Application of Multiple-imputation-particle-filter for Parameter Estimation of Visual Binary Stars with Incomplete Observations

Rubén M. Clavería[1], David Acuña 1, René A. Méndez[2], Jorge F. Silva[1] and Marcos E. Orchard[1]

[1] *Universidad de Chile, Department of Electrical Engineering. Av. Tupper 2007, Santiago, Chile*
rclaveria@ing.uchile.cl
davacuna@ing.uchile.cl
josilva@ing.uchile.cl
morchard@ing.uchile.cl
[2] *Universidad de Chile, Department of Astronomy. Casilla 36-D, Santiago, Chile*
rmendez@u.uchile.cl

ABSTRACT

In visual binary stars, mass estimation can be accomplished through the study of their orbital parameters –Kepler's Third Law establishes a strict mathematical relation between orbital period, orbit size (semi-major axis) and the system total mass. Although, in theory, few observations on the plane of the sky may be enough to obtain a decent estimate for binary star orbits, astronomers must frequently deal with the problem of partial measurements (i.e.; observations having one component missing, either in (X, Y) or (ρ, θ) representation), which are often discarded. This article presents a particle-filter-based method to perform the estimation and uncertainty characterization of these orbital parameters in the context of partial measurements. The proposed method uses a multiple imputation strategy to cope with the problem of missing data. The algorithm is tested on synthetic data of relative position of binary stars. The following cases are studied: i) fully available data (ground truth); ii) incomplete observations are discarded; iii) multiple imputation approach is used. In comparison to a situation where partial observations are ignored, a significant reduction in the empirical estimation variance is observed when using multiple imputation schemes; with no numerically significant decrease on estimate accuracy.

1. INTRODUCTION

Mass is arguably the most important property of a star, since it determines to a great extent the structure and evolution of the celestial body. Visual binaries, defined as gravitationally bound pairs of stars whose components can be individually resolved with the aid of a telescope, are probably the main source of data on stellar masses, since Kepler's Third Law establishes a strict relation between orbital parameters and the system total mass:

$$\frac{a^3}{P^2} = \frac{G}{4\pi^2}(m_1 + m_2), \tag{1}$$

where P, a and G are the orbital period, orbit semi-major axis[1], and the gravitational constant, respectively. Individual masses are denoted m_1 (primary star) and m_2 (companion star). Observations of visual binaries are expressed in terms of the relative position of the companion star with respect to the primary star in the plane of the sky, using either Cartesian or polar coordinates. It must be noted that, unless other measurements are incorporated (for example, radial velocity), observations of visual binaries are samples from the apparent orbit, that is, the projection of the real, three-dimensional orbit in the plane of the sky.

Theoretically, few observations are enough to determine the orbital parameters of a double star. Docobo's analytic method requires as few as three observations of the apparent orbit (Docobo, 1985). However, those must be high quality observations in well-chosen points of the orbit. In (Lucy, 2014), the author uses 15 observations with varying degrees of quality and orbit coverage. Performing simulation-driven testing of estimation methods allows researchers to control most aspects of the algorithm input, from the underlying orbital configuration to the quality and quantity of the observations. However, when data is obtained from real sources, astronomers have to deal with several issues: noisy observations, diversity of sources, incomplete measures, poor orbit coverage. As most algorithms require a full set of measurements, partial observations –i.e.; those having one component missing, either in (X, Y) or (ρ, θ) representation– are often discarded.

[1] Half of the longest diameter of an ellipse. This work is focused on systems whose specific energy is less than zero, so each of their components follows an elliptical path.

As a means to cope with the aforementioned difficulties, this work addresses the task of estimating the orbital parameters and characterizing their uncertainty from a Bayesian standpoint and merging empirical, theoretical, and statistical knowledge (represented by observations, the dynamical model of binary stars, and prior knowledge of orbital parameters, respectively). Moreover, knowledge hidden in incomplete observations is incorporated through the implementation of a multiple-imputation scheme.

Although this research has been motivated by the problem of orbital estimation, the proposed method may be applicable to a variety of parameter estimation problems. Moreover, its implementation only requires the statistical characterization of a merit function (e.g., Mean Square Error). Since this method does not focus on finding minima/maxima of a fitness function, but in assigning a degree of likelihood to each candidate solution according to a Bayesian approach, it provides a sound theoretical framework for uncertainty characterization.

The article is structured as follows. In Section 2, a theoretical background is presented, reviewing the underlying concepts of particle filters, artificial evolution of parameters and the multiple imputation approach. Section 3 presents the dynamical model of binary stars, the observational aspects related to this problem, and the application of the multiple-imputation particle filter to the estimation of orbital parameters. In Section 4, the algorithm is tested on a synthetic data set with some incomplete observations, analyzing its performance in comparison to the scenarios where data is fully available and where partial observations are discarded. Finally, conclusions and future work are presented in Section 5.

2. THEORETICAL BACKGROUND

Real world systems are commonly dynamic, nonlinear, and may involve a high dimensionality relationship between variables. In this regard, state-space models offer a good treatment for these systems; for example, when monitoring critical system components which physical phenomenology may be modeled directly under the state-space form. Moreover, uncertainty due to noisy measurements associated with sensors constraints or other sources of disturbances such as the lack of knowledge about the actual system dynamics, can be incorporated into the state-space form with ease. This allows us to adopt a Bayesian approach, where the main objective is to estimate the underlying probability distribution in order to perform statistical inferences. Since the analytical solutions may be founded under certain conditions, the real problem to be addressed is that of evaluating complex integrals where numerical methods tend to breakdown, even more when high dimensional systems are involved. An alternative to address this problem is the use of particle filters, which is presented in the following section. Finally, we present the multiple imputation particle filter.

2.1. Particle Filters

Due to the use of digital computers for signal processing, it is of interest to develop a Bayesian processor in which measurements arrive sequentially in time. The recursive estimation of the evolving posterior distribution is the so called *optimal filtering* problem. A mathematical framework is provided below for solving this problem using particle filters.

Let $X = \{X_t, t \in \mathbb{N}\}$ be a first order Markov process denoting a n_x-dimensional system state vector with initial distribution $p(x_0)$ and transition probability $p(x_t|x_{t-1})$. Also, let $Y = \{Y_t, t \in \mathbb{N} \setminus \{0\}\}$ denote n_y-dimensional conditionally independent noisy observations. The whole system is represented in state-space form as

$$x_t = f(x_{t-1}, w_{t-1}), \tag{2}$$

$$y_t = g(x_t, v_t), \tag{3}$$

where w_t and v_t denote independent random variables whose distributions are not necessarily Gaussian. Since it is difficult to compute the filtering posterior distribution $p(x_t|y_{1:t})$ directly, Bayesian estimators are constructed from *Bayes' rule*.

Under Markovian assumptions, the filtering posterior distribution can be decomposed into

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t) \cdot p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}. \tag{4}$$

In this context, sequential Monte Carlo methods (SMC) offer an alternative to numerical integration techniques that fail due to high computation. SMC methods, also called particle filters, are stochastic computational techniques designed for simulating highly complex systems in an efficient way. In Bayesian estimation, these techniques simulate probability distributions by using a collection of N weighted samples or *particles*, $\{x_t^{(i)}, \mathcal{W}_t^{(i)}\}_{i=1}^N$, that yields to discrete mass probability distributions, as shown in Equation 5.

$$\hat{p}(x_t|y_{1:t}) \approx \sum_{i=1}^{N} \mathcal{W}_t^{(i)} \delta(x_t - x_t^{(i)}). \tag{5}$$

The weighting process is made by applying the *sequential importance resampling* (SIR) algorithm, which is explained in the following subsections.

