Problems caused by biased data in models of catalog adjustment

F.J. Marco*, Martínez M.J.**, López J.A. *

***Universidad Politécnica de Valencia. IUMPA. Dept. Matemática Aplicada. Camino de Vera S/N. 46022 Valencia. Spain.*

**Universidad Jaume I. IMAC. Dept. Matemàtiques. Campus de Riu Sec. 12071 Castellón. Spain*

marco@mat.uji.es, mjmartin@mat.upv.es, lopez@mat.uji.es

Working with catalogues requires an increasingly needed precision. The comparison between different catalogues has always been a powerful tool for their. To this aim contribute both the sophistication / refinement of the models of adjustment and the most careful treatment of the data. We introduced a mixed method for the study of the relative orientation between the catalogues Hipparcos and FK5 (Marco F. et al. ,2004, A&A. 418), obtaining results comparable to those of other authors. We also introduced the VSH model of first order (rotation+ eformation) but the coefficients were obtained by a simple least squares adjustment because we were interested only in the relative rotation between the catalogues (The corresponding calculations for the spin were published in Marco F. et al. 2009. Pub. of the Astron. Soc. Pacific. 121.). The good distribution of the data (density and homogeneity), makes reliable the coefficients even applying a simple least square method. In particular, the value of $d_{1,0}$ was high enough as to be taken into account in later studies. And that was all. Very recently,.Mignard, F.,& Klioner, S. 2012, A & A, 547, A59, have published a paper to introduce the use of VSH (of arbitrary order) in the comparison of catalogues. After reading it, we verified that their $d_{1,0}$ value was practically the same that we obtained. We considered, then, to re-do the calculations with our method of 2004 (and 2009), being the results corroborated. We can affirm that our method is compatible with the developments in VSH given by Mignard et al.and it makes useful contributions when the catalogues are not homogeneous. For that reason, we sent for publication a paper in this sense. Much to our surprise, the manuscript was rejected after a week with poor (and someone of them false) arguments showing with this that the reviewer had little and badly understood our methodology and our aims. This bad interpretation fits perfectly with the title of the present poster and that is why, avoiding telling the magazine to which it was sent, we believe it is appropriate to reply each argument that served to reject our work. A summary of the our paper is presented in the poster "An accurate and stable mixed method to obtain coefficients in VSH developments of residuals from ICRF2-catalog differences" in this Journées. A revised version has been sent to another journal and it is now under review.

A) This paper deals with the estimation of position-correlated errors of positions or proper motions in astrometric catalogs, which can be considered as implicitly smooth vector fields on the unit sphere surface. The natural choice for such representations are the vector spherical functions constituting an orthogonal basis of tangent vector fields on the sphere. The technique can be used for assessing both systematic and random-correlated components of position or proper motion differences between two astrometric catalogs. The VSH fitting technique has been described and used in a number of recent papers, including the cited Mignard & Klioner 2012. One of the possible applications is an independent assessment of the accuracy of the future Gaia catalog with respect to the very accurate positions of ICRF2 radio sources. The authors propose an alternative technique, which fits the observed discrete vector field with 6 low-degree VSH through the medium of regularized regression with a set of arbitrarily chosen kernels.

Answer A) First, the scalar functions over the sphere with integrable square have a Hilbert's space structure. This space is complete, in the sense that any given function admits a development in surface spherical harmonics, which are orthogonal in sense of the inner product of the space. This property is crucial to give a simple and closed expression of every coefficient of the development. For vectorial fields over the sphere, a similar result exists, but in this case the complete base of Hilbert's space is the set VSH that are orthogonal in the functional sense and complete in the analytical sense. The part of the VSH that is interesting for the developments of vectorial fields over the sphere are the toroidal and the spheroidal that are orthogonal in every point and generate its the tangent space to the sphere. This affirmation is too generic because in every point the tangent point is two-dimensional and only two independent vectors are necessary to generate it. So, considering that for each order (of two indexes in this case) of VSH we have the same property, we can conclude, wrongly, that the VSH of order (k, l) are linearly dependent on those of another order (k', l') because they generate the same plane. We have considered the meaning of functional orthogonality that allows the calculation of coefficients of a development on the basis of the calculation of determined integrals. The fact of using a truncation of order one is casual and it does not alter the philosophy of the proposed method.
We would like to add that the last phrase in comment A) has no sense.

