

# A geometric model defined by a family of Splines for modelling complex systems

Y. Villacampa, F. G. Navarro-González, P. Cerdán & M. Cortés  
*Department of Applied Mathematics, University of Alicante, Spain*

## Abstract

The study of natural systems implies considering new modelling methodologies that are able to produce different relationships to those described with mathematical functions, which are derived from the geometry of Euclid.

In this paper, the authors propose a geometric model defined by families of cubic Splines, which are the basis for a definition of a numerical methodology for the study and modelling of complex systems.

Geometric models are applied in specific cases for types of relationships. One feature of the model is that the polynomials are not represented by their coefficients, because they could be highly dependent of small variations in their values, as it is analyzed in the article. The polynomials will be represented by their values at points considered in their ranges of definition, which will be called nodes.

For each variable, a Spline generated from  $k$  cubic polynomials is defined, so the first objective is the analysis of a family of Splines determined by a set of polynomials.

Finally, the geometric model is determined on practical examples from experimental data and the advantages of using the new methodology, based on the identification of Splines by their values in a number of points, are discussed, compared with the usual definition from the polynomial coefficients.

*Keywords: modelling, complex systems, Splines.*

## 1 Introduction

In the scientific literature there are different symbolic methodologies that we can use to obtain mathematical equations which express a relationship  $y = f(x_i)$



and are generated from experimental data of the variable  $\{x_i\}$ ,  $\{x_{ij}\}$  [1–7]. In this paper a new numerical methodology that could be applied to complex systems is developed, with its extension to  $n$ - dimensional variables. This methodology could be used together with others methodologies.

In the study and modelling of Complex Systems where the number of variables and relationship can be higher, as natural phenomena are, it is very important to obtain numerical methodologies that help us to modelling relations between variables, and to get information about the studied phenomena.

In the nature phenomena, that is, physical world and life, it can be possible that mathematical functions can't solve all the problems because the mathematical functions arise from the geometry of Euclid and its geometric properties in the plane and space.

The linear models are generated in the same way, being different the methods in finding nonlinear models

Furthermore, software [1, 2], may only generate a nonlinear model in each execution of the program; as long as a type of mathematical equation is proposed, which depends of parameters, that are obtained. In the methodologies developed and computationally implemented in the articles [3, 5], families of models can be generated in each execution of the program, by performing transformations to be analyzed as linear models. Moreover, in [6] families of models according to a type of nonlinear models in their parameters have been generated. Related to the numerical methodologies in the articles [4, 7], a method using  $n$ - dimensional geometric models of finite elements has been developed, which it could be used in a complementary way to the methodology generated in this article.

## 2 Polynomial regression

Given the equation  $y = f(x_i)$  and a set of experimental data, it is possible to generate a polynomial regression of degree  $n$ , in particular  $n=3$ , and then to obtain the best polynomial that fits the data. Likewise, a set of polynomials which define a cubic Spline  $\delta x_i = \{P_i^j\}_{j=1}^{k_i}$  could be determined, and that best represent the relationship  $y = f(x_i)$ . In these cases the method calculates polynomials coefficients. This process can be made using the method of least squares or minimizing any error function that could be defined on the set of feasible polynomials. If we choose to express the polynomial using their coefficients, in some cases, stability problems could appear in the generated models, because small variations on the coefficient can produce large fluctuations in the values of the polynomials. These fluctuations can distort the obtained model. To better understand the importance of generating models without making them dependent on the coefficients of the cubic polynomials, in the following figures 1 and 2, two polynomials are shown, together with other polynomials obtained from small perturbations in their coefficients. As can be observed, the changes in the values of the polynomials can be large.

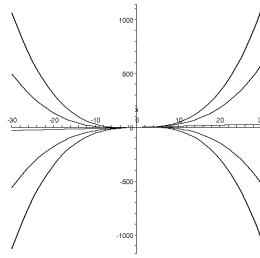


Figure 1: First example of small perturbations in the coefficients.

