within the rooms of their home, one can intuitively think that: (1) their successive postures depend on the room occupied and the time of the day, (2) the activity level of the move and posture, and (3) the main variations of the mean heart rate are conditioned by the posture and activity level. Thus, the conceptual model is defined using a cascade structure, and run in four steps to successively generate time-series corresponding to: (1) the moves of the subject in a given period of time, (2) their successive postures, (3) the sequences of the activity levels, and (4) the values of the mean heart rate. This cascade structure with four levels entails to define the same number of sub-models, one per simulation parameters, involving different modeling and validation techniques, as described on figure 4.

B. Moves

1) Principle of modeling

The model defined for the generation of the moves of a person at home is based on commonsense knowledge about their activities of daily living. The corresponding modeling

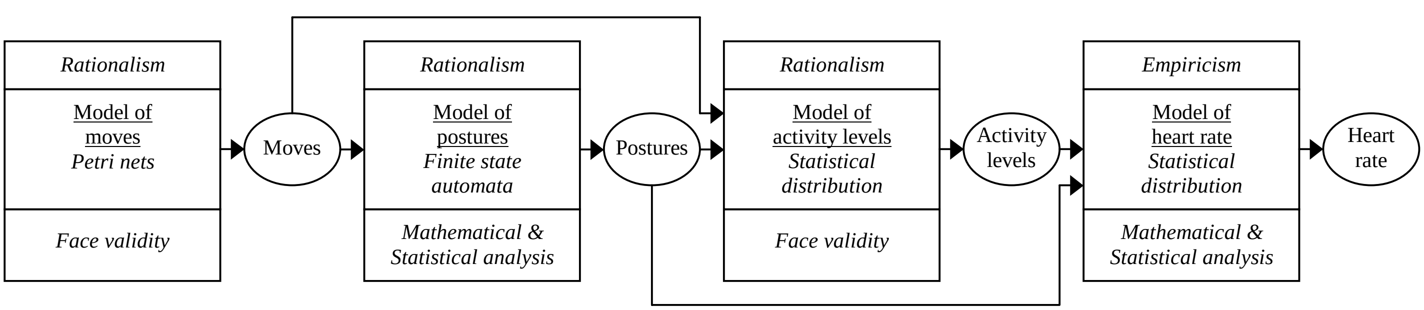

Figure 4. A cascade structure for simulation.
The figure shows in ⬭ the simulated parameters, and in ▭ the submodels used for the simulation, detailing from top to bottom: (1) the technique used for modeling, (2) the model built, and (3) the technique of validation.

techniques are thus the rationalism and logic deductions. The validation is performed with experts using face validity because no experimental data are available for this parameter.

2) Model building

This model has been built and implemented by Virone [13]. It is made up of a set of Petri nets representing the expected moves for a person according to the moment of the day. One day is divided into seven periods: (1) night, (2) getting up, (3) morning, (4) lunch, (5) afternoon, (6) evening, (7) bedtime. The opportunity for the subject to go out off their house is not considered. The house is supposed to have six rooms on the same floor: (1) kitchen, (2) living room, (3) bedroom, (4) bathroom, (5) toilets, and (6) corridor. However, simulated data are pre-processed so that short stays (less than 10 seconds) in any room are not recorded, which requires the corridor to be never occupied considering the results of the simulation.

C. Postures

1) Principle of modeling

The postures of a person represent “static” information about their activity at a given time. Just as in the previous case, the model defined for the generation of the postures is based on commonsense knowledge about the characteristics of a person’s postures in their activities of daily living (rationalism and logic deductions). The validation is also performed using face validity, with experts, because no experimental data are available for this parameter.

2) Model building

The model defined at this stage is based on finite state automata, using the outputs of the previous simulation process (the moves) as input. The same three states are defined for each automaton, corresponding to the possible postures of a person, that is: (1) standing, (2) sitting, and (3) lying down. Considering the context of home health telecare, one assumption is that the “lying down” and “standing” postures are only reachable through a “sitting” posture. When entering or getting out of a room, the person is also supposed to be standing. Thus, the posture is forced to change from lying down to sitting and then standing, or from sitting to standing before a move from one room to another.