B) Methods of statistical estimation with smoothing kernels are well known and used in various applications but they are not relevant for the problem in question.

Answer B) The mathematical methods must be used in particular problems and its utility is given related not only to a consistent theory, but also to a contrastable numerical results. Assessing from the start that they are not relevant for the problem in question is only an opinion. Applying the SH ó VSH to the study of catalogues was usual since someone proposed it and it was accepted. To think that a certain method is only valid for one kind of problem means to go against the philosophy that is implied not only in mathematical methods, but in Mathematics, in general. The affirmation is, in addition, very inconsiderate with the colleagues who realized Referee's labor in our article of 2004 in the same magazine since they should have believed that it was indeed relevant for the problem in question. In fact, there the general problematics was studied and it was applied to the residues Hipparcos-FK5, and a value of $d_{1,0}$ was obtained by the first time as it has been now obtained by Mignard et al. In their article, they mention this result (Section 6.1) as *This is to very large systematic effect given the claimed accuracy of the FK5 below 50 mas* […] *Any transformation of to catalog tied by its construction to the FK5 system and later aligned to the HCRF (HIPPARCOS frame aligned to the ICRF) must absolutely take this large systematic difference into account in addition to the rotation" Since of importance for later studies.* We have not seen in the above mentioned article any mention to our work since, if it is a result to bear in mind, maybe it should have been commented that this result was already obtained by Marco et al. (A&A 418 (2004)). It does not seem that our method is not relevant for the problem in question.
On the other hand, comment B) does have consideration to other colleagues who, in the year 2008, were referees of our article in which, by means of the presented procedure, allowed us to obtain results related to the luni-solar precesión in total agreement with the obtained by other authors. Our work was accepted and published in in the PASP 876 (121), 2009.

C) A need to employ such methods may arise in multivariate statistics when a (usually, nonlinear) model has some intrinsic uncertainties. A simple (but linear) example is if we want to estimate the trajectory h(t) of a free-falling ball using a yardstick and a clock. The measurements of h are taken with some random errors, but if the clock is not very good, the times of observations are given with significant uncertainties too. A smoothing kernel reproducing the PDF of the time measure can then be introduced in the process of estimation, although in this case, there are simpler ways of doing this. This idea is hardly relevant for the astrometric problem.

Answer C) This one is one example in which the kernel non parametrical method is used to find a pdf. Also there are thousands that are used for regression. We will see (by means of an example) what means that is not relevant for the astrometric problem.

D) The positions of stars, at which the vector field is sampled, are uncertain, but their uncertainty (in mas) is completely negligible compared to the characteristic scale of low-degree VSH variations (many degrees). In other words, there is no reason to be worried that the field is sampled not where (sic) we assume it is sampled. Even if we wanted to take into account the uncertainty of the input positions, the kernels would have to be extremely narrow, and the result would be identical within the working precision.

Answer D) We understand that the "uncertainty" (in mas) compared with the scale of the variations of the harmonic (in degrees) is negligible. This argument is very dark since we do not undestand why is necessary to compare variations in the VSH (the vectorial fields of the base?) with the variations in the positions of stars. Any model of adjustment deals with functions (or vectorial fields) whose magnitudes are of very different orders from the residues to analyze. Does this imply that we can take an arbitrary method and as far as the coefficients are small, the results will be ok? Why does the concept of stability exist, then? For how long small variations in the "input" imply small variations in the "output"? The magnitudes of the VSH must be compared, in any case, not with those of the residues to analyze but with the magnitudes of the parameters to estimate.

E) The beauty of the VSH approach is that it handles the 2D vectors at once, we do not have to consider separately the $\Delta\alpha\cos(\delta)$ and $\Delta\delta$ components as scalar fields on the sphere.