2.1.1. Sequential Importance Sampling

Importance sampling is used to simulate samples from a proposed distribution, all to obtain Monte Carlo (MC) estimates of complex expressions such as Eq. 6. The method heavily depends on choosing an appropriate importance distribution. However, this task may be computationally intensive.

$$\hat{f}(x_t) = E_{X|Y}\{f(x_t)\} = \int_X f(x_t) p(x_t|y_{1:t}) dx_t. \tag{6}$$

Drawing N independent identical distributed random samples from $p(x_t|y_{1:t})$, the integral may be approximated by a sum of Dirac delta functions.

$$\hat{f}(x_t) \approx \frac{1}{N}\sum_{i=1}^{N} f(x_t)\delta(x_t - x_t^{(i)}), \quad (7)$$

$$= \frac{1}{N}\sum_{i=1}^{N} f(x_t^{(i)}). \quad (8)$$

These approximations may not hold when it is not possible to sample directly from $p(x_t|y_{1:t})$. The *sequetial importance sampling* (SIS) algorithm avoids these difficulties by drawing samples from an *importance distribution* and approximating the posterior distribution by appropriate weights. Weights are recursively defined as

$$w_t^{(i)} = w_{t-1}^{(i)} \cdot \frac{p(y_t|\tilde{x}_t^{(i)}) \cdot p(\tilde{x}_t^{(i)}|x_{t-1}^{(i)})}{\pi(\tilde{x}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, y_{1:t})}, \quad (9)$$

where $\{\tilde{x}_t^{(i)}\}_{i=1}^{N}$ is a set of N random samples drawn from the importance distribution $\pi(\tilde{x}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, y_{1:t})$. Also, by defining normalized weights

$$\mathcal{W}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^{N} w_t^{(i)}}, \quad (10)$$

then the posterior distribution can be approximated by the expression described in Equation 5.

2.1.2. Resampling

When the updating process begins, a tendency to increase the variance of particles is seen, setting negligible weights to some of them. These particles become useless as they track low probability paths of the state vector. In order to solve this problem, a *resampling* step is incorporated, which leads to the SIR algorithm.

An analytical expression for measuring how degenerated are the particles is given by the *effective particle sample size* showed in Equation 11.

$$N_{eff}(t) = \frac{N}{1 + Var_{p(\cdot|y_{1:t})}(w(x_t))}. \quad (11)$$

As it is not possible to calculate N_{eff}, an estimate is given by

$$\hat{N}_{eff}(t) = \frac{1}{\sum_{i=1}^{N}(\mathcal{W}_t^{(i)})^2}. \quad (12)$$

In other words, the resampling step consists of removing small weighted particles while retaining and replicating those of large weights. Thus, whenever $\hat{N}_{eff} \leq N_{thres}$, with N_{thres} a fixed threshold, the depletion of the particles is imminent and resampling must be applied.

Algorithm 1 SIR Particle Filter

1. Importance Sampling
   for $i = 1, \ldots, N$ **do**
   • Sample $\tilde{x}_t^{(i)} \sim \pi(x_t|x_{0:t-1}^{(i)}, y_{1:t})$ and set $\tilde{x}_{0:t}^{(i)} \triangleq (x_{0:t}^{(i)}, \tilde{x}_t^{(i)})$
   • Compute the importance weights $w_t^{(i)} = w_{t-1}^{(i)} \cdot \frac{p(y_t|\tilde{x}_t^{(i)}) \cdot p(\tilde{x}_t^{(i)}|x_{t-1}^{(i)})}{\pi(\tilde{x}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, y_{1:t})}$
   • Normalize $\mathcal{W}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^{N} w_t^{(i)}}$
   end for

2. Resampling
   if $\hat{N}_{eff} \geq N_{thres}$ **then**
   for $i = 1, \ldots, N$ **do**
   • $x_{0:t}^{(i)} = \tilde{x}_{0:t}^{(i)}$
   end for
   else
   for $i = 1, \ldots, N$ **do**
   • Sample an index $j(i)$ distributed according to the discrete distribution satisfying $P(j(i) = l) = \mathcal{W}_t^{(i)}$ for $l = 1, \ldots, N$
   • $x_{0:t}^{(i)} = \tilde{x}_{0:t}^{j(i)}$ and $w_t^{(i)} = \frac{1}{N}$
   end for
   end if

In general, the SIR particle filter is divided into two steps. Firstly, a *prediction* is done using the state transition model to generate the prior distribution $p(x_k|x_{k-1})$. Then an *update* step is done to modify the particle weights through the likelihood $p(y_k|x_k)$. If the resulting particles are degenerated, a $resampling$ step is added, as it was shown previously.

2.1.3. Online Parameter Estimation with Particle Filter

In the context of state estimation, it is sometimes necessary to handle an online estimation scheme for a model parameter vector. Although parameters are considered as fixed in a wide range of problems, this approach is not always adequate, since parameters of a system might be time-varying or simply unknown.

To understand the problems of parameter estimation outside a Bayesian context, let θ be a vector parameter. The maximum likelihood estimate of θ is obtained by maximizing the log-likelihood function (Kitagawa & Sato, 2001):

$$l(\theta) = log[L(\theta)] = \sum_{t=1}^{T} log[p(y_t|y_{1:t-1}, \theta)], \quad (13)$$

where the term

$$p(y_t|y_{1:t-1}, \theta) = \int p(y_t|x_t, \theta)p(x_t|y_{1:t-1}, \theta)dx_t \quad (14)$$

needs to be approximated via Monte Carlo methods.

The maximization of Expression 13 for the estimation of θ is not always direct, and approximations over Eq. 14 make this method impractical, due to the high computational costs involved if parameter estimation is intended for every time step. Thus, a different perspective is necessary to approach the online parameter estimation problem. This idea is attacked through the artificial evolution of parameters.

The first ideas about introducing random disturbances to particles were proposed by (Gordon, Salmond, & Smith, 2002). In their work, the authors propose to introduce random disturbances to the positions of particles (called *roughening penalties*) in order to combat degeneracy. This idea has been extended in order to estimate online a vector of fixed model parameters, which is referred to as *artificial evolution*(Liu & West, 2001). Artificial evolution of parameters is a simple and powerful idea, nevertheless, it requires careful handling because of the inherent model information loss given by the consideration of time-varying parameters that are fixed.

Instead of estimating the vector parameter θ through maximum likelihood, the Bayesian framework may be introduced to estimate θ online. This is achieved by augmenting the state vector x_t with unknown parameters θ as:

$$x_t = \begin{bmatrix} x_t \\ \theta_t \end{bmatrix}. \tag{15}$$

Where $\theta_t = \theta$ implies the consideration of an extended model where parameters are time-varying. Then, an independent, zero-mean normal increment is added to the parameter at each time step (Liu & West, 2001):

$$\theta_t = \theta_{t-1} + \epsilon_t, \tag{16}$$

$$\epsilon_t \sim \mathcal{N}(0, W_t), \tag{17}$$

where W_t is a variance matrix and θ_t and ϵ_t are conditionally independent given Σ_t. The key motivation is that the artificial evolution of parameters gives new values for each iteration, and thus, weight assignment in Particle Filters considers the likelihood of the state and parameter values.