A finite state automaton is characterized by transition probabilities between states. These probabilities are defined by taking account of the room occupied by the subject. An additional criteria in case the room is supposed to generate time-dependent behaviors (like when considering the bedroom, but not the toilets for instance) is the moment of the day (one of the seven moments defined for the model of moves). The time to transition depends on the room occupied, and also on the moment of the day in some cases, as for transition probabilities. It is randomly determined for each transition time according to a gaussian distribution.

D. Activity levels

1) Principle of modeling

The activity level of a person represents “dynamic” information about their activity at a given short period of time (1 minute). In the context of home health telecare, we assume that the activity level accurately represent the efforts put in.

Four types of movements, which can be identified using information about the moves and postures, are considered because of the different features they intuitively induce on the activity level: (1) movements when lying down, (2) movements when sitting, (3) movements when standing, and (4) walking. The last one is a subset of movements when standing, and is identified as standing and moving from a room to another. We expect the activity level to be, on average, higher when the movements’ group number is higher. The model of simulation is based on a random generation of activity levels at any time according to the distribution characterizing the type of movement observed at the same time. The corresponding conceptual modeling techniques are thus the rationalism and logic deductions.

The model validation is performed using statistical analysis on experimental values from the modeling data set: we must study the distributions of activity levels according to the type of movements to check the intuitive assumptions and determine their features.

2) Model building

The model building starts with four statistical analysis of activity levels, one per type of movements that can be observed. This study confirms the intuition with an average increase in activity levels with the movements’ group number: a mean level of 0.60 when lying down, 1.58 when sitting, 3.46 when standing, and 4.82 when walking. The distributions appeared to match a mix of (1) a normal distribution at low levels, and (2) an exponential distribution at higher levels (figure 5). The interpretation of such results may be that the activity levels are usually concentrated about a mean value considering a given type of movements, but it is possible to occasionally get some higher values. The features of each distribution are calculated from the empirical mean and variance.

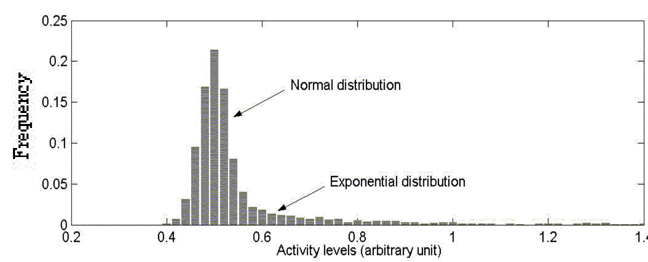

Figure 5. A mix of normal and exponential distributions to model the distribution of activity levels considering a given type of movements (this sample is when lying down)

Given that the information about the type of movement performed at a time is extracted from the sequences of moves and postures previously generated, a value of activity level is randomly generated according to the appropriate distribution at any time.

E. Heart rate

1) Principle of modeling

The model defined for the generation of mean heart rate values is based on empiricism, relying on academic knowledge that especially shows: (a) an inter-subject variability in heart rate values, and (b) a dominant sensitivity of this parameter to the posture and activity level. At least part of the subject's specificity in their heart rate variations lies in the features of their resting values, which we therefore consider in the simulation process: (1) the generation of a time-series for the resting heart rate, and (2) the introduction of some aleas on these values according to the posture and activity level at the same time. That also gives the opportunity to generate data sequences corresponding to different personal profiles.

The model validation is performed using statistical analysis on experimental values from the modeling data set to check and quantify the academic knowledge integrated in the conceptual model building. Given that the modeling data set is made up of records from several people and that everyone has specific features in their heart rate variability, we propose the normalization of the mean heart rate values of each subject by considering the aleas from the specific variation of their resting heart rate. Thus, the model building includes study of: (1) the characteristics of variation of the resting heart rate for any subject, (2) the temporal correlation between activity level and heart rate, (3) the relation between the aleas from the resting heart rate and the activity level considering each possible posture, and (4) the distribution of mean heart rate values according to a given posture and activity level.

