Inverse problems of NEO photometry: Imaging the NEO population

Mikko Kaasalainen[1] and Josef Ďurech[1,2]

[1]Department of Mathematics and Statistics, P.O.Box 68,
FI-00014 University of Helsinki, Finland
email: mikko.kaasalainen@helsinki.fi

[2]Astronomical Institute, Charles University in Prague, V Holešovičkách 2,
18000, Prague, Czech Republic
email: durech@sirrah.troja.mff.cuni.cz

Abstract. The physical properties of NEOs and other asteroids are mostly obtained with photometry. The resulting models describe the shapes, spin states, scattering properties and surface structure of the targets, and are the solutions of inverse problems involving comprehensive mathematical analysis. We review what can and cannot be obtained from photometric (and complementary) data, and how all this is done in practice. The role of photometry will become completely dominating with the advent of large-scale surveys capable of producing calibrated brightness data. Due to their quickly changing geometries with respect to the Earth, NEOs are the population that can be mapped the fastest.

Keywords. data analysis; photometry; infrared; scattering; asteroid, surfaces

1. Introduction

Photometry is the main source of information on the NEO population and the other asteroids as a whole. Traditional dense lightcurves have been obtained for a large number of targets over several decades, and this mode of observation will continue in the future. However, a real paradigm change will be brought about by surveys such as Pan-STARRS and LSST that will produce colossal photometric databases that, though sparse in time, can readily be used for obtaining the physical characteristics of asteroids everywhere (Kaasalainen 2004; Ďurech *et al.* 2006, 2007). This is a unique opportunity to map NEOs (and other asteroids) both as individual targets and as a population: no other observing mode can reach such a vast number of targets. The survey datasets are efficiently enriched by any additional dense photometric or other observations.

The analysis of sparse photometric datasets will very soon become an automated industry, resulting in tens of thousands of asteroid models, a large portion of them NEOs. The computational effort in this is considerable in both computer and human time, which means that most of the targets will not be analyzed with close scrutiny: we will have to trust the computer. This, again, means that we have to have a good understanding of the reliability of our models, and this is impossible without a thorough understanding of the mathematical nature of the inverse problem(s) involved. Indeed, the inversion of photometric data involves some profound mathematical truths, and the effect of these theorems is quite visible in all parts of the inversion process. Very important concepts are the uniqueness and stability of the solution, the parameter spaces, the so-called inverse crimes in simulations and error prediction, and the domination of systematic errors over random ones.

151

This paper is organized as follows. In §2 we summarize and discuss some main facts about photometric inversion from both theoretical and practical points of view, and investigate the role of systematic errors. Some particular properties of NEO (and other asteroid) observations and modelling are discussed in §3, where we also introduce a convenient way of producing synthetic asteroids and a useful method for initial period scanning of sparse datasets. In §4 we discuss the sparse observation mode in combination with dense lightcurves and/or other data such as thermal infrared or radar. We sum up and discuss future observations in §5.

2. Inversion of photometric data

Asteroid lightcurve observations and their analysis have perhaps three main components. In historical order these are:

I. Period analysis for almost 2000 asteroids, including the detection of many binary systems.

II. Full physical (spin, shape and scattering) modelling from combined datasets, also with data other than photometric (including, e.g., radar, stellar occultations, thermal infrared, and adaptive optics).

III. A vast quantity of physical models using accurately calibrated photometry from large surveys (Pan-STARRS, LSST, etc.) as the main database.

Item I has resulted in statistically important catalogues by Pravec, Harris and others (Pravec *et al.* 2007; Harris, private communication): despite the inevitable observational selection effects, we are beginning to have some idea of the period distribution of asteroids. Item II has produced the first reasonably large (more than 100 objects) catalogue giving us some idea what asteroids are really like: how their spin axes are distributed in space, what kinds of shapes and irregularities they exhibit, what their actual (spin/shape corrected) solar phase curves are like, what we can say about their surface properties, etc. We now have several ground truths from space missions, laboratory studies, etc. from which we know that photometric modelling gives a good global portrait of the target. For example, the Keck adaptive optics images of several asteroids coincide, within uncertainties of the two methods, with the predicted plane-of-sky images from photometry-based models determined earlier (Marchis *et al.* 2006; the same level of correspondence is also obtained between the Hubble Space Telescope images of Storrs *et al.* (1999) and the photometry-based predictions). Similar shape and pole agreement is also obtained with the laboratory model of Kaasalainen *et al.* (2005), or the radar and fly-by target models in Kaasalainen *et al.* (2001). Indeed, we can say that a good set of photometric data essentially enables us to image an asteroid crudely.

Similarly, we know that combining thermal infrared observations with these models yields, e.g., accurate estimates of surface regolith properties (Mueller *et al.* 2005). The models can even be used for getting a colour map of the surface using data at different wavelengths and thus gain some insight on mineral distributions (Nathues *et al.* 2005). The spin properties can reveal evolutionary surprises, in particular in connection with the YORP effect from thermal radiation (Vokrouhlický *et al.* 2004; Bottke *et al.* 2006 and references therein). While the level of detail from groundbased observations cannot reach that of in situ space missions, the latter are going to remain few in number. Photometry alone has the chance to give us a well representative and statistically significant coverage of the physical properties of asteroids and asteroid populations such as NEOs. Photometric inversion is now a routine process, so obtaining the models is straightforward. We should also like to note that, from the mathematical point of view, lightcurve inversion is of considerable interest as it is one of the rather few difficult inverse problems that

now have rigorous uniqueness and stability results and a robust, well-converging practical numerical procedure.

2.1. *Fundamentals of the problem*

There are a few fundamental theorems on lightcurve inversion that we review here in a nutshell as they are the key to understanding the potential and limitations of photometric inversion. For details, see Kaasalainen *et al.* (2001, 2003) and Kaasalainen & Lamberg (2006) and references therein. With ω and ω_0 we denote the viewing and illumination directions on S^2 (unit sphere), and L is the measured brightness (intensity).

THEOREM 2.1 (UNIQUENESS THEOREM). *$L(\omega_0, \omega)$-data on $S^2 \times S^2$ uniquely determine the curvature of a convex surface. In other words, different viewing and illumination geometries break the Russell degeneracy $\omega = \omega_0$, i.e., the ambiguity in modelling a target from the areas of its simple projections.*

In fact, many asteroid observations such as photometric, radar, interferometric, or adaptive optics ones, are mathematically close cousins: they are essentially representations of generalized projection operators. This is why the inverse problems of these data modes are closely related and can be studied with a common approach (Kaasalainen & Lamberg 2006). Uniqueness theorems can also be shown for, e.g., radar analysis.

It can also be shown that the curvature function of a convex surface $\mathcal{B}$ uniquely determines its shape $\mathbf{x}_\mathcal{B}$ (up to a translation of $\mathbf{x}$), and that the shape construction converges (the so-called Minkowski problem, see Lamberg & Kaasalainen 2001).