Answer E) The method VSH handles vectors in 2D in its theoretical formulation. In his later implementation it must consider individual residues as unidimensional in the process of adjustment. We also consider vectorial fields (it is true, though, that the method is also applicable to scalar fields and, therefore, more general than the VSH) and, in fact, after having compared the coefficients calculated by us in (2004), with those of Mignard et al (2012) and with ours again (2013), they coincide. We do not discuss the beauty of the method proposed by Mignard et al, but we have not understood why ours cannot be equally beautiful, since it deals elegantly with the cotinuous problem, with an analytical and statistical approach, providing final algorithms where, indeed and as in the VSH, functions and integrals are necessary discretized. We insist that the procedure has been supported by other Referees that must be thought with identical authority and knowledge of the topic that the one that made the comments in blackboard.

F) The proposed "mixed" method loses this advantage right away, because the kernels are scalar functions […] The only reason the authors were able to do some practical computations with this method was that the problem was restricted to the first-degree VSH, that is, to the three rigid rotations and three dipoles. For the higher-degree VSH terms, the projections onto the local basis vectors (that is, the local East and North direction) are not mutually independent functions of α, δ. Effectively, one would have to introduce 2D kernels, and a separate set for each particular VSH. My prediction is that this technique cannot go beyond the first degree. The scientific value of such a limited, but numerically cumbersome, method is nil, as the astronomical phenomena of interest, such as the differential rotation of the Galaxy, local expansion, show up in the second- and higher degreed of the proper motion field.

Answer F) In the first place, we must say that there are scalar and vectorial kernel (see "Kernel Regression Estimation for Random Fields" M. Carbon, C. Francq, L.T. Tron). We have used the adjustments for $V^{\alpha}(\alpha, \delta)$ and $V^{\delta}(\alpha, \delta)$ with the scalarly developed models and, thereby appear, for first order, decoupled equations for ε_z and for $d_{1,0}$ whereas other equations appear coupled in pairs for the variables $\{\varepsilon_x, d_{1,-1}\}$ and $\{\varepsilon_x, d_{1,1}\}$. This way of obtaining the coefficients is not the aim of the method. Only for information, we would like to comment that kernels and methods of kernel regression do exist for several dimensions. Nevertheless the low technical difficulty from a mathematical point of view, makes not necessary the use of such methods. Let's see how our method is adapted to the calculation of harmonics of arbitrary orders. We consider the vectorial field $V(\alpha, \delta) = V^{\alpha}(\alpha, \delta)\mathbf{e}_\alpha + V^{\delta}(\alpha, \delta)\mathbf{e}_\delta = (\Delta\alpha\cos\delta)\mathbf{e}_\alpha + (\Delta\delta)\mathbf{e}_\delta$ being $V^{\alpha}(\alpha, \delta)$ and $V^{\delta}(\alpha, \delta)$ the scalar fields of the residues and $\mathbf{e}_\alpha$, $\mathbf{e}_\delta$ the unitary vectors in the tangent plane and in the directions of the right ascension and declination, respectively. Their expressions, in Cartesian coordinates, are known:

$$e_\alpha = \frac{1}{\cos\delta}\frac{\partial X}{\partial \alpha} = \begin{bmatrix} -\sin\alpha \\ \cos\alpha \\ 0 \end{bmatrix}, \quad e_\delta = \frac{\partial X}{\partial \delta} = X \times e_\alpha = \begin{bmatrix} -\cos\alpha\sin\delta \\ -\sin\alpha\sin\delta \\ \cos\delta \end{bmatrix}$$

On the other hand, and provided that we are in the surface of the unitary sphere, the only vectorial spherical involved are the spheroidal ones $S_{l,m}$ and the toroidal $T_{l,k}$. We suppose that the field V admits a development:

$$V(\alpha,\delta) = \sum_{k\geq 1}\sum_{l=-k}^{k}\left[t_{l,m}T_{l,m} + s_{l,m}S_{l,m}\right]$$