2) Model building: normalization of heart rate values

The normalization of heart rate values is performed according to the features of the resting variations. Considering a physiological parameter such as the mean heart rate, these variations very nearly follow a circadian rhythm (period of 24 hours). We decided to analyze the circadian rhythm of every subject using the cosinor technique [14,15], in which collected over 24 hours are represented by the best cosine function using the "least squares" calculation. More powerful procedures have been perfected to overcome the main limitation of this technique resulting from the assumption that the studied rhythm fits a sine curve. But given that, in our context, few data are available for each subject when resting, the results of any circadian rhythmicity analysis may be inaccurate anyway. The resultant characteristics produced by such an analysis are as follows: (1) the MESOR, M (Midline Estimating Statistic Of Rhythm) (average level around which the oscillation occurs), (2) the amplitude, A (measure of the extent of rhythmic change), and (3) the Acrophase, $A\phi$ (measure of the time at which the fitted cosine reaches its maximum value), ϕ being the phase, expressed in trigonometric units. The time-variations of the heart rate values, HR(t), are then following the equation (1), where t is the time expressed in hours.

$$\mathrm{HR}(t) = M + A.\sin((2\pi/24)t + \phi). \quad (1)$$

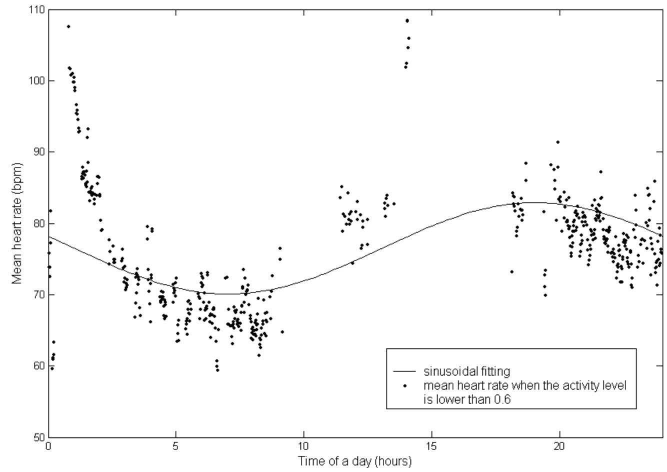

Figure 6. Circadian rhythmicity analysis for a typical subject

This cosinor method is applied on experimental heart rate records for each subject, after selecting the values which correspond to a low activity level (i.e. lower than 0.6). Given that the heart rate varies with posture, even when the patient is really quiet, the values are adjusted considering the posture observed at the same time: values are divided by 1.1 when the person is sitting, and by 1.25 when standing. The analysis showed that, for a sinusoidal circadian rhythm, average values were approximately: (1) $M \approx 70$ bpm, (2) $A \approx 6$ bpm, and (3) $A\phi \approx 16$h. Figure 6 shows one sinusoidal circadian rhythm estimation.

3) Model building: aleas on the heart rate from the resting values

In the study of the alea on heart rate from the resting values, a first step is the analysis of the temporal correlation between the activity level and heart rate. Indeed, considering the need for time to recover after any activity, commonsense knowledge lets us think that the heart rate at a given time may not depend only on the activity level at the same time. We therefore need to determine the time interval preceding a measure of heart rate during which the activity level has an influence on the measured heart rate. An analysis of intercorrelation between these two parameters shows a peak when the two time-series are in phase, indicating that the temporal relation can be described as (2), where HR(t) represents the value of heart rate at time t, and ACT(t) the one of the activity level at the same time:

$$\mathrm{HR}(t) = f(\mathrm{ACT}(t), \mathrm{ACT}(t\text{-}1), \mathrm{ACT}(t\text{-}2),\ldots). \quad (2)$$

Moreover, considering the correlation between heart rate and mean values of activity levels, the best correlation is obtained when activity levels are averaged over the two minutes preceding heart rate determination, so that, at last, the equation describing the temporal relation is as (3).

$$\mathrm{HR}(t) = f(\mathrm{mean}(\mathrm{ACT}(t), \mathrm{ACT}(t\text{-}1))). \qquad (3)$$

Thereafter, the relation (3) is detailed using a statistical and mathematical analysis on experimental data. Given that heart rate values depend on posture, whatever the level of activity, the modeling data set is cut, according to the observed posture, into three subsets, which are then analysed independently. To face the effect of saturation in the values of heart rate with high activity levels, we decided to study the relation between the aleas on heart rate and the logarithm of the mean activity level over the two minutes preceding any measure of heart rate. The results, presented on figure 7, show a quasi-linear relation between these two parameters. We must note that there are only few data available for moderate to high activity levels, that is over a value of 1.5 for the logarithm, and even less when the person is lying down. A linear regression over average value gives an estimate of the parameters linking activity level and heart rate. The slope and ordinate values of the linear fit increase with the effort required by the posture. The variability of heart rate values can also roughly be described as a linear function of the mean activity levels, whatever the posture. Considering now a given posture and activity level, aleas on heart rate are distributed along a normal curve, as shown on figure 8.