2.2. Multiple imputations

Missing data is a problem that may be treated mainly from two perspectives. On the one hand, *single imputation* techniques fill the incomplete data set imputing single values at each missing datum. The advantage of this perspective is that it allows standard complete data methods to be used. However, these techniques fail due to the lack of uncertainty characterization of both, the sampling variability and the uncertainty associated with the imputation model. On the other hand, the idea of *multiple imputations* retains the advantages of single imputation techniques and also accounts for the uncertainty of the missing mechanism. Multiple imputations (Rubin, 1987) consist of creating multiple complete data sets imputing m values for each missing datum so that sampling variability around the actual values is incorporated for performing valid inferences. Nevertheless, multiple imputations has disadvantages like the need of drawing more imputations and larger memory space for storing and processing multiple-imputed data sets.

It is important to choose the right number of imputations (Graham, Olchowski, & Gilreath, 2007). Obviously, the computational cost is higher as the number of imputations increases. In this regard, (Rubin, 1987, p. 114) shows that an approximation of efficiency for an estimate is given by

$$(1 + \frac{\gamma}{m})^{-1/2}, \tag{18}$$

in units of standard errors, where m is the number of imputations and γ is the fraction of missing information in the estimation. Consequently, excellent results may be obtained using only few imputations ($m = 3, 4, 5$).

2.3. Multiple Imputation Particle Filter

Originally introduced by (Housfater, Zhang, & Zhou, 2006), the Multiple Imputation Particle Filter extends the PF algorithm by incorporating a multiple imputation (MI) procedure for cases where measurement data is not available, so that the algorithm can include the corresponding uncertainty into the estimation process. The main statistical assumption in this approach is that the missing mechanism is *Missing at Random (MAR)*, thus, it does not depend on the missing measures given the observed ones.

For readability, a change in notation is necessary. As it was stated in (Housfater et al., 2006), lets denote now the measurements as a partitioned vector $U_t = (Z_t, Y_t)$, where Z_t corresponds to the missing part and Y_t is from now on the observed part. Then, the MI PF algorithm performs the same as the SIR PF except that there are missing measures. In this case, a MI strategy is adopted.

An imputation model expressed as a probability distribution ϕ is required for drawing m samples -imputations-, that is

$$z_t^j \sim \phi(z_t | y_{1:t}), \tag{19}$$

where $j = \{1, \ldots, m\}$ denotes the imputation index. Similarly to importance sampling, each imputation is associated with a weight p_t^j holding the condition $\sum_{j=1}^{m} p_t^j = 1$. The filtering posterior distribution is (Liu, Kong, & Wong, 1994):

$$p(x_t | y_{1:t}) = \int p(x_t | u_{1:t-1}, y_t) p(z_t | y_{1:t}) dz_t. \tag{20}$$

By performing a Monte Carlo approximation,

$$p(x_t | y_{1:t}) \simeq \sum_{j=1}^{m} p_t^j p(x_t | u_{1:t-1}, u_t^j), \tag{21}$$

where $u_t^j = (z_t^j, y_t)$ are complete data sets formed from imputed values. Additionally, by particle filtering each of these data sets yields

$$p(x_t|u_{1:t-1}, u_t^j) \approx \sum_{i=1}^{N} w_t^{(i,j)} \delta(x_t - x_t^{(i,j)}), \qquad (22)$$

where the indexes i and j indicate the particle and the imputation, respectively. Thus, an approximation of the desired posterior distribution is

$$p(x_t|y_{1:t}) \approx \sum_{j=1}^{m} \sum_{i=1}^{N} p_t^j w_t^{(i,j)} \delta(x_t - x_t^{(i,j)}). \qquad (23)$$

3. MULTIPLE-IMPUTATION-BASED UNCERTAINTY CHARACTERIZATION FOR ORBITAL PARAMETERS ESTIMATION

3.1. Binary Stars Dynamics

Under some basic assumptions, the dynamics of a binary stellar system is reduced to the two-body problem, whose solution is the well-known Keplerian orbit. Keplerian orbits describe the motion of orbiting celestial bodies in terms of an ellipse, parabola or hyperbola. The specific geometry of the orbit depends on the underlying physical properties (i.e., energy) of the system under study. In this work, we are interested in systems with periodic elliptical orbits.

An ellipse is fully described by its eccentricity, e, and its semi-major axis, a. It can be demonstrated that both the individual bodies of the binary system and the relative position vector follow elliptical paths with the same eccentricity and period (P), but different phase angles (companion star is always $180°$ ahead of primary star). Figure 1 illustrates the individual paths of the binary system (primary star in blue, companion star in red); the relative position is represented by the black line that joins both components. Individual semi-major axes comply the following relations:

$$a = a_1 + a_2, \qquad (24)$$

$$a_1 \cdot m_1 = a_2 \cdot m_2, \qquad (25)$$

where a_1, a_2 and a are the semi-major axis of the primary star, the companion star and the relative orbit, respectively. Thus, determination of the parameters of the relative orbit between the primary and companion star is enough to calculate the total mass of the system.

For a given epoch τ, the value of the relative position between stars is computed as follows. Let be T the epoch when separation between primary and companion star reaches its minimum value. This value is known as *time of periastron passage*. Then, the expression known as Kepler's equation

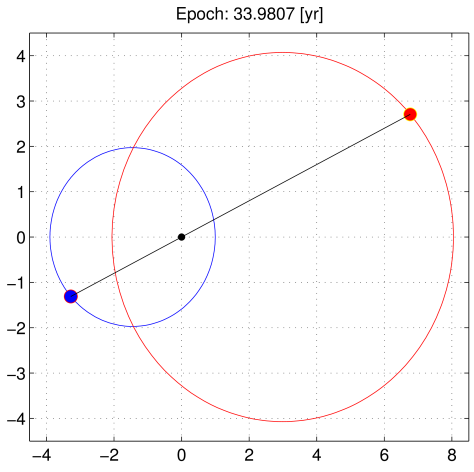


Figure 1. Elliptical orbits of a binary system.

can be written:

$$M = \pi(\tau - T)/P = E - e \sin E. \qquad (26)$$

where terms M and E are called *mean anomaly* and *eccentric anomaly*, respectively[2]. Equation 26 doesn't have analytical solution, and must be solved through numerical methods (in this work we use Newton-Raphson algorithm). Once $E(\tau)$ is obtained, the term known as *true anomaly* is directly determined by:

$$\tan\frac{\nu}{2} = \sqrt{\frac{1+e}{1-e}} \tan\frac{E}{2}. \qquad (27)$$

True anomaly ν corresponds to the angle between the main focus of the ellipse and the companion star, provided that the periastron is aligned with the X axis and the primary star occupies the main focus of the ellipse (Figure 2). Evaluating $\nu(\tau)$ in the following expression:

$$r(\nu) = \frac{a(1-e^2)}{1 + e\cos\nu}, \qquad (28)$$

yields the position (r, ν) (polar coordinates) at a given instant τ.