Due to the functional ortogonalidad, we have that:

$$t_{l,m} = \frac{\int_{S^2} VT_{k,l}}{\left\|T_{k,l}\right\|^2}, \quad s_{l,m} = \frac{\int_{S^2} VS_{k,l}}{\left\|S_{k,l}\right\|^2}$$

Here the denominators are calculated of exactly, and for the numerators an estimation is obtained using the proposed method. For the calculation of the components of the vectorial field V over points regularly spreaded over the sphere we can use the simple method of kernel regression or a method of local kernel polynomial regression. Computationally, the first is more economic and, in addition, it is sufficient for the problem that we are studying. It is important to emphasize that, once the adjustment has been established for V over the grid, this grid is also used for the numerical integration of the numerators. Thus, we do not manage to see what are the problem to calculate higher order estimations. Let's see this briefly without entering in details. With respect of the calculation, for any orders (except order 0 and 1), of the spheroidal and toroidal harmonics, we have to bear in mind that, from the definition of $Y_{k,l}$, the partial derivatives with respect to α are immediate and only the partials with respect to δ are more difficult, actually, they require the calculation of the partial derivatives of the Legendre associate functions $P_{k,l}(x)$ with $x=\sin\delta$ (for $l=0$ we obtain the Legendre polynomials).
With regard to the scalar expression used, we have said that there are coupled equations. Let's see what happens for higher orders. The functions that appear are the gradients of the surface spherical harmonics and only couples in pairs appear. So that the generalization for the procedure that strictly has followed in our article is immediate.

G) Listed among the advantages of the technique, is the possibility of replacing the observed sample field with a (high-density?) grid of reference points, thus avoiding the non-orthogonality issue -- or at least, this is how I understood this paragraph i) on page 3. But the non-orthogonality of discretized VSH is really a trifle compared to the daunting prospect of integrating and normalizing the kernel products in the Hilbert space of spherical functions. This is to be done numerically on the computer anyway, which inevitably involves another discretization.

Answer G) The density of points of control that have been taken over the sphere, the same that Mignard et al. (2012) propose, has 5050 points (and this amount does not increase proportionally to the number of known positions, on the contrary). The not ortogonalithy of the discrete method it cannot be qualified of a"trifle" since this not ortogonalithy has two big problems:
G1) When the order of the model of adjustment rises, the previously computed coefficients are not valid. It is necessary to restart every time, in order that the implementation is coherent with the initial exposition.
G2) The non ortogonalithy turns the problem unstable and the comment of G1) is one of its consequences.
We remark that the adjustments of residues are done once for the whole process and the integration takes, for every coefficient, less than one tenth of second in a portable computer. We do not understand to what refers with "to the daunting prospect of integrating and normalizing the kernel products in the Hilbert space of spherical functions ". Only one discretization is needed and the normalization is carried out directly because the denominators are already known.

H) Why would the VSH terms weighted with the kernel functions centered on the nearby stars be still orthogonal? This is not obvious to me, and I very much doubt that this is the case.

Answer H) We believe that the Referee has confused the algebraic ortogonalithy in a point with the functional ortogonalithy over the whole sphere. We always work with the latter concept of ortogonalithy. The weight, in every point of sphere, is provided in agreement by a choice of bandwidth for the considered kernel, using the formulae that the authors Wand&Jones and Simonoff, propose and that are sufficiently accurate to work in the celestial sphere, because there is no frontier (where problems may appear).

2i) The non-orthogonality of discretized, or sampled, VSH over a given star field is not a big technical problem because in practical applications, the VSH decomposition is done on a finite number of harmonics, and with modern computers, one should simply solve the complete least-squares problem and derive the full covariance matrix, with the nonzero off-diagonal covariances. I have done this myself easily on my laptop for up to 140 VSH. The point is, the size of the normal matrix to be inverted is defined by the number of fitting VSH, not by the number of stars. So, even though you cannot compute the coefficients by the direct dot-product of the two vector fields, it is not at all critical.

