Objectivity and subjectivity in the analysis of 3D movement data

II. The design of the space-time windows and the subsequent analysis

P. Loslever, F. Barbier, S. Bouilland, E.M. Laassel
Laboratory of Human and Industrial Automatics, Mechanics and Informatics of Valenciennes University
LAMIH, Valenciennes, France
Pierre.Loslever@univ-valenciennes.fr

***Abstract*— This part II proposes two examples where the space-time windowing is useful: gait and load lifting. Giving the main importance of the windowing and graphing result (for a very first analysis of numerous multidimensional signals), the discussion tackle subjective aspects with and without windowing.**

Keywords-component; data chacterization; data summarizing; gait; load filfing, fuzzy windowing; graphics

I. Introduction

Let us suppose an experiment or observation design where one (or several) 3D measurement systems is (are) used, e.g. a 3D imaging system and/or two force-plates (in some cases, physiological signals can present also). Suppose there are F factors (individual factor belongs to the factor set) and V time variables (may be recorded at different sampling frequencies), yielding a total of N multidimensional signals with V components. Here are two generic examples:

Gait. An observation design is organized in order to evaluate a rehabilitation program for a given pathology. There are F=2 factors, the first one indicating the patient i (i=1, …I, e.g. I=30), the second one indicating the week w (w=1, …W, e.g. W=10). There are V variables, V_1 coming from a 3D imaging system (recorded with the sampling frequency fs_1), V2 from forces plates (with $fs_2=fs_1$) and V3 from EMG measures (with $fs_3>> fs_1$). Globally, there are N=I*W multidimensional signals with $V=V_1+V_2+V_3$ components, e.g. N=300 and V=50.

Load lifting. An experiment design is organized in order to compare L lifting modes, e.g. free and isokinetic, L=2, with L’ levels for each lifting mode, e.g. L’=3. Thus F=3 and, with a full experimental design, N=I*L*L’ (e.g. N=30*2*6=360). We suppose that there are $V=V_1+V_2$ variables, V_1 corresponding to the kinematic variables, V_2 to the kinetic variables, e.g. V=100+100=200.

More generally, in addition to the difference between experiment and observation designs, many other distinctions can be made, i.e. the presence vs. absence of

1) cyclic aspect (in the normal gait case, there is periodicity),

2) identical duration for all the N the multidimensional signals,

3) identical sampling frequency for all the N signals,

4) identical uncertainty for all the V variables, e.g. with a 3D imaging system, the 3 directions yields different uncertainty levels.

These 4 points are important and will mainly condition the design of space time windows and the subsequent analysis. The next 2 sections focus on different distinctions in the windowing (II) and in the subsequent analysis (III). Section IV proposes some actual examples (from studies performed by our laboratory). The final section is a general discussion about our way to consider data analysis compared to some ways encountered the literature.

II. Different distinctions in the windowing

In the following we suppose that the time data are recorded while respecting the sampling theorem of Shannon with a good filtering process). Moreover we give only different ways for distinguishing the windowing according to the methodological point of view (actual examples will be given in section III). We will also suppose that the main aims of the statistical analysis are to show together 1) the influence of the F factors onto the V variables, and 2) the connections between the V variables.

A. *Space windowing (SW)*

1) Distinction according to the considered data set. Here are three main options:
* SW adapted to each variable v (v=1, …, V). The data set corresponds to all the time values of the N recording situations, yielding the range [min(v), max(v)],
* SW adapted to each record, i.e. to each multidimensional signal n (n=1, …, N). The range is [min(v,n), max(v,n)],
* SW adapted to each individual i (i=1, …, I). Such a case occurs when there are several multidimensional signals for a given individual. The range is [min(v,i), max(v,i)].

If these 3 main options are combined, many other ranges can be imagined, e.g. when 2 variables v and v' are obtained using the same measurement device and have quite identical ranges [min(v),max(v)] and [min(v'),max(v')], the windowing can be adapted to the overall range [min(min(v),min(v')), max(min(v),min(v')].