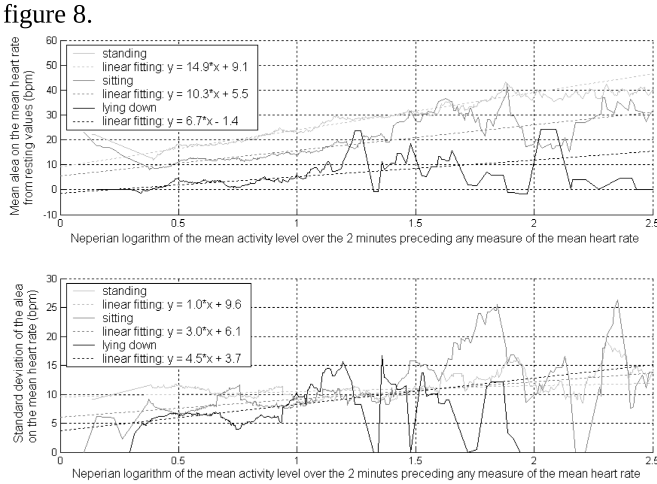

Figure 7. Mean (at the top) and standard deviation (at the bottom) of aleas on the mean heart rate according to the mean values of activity level over the two minutes preceding any measure of the mean heart rate

The simulation of heart rate values consists of: (1) the generation of mean resting values given sinusoidal features for the circadian rhythm (one value every minute); and (2) the introduction of aleas on these values according to the posture and the mean activity level over the two minutes preceding any heart rate value. These two pieces of information indicate those linear features that should be considered for the relation between heart rate and activity, so that we can get the appropriate values for the mean heart rate and its standard deviation. These characteristics (mean, standard deviation) are used to randomly generate a value of alea according to the corresponding normal distribution. Finally, the values of heart rate are calculated by adding circadian features of heart rate variations for the "simulated" subject.

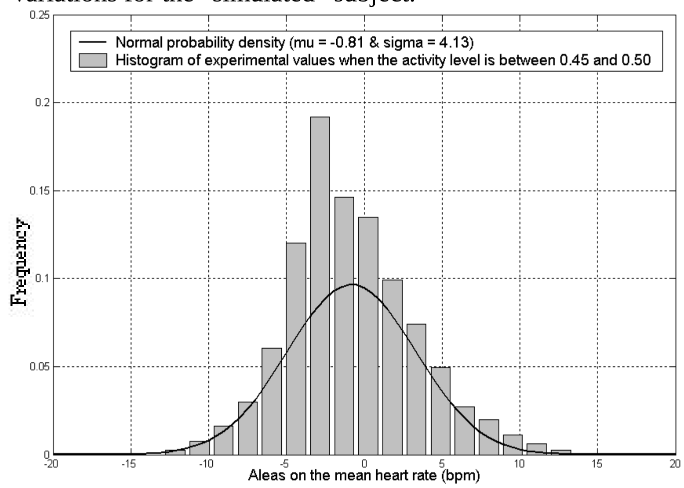

Figure 8. Distribution of aleas on heart rate values when the person is lying down and the activity level is between 0.4 and 0.5.

A sample of the sequences of data produced by the model of simulation is shown on figure 9.