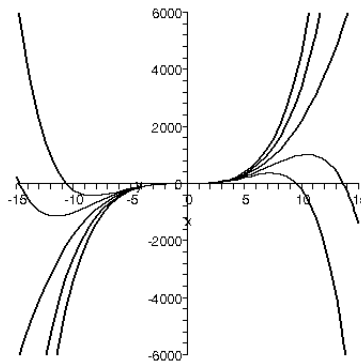


Figure 2: Second example of small perturbations in the coefficients.

For this reason, in this article the authors have studied a modification of the polynomial regression, identifying the polynomial not by its coefficients, but by its values on a finite number of points called nodes. The theoretical basis to generate models is to consider a geometric model defined by a family of Splines, identifying each of them by its value in a finite number of points

### 3 Geometric model

For each variable  $\{x_i\}$  a Spline defined using  $k_i$  cubic polynomials is considered. So, the first objective is to study a Spline  $\delta x_i$  defined on  $\{x_i\}$ , with a set of polynomials,  $\delta x_i = \{P_i^l\}_{l=1}^{k_i}$ .

Considering the variable  $x$ , the model is defined by  $y = \delta x$ , so  $\delta x = \{P_i^t\}_{i=1}^{k_i}$ . The cubic polynomials aren't represented by their coefficients but by their values on the points:  $\delta x \equiv \{v_i\}_{i=1}^m$ .

Let us consider the independent variable  $x$ , with values in the interval  $[-1, 1]$ . This interval can be split in two intervals, defining in each of them a polynomial. So, there are two cubic polynomials, one in the interval  $[-1, 0]$  and the other in  $[0, 1]$ . The Spline is always determined by a finite number of intervals in which the polynomial is defined. Although initially it starts with 2 intervals, subsequently the results are generalized to any number of polynomials. The number of intervals and therefore the number of polynomials, always implies a greater or lesser complexity in the geometric model and therefore in the model.

Initially in the interval,  $I$ , we have two polynomials defined as:

$$P = d_o + x(c_o + x_o(b_o + xa_o)) \text{ if } x > 0$$

$$P' = d_i + x(c_i + x(b_i + xa_i)) \text{ if } x < 0$$

If we identify the polynomials by their coefficients there would be eight variables, however we identify them by the values in the nodes.

### 3.1 Nodes

In the interval  $[-1, 1]$ , we choose a finite number of points called nodes.

In this methodology nodes are selected from the number of polynomials defined on the Spline

In a first discretization the interval is divided in two intervals  $[-1, 0]$  and  $[0, 1]$ , in which the polynomials are defined. The nodes are  $n_1 = -1$ ,  $n_2 = 0$  and  $n_3 = 1$ . Let be  $v_1, v_2, v_3$  respectively the values taken by the Spline at these nodes. From the family of all possible Splines generated by cubic polynomial, the first discretization gives only three nodes. A representation of the relationship  $y = f(x_i)$  is a function obtained from the experimental data  $\{x_{ij}\}$  and defined by:

$$Zm = \theta(v_1, v_2, v_3, x).$$

A discretization of the interval  $[-1, 1]$ , in  $n$  subintervals would give  $n+1$  nodes and the representation of the relationship will be defined by:

$$\theta(v_1, v_2, v_3, \dots, v_{n+1}, x)$$

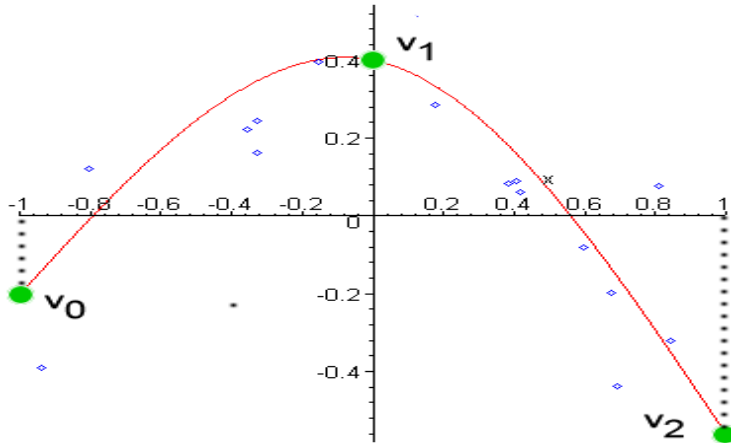


Figure 3: Nodes.