2) Distinction according to the confidence you have in the min and max values

* with a high confidence level, SW is adapted to the full range, i.e. to one of the 3 full ranges mentioned above,
* with a low confidence level, k% and k'% of values situated on the left and right sides on the magnitude histogram are removed first (parallel with the classic notion of trimmed arithmetic mean in order to have a more robust indicator than the arithmetic mean.

Here again many other options can be imagined e.g. k%=k'%, if the magnitude histogram is symmetric or k%=0 and k'%>0 if the minimum value is 0 (when the variable is the magnitude of a force, for instance).

3) Distinction according to the mathematic cutting criterium

* the criterion is spatial only, e.g. S space windows with identical width,
* the criterion is temporal only, e.g. S space windows with identical frequencies.

Here again, many other options exist, e.g. a subset of space windows correspond to negative values and a subset of SW to positive values (for forces, speed, accelerations, …).

4) Distinction according to a value of variable belongs to one or more than one windows

* the first case corresponds to the classic set theory,
* the second case corresponds to the fuzzy set theory. The membership function can have many patterns: triangular (either symmetric or not), trapezoidal (either symmetric or not), Gaussian, …

In addition to these 4 option sets, some other options can be put forward, e.g. the number S of space windows.

B. Time windowing (TW)

Options look like those of space windowing. A typical distinction with time is due to the absence or presence of cyclic phenomena, Fig 1

III. SUBSEQUENT ANALYSIS

Keeping in mind the notion of statistical analysis path presented in part I, it is worth noting that the space-time windowing yields homogeneous data, i.e. membership value averages (mva), generally speaking, and frequencies in the particular case of a crisp windowing, Thus, the classic coding stage inherent to a multivariate approach is not required. The data shaping stage can yields a two entries-table where the R rows correspond to all the time windows of the N signals and the C columns to the space windows of the V variables.

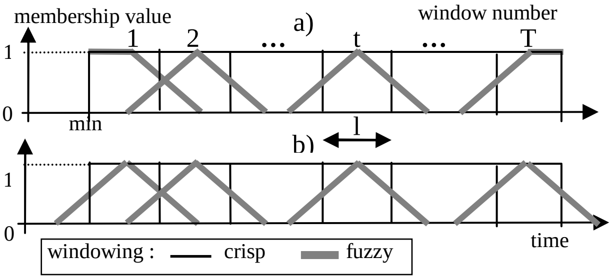

Figure 1. windowing according to absence (a) or presence (b) of cycle.

Many other tables can be considered from this initial table and the data included in such tables can be shown graphically using gray levels [1], see examples in IV.

With the main aims of the statistical analysis presented in section II, the next stage can be the multiple correspondence analysis (MCA) [2]. The principle is identical to that of the principal component analysis (PCA). The main difference is in the distance choice: in PCA there 2 distance models, one for the rows, one for the columns, in MCA there is a unique distance model. It is inspired from the chi-suare metric, see [3] for comparisons between PCA and MCA. Some other specific multivariate methods can be used, such as the hierarchical clustering or the histogram modal analysis in order to find automatically if the N multidimensional signals present classes [4]. Let us consider the two generic examples suggested in the introduction.

IV. SOME EXAMPLES

Gait. We which to generalize a study about the clubfoot gat performed in a clinical context [5]. The variables come from 3D kinematic data (Vicon), kinetic data (2 forces-plates) and physiological data (EMG). One of the main interest for considering here this example is due to the presence of the two sides of the gait, i.e. there are 2 ways to consider the set $\mathbb{V}$ of variables (notation: V=card($\mathbb{V}$), where the card function denotes the number of elements of the set $\mathbb{V}$):

1) a step is described using two sets of variables, one for left side, one for the right side, i.e. sets $\mathbb{V}_L$ and $\mathbb{V}_R$ with $V_L=V_R$ Thus the total number of variables is $V= V_L + V_R=2*V'$ (in [5], V'=37),

2) a step is described using a set of two observations, one for each side, thus the total number of variable is V=V'.