3.2. Observational aspects

In the previous subsection, the reference system has been chosen so that the plane of motion of the binary system (orbital plane) coincides with the XY plane. However, the orbit of a real system may have an arbitrary orientation in the space. What telescopes show is the projection of the real orbit in the reference plane (plane of the sky), known as the *apparent orbit* (Figure 3 provides a graphical insight of the aforementioned orbits). In order to explain the mathematical relation

[2]The geometrical meaning of this terms, although important in the context of Astronomy or Physics, is not relevant in this article, and therefore won't be explained in detail.

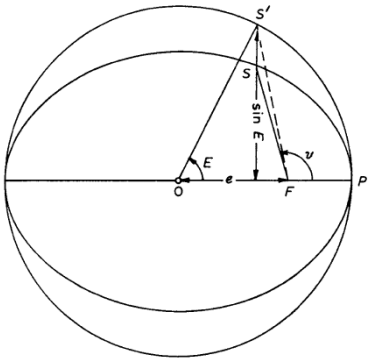


Figure 2. Diagram with mean and eccentric anomaly. F is the focus of the ellipse from which the angle ν is measured; S is an arbitrary point of the orbit and S' its perpendicular projection to auxiliary circumference (i.e., that with same center O and radius equal to a).

between real and apparent orbit, the following quantities must be defined:

- i or *Orbital inclination*: The angle between the reference plane and the orbital plane.
- ω or *Argument of the periapsis*: The angle from the body's ascending node to its periapsis, measured in the direction of motion.
- Ω or *Longitude of the ascending node*: The angle from a reference direction (usually the north celestial pole), called the origin of longitude, to the direction of the ascending node, measured in the reference plane.

Let $\mathcal{O} = \{T, P, e, a, \omega, \Omega, i\}$ be an arbitrary orbital configuration. Then, the apparent orbit is computed as follows:

- Calculate the Thiele-Innes constants:

$$\begin{aligned} A &= a(\cos\omega\ \cos\Omega - \sin\omega\ \sin\Omega\ \cos i), \\ B &= a(\cos\omega\ \sin\Omega + \sin\omega\ \cos\Omega\ \cos i), \\ F &= a(-\sin\omega\ \cos\Omega - \cos\omega\ \sin\Omega\ \cos i), \\ G &= a(-\sin\omega\ \sin\Omega + \cos\omega\ \cos\Omega\ \cos i). \end{aligned} \tag{29}$$

- For any given epoch τ, determine the eccentric anomaly E. With this value, calculate the auxiliary variables (x, y).

$$\begin{aligned} x(E) &= \cos E - e, \\ y(E) &= \sqrt{1-e^2}\sin E. \end{aligned} \tag{30}$$

- Finally, obtain the position in the plane of the sky by evaluating the following expressions:

$$\begin{aligned} X &= Bx + Gy, \\ Y &= Ax + Fy, \end{aligned} \tag{31}$$

  which define a point in the apparent orbit.

Thiele-Innes constants provide not only a straightforward way to calculate the apparent from a set of orbital parameters, but also an alternative mathematical characterization of a binary system. Indeed, parameters $\{a, \omega, \Omega, i\}$ can be substituted by $\{A, B, F, G\}$ since they are mutually mapped by the following expressions:

$$\begin{aligned} tan(\omega + \Omega) &= \frac{B-F}{A+G}, \\ tan(\omega - \Omega) &= \frac{-B-F}{A-G}, \\ a^2(1+\cos^2 i) &= A^2 + B^2 + F^2 + G^2, \\ a^2\cos^2 i &= AG - BF. \end{aligned} \tag{32}$$

However, this representation is not free of ambiguity –given a value of a, different angles ω, Ω, i may yield the same Thiele-Innes constants. In spite of this, Thiele-Innes representation is widely used by astronomers, since some algorithms can take advantage of the linear dependence of the constants with respect to the values of (x, y).

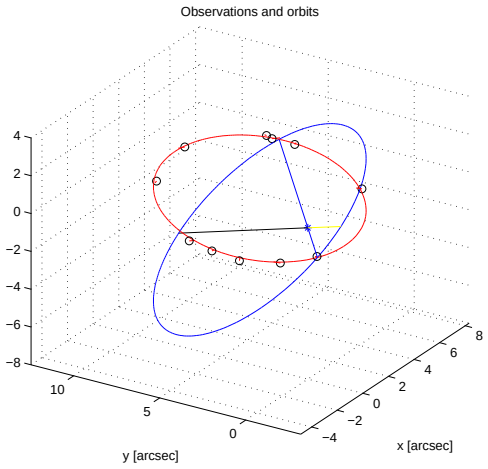


Figure 3. Real orbit in 3D space (blue line), observations (black circumferences) and apparent orbit (red line) in the plane of the sky.

3.3. Particle-filter-based Estimation of Parameters

Although Particle Filter was conceived as a method for online state estimation, it allows online parameter estimation as well, by means of Artificial Evolution of Parameters (explained in Section 2.1.3). Since in the study of visual binaries the object of interest is the orbital configuration (rather than the specific position of the stars at each epoch)[3], this method aims to characterize the observed object in terms of a vector of orbital parameters and its uncertainty, represented by the posterior probability density function given the observations.

Observations of visual binaries are not acquired at a constant,

[3] It is the set of parameters what enables the estimation of the mass and energy of the system.

less-than-one-second sampling rate. In fact, most of them are obtained sparsely within long time intervals (e.g., 15 measurements in 20 years). Thus, real-time estimation is not required and there is no impediment to using all the available observations at once, instead of the one-measurement, one-iteration approach that is usually employed in Bayesian filtering.

In order to adapt the Particle Filter to this specific problem, we propose a design that dispenses with state estimation (namely, relative position (X, Y) in the plane of the sky at each epoch τ) and focuses in parameter estimation using the Artificial Evolution of Parameters approach. That design is summarized in the following points:

- **State definition**: Initially, each particle consists of a vector of orbital parameters:

$$x_t^{(\mathrm{i})} = (T_t^{(\mathrm{i})}, P_t^{(\mathrm{i})}, e_t^{(\mathrm{i})}, a_t^{(\mathrm{i})}, \omega_t^{(\mathrm{i})}, \Omega_t^{(\mathrm{i})}, i_t^{(\mathrm{i})}). \tag{33}$$

  However, Thiele-Innes representation is preferred, as parameters A, B, G, F can be estimated by a least-squares method provided that both parameters T, P, e and the set of observations $\{\tau_k, X_k, Y_k\}_{k=1,\dots,N}$ are known[4]. Thus, each particle is reduced to $x_t^{(\mathrm{i})} = (T_t^{(\mathrm{i})}, P_t^{(\mathrm{i})}, e_t^{(\mathrm{i})})$. Variable i is used as an index for the particles and t indicates the current iteration of the filter.

- **Evolution equation**: Since none of components of a particle is, properly speaking, a state variable[5], they evolve according to the Artificial Evolution of Parameters approach:

$$x_{t+1}^{(\mathrm{i})} = x_t^{(\mathrm{i})} + \epsilon_t^{(\mathrm{i})}, \tag{34}$$

  where $\epsilon_t^{(i)}$ is the artificial evolution noise. Each realization of ϵ is drawn from a multivariate Gaussian distribution of the same dimension of x_t.