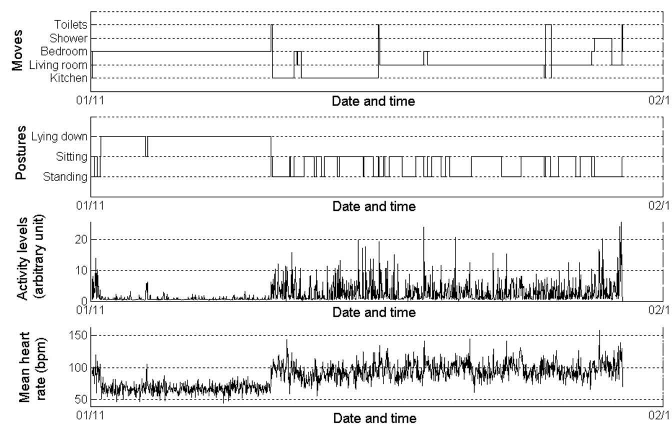

Figure 9. Sequences extracted from the simulation data set

VI. Experimentation and operational validation

This paragraph is dedicated to the experimentation of the model of simulation described above, once implemented using Matlab. The context of experimentation and validation, the results of experimentation, and their validation are presented successively.

A. Context of experimentation and validation

The sequences of data produced by the implementation of the simulation model (moves, postures, activity levels, and heart rate) need to be validated by comparison with time-series observed on a real system, which were unfortunately not available in our context. Concerning moves and postures of a person at home, we do not have any experimental record, so that the validation of the corresponding times series is performed by face validity, with knowledgeable people. In the case of the two remaining parameters (activity levels and heart rate), we have put aside for operational validation a data set quite close to our experimental conditions, called validation data set (see IV.C.1). The idea is then to rearrange these data to get new sequences that are suited to the context of experimentation. Since the values of these parameters are closely related to moves and postures of the person, this manipulation of the validation data set is made according to the sequences of moves and postures produced by the simulation (real records of moves and postures are not available). Thus, it is relevant to compare simulated and so called pseudo-validation data sets.

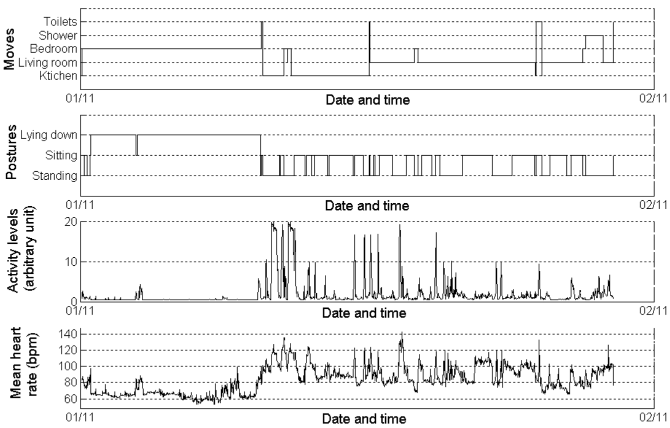

Figure 10. Sequences extracted from the pseudo-validation data set

The method used to rearrange the validation data set is based on the annotation of the pairs of the validation data set (activity level, heart rate) with the type of movement observed at the same time (lying down, sitting, standing, or walking). To get ride of the experimental subjects' specificity, the values of heart rate are normalised according to the sinusoidal circadian rhythm defined for each subject, so that we now have real sequences of (activity level, heart rate alea) pairs. Then, an appropriate pair is selected from these sequences every minute throughout the duration of the simulated sequences of moves and postures setting down the experimental context. Considering all real pairs available (activity level, heart rate alea), the criteria involved in the selection of relevant pairs at a given time are: (1) the type of movement associated with the real pair of values matches the one observed according to the simulated values of moves and postures at the current time; (2) the time of the real record is quite close to the current time; (3) considering the values already selected for the pseudo-validation data set and the real data, the activity levels over the two minutes preceding the current time or the time associated with the real pair considered (if existing) are close. A pair is then randomly chosen among all relevant pairs selected. However, once a pair is selected, the next pair in the real time-series is preferably selected, if relevant, in order to get as many real sequences of data as possible. Finally, the values of heart rate are calculated by adding the circadian features of the "simulated" subject. A part of the pseudo-validation data set obtained in that way is shown in figure 10.