The main advantage of the first approach is the possibility to analyze more carefully the coordination problem and the main advantage of the second approach is the possibility to show more easily the differences between the two sides. In fact both approaches should be tested but only the second one is presented here.

Another main reason for considering this gait example is the presence of EMG data in which one have a poor confidence. Thus, instead of considering EMG data through variables with quantitative scales, one can 1) use these data as qualitative and 2) give these data an illustrative status in MCA.

Given this qualitative aspect and due to the high number of variables (V=37) and the low number of individuals (I=12), only S=3 fuzzy space windows are considered, see Fig.2 for two examples. The time windowing is performed using the principle of Fig. 1.b with T=51: the window n°1 is situated around the beginning of the step and the window n°51 round the beginning of the next step.

Given this space-time windowing, a membership value average (mva) is computed for each space-time window [3] and all the mva are placed within a table with C=37*3=111 columns and R=12*2*51=1224 rows. When one read the MCA output, we first focus on the space windows, i.e. the most discriminant ones, their relationships and windows with "curious" positions, which can correspond to measurements with imperfection [3]. Then we focus on the time windows, which allows to show intra and inter-individual differences (for each individual and each side, the 51 time windows are joined yielding a time trajectory. For instance, Fig. 3 shows two individuals with rather low and rather high differences between the two legs. Other possible uses of MCA are presented in the discussion.

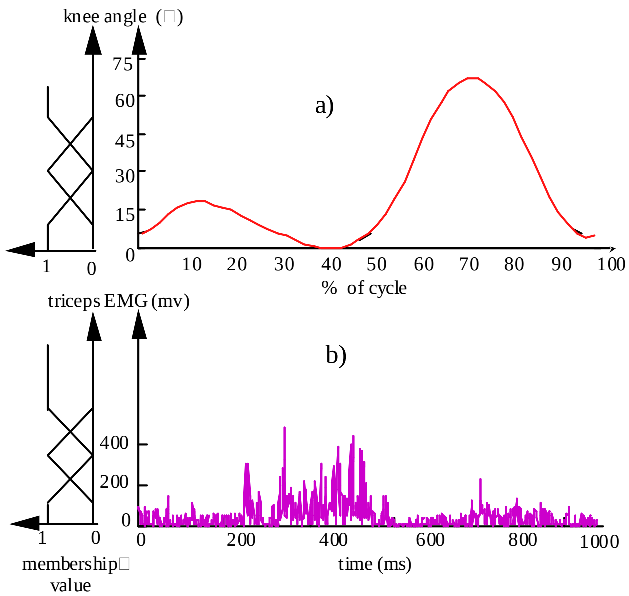

Figure 2. Two space windowing patterns in the gait study

In this example, the clinical context involves to be interested to each patient. Thus there is a picture for each of the I individuals, Fig. 3 (see [6] for other multivariate picture). In experimental context, in most cases, one which to show results more globally, e.g. to show the influences of the experimental factors onto averages computed over the individuals. Such a case is presented below.

Load lifting. The variables come from 3D kinematic data (obtained with SAGA3, designed in the LAMIH), kinetic data (one force-plate, extensometric sensors placed on the load handle and a specific biomechanical model with ADAMS) [3]. There are V=972 and I=13 "normal" subjects (without back pain). One of the main interest for considering here this example is in the very large number V of variables. Once obtained the mva [3], several MCA were performed:

* MCA without taking into account the time factor (its chronologic aspect). Thus 4 MCA are performed 1) with kinematic variables a) in the sagittal plane, b) in the frontal plane; 2) kinetic variables, 3) kinematic and kinetic variables that played a main role in the 3 previous analyses (V'=89 variables);

* MCA with taking into account the time factor with the 89 variables.

The factor planes being rather complex, a specific graphic method was designed for showing the time excursion of the individual data: fuzzy time indexed magnitude histograms based on grey levels. An example is shown Fig. 5. Giving that at t=0, the individual has to pull up the handle of the isokinetic test machine (Cybex), the time range before 0 corresponds to the initial state of the individual. With such a graphic principle, the lower the variation, the darker the picture where the data fall, the more evident the white line becomes. Fig. 5, shows to main results: when the speed imposed by the isokinetic machine increases

* before the individuals reach this speed, the maximal speed decreases,

* the inter-individual dispersion decreases. See [3] for further details.