- **Weight update routine**: In this scheme, all available observations are used at once and no new measurements are received at the beginning of a iteration. Because of this, a novel weight update criterion is proposed. For each particle $x_t^{(i)}$, we calculate the Mean Squared Error with respect to the set of observations $\{\tau_k, X_k, Y_k\}_{k=1,\dots,N}$:

$$\mathcal{Y}_t^{(i)} = \frac{1}{N}\Big(\sum_{k=1}^{N} \frac{1}{\sigma_x^2(k)}[X_k - X_{comp}^{k,t,i}]^2 + \sum_{k=1}^{N} \frac{1}{\sigma_y^2(k)}[Y_k - Y_{comp}^{k,t,i}]^2\Big), \tag{35}$$

  where the triplet (k, t, i) indicates epoch τ_k, filter iteration t and particle index i. Ordered pairs $(X_{comp}^{k,t,i}, Y_{comp}^{k,t,i})$ are the relative positions at epoch τ_k computed according to the T, P and e values of particle $x_t^{(i)}$.

  Assuming that observations in the plane of the sky are affected by Gaussian noise (that is, $X_{obs} = X_{real} + n$, where $n \sim \mathcal{N}$), variable $\mathcal{Y}$ must follow a Gamma distribution (details in Appendix A). Thus, by evaluating the value of $\mathcal{Y}_t^{(i)}$ in the p.d.f. of Gamma distribution, we obtain a measure on how likely is the particle $x_t^{(i)}$. That enables to perform the weight update step as follows:

$$w_t^{(i)} = w_{t-1}^{(i)} \cdot \mathcal{L}_{\mathcal{Y}}(x_t^{(i)}|\mathcal{Y}_t^{(i)}), \tag{36}$$

  where the likelihood function $\mathcal{L}_{\mathcal{Y}}$ has the analytical form of Gamma p.d.f.:

$$\mathcal{L}_{\mathcal{Y}}(\Theta|\mathrm{x}) = P_{\mathcal{Y}}(\mathrm{x}|\Theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}}\mathrm{x}^{\alpha-1}\mathrm{e}^{-\frac{\mathrm{x}}{\theta}}. \tag{37}$$

  Equation 37 uses the traditional notation of Statistics: likelihood function has parameters Θ (in uppercase in order to avoid confusion with parameter θ of Gamma distribution) and outcomes x (observations). In this case, $\mathrm{x} = \mathcal{Y}_t^{(i)}$ and $\Theta = x_t^{(i)}$. Values of parameters α and θ are specified in Appendix A;

Algorithm 2 summarizes the implementation of the method proposed.

Algorithm 2 Particle Filter for Estimation of Orbital Parameters

for $t = 1 : N_{it}$ **do**
  for $i = 1 : N_{part}$ **do**
    if $t = 1$ **then**
      • Particle initialization
      $x_t(i) \sim initDist$
    else
      • Evolution equation
      $x_t(i) \leftarrow x_{t-1}(i) + W_t(i), W_t(i) \sim \mathcal{N}(0, \Sigma_\epsilon)$
      • Thiele-Innes constants calculation ($\mathcal{E}$: epochs)
      $[B, A, G, F] \leftarrow leastSquaresEstimate(\dots$
      $\mathcal{E}, X_{obs}, Y_{obs}, \mathcal{T}_t(i), P_t(i), e_t(i), W_{obs})$
      • MSE calculation (W_{obs}: observation weight)
      $\mathcal{Y}_t(i) \leftarrow MSE(\mathcal{E}, X_{obs}, Y_{obs}, \cdots$
      $W_{obs}, x_t(i), B, A, G, F, W_{obs})$
      • Weight update
      $w_t(i) \leftarrow w_{t-1}(i) \cdot \mathcal{L}_{\mathcal{Y}}(\mathcal{Y}_t(i))$
    end if
  end for
  • Weight normalization
  $w_t(1 : N_{part}) \leftarrow w_t(1 : N_{part}) / \sum_{i=1}^{N_{part}} w_t(i)$
  • Resampling
  if $1/\sum_{i=1}^{N_{part}} w_t^2(i) < R_{th}$ **then**
    $[x_t(1 : N_{part}), w_t(1 : N_{part})] \leftarrow resampling(\dots$
    $x_t(1 : N_{part}), w_t(1 : N_{part}))$
  end if
end for
• Final values of particles and weights
return $x_N(1 : N_{part})$, $w_N(1 : N_{part})$

[4]The least-square method for the estimation of Thiele-Innes constants is explained in detail in (Lucy, 2014).

[5]In the sense of the state-space representation widely used in Control Theory.

3.4. Implementation of a Multiple Imputation Strategy

In Section 3.3 a Particle Filter-based approach for estimation of orbital parameters is proposed, but nothing is said about incomplete measurements. This section puts forward a strategy to address that problem.

Astronomers must often deal with the problem of partial observations. Given an image of a visual pair (obtained by means of an optical telescope or interferometric techniques), the two stars are expected to be individually resolved, so their positions in the plane of the sky can be clearly identified. The most usual situation of incomplete observations occurs when the bodies of a binary system are too close to be individually resolved in an image. In this cases, a blob-like shape appears instead of a pair of sharp points, but a slope is still noticeable. In mathematical terms, it can be said that θ is known but ρ isn't. In this article, Cartesian coordinates are used instead of (ρ, θ) representation, so incomplete observations have the form of $(X,\ missing)$ or $(missing, Y)$. Coping with polar coordinates representation is part of the future work.

The estimation method detailed in Section 3.3 requires a set of complete observations to compute Equation 35. If partial observations (i.e., only X or only Y is known for a given epoch) are incorporated into the sum, the statistical characterization of term $\mathcal{Y}$ could not be accomplished in the way presented in Appendix A. Multiple Imputation Particle Filter, summarized in Algorithm 3, is used as a means of integrating the knowledge contained in incomplete observation in the estimation routine.

The basic idea of the Multiple Imputation Particle Filter is to *fill* the set of observations by imputing values where either X or Y is missing. Thus, there are as many complete data sets as imputations generated, and those sets are evaluated in a similar way to that presented in Section 3.3: for each particle, the $\mathcal{Y}$ value is computed and then weight update is performed according to Equation 36. Let m denotes the number of imputations (and number of different data sets as well) and N the number of particles at the beginning of an iteration. Since each combination of particle and imputation results in a different vector of orbital parameters (P, T, e) and a different weight, the total number of particles would increase to $m \times N$. In order to avoid an exponential increase in the size of the particle population, number of particles is fixed at N by adding a reduction stage in which Rubin's rule is applied(Rubin, 1987).

According to multiple imputation theory, imputed values must be drawn from a proposal distribution ϕ. In this case, in absence of additional information of star position at certain epoch of observation, values of X, Y are drawn from the prior distribution $p(x_t|y_{t-1})$ (in the context of Bayesian filtering, the posterior distribution obtained by the filter after the iteration $t-1$ is used as prior in the next iteration).