#### 4 Representation models of the relationship

The main objective of the geometrical model is to obtain a new numerical methodology to model systems using representation models. This requires defining an optimization problem that allows determining the values of the Spline in the nodes.

If the relationship is defined by  $y = f(x_i)$  and the following experimental data  $\{x_{ij}\}$  are known, for the model represented by:

$$Zm = \theta(v_1, v_2, v_3, \dots, v_{n+1}, x),$$

the error function to optimize is defined as:

$$E = \sum_i (Zm - z_i)^2$$

The next step is to calculate the values,  $v_i$  that minimize the error function. It will be necessary to determine initial values to start searching the solution of the system determined by the annulment of partial derivatives.

#### 5 Computational implementation of the geometric models

The input variables of the problem are the set of experimental data. The objective is to select the set of nodes  $\{n_i\}_{i=1}^{N+1}$  and to determine the values of the polynomials  $\{v_i\}_{i=1}^{N+1}$ . From these values, the final output of the algorithm is the set of coefficients,  $\{(a_i, b_i, c_i, d_i)\}_{i=1}^P$  with which are generated the Splines.

The scheme of the used algorithm is:

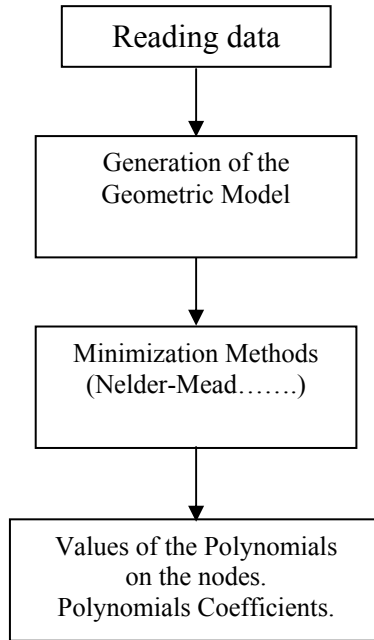


Figure 4: Scheme of the algorithm.

The geometric model starts with the number of Splines to be used, and the nodes  $\{n_i\}_{i=1}^{N+1}$  are set so that the number of points assigned to each interval is similar.

Obtaining polynomials from a set of values in the nodes  $\{v_i\}_{i=1}^{n+1}$  is performed by the following algorithm:

```

void CalculateSpline (double *a, double *b, double *c, double *d, double *v)
{
    int n=NumSplines;
    //System creation
    double e[n];
    double f[n+1];
    double g[n];
    double z[n+1];
    double h[n];

    for (int j=0;j<n;j++) {h[j]=XLim[j+1]-XLim[j];}
    //Initial values
    f[0]=1;g[0]=0;z[0]=0;
    for (int j=1;j<n;j++)
    {
        e[j-1]=h[j-1];
        f[j]=2*(h[j-1]+h[j]);
    }
}
  