COROLLARY 1. *$S^2 \times S^2$-data uniquely determine the shape of a convex surface: the mapping $L(\omega_0, \omega) \to \mathbf{x}_\mathcal{B}$ is unique.*

THEOREM 2.2 (STABILITY THEOREM). *The mapping $L(\omega_0, \omega) \to \mathbf{x}_\mathcal{B}$ is continuous for convex bodies in usual topologies (the inverse problem is conditionally well-posed in the sense of Tikhonov).*

We call this *Minkowski stability*; its role is now understood to be important in protecting the solution not only from random noise, but also from systematic errors in both data and modelling.

It should be noted here that the above theorems assume that the intrinsic darkness (albedo) of the surface is uniform. So far, probe data from asteroids have shown no striking albedo contrasts over large areas. Also, noticeable violation of the convexity condition can be used as an indicator of albedo asymmetry in inversion, and so far only a couple of asteroids have displayed moderate asymmetry in this sense (Kaasalainen *et al.* 2003). Significant albedo variegation without asymmetry would be physically rather implausible, so albedo uniformity is apparently quite well satisfied on global resolution level. Furthermore, Minkowski stability applies to *shape* determination, not to attributing brightness variations to albedo variegation on the surface, so the former is much safer than the latter. In other words, visible instability in the inferred albedo map is replaced by minor changes in the inferred shape (one can imagine a polyhedron modification where the facet areas are changed here and there, but the overall shape changes little due to the basic nature of the Minkowski problem).

There is a fundamental difference between the stability and reliability of convex and nonconvex modelling, and between the parameter spaces the two employ. The convex modelling is performed in the parameter space describing the Gaussian image of a shape, and this image is then transformed into shape information in radius space (Kaasalainen & Lamberg 2006 and references therein). Nonconvex inversion is performed in the radius

space ($\mathbb{R}^3$) directly, which makes the whole process much more vulnerable and ambiguous. On conjectural level, it appears that the photometric $L(\omega_0, \omega)$ data of at least a starlike nonconvex body (at least when observed at solar phase angles α sufficiently high) determine its shape as robustly as with a convex body, but *only if* the scattering function is known accurately and the noise level is low (Kaasalainen *et al.* 2001; Ďurech & Kaasalainen 2003). In all such simulations we have carried out, a nonconvex model efficiently converges towards the correct minimum of χ^2 (with suitable regularization of radius variation), faithfully displaying the same features as the target even with model resolution (discretization of the problem) much lower than that of the simulation model. This seems to be a mathematically interesting extension of the uniqueness theorem for convex bodies. However, accurate photometry of real targets of known nonconvex shapes but with inadequately known (and modelled) scattering functions does not yield similar convergence (and the fit remains worse than with a convex model). This is an example of how taking the numerical conjecture directly as a model of reality would be an "inverse crime" (Kaipio & Somersalo 2005) leading to overoptimistic results: nonconvex photometric inversion is sensitive to (systematic) errors and the insufficiency of the scattering model, and in reality the scattering behaviour is never known well enough for this purpose. On the other hand, due to Minkowski stability, *convex* inversion is quite stable against the incorrectness of the scattering model (including slight albedo variegation) or other systematic errors.

We discuss the pole longitude ambiguity theorem with some detail here as the concepts are necessary in later discussion in the paper. Let $\mathbf{x} = (x, y, z)$ denote a vector in a coordinate system fixed to the target (i.e., rotating with it, z-axis aligned with the rotation axis), and $\mathbf{x}'$ a vector in a nonrotating system (denoted by primes) where the rotation vector points at the direction given by the spherical coordinates (θ', φ') (e.g., ecliptic or equatorial coordinates; rotation is in the positive direction around this vector, with period P). Then $\mathbf{x}'$ and $\mathbf{x}$ are related by

$$\mathbf{x} = \mathsf{R}\,\mathbf{x}', \tag{2.1}$$

where

$$\mathsf{R} = \mathsf{R}_z(\phi_0 + \Omega(t - t_0))\mathsf{R}_y(\theta')\mathsf{R}_z(\varphi'), \tag{2.2}$$

where t is the time, $\Omega = 2\pi/P$, ϕ_0 and the epoch t_0 are some initial values, and $\mathsf{R}_i(\phi)$ is the rotation matrix corresponding to the rotation of the coordinate frame through angle ϕ counterclockwise about the positive i-axis. In particular, $\mathsf{R}_z(\phi)$ is

$$\mathsf{R}_z(\phi) = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.3}$$

THEOREM 2.3 (AMBIGUITY THEOREM). *If the viewing and illumination directions ω' and ω_0' in a nonrotating frame remain in the same plane at all times of observation t_i, the infinite-distance observations of a body $\mathcal{B}$, with surface points $\mathbf{b} = (x, y, z)$ and rotation vector $\beta'(\theta', \varphi')$ given in a coordinate system whose $x'y'$-plane is the invariant plane, are indistinguishable from those of a body $\hat{\mathcal{B}}$ with $\hat{\mathbf{b}} = (x, y, -z)$ and $\hat{\beta}' = \beta'(\theta', \varphi' + \pi)$. We call this ecliptic degeneracy.*

Proof: We choose the invariant plane defined by ω' and ω_0' as the $x'y'$-plane of the nonrotating system, so the z'-coordinates of the viewer and the illumination source are zero. From (2.3) we have

$$\mathsf{R}_z(\varphi' + \pi)\mathbf{x}'|_{z'=0} = -\mathsf{R}_z(\varphi')\mathbf{x}'|_{z'=0}, \tag{2.4}$$

so (2.2) yields

$$\mathbf{x}(\varphi' + \pi, \mathbf{x}'|_{z'=0}) = -\mathbf{x}(\varphi', \mathbf{x}'|_{z'=0}). \tag{2.5}$$

Since ϕ_0 is arbitrary, we can set $\hat{\phi}_0 = \phi_0 + \pi$ (as $(x, y) \to (-x, -y)$ corresponds to a trivial shape rotation of π in the xy-plane). Therefore a vertical mirror-image shape $\hat{\mathcal{B}}$ ($z \to -z$) with a rotation direction changed by $\varphi' \to \varphi' + \pi$ has the same viewing and illumination directions with respect to the body shape as $\mathcal{B}$ and thus yields exactly the same observations as those of $\mathcal{B}$. □

This ambiguity property affects all data that are not two-dimensionally resolved in a plane projection, i.e., in this sense equivalent to observations made at infinity. Thus it appears also with radar data in addition to photometric observations. The coplanarity of ω' and ω'_0 is often the approximate case as many asteroids move close to the plane of the ecliptic. For such targets, only observations with resolved plane-of-sky projections can properly remove the spin direction ambiguity.