Answer I) To solve a possible bad conditioned system is not too suitable from a computational point of view. Certainly, the direct inversion (as the Referee seems to allude) is not recommended. Maybe, a QR decomposition might contribute to provide the maximum of stability to the results. The number of operations is problematic when there is accumulation of errors. In our procedure this danger does not exist since only an initial adjustment and later a numerical integrations are carried out. In addition, in the classic method of least squares minimums, as it has been commented in 2G), appears the problem of the stability in the calculation of the coefficients. All the calculations that we do have been carried out using a simple portable, and the resolution of the complete problem takes only a few seconds. Another potential source of mistakes in the calculation of the coefficients by means of the classic least squares method appears when we raise the order of the harmonics. As we have mentioned, the non ortogonalithy of the method used by Mignard et al. (2012), produces that an order for the development should be selected previously. Let's imagine that we choose an order 4, for example, and that the values of the coefficients are accepted. To obtain estimations of higher order all the calculations must be repeated. The better thing to do would be to substract to the initial residues the part of the accepted signal. With the resultant residues, it is possible to do an adjustment of higher order. The results could be considered for statistical contrast, if they are zero. Otherwise there is an imbalance in the mean quadratic error, between bias and variance and the null hypothesis, in this case, is false. Our method allows to calculate the coefficients independently, only once and in the desired order. Finally, the limitations that the discrete set of points must be considered, preventing from going beyond a certain order of development trustworthy.

2J) "The statistical information is completely included in the kernel non-parametric regression adjustment." What does it mean, actually? Does it imply that the choice of kernel should be driven by some sort of statistical information about a particular measurement? Is some sort of prior inference involved in the selection of kernel?

Answer J) The justification for the use of different kernels may be seen in the scientific literature. The statistical information of the data, is given by his function of density that, is not known. A possibility is the use of a kernel estimation for the pdf. One must not confuse this with the kernel regression used later in the estimation of the vectorial field $V(\alpha, \delta)$. It is interesting to observe, in the latter sense, the property that fulfills Nadaraya Watson's method (unidimensional or over the sphere, it does not matter). If Z is a random variable, then the estimation by means of the above mentioned method in a point x is the value ß that minimizes

$$\frac{1}{nh}\sum_{i=1}^{n}(z_i-\beta)^2 K\left(\frac{z-z_i}{h}\right)$$

This is a particular case for the sphere of the search of the β value that minimices :

$$Var\left(Z\,|\,(\alpha,\delta)\right) = \int_D \underbrace{(Z-\beta)^t(Z-\beta)}_{\text{Infinitessimal pdf}}\underbrace{\hat{f}_Z(z)\,dz}_{\text{pdf element}}, \quad \beta = E[Z] \text{ and } \hat{f} \text{ is the estimator by kernels of the pdf}$$

K) The following text seems to suggest that a known bias in the measurement can be captured in the kernel smoothing (how?). But what astrometrists do when they know there is a bias in their measurements (which happens rarely), they subtract the bias right away! These statements are in obvious conflict with the arbitrary and unjustified choice of a particular function (Epanechnikov) for the kernel.

Answer K) The term "bias" refers to the use that is given to measure the difference between the expectation of the estimator and the value of the parameter to estimate. This means that if, for example, the mean of the data $\Delta\alpha\cos\delta$ is not null (as in the considered case), it is difficult that a theoretical adjustment model, whose expectation over the whole sphere is null, is able to reflect this deviation, which will indicate that we are using a biased model. How can this allow the application of Gauss-Markov's Theorem, which is the base of the method of the least squares, with biased data and an unbiased model?

L) How would the results shown change if a Gaussian kernel is selected? If a kernel of different width is selected?

Answer L) The choice of the kernel has very little impact. On the other hand, we understand that "kernel of different width " really means "if we took different *h* for a given kernel ". In this case, we must assert that the length of *h* is computed using methods well stablished in the bibliography.

Part of this work was supported by grant P1-1B2009-07 and a grant P1-061I455.01 from Bancaja