```

```

        g[j]=h[j];
        z[j]=3*((v[j+1]-v[j])/h[j]-(v[j]-v[j-1])/h[j-1]);
    }
    f[n]=1;z[n]=0;e[n-1]=0;
    //System solution
    for (int k=1;k<=n;k++)
    {
        e[k-1]=e[k-1]/f[k-1];
        f[k]=f[k]-e[k-1]*g[k-1];
    }
    for (int k=1;k<=n;k++)          {z[k]=z[k]-e[k-1]*z[k-1];}
    c[n-1]=z[n-1]/f[n-1];
    for (int k=n-2;k>=0;k--)      {c[k]=(z[k]-g[k]*c[k+1])/f[k];}
    //Calculating the rest of parameters
    for (int j=0;j<n-1;j++)
    {
        a[j]=v[j];
        b[j]=(v[j+1]-v[j])/h[j]-(2*c[j]+c[j+1])*h[j]/3;
        d[j]=(c[j+1]-c[j])/(3*h[j]);
    }
    a[n-1]=v[n-1];
    b[n-1]=(v[n]-v[n-1])/h[n-1]-(2*c[n-1])*h[n-1]/3;
    d[n-1]=(-c[n-1])/(3*h[n-1]);
}

int RectGeneration()
{
    map<double,int>::iterator it;
    map<double,int>::iterator it2;
    //Sizing the window
    double window_size=(double) NumPoints/(double) NumSplines;
    XLim=new double[NumSplines+1];
    int interval_cont=0;
    int cont_lim=1;
    int cont=0;
    XLim[NumSplines+1]= pointsX [0].begin()->first;

    for (it=pointsX[0].begin();it!= pointsX [0].end();it++)
    {
        cont+=it->second;
        it2=it;it2++;
        if ((cont+it->second)>floor(cont_lim*window_size))
        {
            if (it2== pointsX [0].end())
            {
                XLim[cont_lim]=it->first;

```



```
                break;
            }
            XLim[cont_lim]=(it->first+it2->first)/2;
            cont_lim++;
        }
    }
    XLim[NumSplines]=( pointsX [0].rbegin()->first);
    interval_cont+=cont_lim;
    return interval_cont;
}
```

## 6 Applications

In the following examples, several geometric models have been determined and the representation models have been calculated, comparing the results with others methodologies. Furthermore, models whose equations are known have been selected, to compare the results with those obtained with the methodology developed in this article. Others models were generated from the experimental data.

The new methodology has been compared with others using data obtained from [www.statsci.org](http://www.statsci.org). The selected examples are the following:

### 6.1 Age and eye lens weight for rabbits in Australia

The European rabbit *Oryctolagus cuniculus* is a major pest in Australia. A reliable method of age determination for rabbits caught in the wild would be of importance in ecological studies. In this study, the dry weight of the eye lens was measured for 71 free-living wild rabbits of known age. Eye lens weight tends to vary much less with environmental conditions than does total body weight, and therefore may be a much better indicator of age.

The rabbits were born and lived free in an experimental 1.7 acre enclosure at Gungahlin, ACT. The birth data and history of each individual were accurately known. Rabbits in the enclosure depended on the natural food supply. In this experiment, 18 of the eye lenses were collected from rabbits that died in the course of the study from various causes such as coccidiosis, bird predation or starvation. The remaining 53 rabbits were deliberately killed, immediately after being caught in the enclosure or after they had been kept for some time in cages. The lenses were preserved and their dry weight determined.

Table 1: Description of model variables.

Variable	Description
Age	Rabbit age in days
Lens	Dry weight of the ocular lens in mg.



The primary source of data is [8]. They concluded that the weight of eye lens is a reliable indicator of the rabbit's age up to about 150 days. The deterministic component of their model is:

$$\text{Lens} = \alpha \exp \{ -\beta / (\text{Age} + \gamma) \}$$

whose representation is:

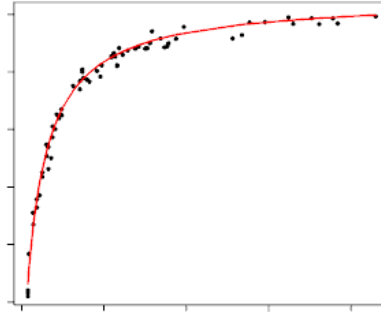


Figure 5: Age and eye lens weight for rabbits in Australia. (Theoretical model.)

Applying the new methodology, a model by a Spline with eight intervals and a determination coefficient  $R^2=0,987$  has been generated. Representing the spline, a graph similar to the theoretical model curve is obtained.

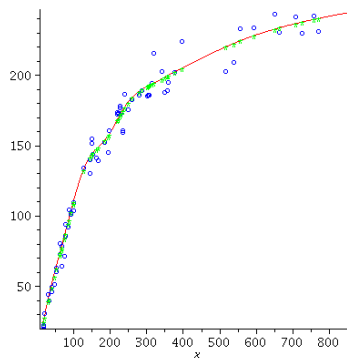


Figure 6: Age and weight of the eye lens for rabbits in Australia. (Numerical model.)