Both Russell and ecliptic degeneracies break fast: asymmetric shapes can easily produce asymmetric lightcurves at solar phases less than $10°$ (i.e., asymmetry or nonzero first Fourier harmonic of the lightcurve does *not* imply albedo variegation), and ecliptic observation latitudes higher than $10 - 15°$ can already be sufficient for a unique pole solution (depending on the dataset).

2.2. *Surface scattering properties*

The main signature of surface scattering properties is the asteroid's phase curve, particularly its behaviour near zero phase angle. Proper definition of a phase curve necessarily includes pole and shape modelling, as then we can plot both the disk-integrated phase curve for the whole body (with equatorial viewing geometry for standard reference) and the phase curve for the surface material alone as if it were measured from a sandbox (Kaasalainen *et al.* 2001, 2004). If the size of the target is known, we obtain a scale factor for the latter curve, automatically defining the intrinsic darkness of the material (rather than use less well-defined concepts such as geometric albedo). Proper modelling also automatically produces, e.g., the "amplitude-phase relationship" which is thus actually a redundant concept. Often the phase curves are not very well described by the H-G system; usually it is better to take the phase curves as such (as they are readily reproduced from a limited set of parameters) and derive from them any particular values needed.

In this context, we should discuss the choice of the scattering model in some detail, rather than take Hapke's or other models for granted. From the general modelling point of view, we actually should not use a ready-made scattering model: instead, we should let it be constrained only by basic principles. Thus, the scattering model S should be of the general form $S(\mu, \mu_0, \alpha)$ (with the natural assumption that scattering from the surface material is invariant in the rotation of the surface patches in the tangent plane) where μ, μ_0 are the customary normal cosines of the emergent and incident rays. (Using α here is essentially equivalent to using the azimuthal angle between the two rays.) With some physical guidance, a general form for such a model is

$$S = f(\alpha)S_0 = f(\alpha)\frac{\mu\mu_0}{\mu + \mu_0} \sum_{ijk} b_{ijk}\mu^i \mu_0^j \cos(k\alpha), \tag{2.6}$$

with $b_{00k} \equiv 0$, $b_{ijk} \equiv b_{jik}$ due to ray reciprocality, and $\cos(k\alpha)$ due to symmetry.

The Lommel-Seeliger and Lambert combination typically used in lightcurve inversion (Kaasalainen *et al.* 2001) is thus equivalent to using only the lowest-order coefficients b_{000} (set to unity) and $b_{100} = b_{010}$. It does not have to have any particular physical

justification: we just use a suitable truncated series of the general scattering form. As it happens, it already mimics "physical" scattering models very well, as discussed in Kaasalainen *et al.* (2001). In particular, the separation of a phase function $f(\alpha)$ is physically consistent as by far the most of α-dependent variation of scattering is in this part. From the inversion point of view, this is enhanced by Minkowski stability: the result is not very dependent on the scattering model. Thus there is no practical necessity to carry the scattering series expansion further; indeed, many further coefficients would cause instability as the disk-integrated data do not contain proper information on them. However, it is interesting to note that just by including the next two coefficients, b_{001} and $b_{101} = b_{011}$, i.e.,

$$S_0 = \mu\mu_0 \left[\frac{1 + c_1 \cos\alpha}{\mu + \mu_0} + c(1 + c_2 \cos\alpha) \right], \tag{2.7}$$

we can already get a virtually exact match with the four-parameter Hapke model, and by adding the next few orders we get the same for the five-parameter rough-surface Hapke model (Helfenstein & Veverka 1989). This underlines the fact that any scattering behaviour can be expected to be modellable with a short scattering series, even when physical scattering models are not sufficient or are ambiguous. Disk-resolved data and other additional information can be used in determining the coefficients. Also, solving for b_{ijk} is essentially a linear problem, which is easier than, e.g., the determination of Hapke parameters.

2.3. *Systematic errors*

The scattering behaviour of asteroid surfaces can be studied accurately with in situ measurements; so far, only (433) Eros and possibly Itokawa have been studied extensively enough for this purpose. Li *et al.* (2004) derived a set of Hapke parameters for Eros from disk-resolved data. It is natural to ask whether the exactly known shape model (Konopliv *et al.* 2002), combined with the determined Hapke model, can reproduce the observed lightcurves (Lagerkvist *et al.* 2001) – in other words, can the whole of Eros' surface be described by a uniform scattering model? The answer is a conditional yes. Most lightcurves are fitted perfectly; however, some show clear discrepancies. Upon closer inspection, the discrepancies are due to systematic data errors. An example of this is shown in Fig. 1. The first and last lightcurves are fitted exactly, whereas the middle ones, essentially in the same geometries, are not. Thus the fit deviation is caused by persistent systematic errors in the data, even though the random noise level is low. This is a good example of the domination of systematic errors over random ones.

Another example of systematic (modelling) errors in inversion is presented by Ida and Gaspra. As discussed in Simonelli (1995, 1996), no single scattering model over the surface can be found that would fit the data with the known shape models (down to the noise level and without any systematic offset etc. effects). Scattering apparently varies over the surface. On the other hand, we get an excellent fit with our low-order scattering model and convex shape. This underlines both the role of Minkowski stability and the fact that local (nonconvex) shape details cannot be obtained from disk-integrated data. In short, systematic errors in both data and the model set a resolution limit to our modelling.

How much can we expect our result to be off due to insufficiently well modelled (or indeed not accurately modellable) scattering? While the effect of random noise is easy to study, unknown systematic effects are much harder to predict. Thus, it is important to search for ground truth whenever possible. In this respect, the laboratory study of Kaasalainen *et al.* (2005) was particularly useful as there we modelled a target with a decidedly incorrect scattering law. Even so, the shape obtained was surprisingly accurate, and the worst pole estimate differed by less than ten degrees of arc from the correct pole
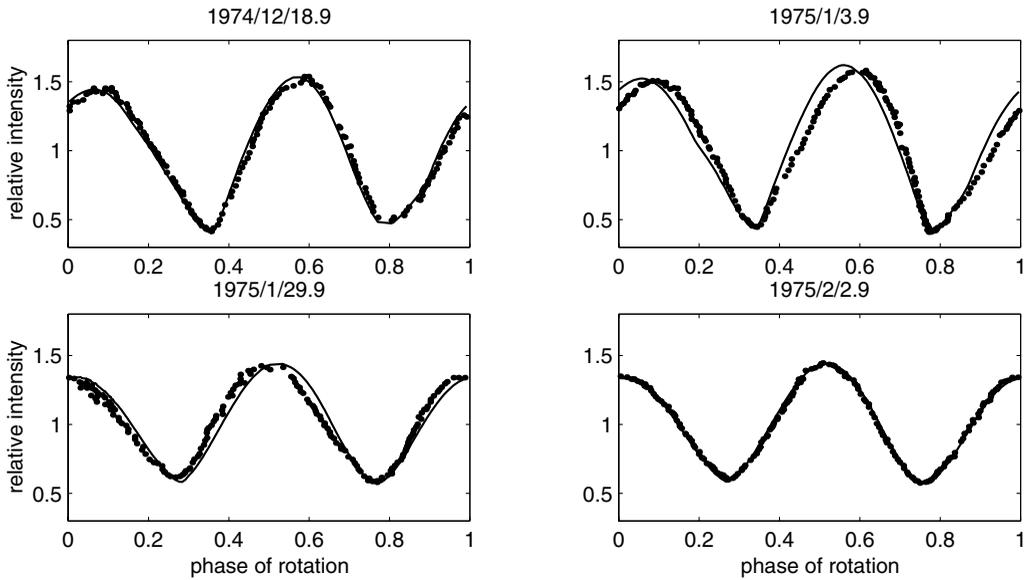
Figure 1. Examples of simulated lightcurves of the probe-based Eros model against observations: two lightcurves have systematic errors.