Algorithm 3 Multiple Imputation Particle Filter

1. MI Importance Sampling
   for $j' = 1, \ldots, m$ **do**
      for $i = 1, \ldots, N$ **do**
         • Sample $\tilde{x}_t^{(i,j')} \sim \pi(x_t|x_{0:t-1}^{(i)}, \tilde{y}_{t-1}^{j'})$ and set $\tilde{x}_{0:t}^{(i,j')} \triangleq (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i,j')})$
      end for
   end for
   • Compute m imputations $y_t^j \sim \phi(\{\tilde{x}_t^{(i,j')}, w_t^{(i,j')}\}, \eta_t)$ and its associated weights p_t^j.
   • Reduce the particle population from $N \cdot m$ to N. $\tilde{x}_t^{(i,j)} \rightarrow \tilde{x}_t^{(i)}$
   • Defining the importance weights $w_t^{(i,j)} = w_{t-1}^{(i)} \cdot \frac{p(y_t^j|\tilde{x}_t^{(i)}) \cdot p(\tilde{x}_t^{(i)}|x_{t-1}^{(i)})}{\pi(\tilde{x}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, \tilde{y}_{1:t}^j)}$
   for $i = 1, \ldots, N$ **do**
      • Apply Rubin's rule
      $w_t^{(i)} = \sum_{j=1}^m w_t^{(i,j)}$
      $x_t^{(i)} = \frac{1}{w_t^{(i)}} \sum_{j=1}^m w_t^{(i,j)} x_t^{(i,j)}$
   end for
   for $i = 1, \ldots, N$ **do**
      • Normalize
      $\mathcal{W}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$
   end for

4. EXPERIMENTAL RESULTS

This article uses model-based, computer-generated data to test the algorithms proposed, being the incorporation of real data an important part of the future research. Both data synthesis and filtering algorithms were implemented in MATLAB. The specific experimental setup is detailed as follows:

- Orbital parameters of the visual binary known as Sirius were used to generate synthetic data, by means of equations 29, 30 and 31. Specific values of Sirius system are specified in Table 1. Observation noise has standard deviation $\sigma = 0.075\ [arcsec]$ for both X and Y axes.
- An alternative representation of parameter T is used, in order to bound its range of acceptable values. Let be t_0 the time when the observation campaign begins (it may be the epoch of the first observation, but an arbitrary value may be chosen as well). Instead of being expressed as an explicit year (year 1894.130 in the case of Sirius, for example), time of periastron passage is represented as a fraction of the orbital period P, with value $T_{alt} = 0$ being t_0 and $T_{alt} = 1$ corresponding to the year $t_0 + P$. Through different combinations of P and T_{alt} (in fraction representation), arbitrary values of T (in the explicit year representation) can be achieved, so this re-expression does not leave out any feasible solution. In other words, both representations are equally expressive, but the fractional one is more convenient in terms of the search algorithm: the random walk induced by

the Artificial Evolution of Parameters is performed in the range $[0, 1)$ instead of the possibly wider range of years $[t_0, \infty)$. From this point on, the alternative representation T_{alt} will be referred to simply as T.

- Particle Filter is run with $N = 500$ particles during $N_{it} = 40$ iterations, which proved to produce good results in practice [6]. However, more rigorous termination criteria (e.g., convergence conditions) must be investigated in future work. In Section 4.2 the multiple imputations approach is applied, with the first 20 iterations being run without taking into account the incomplete observations (no data is imputed), and the following ones with $m = 20$ imputations per cycle. This may be seen as a way of calculating a decent prior for the parameters, before the application of the Multiple Imputation Particle Filter. Particles are sampled from the a valid prior (obtained from the last iteration of the filter). In other words, the prior p.d.f. is used as importance sampling distribution. The transition model is set according to the Artificial Evolution approach (Equation 34). In MI PF, imputations are also sampled from the current prior.

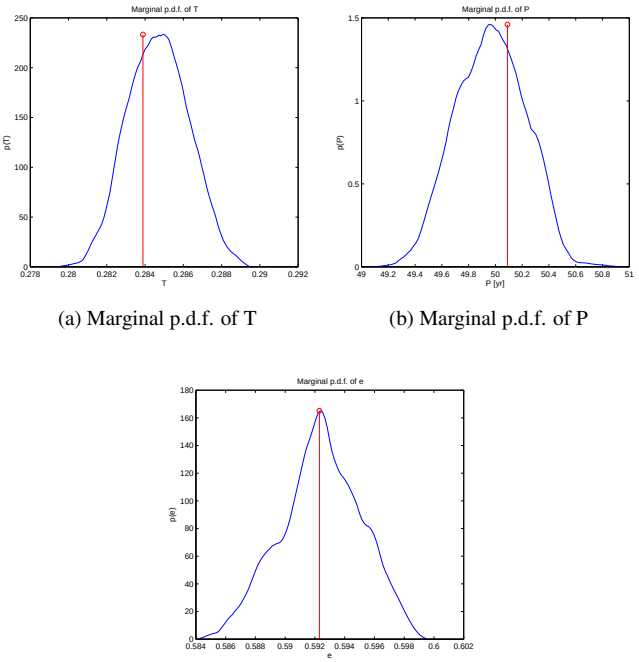
4.1. Uncertainty characterization of orbital parameters

The scope of this section is limited to show the results of particular realizations of the algorithm, in order to deliver an intuitive idea of how the method performs. The evaluation of the method according to statistical criteria is performed in Section 4.2.

Figure 4 depicts the marginal p.d.f. of parameters (T, P, e) obtained after one execution of the method proposed with complete data and, therefore, no imputations. That scenario is fully detailed in Section 4.2 and illustrated in Figure 5. Results obtained in the Data Discarding and the Multiple Imputation scenarios (Figures 6a and 6b, respectively) are quite similar in form, but with obvious differences in accuracy and precision. Figure 4 shows that, although close to the real values, the estimates obtained may present some degree of bias[7].

What allows Particle Filter to use as few as 500 particles is the dimensionality reduction accomplished by using Thiele-Innes representation. Since parameters (a, ω, Ω, i) are not directly estimated with Artificial Evolution of Parameters, their probability density functions are not shown in Figure 4, although they can be estimated through the application of the expressions shown in Equation 32 to the (A, B, F, G) values calculated for each particle.

The main conclusion of this section is that uncertainty characterization of orbital parameters can be achieved with our Particle-Filter-based implementation. On the one hand, it



(a) Marginal p.d.f. of T

(b) Marginal p.d.f. of P

(c) Marginal p.d.f. of e

Figure 4. Marginal p.d.f. of parameters (T, P, e). Red bars denote real parameter values whereas blue lines indicate their estimated p.d.f., obtained by applying the Epanechnikov kernel on set of particles.

may be possible that other algorithms (gradient-based methods, for example) may perform better in terms of accuracy, since Particle Filters have at least two key limitations: 1. it is a sample-based representation, which may fail to characterize the real p.d.f., specially when a limited number of particles is used; 2. it utilizes the concept of Artificial Evolution, which limits the achievable precision, since particles keep being perturbed even when they have reached a point of high (or even maximum) likelihood (an adaptive variance of the evolution noise may be investigated in future work). On the other hand, it must be noted that most methods found in the literature do not offer uncertainty characterization of orbital parameters, but a single point estimate.