## 6.2 Blood sulphate in a baboon named Brunhilda

The observed responses are Geiger counter counts (times  $10^{-4}$ ) used to measure the amount of radioactively tagged sulphate drug in the blood of a baboon named Brunhilda after an injection of the drug [9, 10].

In the data source web, no theoretical model is suggested. Applying the new methodology, a model with a spline with eight intervals has been generated. The determination coefficient is  $R^2=0,995$ :

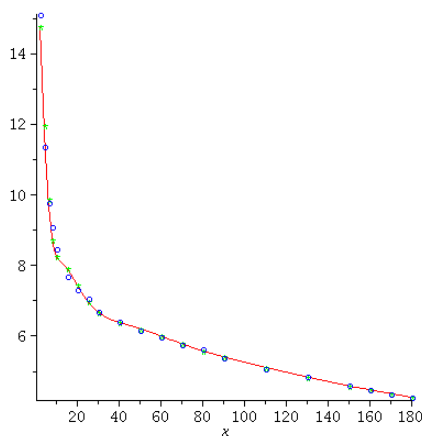


Figure 7: Blood sulphate in a baboon named Brunhilda. (Numerical model).

## 7 Conclusions

In this article a geometric model formed by a family of Splines has been defined, to obtain a new methodology for the study and modelling of Complex Systems. To do this, initially, a function of one variable has been considered, defining a geometric model and a new methodology to generate models has been proposed.

The results of the modelling methodology proposed in this article have been compared with those obtained by other methodologies. On the one hand it compares with models theoretically defined and on the other hand with models obtained from the experimental data.

In the future the authors will develop the new methodology, generalizing the methodology implemented in this article, to dimension  $n$ . For this, it would be necessary to analyze the resolution of the optimization problem with more complexity, in order to apply to model Complex Systems. This will involve having a methodology that will be used in a hybrid form with others methodologies to apply in higher Complex Systems.

## References

- [1] Splus. (1997). MathSoft, Inc. Seattle, Washington. USA.
- [2] Spss. (1999). Inc. Software. Chicago. USA.
- [3] Villacampa, Y.; Cortés, M.; Vives, F.; Usó, J.L.; Castro, M. A. (1999). A new computational algorithm to construct mathematical models. *Advances in Ecological Sciences. Ecosystems & Sustainable Development. Serie II.* 185312 687X – pp. 323–332.



- [4] Villacampa, Y.; Navarro-González, F.J.; Llorens, J. (2009). A geometric model for the generation of models defined in Complex Systems. A geometric model for the generation of models defined in Complex Systems. 978-1-84564-194-8 – pp. 71–82.
- [5] Cortés, M.; Villacampa, Y.; Mateu, J.; Usó, J.L. (2000). A new methodology for modelling highly structured systems. *Environmental Modelling & Software* (1364-8152). V.15, nº 2, pp. 461–470.
- [6] Verdú, F and Villacampa, Y. (2008). A Computational algorithm for the multiple generation of nonlinear mathematical models and stability study. *Advances in Engineering Software*. Vol. 39, Issue 5, pp. 430–437.
- [7] F. J. Navarro-González, Y. Villacampa A new methodology for complex systems using n-dimensional finite elements. *Advances in Engineering Software* 48 (2012) 52–57.
- [8] Dudzinski, M. L., and Mykutowycz, R. (1961). The eye lens as an indicator of age in the wild rabbit in Australia. *CSIRO Wildlife Research*, 6, pp. 156–159.
- [9] Jennrich, R. I., and Bright, P. B. (1976). Fitting systems of linear differential equations using computer generated exact derivatives. *Technometrics* **18**, pp. 385–392.
- [10] Jennrich, R. I. (1995). *An Introduction to Computational Statistics*. Prentice-Hall, Englewood Cliffs, New Jersey. Section 8.2.1.