direction, while the best one was only a few degrees away. This seems to set an upper "worst case" limit – again Minkowski stability comes to our rescue. Accuracy from the ground truth by space probes is consistently good: the poles in Kaasalainen *et al.* (2001) were only a few degrees away from the in-situ-determined ones, while the pole for Itokawa (Kaasalainen *et al.* 2006) was only one degree away from the Hayabusa-determined one (Demura *et al.* 2006). Thus we can usually safely set a standard error of five or ten degrees of arc to pole estimates from lightcurve inversion.

Systematic errors can also muddle the details of asteroid models even with data sources that are expected to be more informative than photometry. For example, delay-Doppler radar techniques ostensibly provide an excellent chance of obtaining detailed shape information. However, ground truth from Itokawa shows that the shape inversion of the apparently detailed high signal-to-noise radar plots did not yield the most obvious global nonconvex feature ("otter's head") of the asteroid – an indication of the feature was visible, but its actual nature was not revealed (compare the figures in Ostro *et al.* 2005 and Demura *et al.* 2006). In this case this was mostly due to the very limited observing geometries, but nevertheless Minkowski stability in convex inversion is apparently a rather unique (and lucky) phenomenon as a guard against systematic errors in data and the model. With partially disk-resolving data sources such as radar, there is the danger of producing overdetailed models with features that are not actually there.

3. NEO observations and modelling

The foremost characteristic of NEOs is the quickly changing and wide-ranging observing geometry. The spin solution seldom has the ecliptic degeneracy as the observational ecliptic latitudes are often sufficiently high; just one or two apparitions may already offer suitably varying geometries for modelling, and the solar phase angle α reaches large values. Typical NEO photometric observations and their modelling are presented in Kaasalainen *et al.* (2004).

3.1. *The effect of solar phase angle*

The larger the phase angle α, the further away the target is from the Russell degeneracy $S^2 \times S^2 \to S^2$, i.e., the larger the shading effects, the more information there is on the shape (and spin). In principle, photometry at large phase angles obviously carries information on nonconvex shape features. In practice, this information is seldom obtainable as the phase angle should be unrealistically high, especially since the scattering properties are not modellable accurately enough.

It is possible to express a rough relationship between the size of nonconvex features of a body and the minimum solar phase angle $\alpha_{\min}$ needed for the photometric detection of those features (Ďurech & Kaasalainen 2003). A convenient measure of the nonconvexity degree of a body is the dimensionless quantity of *nonconvexity measure* $\mathcal{V}_{\mathrm{nc}}$, $0 \leqslant \mathcal{V}_{\mathrm{nc}} < 1$ defined as

$$\mathcal{V}_{\mathrm{nc}} = 1 - \frac{V}{V_{\mathrm{ch}}}, \tag{3.1}$$

where V is the volume of the body and V_{ch} the volume of its convex hull. For example, for Gaspra $\mathcal{V}_{\mathrm{nc}} = 0.05$, while for Castalia (radar), Eros, Ida, and Kleopatra (radar) the corresponding values are, respectively, 0.08, 0.15, 0.18, and 0.36. For two ellipsoids in contact with each other, $\mathcal{V}_{\mathrm{nc}} = 0.2$, while for two separated ellipsoids with an ellipsoid-size gap $\mathcal{V}_{\mathrm{nc}} = 0.5$.

If a set of lightcurves can be explained with a convex shape down to the noise level, this set obviously contains no information on nonconvexities; $\alpha_{\min}$ is defined as the phase angle at which a convex model no longer fits the data as well as a nonconvex one. A large set of both synthetic and real shape models displays a clear correlation between $\alpha_{\min}$ and $\mathcal{V}_{\mathrm{nc}}$ (Ďurech & Kaasalainen 2003); we roughly have

$$\alpha_{\min} = 120° - 220° \mathcal{V}_{\mathrm{nc}}. \tag{3.2}$$

Usually only small asteroids have $\mathcal{V}_{\mathrm{nc}} > 0.2$, so it is clear that very few MBAs can show any photometric information on nonconvexities. Even strongly nonconvex NEOs seldom have $\mathcal{V}_{\mathrm{nc}} > 0.3$, so photometry cannot detect nonconvexities on most NEOs either.

Surface undulation gives more pronounced shadowing effects at larger phase angles, as is well known from, e.g., the role of the surface roughness parameter in the Hapke model. However, simple crater simulations (Kaasalainen *et al.* 2004) show that statistical cratering mostly yields the same correlation effect as $\alpha_{\min}$ for separate nonconvex features. Only datasets of very densely cratered surfaces at high solar phase angles ($\alpha > 90°$) cannot be explained with a locally smooth surface (i.e., a matte-like surface containing only small-scale roughness included in the scattering law).

3.2. *Modelling considerations*

More complex dynamical behaviour than constant-period principal-axis rotation is found in some NEOs due to their small size. Any well-modellable dynamics is straightforward to analyze in the lightcurve inversion scheme: we just modify the rotation equation (2.2) suitably. Precessing motion (and its photometric distinguishability from binary motion) is discussed in Kaasalainen (2001, *et al.* 2003) and Pravec *et al.* (2005, 2007). YORP effect causes an essentially linear change in the angular speed Ω (for some range of time) due to thermal torque (Vokrouhlický *et al.* 2004, Bottke *et al.* 2006):

$$\Omega(t) = \Omega(t_0) + \upsilon \Delta t, \tag{3.3}$$

where the effect parameter is υ, and $\Delta t = t - t_0$. Thus the period change rate is