The main aim for considering these two examples was to show that a multivariate and multifactor data analysis require several stages, each stage with many choices, thus a part of subjectivity. Let us discuss this aspect.

V. DISCUSSION AND CONCLUSION

First let us discuss the reason for cutting the range of quantitative variable (thus yielding an ordinal scale) instead of keeping the initial scale. Three main reasons can be put forward. The first one is the well know performance of the magnitude histogram which also requires a space windowing. This statistical tool allows to show whether there are 1) classes (thanks to the mode, given that classes can correspond to bad data), 2) Laplace-Gauss presence (which is "more or less" required by most parametric statistical tests), 3) abnormal too low or too high values (thanks to the sides of the histogram), which could correspond, here again, to bad data (with a technical and/or human origin). The experiences of the authors is that with a large number of large multidimensional signals (N>>50 and V>>30) often produces by studies using 3D measurement systems, the percentage to have doubtful data is rather high (more than 50% !). Thus performing a windowing and then analyzing the mva through a multivariate method will allow us find quickly doubtful data (e.g. with MCA, for which rows and which columns such data are present)

The second reason is that if averages have to be computed, the space windowing loss less information (see part I). Let us give a simple example:

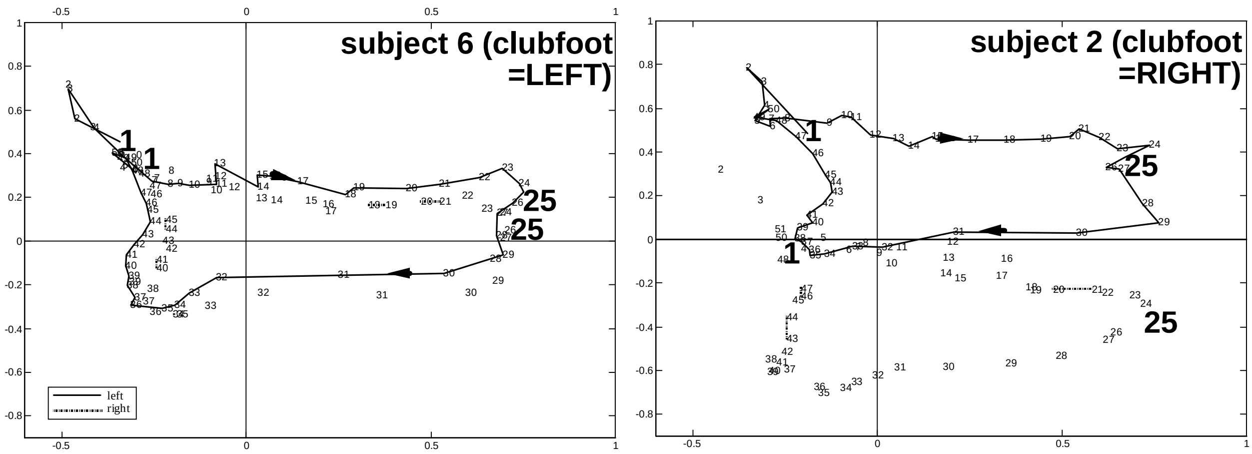

Figure 3. example of time trajectories yielded by the MCA of gait data. Case of patients with a) small and b) large differences between the 2 legs