Another advantage of this method is that, in presence of multiple feasible solutions[8] (that appear in the form of multiple and not mutually close high likelihood points in the parameter space), Particle Filters will incorporate them possible modes of the posterior p.d.f. instead just of discarding them. Gradient-based methods usually converge to the maximum/minimum solution found during the execution (regardless of whether it is a global maximum/minimum or not), discarding other feasible solutions. However, like gradient-

[6] This means that an increase in either N or N_{it} shows no significant improvement in terms of accuracy or precision of estimation.

[7] Regardless of how that estimate is being calculated –both maximum likelihood points and average values of the p.d.f. may yield values which are different from the real parameter.

[8] It is not infrequent that a number of different orbital configurations fit the same set of observations, specially when the binary system is poorly observed in terms of orbital coverage.

based methods, Particle Filters do not guarantee that all feasible solutions will be explored, since it is conditioned by the initial particle distribution.

The importance of keeping multiple feasible solutions relies in the fact that, when estimating orbital parameters, points of high likelihood do not necessarily coincide with actual parameter values. In fact, sometimes the orbital parameters estimates change dramatically when new information is incorporated (like other points of the orbit or radial velocity data). In this regard, future work will be focused on the use the uncertainty characterization to determine, given the information available, what kind of observations is needed to improve the quality of the estimation (in terms of accuracy and precision, but also in terms of amount of feasible solutions considered in the analysis).

4.2. Incomplete observations and MI PF

This section compares the performance of the method under three different scenarios, which are detailed as follows:

1. Full data set is available (Figure 5). The data set consists of $N_{obs} = 11$ observations, with the first one (first epoch) approximately at $(X_1, Y_1) = (6.8, 7.8)$ $[arcsec]$ and the last one at $(X_{11}, Y_{11}) = (4.2, 10.2)$ (this system traces a clockwise orbit). Algorithm 2 is executed.
2. Data Discarding (Figure 6a). Observations at epochs 10 and 11 are incomplete: $(X_{10}, Y_{10}) = (missing,\ Y)$, $(X_{11}, Y_{11}) = (X,\ missing)$. Those data records are discarded, and then Algorithm 2 is executed.
3. Multiple Imputations (Figure 6b). Same data set as the previous point, but Multiple Imputation Particle Filter is executed (Algorithm 3).

Figures 5, 6 depict the estimated orbit for particular, though representative, executions of the algorithm under certain scenarios but are not valid as an evaluation of the methods proposed. Table 1, on the other hand, presents the average and standard deviation values obtained after 10 repetitions of each of the scenarios described previously (standard deviation is shown between parentheses).

In Figure 5, actual and estimated orbits are almost overlapped, which suggests that the estimation is quite satisfactory. Although that figure provides just a visual insight of the method output (moreover, a particular realization of it), it is also true that the "Complete Data" row in Table 1 supports the idea that, just as theory predicts, better results are obtained when more information is available. In particular, results are better in terms of precision (lower standard deviations), since the average value is worse than some of the values obtained in Scenarios 2 and 3 (check parameter P, for example). The quality of those results may be originated by particular values associated with this specific set of realizations, and by no means implies that this particular setup generates more biased estimates.

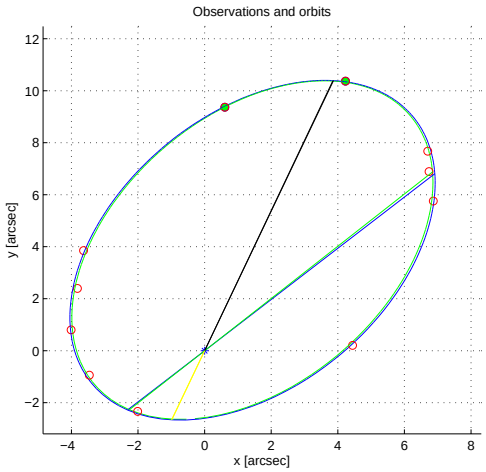


Figure 5. Orbit estimation with complete data set. Real orbit in blue, orbit estimated in green, observations as red rings. Observations affected by data loss in scenarios 2 and 3 are represented as green-filled circles.

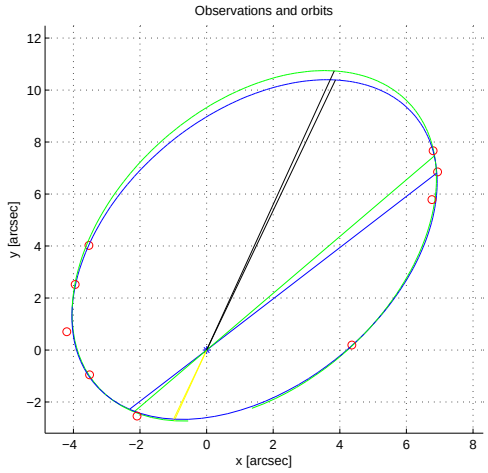
By contrast, the orbit estimate in Figure 6a shows significant differences with the actual orbit in the fraction that is not covered by observations, whereas it is quite more concordant where complete measurements are available. When the multiple imputation approach is used (Figure 6b), the estimated orbit shows significant coherence with the actual orbit (though not as much as in Figure 5), suggesting that the incorporation of partial measurements improves the quality of estimates. Table 1 supports this statement, since the estimation variance decreases when using the Multiple Imputation scheme (for most of the parameters), although not achieving the values of the "Complete Data" scenario.

Although according to Table 1 the "Data Discarding" scenario may not, in average, produce biased results, it presents higher estimation variance, specially for parameters P and a. That occurs because, by ignoring partial measurements, the method has fewer restrictions and therefore more degrees of freedom. Thus, a larger number of orbital configurations can *satisfy* the observations available (as in Figure 6a), whereas in the study cases 1 and 3 those combinations are discarded.

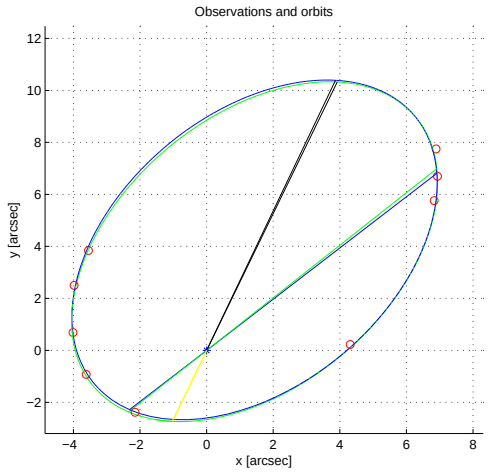
The effects of having a larger number of incomplete observations would depend on the number of complete observations. Although a large number of observations would theoretically provide more precision and accuracy on estimates, it is also true that a few, well placed, complete observations would be enough to obtain decent estimates of the orbit. Docobo's analytic method uses 3 measurements (although those measurements are required to be: i. reliable; ii. well spaced along the orbit); in (Lucy, 2014), 15 observations with more than 40% of orbital coverage provide satisfactory orbital estimates. Re-

Table 1. Average value of estimated parameters (standard deviation in parentheses)

	T	$P\ [yr]$	e	$a\ [arcsec]$	$\omega\ [rad]$	$\Omega\ [rad]$	$i\ [rad]$
Real Parameters	0.2839	50.09	0.5923	7.5000	2.5703	0.7779	2.3829
Complete data	0.2820 (0.0033)	50.4178 (0.5571)	0.5961 (0.0045)	7.5030 (0.0289)	2.5639 (0.0234)	0.7735 (0.0201)	2.3766 (0.0075)
Data Discarding	0.2859 (0.0144)	49.8637 (2.3624)	0.5909 (0.0153)	7.4874 (0.1129)	2.5669 (0.0239)	0.7717 (0.0213)	2.3823 (0.0128)
Multiple Imputations	0.2830 (0.0059)	50.2793 (0.9475)	0.5949 (0.0122)	7.5035 (0.0354)	2.5670 (0.0397)	0.7740 (0.0316)	2.3770 (0.0072)



(a) Data discarding



(b) PF with multiple imputations

Figure 6. Comparison of observations, real and estimated orbit. a) PF with discarding of incomplete observations. b) MI PF

garding incomplete measurements, they are relevant as long as they provide information not supplied by complete observations. For example, incomplete measurements around the periastron or being the only available datum in an orbital segment not covered by complete observations (the latter being the case addressed in this work) would be more valuable that a partial observation the "repeats" the information provided by the complete data. Thus, the ratio of complete vs. incomplete data is not as important as the relevance[9] of the observations available (both partial and complete).