$$\dot{P} = -\frac{\upsilon}{2\pi} P^2 \qquad [\Rightarrow P(t) = (\frac{\upsilon}{2\pi} \Delta t + P(t_0)^{-1})^{-1}]. \tag{3.4}$$
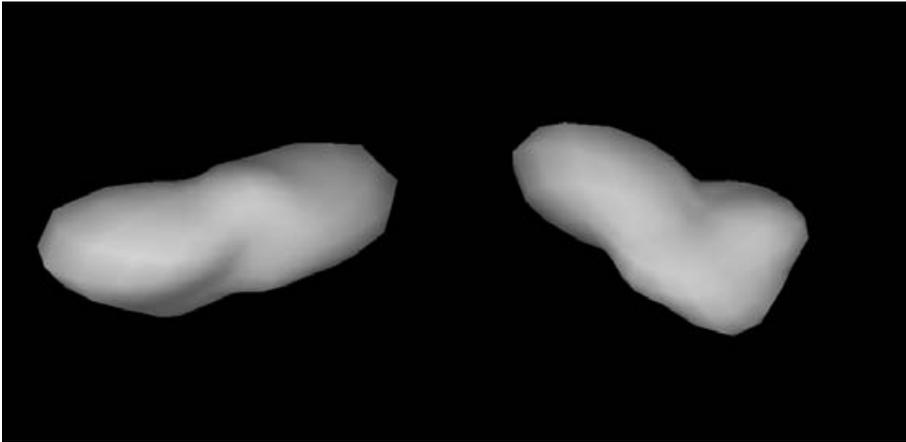
Figure 2. Offspring of Golevka and Eros (left), and Ida and Gaspra (right).

This causes a phase lag quadratic in time, described by just the one additional parameter υ that can be directly included in inversion. To be detected, the phase lag $\delta\phi$ due to YORP must be larger than the linear phase shift due to period uncertainty ΔP (estimated at an epoch earlier than the comparison time Δt) and the difference between YORP base period $P(t_0)$ and the best-fit constant period P_c:

$$\delta\phi = \frac{1}{2}|\upsilon|(\Delta t)^2 > \Delta t \Big[\frac{2\pi}{P_c^2}\Delta P + \Big|\Omega(t_0) - \frac{2\pi}{P_c}\Big|\Big]. \tag{3.5}$$

The latter contribution to the linear shift occurs since the constant-period fit tries to compensate for YORP effect by finding a period slightly different from $P(t_0)$ such that the phase shifts are minimized over the timeline used for period determination. The phase lag as such is basically just a visual effect. The ultimate test and measurement of YORP effect is the inclusion of υ in (2.2) in inversion: if the resulting χ^2 is clearly better with a distinctly nonzero υ, the change in rotation is properly measurable.

As demonstrated by the colour map of Eunomia in Nathues *et al.* (2005), the use of various filters can provide information on the global variegation of surface material. Surveys such as Pan-STARRS will obtain data through a number of filters, so even sparse datasets should be sufficient for at least indicating colour variegation.

Since even very small NEOs are observable, the target shapes can vary considerably, as can be seen from the variety of models in Kaasalainen *et al.* (2004). In modelling synthetic datasets, this should be taken into account. Arbitrary synthetic shapes can be produced in many ways; here we introduce a rather natural means of obtaining quasiasteroidal shapes. Rather than use random-generated realizations of statistical or fractal measures for describing surface undulation and global shape variation, we can draw from the supply of existing real shape models and modify these. The results look usually more realistic than synthetic random models. For example, we can combine the shapes of two asteroids in a genetic fashion; with such asteroid breeding from an initial population, we can obtain as large and as varying an offspring population of models as desired. In Fig. 2 we show two such examples from pairing Eros and Golevka (Konopliv *et al.* 2002; Hudson *et al.* 2000) or Ida and Gaspra (Thomas *et al.* 1994, 1996). These were generated by expressing the original shapes (with the same volume and centroids at origin) as spherical harmonics series, and then picking coefficients for a new series by randomly switching between the two series. Coefficients can also be "mutated" randomly. Bodies generated in this manner present typical-looking features – indeed, one can breed a series of varying shapes with

some common main characteristics, if wanted. Morphological characteristics, such as the "banana" shape of Eros, are easy to preserve in this way.

3.3. *Sparse photometry and surveys as future basis*

It is at first sight almost counterintuitive that sparse sequences are sufficient for asteroid modelling as shown in Kaasalainen (2004) and Ďurech *et al.* (2006, 2007). After all, the sampling interval is for most targets much longer than the sidereal period, so ordinary time-series methods such as Fourier or power spectrum analysis (used for initial period estimation of traditional lightcurves) are completely useless. The reason for the sufficiency is that the underlying well-defined mathematical model is highly constraining: only a certain type of an object can create a given sparse sequence as long as there are enough points at various observing geometries. In fact, sparse sequences are handled just like the ordinary lightcurve inversion problem: the mathematical model takes care of filling the gaps. Now we just have to scan a wide range of potential periods as the rotation period is not apparent in the data before the actual modelling.

The YORP effect can easily be included in the inversion just like for dense lightcurves. For the latter, one can directly see the phase lag in plots; for sparse sets, there have to be enough data points in different apparitions so that the lagged phases affect the fit χ^2 properly, i.e., the inclusion of the YORP parameter υ significantly improves the fit. One may also see an indication of YORP if the constant-period fit is good for one time interval while the deviations grow larger further away from it. For certain targets, υ should possibly be included at the outset as otherwise the effect might go unnoticed (there could still be a clearly best period for the dataset, and the increased deviation would be taken for noise).

A typical requirement for a sparse dataset is some 100 well-distributed data points over five years for main-belters, while for NEAs even less is sufficient (Kaasalainen 2004). The calibration accuracy should be at least around 0.05 mag, so the new surveys can indeed meet the requirements. The surveys can make use of data at smaller solar elongations than those typical for ordinary lightcurves since only one point is needed at a time; thus the geometry coverage is wider (i.e., the observational gaps between apparitions are narrower). This is why the observation geometry range of groundbased surveys is not really very much smaller than that of satellite-based ones.

3.3.1. *Fast initial period scanning*

The sparse datasets have to be scanned for all potential period ranges; in certain cases there are some indications of the range (fast or slow rotators; see Ďurech *et al.* 2007), but basically the sparse data do not show any period signatures prior to modelling. Doing full modelling for each possible period "slot" roughly separated by the interval $\Delta P \approx (1/2)P^2/T$, where P is the trial period and T the timeline of the dataset, is somewhat time-consuming. The period ranges can effectively be initially scanned by using a simplified model to detect the most likely values or value ranges. Indeed, in many cases even using just a brick to model the asteroid will highlight the best few period locations. We have found that a practical model for initial scanning is simply a geometrically scattering ellipsoid (Ostro and Connelly 1984) with linear-exponential phase function $f(\alpha)$ (Kaasalainen *et al.* 2001). Thus the model reads (with an arbitrary scaling factor absorbed in $f(\alpha)$ if desired)