B. Discussion about the quality of the simulation

The assessment of the quality of the simulation results is based on the comparison between the simulation and pseudo-validation data sets (figures 9 and 10). Before any objective approach using statistical tests and procedures, a first step is to make a subjective judgement on the overall aspect of the signals. As already noticed, the validation of the sequences of moves and postures is performed by face validity, with knowledgeable people. Considering activity levels and heart rate, one can immediately notice a higher variability in the simulated values than in the validation ones. However, the distributions of activity levels in both cases of simulation and validation look similar, due to the way in which these values were modeled. That means that features other than statistical characteristics of activity levels according to the movements of a person may have been integrated in the model to get a more realistic aspect for the simulated sequences. An intuitive idea is the need for information about the temporal arrangement of these sequences. As a matter of fact, the validation data set is preferably built by including real data that follow one another in time in the experimental sequences, rather than individual, randomly chosen, pairs of values. Figure 11 shows the aspect of a validation data set that could be obtained without this criterion for the rearrangement of the experimental data, that is without any temporal constraint. The aspect of these signals is much closer to that of the simulated data, especially when looking at the heart rate variability (directly related to the activity level). This confirms the need for a temporal constrainst in the modeling of activity levels to get more realistic sequences of data. In the next paragraph, we will discuss the really need for taking into account the temporal component.

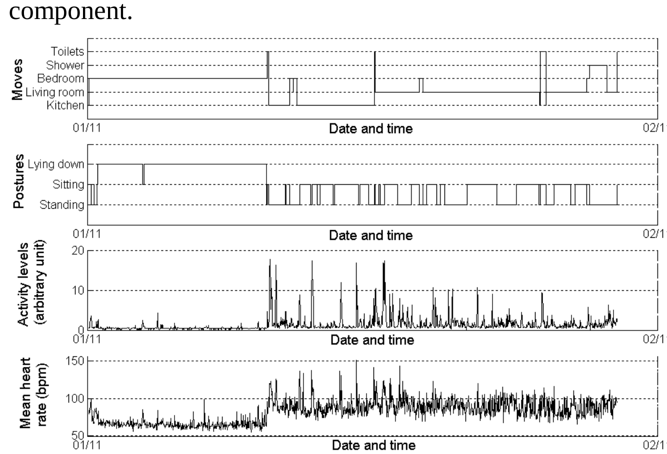

Figure 11. Sequences extracted from the pseudo-validation data set obtained with an individual and random selection of pairs (activity level, heart rate)

VII. DISCUSSION, CONCLUSION AND PERSPECTIVES

This paragraph is built around three major points: (1) the validity of the experimental data set, (2) the simulation model building, and (3) the results of its experimentation.

The experimental data used for the simulation process may not be completely satisfactory since they have been recorded from a restricted class of people: young, normotensive, and healthy, male or female, between 20 and 30 years old. Furthermore, this class of individuals does not fit the one targeted by home health telecare projects, that is elderly people living on their own. Moreover, the information about the activities performed by the subjects at any time are given by subjective annotations written down by the subjects themselves. Then, for instance, activities like “walking quickly” and “walking slowly” may be difficult to discriminate in the same manner for every subject. Finally, data corresponding to activities of high intensity and for the four minutes following their end has been removed. That might not be enough in some case when the subject is particularly affected by a hard task. However, it was not desirable to cut down many experimental data. All these reasons may already bias the quality of the model of simulation for the intended purpose.

Other remarks can be made on the model building itself, especially concerning the simulation of heart rate values. Activities and postures have been considered as the main factor influencing the heart rate values. That is probably true in general, but many other factors may have an influence as well, such as many external factors (outside temperature, use of medicines, stress, external events such as phone ringing or door banging, etc.) or the organism productivity (vegetative activity for instance). The methods for estimating the circadian rhythms and normalizing the heart rate values are also far from perfection given the underlying inaccuracies and assumptions (threshold on the activity level to define the resting heart rate values, adjustment of the heart rate values according to the posture, assumption of a sinusoidal circadian rhythm).

In spite of the imprecision described above, the results produced by the model are not so inappropriate from the point of view of the simulation’s purpose. Indeed, the sequences of simulated data are expected to be used for the study of mean- to long-term critical trends in a person’s behaviour, that is a “high level” analysis of temporal data. For instance, preserving the statistical properties of time-series while respecting the shape of signals over a day might be enough to complete such a study. The aim is not to simulate precisely the outputs of sensors, as observed on the real system, but to provide large sets of data that are relevant for the decision-making purpose. As already mentioned in this paper, a major point is to well define the purpose and context of use of the simulation results. We then need to refine the purposes of the decision-making process in order to precise the type of temporal information that is really required within the time series. Then we will know whether or not we need another cycle to refine the simulation process so that it generates more appropriate data sequences.