From a computational point of view, the MI PF adds significant increase to execution times, since the particle population grows from N to $N \times m$, increasing the number of operations performed on each iteration. All algorithms were run on a personal computer with the following specifications: i5 processor (2.6 GHz), 8GB RAM, Windows 8 Operating System. Average execution times are on the order of 10 seconds for "Complete Data" and "Data Discarding" cases, whereas MI PF has execution times on the order of 3 minutes. Although computational costs of these methods are not especially large, switching from the unoptimized MATLAB code used in this work to a more computationally efficient implementation would bring some benefit when dealing with real data (e.g., data from an astronomical catalogue or an astronomical campaign, where the objects of interest may not be one but many binary stars).

5. CONCLUSION AND FUTURE WORK

A new particle-filter-based method for estimation of orbital parameters is presented. Also, a Multiple-Imputation Particle Filter is proposed to incorporate incomplete observations, as an extension of the method described in Algorithm 2. The contribution of this article is twofold: first, it adds uncertainty characterization to the problem of orbital parameters estimation; secondly, it provides a strategy to cope with incomplete observations. Preliminary results suggest that the incorporation of incomplete observations can increase the precision of the estimation without noticeable decrease in the accuracy. Potentially, it may help discard feasible solutions that would be otherwise included in the posterior p.d.f. Furthermore, the methodology here presented may be applied to other systems, as long as statistical characterization of a merit function (Mean Square Error in this work) can be accomplished.

The long-term aim of this research is the incorporation of uncertainty characterization in the planning of astronomical campaigns –given an orbit with certain parameters and the uncertainty associated, which epoch of observation would be useful in order to reduce estimation variance? Would certain observation contribute to discard a feasible solution? However, there are a number secondary aspects to be addressed as well: investigating alternative convergence criteria for the

[9]Evaluated as their informative potential. High orbital coverage and observational precision of individual measurements are the traits desirable of a visual binary data set.

particle filter, incorporating adaptive evolution noise, studying a mapping between the observational error of polar and Cartesian representations.

ACKNOWLEDGEMENTS

This work has been partially supported by FONDECYT Chile Grant Nr. 1140774 and Nr. 1151213.

We also thank the Advanced Center for Electrical and Electronic Engineering, AC3E, Basal Project FB0008, CONICYT.

REFERENCES

Candy, J. (2009). *Bayesian signal processing: Classical, modern and particle filtering methods*. Wiley.

Crisan, D., & Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, *50*(3), 736-746.

Docobo, J. (1985). On the analytic calculation of visual double star orbits. *Celestial mechanics*, *36*(2), 143–153.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, *10*(2), 197-208.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (2002). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *Iee proceedings f (radar and signal processing)* (Vol. 140, pp. 107–113).

Graham, J., Olchowski, A., & Gilreath, T. (2007). How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206-213.

Housfater, A., Zhang, X., & Zhou, Y. (2006). Nonlinear fusion of multiple sensors with missing data. *IEEE International Conference on Acoustics, Speech and Signal Processing*, *4*, 961-964.

Kitagawa, G., & Sato, S. (2001). Monte carlo smoothing and self-organising state-space model. In *Sequential monte carlo methods in practice* (pp. 177–195). Springer.

Liu, J., Kong, A., & Wong, W. (1994). Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association*, *89*(425), 278-288.

Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential monte carlo methods in practice* (pp. 197–223). Springer.

Lucy, L. (2014). Mass estimates for visual binaries with incomplete orbits. *Astronomy & Astrophysics*, *563*, A126.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

Torres, G., Claret, A., & Young, P. A. (2015). Capella (α-aurigae) revisited: Binary orbit, physical properties, and evolutionary state. *The Astrophysical Journal*, *807*(26), 15pp.

APPENDIX A

This appendix shows a proof for the likelihood function used in Section 3.3. The restriction on the observation weights aims at simplifying the statistical characterization of the Mean Squared Error $\mathcal{Y}$ of Equation 35.

- Let be $n_x \sim \mathcal{N}(0, \sigma_x)$ and $n_y \sim \mathcal{N}(0, \sigma_y)$. We assume that the observed position satisfies the following equations:

$$X_{obs} = X_{real} + n_x$$
$$Y_{obs} = Y_{real} + n_y,$$

  so differences between observed and real positions are Gaussian distributed.

- Recall the equation for Mean Squared Error:

$$\mathcal{Y}_t^{(i)} = \frac{1}{N}\Big(\sum_{k=1}^{N} \frac{1}{\sigma_x^2(k)}[X_k - X_{comp}^{k,t,i}]^2 + \sum_{k=1}^{N} \frac{1}{\sigma_y^2(k)}[Y_k - Y_{comp}^{k,t,i}]^2\Big),$$

  Each difference is squared and weighted by the inverse of its observation standard deviation σ_x or σ_y (assummed to be known). Rearranging the term $\frac{1}{\sigma^2}(P_{obs} - P_{comp})^2$ as $\big(\frac{1}{\sigma}(P_{obs} - P_{comp})\big)^2$, it is easy to see that, for a given epoch τ_k, term $\frac{1}{\sigma}(P_{obs} - P_{comp}) \sim \mathcal{N}(0, 1)$, being $P = X, Y$ and $\sigma = \sigma_x, \sigma_y$. As $\frac{1}{\sigma}(P_{obs} - P_{comp})$ follows a standard normal distribution, its square yields a χ_1^2 random variable.

- For a given k index, the expression within the sum ($\frac{1}{\sigma_x^2}(X_{obs} - Y_{comp})^2 + \frac{1}{\sigma_y^2}(X_{obs} - Y_{comp})^2$) is the addition of two χ_1^2, which follows a χ_2^2 distribution.

- Thus, $\mathcal{Y}$ is reduced to the arithmetic mean of N random variables with χ_2^2 distribution.

$$\mathcal{Y} = \frac{1}{N}\sum_{k=1}^{N} R_k, \text{ with } R_k \sim \chi_{deg}^2,\ deg = 2. \qquad (38)$$

  It can be proved that that expressions constructed as Equation 38 follow a Gamma distribution with parameters:

$$\alpha = N \cdot \frac{deg}{2},\ \ \theta = \frac{2}{N}.$$