$$L = f(\alpha)\Big[\sqrt{\mathbf{e}^T\mathsf{M}\mathbf{e}} + \frac{\mathbf{e}^T\mathsf{M}\mathbf{s}}{\sqrt{\mathbf{s}^T\mathsf{M}\mathbf{s}}}\Big], \tag{3.6}$$

where

$$f(\alpha) = A_0 \exp\left(-\frac{\alpha}{D}\right) + k\alpha + 1, \tag{3.7}$$

with A_0, D, k as free parameters, and

$$\mathsf{M} = \begin{pmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/b^2 & 0 \\ 0 & 0 & 1/c^2 \end{pmatrix}, \tag{3.8}$$

and

$$\mathbf{e} = \mathsf{R}\mathbf{e}', \quad \mathbf{s} = \mathsf{R}\mathbf{s}', \tag{3.9}$$

where $\mathbf{e}'$ and $\mathbf{s}'$ are, respectively, the unit vectors of the Earth's and the Sun's positions in an inertial coordinate system (equatorial or ecliptic), and R is given by (2.2). The model has only nine parameters: three for $f(\alpha)$, two for semiaxes a, b (setting $c = 1$), two for the pole direction, and P, ϕ_0 for period and initial rotational phase (of a-axis). These parameters are optimized very fast by the Levenberg-Marquardt algorithm (Press *et al.* 1990) as both the model and the gradients w.r.t. parameters are very simple to evaluate in the fully analytical form above. (Using a more realistic scattering model for the ellipsoid would be useless as then the numerical computation of the surface integral would take just as long as for the proper full model.) To ensure positive values for a, b as well as realistic shapes ($a, b > c = 1$) it is useful to write the semiaxes in the form

$$a = 1 + \exp(p), \quad b = 1 + \exp(q), \tag{3.10}$$

where p, q are the parameters to be optimized. It is interesting to note that though the best fit from this initial model is, of course, much worse than that obtained with a full model, the fit for incorrect periods is also considerably worse than with a full model. In other words, the level of the typical chi-square "thicket", below which the best period must stand out, increases, making it possible for the good period(s) to show even if their actual fits (and the rest of the simplified model) are bad. Of course this does not apply to all targets, but the majority can be initially scanned in this manner (Ďurech *et al.* 2007).

4. Combined datasets

4.1. *Sparse and dense photometry*

Of course one cannot expect to get too much out of a handful of data points, especially if a number of these are noisier than expected (there are bound to be several outliers no matter what the engineers say). The models from sparse photometry are rough and in many cases (mildly) nonunique. This is where follow-up observations and observer networks come in. Even just one additional dense lightcurve would be of great help in at least the following cases:

1. *There are more than one possible periods fitting the sparse data.* This happens if the number of sparse data points is subcritical. The number is a complicated function of survey strategy and technical choices and can thus vary a lot. This may also happen with faint targets for which data are noisy, or with very spheroidal targets. Also, since the number of objects is so large, each target will only be given some standard computer time for analysis. Most of this time is spent in period scanning, and some targets will run out of the allocated pipeline time for period search as there will be both very fast and very slow rotators. Such targets will thus be flagged with "period not found" and saved
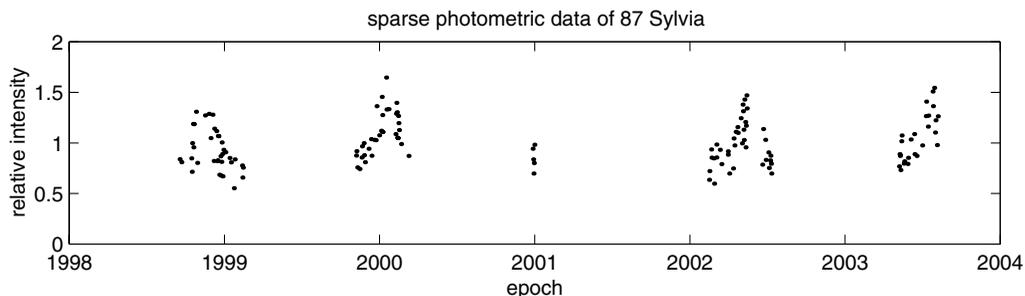
Figure 3. Sparse USNO dataset of (87) Sylvia from five apparitions (143 epochs). The observed intensity is reduced to unit distance from the Earth and the Sun.

for later analysis. Additional lightcurves will help to determine what the actual period region is.

2. *Even if the period is known, there may be more than one independent pole solutions.* This, again, happens with slightly too small sparse datasets. Here one should note that for objects moving close to the plane of the ecliptic, the ecliptic degeneracy cannot be removed by photometric means, regardless of the method.

3. *There are sparse data points that just cannot be fitted.* This usually means that the target is a binary (mutual events affect some data) or a tumbler (or otherwise somehow strange and thus interesting), or the points are just outliers. A dense lightcurve can help in clearing the matter. We expect several targets to be flagged for follow-up observations in this manner.

4. *Quality and reliability check.* Even if everything seems to be fine with the sparse data analysis and we get a full model, we must do random checks to make sure that essentially the same model pops out from the combined sparse and dense datasets. If the models are different, we have overlooked something.

5. *More detail needed.* If the object seems to be strangely shaped, we need more data points to get additional details.

The following example well portrays the power of combined databases, even when the survey data are very noisy. Even a limited additional dataset thus greatly boosts the value of the survey set. In this case, it actually makes the inversion possible in the first place as the sparse set alone is insufficient for modelling due to noise. The sparse set of asteroid 87 Sylvia (Fig. 3) was extracted from the astrometric database available at the Asteroids Dynamic Site (AstDys); the data were obtained at the U.S. Naval Observatory, Flagstaff. The dense lightcurves of the same asteroid in Fig. 4 essentially correspond to just one observation geometry, simulating a typical additional dataset for a target. These were chosen from the Uppsala Asteroid Photometric Catalogue (Lagerkvist *et al.* 2001). Lightcurve inversion of the full set of all available 40 lightcurves of Sylvia leads to the shape model shown in Fig. 5, left. The model derived from only the four lightcurves and sparse data (Fig. 5, right) is somewhat different as expected as the USNO data are obviously very noisy, more than 0.1 mag on average. Nevertheless, the model portrays many similar global characteristics. Above all, the pole directions of the two models are only seven degrees of arc apart and the rotation periods are the same. The inversion would not have been possible with either dataset alone.