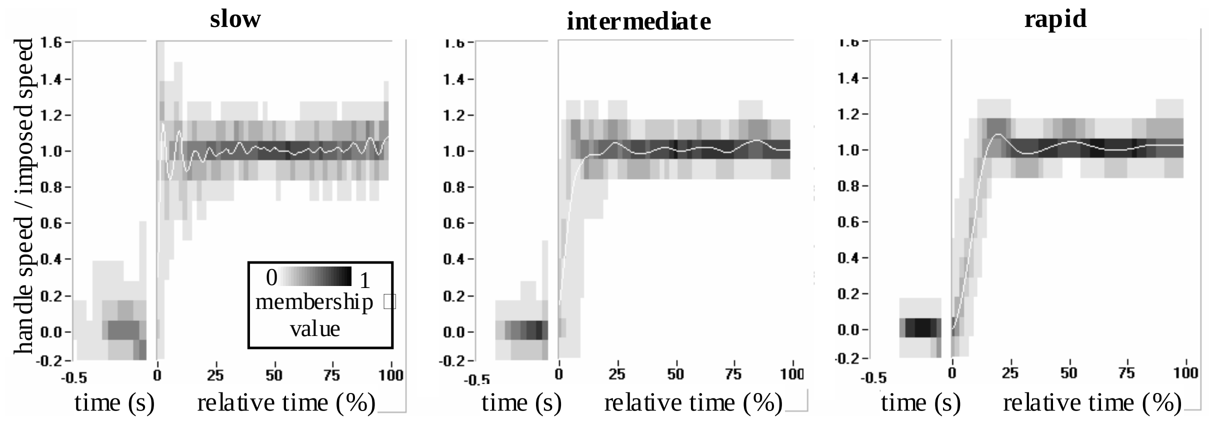

Figure 4. example of mva represented using gray level (for each space-time window, the mva is ciomputed when averaging over the individuals; the white line correspond the the "classic" average, i.e. computed from the signals)

if the range of a variable is [0, 100] and if one has to summarize 2 values, e.g., 0 and 100, the classic arithmetic mean gives 50, while a windowing with 3 windows gives the membership value triplets {1, 0, 0} and {0, 0, 1} respectively, which yields the mva set {1/2, 0, 1/2}, and thus keeps the idea that one value was rather low, one was rather high.

A third reason is the possibility to show complex relationships, whatever all the variables are quantitative or some are quantitative and some are qualitative (for instance presence/absence of a given action).

Keeping in mind these 3 reasons, it is worth noting that the multiples choices inherent to the space windowing can be seen very subjective. From our experience, here is a main rule: for a very first analysis and if the number V of variables is large (V>20), the number S of space windows can be 3. This choice seems a good compromise between the accuracy and complexity reasons, as 5% is a compromise between risk of types I and II in hypothesis test method. This "classic" value could be seen as very subjective also, as well as the choice between a parametric vs. non parametric version of a test (if the choice is not justified, as it is often the case).

If V<20, S and the membership function patterns can be supplied by the domain expert (biomecanician, physician, ergonomist, …). Such an approach seems less subjective than using the principal component analysis (PCA) with its normalized version (the V variables are standardized using the arithmetic and standard deviation) without justifying such a choice. Once main results have been found using our exploratory approach, they can be tested using inference approach.

References

[1] J.. Bertin, Le graphique et le traitement graphique de l'information, Paris: Flammarion, 1977.

[2] J.P. Benzecri, Correspondance analysis handbook, NY: Marcel Dekker, 1992.

[3] P. Loslever, S. Bouilland, "Multidimensional signal exploration using multiple correspondence analysis. An example of a load lifting study"., IEEE transactions on Neural systems and rehabilitation engineering, vol. 11(3), pp. 301-310, 2003.

[4] P. Loslever, F.X. Lepoutre, A. Kebab, H. Sayarh, "Descriptive multidimensional statistical methods for analyzing signals in a multifactorial biomedical database", Med. Biol. Eng. Comp, vol. 132, pp. 133–146, 2001.

[5] P. Loslever, E.M. Laassel, C. Guibal, A. Dimeglio, "Organisation, caractérisation et analyse multidimensionnelle des données de la marche: application à la marche du pied bot varus equin", Innovovation et technologie en biologie et médecine, vol. 24, pp. 1–12, 2003.

[6] P. Loslever, F. Barbier, "Multivariate graphical presentation for gait rehabilitation study", Gait and Posture, vol. 7, pp. 39-44, 1998.