ACKNOWLEDGMENT

I would like to acknowledge the Ministry of Foreign Affairs (France) and the Ministry of International Relations (Québec) for providing their support within the context of the Quebec French HIT project (“Habitats Intelligents pour la Télésanté”). I am particularly grateful to Pr. Bernard Lefebvre for his welcome, good advice, and training during several stays at UQAM in the framework of the HIT project. The authors would also like to thank the French Ministry of Education and Research for their financing of the TISSAD project, as well as the township community “La Métro” for their financing in building the HIS (“Habitat Intelligent pour la Santé”).

REFERENCES

[1] V. Rialle, F. Duchêne, J. Demongeot, N. Noury, and L. Bajolle, “Health "smart" home: Definition and applications of Information Technology to patients at home,” *Telemedicine Journal and E-Health*, vol. 8(4), pp. 395-410, Winter 2002.

[2] R.E. Shannon, “Introduction to the Art and Science of simulation,” in *Proc 30th Winter Simulation Conference*, Washington, D.C., United States, pp. 7-14, 1998.

[3] W.D. Kelton, “Perspectives on Simulation Research and Practice,” *ORSA Journal on Computing*, vol. 6, pp. 318-328, 1994.

[4] L. Chwif, R.J. Paul, “On simulation model complexity,” *Proc 32nd conference on Winter simulation*, Orlando, Florida, pp. 449-455, 2000.

[5] T.I. Ören, “Simulation: Taxonomy,” in *Systems and Control Encyclopedia*, M.G. Singh, Ed. Oxford (England): Pergammon Press, 1987, pp. 4411-4414.

[6] R.G. Sargent, “Validation and verification of simulation models,” in *Proc 31st conference on Winter simulation*, Phoenix, Arizona, United States, pp. 39-48, 1999.

[7] L.R. Ramos, E.J. Simoes, M.S. Albert, “Dependence in activities of daily living and cognitive impairment strongly predicted mortality in older urban residents in Brazil: a 2-year follow-up,” *J Am Geriatr Soc*, vol. 49(9), pp. 1168-1175, 2001.

[8] K. Cameron, K. Hughes, K. Doughty, “Reducing fall incidence in community elders by telecare using predictive systems,” in *Proc 19th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, Chicago, USA, Oct.-Nov. 1997.

[9] R. Washburn, K. Smith, A. Jette, C. Janney, “The Physical Activity Scale for the elderly (PASE): development and evaluation,” *J Clin Epidemiol*, vol. 52(4), pp. 153-162, 1993.

[10] P.H.P. Peeters, “Design criteria for an automatic safety-alarm system for elderly,” *Technol Health Care*, vol. 8, pp. 81-91, 2000.

[11] G. Williams, K. Doughty, K. Cameron, D.A. Bradley, “A smart fall and activity monitor for telecare applications,” in *Proc 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chang and Zhang Eds. Piscataway, NJ: IEEE Computer Society Press, 1998, pp. 1151-1154.

[12] H. Monod et M. Pottier, “Adaptations respiratoires et circulatoires du travail musculaire,” *in Précis de physiologie du travail, Notions d’Ergonomie*, 2nd ed., J. Scherrer et al., Ed. Paris: Masson, 1981, pp. 159-204.

[13] G. Virone, N. Noury, B. Lefebvre, J. Demongeot, “Modeling and Computer Simulation of Human Physiological and Physical Activities with Direct Applications,” *submitted at IEEE Trans Biomed Eng*, 2003..

[14] L. Bourdon, A. Buguet, M. Cucherat, M.W. Radomski, “Use of a Spreadsheet Program for Circadian Analysis of Bilogical/Physiologial Data,” *Aviation, Space, and Environmental Medicine*, vol. 66(8), pp. 787-791, August 1995.

[15] F. Halberg, “Chronobiology,” *Ann Rev Physiol*, vol. 31, pp. 675-725, 1969.