There are currently almost 2000 asteroids for which one or more lightcurves have been observed – thus we can expect that many of these can be modelled after only a few years of survey photometry. For targets with limited datasets it will probably be useful to run cross-check analyses by using the dataset geometries and a synthetic model; this will give
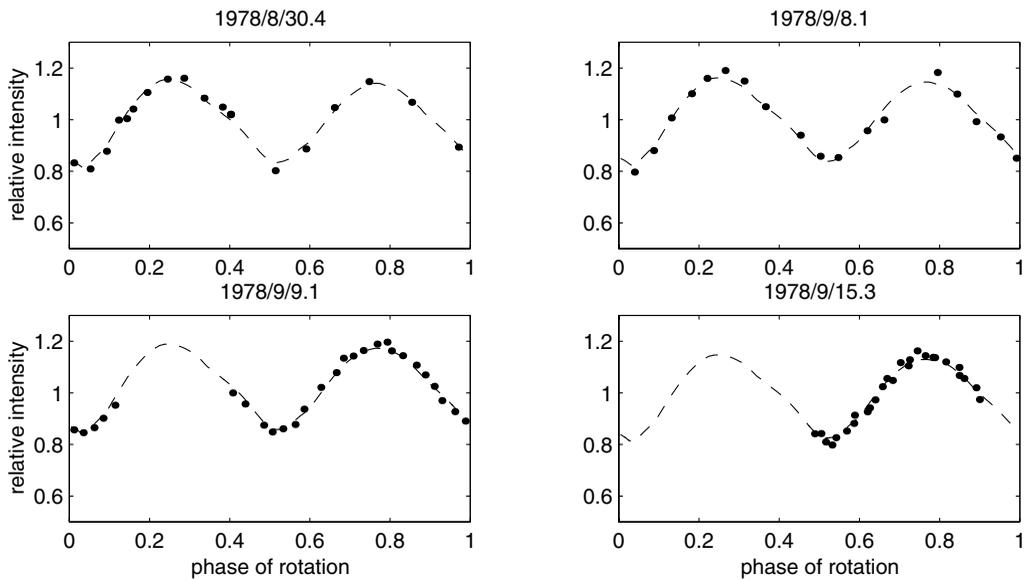
Figure 4. Dense lightcurves of Sylvia as an additional dataset. The dashed curve is the fit from the shape model.

Figure 5. Left: shape model (pole-on view) of 87 Sylvia from all dense lightcurves (Uppsala catalogue), right: the model from the noisy sparse+dense data. The spin solutions are essentially the same.

an indication of how much the result of the analysis might be off (even when the result appears to be unique and well-behaving).

4.2. *Multidatainversion with generalized projection operators*

Multidatainversion is usually based on photometry combined with any other data sources. Typically, photometry (sparse or dense) already provides a good estimate of the parameters, and the complementary source is employed to yield a detailed solution. Most often the improvement lies in revealing nonconvex features of the body or in removing the ecliptic degeneracy.

The general principle of multidatainversion is to form a joint goodness-of-fit $\chi^2_{\rm tot}$ that combines the χ^2 from the main source with the χ^2_i of the complementary sources i, multiplied by suitable weights λ_i. Thus we have

$$\chi^2_{\rm tot} = \chi^2 + \sum_i \lambda_i \chi^2_i. \tag{4.1}$$

The weights are adjusted in minimizing $\chi^2_{\rm tot}$ with the condition that each χ^2_i as well as χ^2 be acceptable. This condition usually leads to a certain degree of nonuniqueness in the solution as there may be several feasible sets of weights that fulfill the condition and
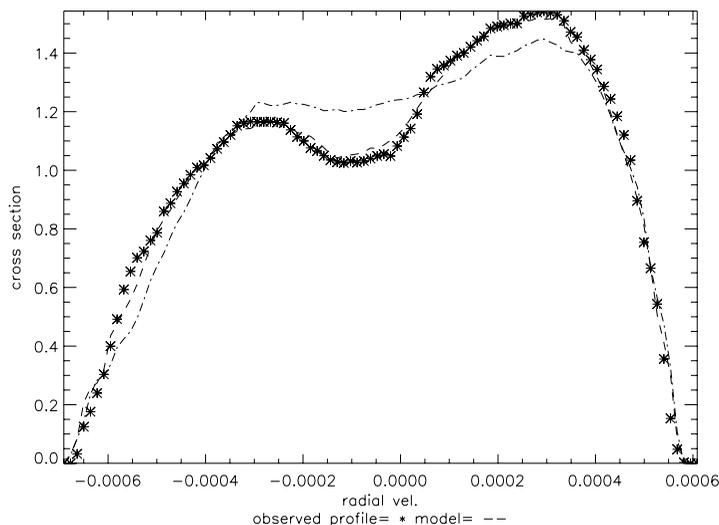
Figure 6. Doppler spectrum of the target (asterisks), with fits from convex model based on photometry alone (dot-dash) and the nonconvex model based on combined photometry and radar (dashed line).
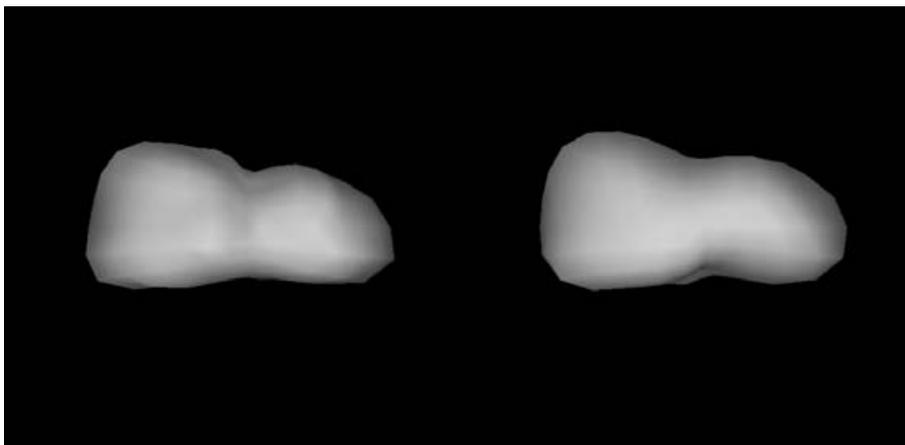
Figure 7. View of the target (left) and the nonconvex model (right) in the same direction in which the Doppler spectrum was computed.

lead to virtually equal values of χ^2_{tot}. Also, the exact values of the maximal allowed χ^2_i are usually not well defined in practice: these depend not only on the noise levels of the individual sources, but also on systematic effects such as the expected reliability of a source. Furthermore, insufficiency of the model affects fits for separate data sources differently. In practice, the multidatainversion results from generalized projection operators appear to be quite stable; one reason for this is the stability of the solution from the main source, photometry.

Here we demonstrate the important role of even small datasets of complementary information in multidatainversion. In Fig. 6 we show the simulated Doppler radar spectrum (in arbitrary units) of the target on the left in Fig. 7. The viewing direction ω is the same in both cases (note that the sign convention of the Doppler velocity makes the spectrum look like a "mirror image"). The convex inversion of simulated photometric

data produces a model close to the convex hull of the target, giving an excellent fit to the data. Thus the data do not contain proper information on the sizable nonconvex features of the target. However, the convex model gives a poor fit to the observed Doppler spectrum (dot-dash line), so even one spectrum already contains significant information when added to other data. Using a spherical harmonics series for the radius of the (starlike) model and Levenberg-Marquardt algorithm to minimize χ^2_{tot}, we obtained the nonconvex model displayed on the right in Fig. 7. The generalized projection integrals were computed with a densely tessellated polyhedral surface representation. While still giving an excellent fit to photometric data, the model also fits the Doppler spectrum very well (dashed line), and well reproduces the main features of the target (with lower resolution: largest degree and order l, m for the function series are both 6).

As shown in Kaasalainen & Lamberg (2006), similar multidatainversion can be performed with, e.g., interferometric data. For NEOs, radar is the most important additional data source. With survey photometry analysis, we can expect to detect interesting targets (binaries, tumblers) that need radar (as well as dense lightcurve) observations for complete analysis. Such synergy will make all observations much more time-efficient than now. Another important additional data source is thermal infrared. Combining photometry-based models with IR, we can get very good estimates of the sizes and regolith properties of NEOs, as Hayabusa ground truth from Itokawa well confirms (Mueller *et al.* 2005, 2007).

5. Conclusions

Photometric datasets at well-covered observing geometries essentially allow us to image asteroids and to obtain their spin states (including tumbling, YORP effect, etc.) and scattering characteristics. Observations of NEOs offer widely ranging and quickly changing geometries, so data sufficient for modelling are usually obtained faster than for other asteroids. The future of photometric observations and analysis lies in sparse datasets obtained in surveys, and often also in their combination with other datasets. Systematic errors dominate over random noise, which should be remembered in the analysis and error estimation. The most time-consuming part of sparse data analysis is the period determination, but it can often be greatly speeded up with a simplified model for initial period scanning.

We will obtain thousands of good NEO models from surveys such as Pan-STARRS and LSST. With such numbers, the selection effects and biases are smaller than hitherto as the surveys simply record everything. Obtaining additional data (radar, dense lightcurves, etc.) will also be more time-efficient as we know what to observe and when, and even small amounts of additional data will improve sparse photometry considerably. This makes observer networks very important in coordinating databases and follow-up work.

Software packages, manuals and links for lightcurve inversion (and links for downloading) can be found at `www.rni.helsinki.fi/~mjk`. While the programs are straightforward to use, the user must have some knowledge of what they require and produce, and how to interpret the result. This paper and some of the references here should be sufficient for the purpose, but experience is the only good teacher. We hope that the open software will be developed further by various users.

References

Bottke, W.F., Vokrouhlický, D., Rubincam, D.P. & Nesvorný, D. 2006, *Annu. Rev. Earth Planet. Sci.* 34, 157

Demura, H. & 19 colleagues 2006, *Science* 312, 1347

Ďurech, J. & Kaasalainen, M. 2003, *A&A* 404, 709

Ďurech, J., Grav, T., Jedicke, R., Kaasalainen, M. & Denneau, L. 2006, *Earth, Moon & Planets*, in press

Ďurech, J., Scheirich, P., Kaasalainen, M., Grav, T., Jedicke R. & Denneau, L. 2007, this volume

Helfenstein, P. & Veverka, J 1989, in: R.P. Binzel *et al.* (eds.), *Asteroids II* (Tucson: University of Arizona Press), 557

Hudson, R. & 26 colleagues 2000, *Icarus* 148, 37

Kaasalainen, M., Torppa, J. & Muinonen, K. 2001, *Icarus* 153, 37

Kaasalainen, M. 2001, *A&A* 376, 302

Kaasalainen, M., Mottola, S. & Fulchignoni, M. 2003, in: W.F. Bottke, A. Cellino, P. Paolicchi & R.P. Binzel (eds.), *Asteroids III* (Tucson: University of Arizona Press), p. 139

Kaasalainen, M. & 21 colleagues 2004, *Icarus* 167, 178

Kaasalainen, M. 2004, *A&A* 422, L39

Kaasalainen, M. & Lamberg, L. 2006, *Inverse Problems* 22, 749

Kaasalainen, M. & 20 colleagues 2006, Proceedings of the 1st Hayabusa Symposium, in press

Kaasalainen, S., Kaasalainen, M. & Piironen, J. 2005, *A&A* 440, 1177

Kaipio, J. & Somersalo, E. 2005, *Statistical and computational inverse problems*, Springer, New York

Konopliv, A., Miller, J., Owen, W., Yeomans, D., Giorgini, J., Garnier, R. & Barriot, J. 2002, *Icarus* 160,289

Lagerkvist, C.-I., Piironen, J. & Erikson, A. 2001, Asteroid Photometric Catalogue, 5th update. Uppsala Univ. Press, Uppsala.

Lamberg, L. & Kaasalainen, M. 2001, *J. Comp. Appl. Math* 137, 213

Li, J., A'Hearn, M. & McFadden, L. 2004, *Icarus* 172, 415

Marchis, F., Kaasalainen, M., Hom, E., Berthier, J., Enriquez, J., Hestroffer, D., Le Mignant, D. & de Pater, I. 2006, *Icarus* 185, 39

Mueller, T., Sekiguchi, T., Kaasalainen, M., Abe, M. & Hasegawa, S. 2005, *A&A*, 443, 347

Mueller, T., Sekiguchi, T., Kaasalainen, M., Abe, M. & Hasegawa, S. 2007, this volume

Nathues, A., Mottola, S., Kaasalainen, M. & Neukum, G. 2005, *Icarus* 175, 452

Ostro, S.J. & Connelly, D. 1984, *Icarus* 57, 443

Ostro, S.J. & 12 colleagues 2005, *Meteor. Planet. Sci.* 40, 1563

Pravec, P., Harris, A., Scheirich, P. *et al.* 2005, *Icarus* 173, 108

Pravec, P., Harris, A. & Warner, B. 2007, this volume

Press, W., Flannery, B., Teukolsky, S. & Vetterling, W. 1990, *Numerical Recipes*, Cambridge University Press

Simonelli, D., Veverka, P., Thomas, P. & Helfenstein P. 1995, *Icarus* 114, 387

Simonelli, D., Veverka, P., Thomas, P., Helfenstein, P. & Carcich B. 1996, *Icarus* 120, 38

Storrs, A., Vilas, F., Landis, R., Wells, E., Woods, C., Zellner, B. & Gaffey, M. 1999, *Icarus* 137, 260.

Thomas, P., Veverka, J., Simonelli, D., Helfenstein, P.,Carcich, B., Belton, M., Davies, M. & Chapman, C. 1994, *Icarus* 107, 23

Thomas, P., Belton, M., Carcich, B, Chapman, C., Davies, M., Sullivan, R. & Veverka, J. 1996, *Icarus* 120, 20

Vokrouhlický, D., Čapek, D., Kaasalainen, M. & Ostro, S.J. 2004, *A&A* 414, L21