

# Determining the parameters of high amplification microlensing events by means of statistical machine learning techniques

Elena Fedorova

Astronomical Observatory of National Taras Shevchenko University of Kyiv, Observatorna str.3, Kiev 04053, Ukraine  
email: [efedorova@ukr.net](mailto:efedorova@ukr.net)

**Abstract.** Strong gravitational microlensing (GM) events provide us a possibility to determine both the parameters of microlensed source and microlens. GM can be an important clue to understand the nature of dark matter on comparably small spatial and mass scales (i.e. substructure), especially when speaking about the combination of astrometrical and photometrical data about high amplification microlensing events (HAME). In the same time, fitting of HAME lightcurves of microlensed sources is quite time-consuming process. That is why we test here the possibility to apply the statistical machine learning techniques to determine the source and microlens parameters for the set of HAME lightcurves, using the simulated set of amplification curves of sources microlensed by point masses and clumps of DM with various density profiles.

**Keywords.** gravitational microlensing, high amplification microlensing events, statistical machine learning, etc.

---

## 1. Introduction

It is considered nowadays as almost undoubted that the Universe is composed of 26.8% non-baryonic dark matter (DM) and 68.3% dark energy (represented by  $\Lambda$  term), and just 4.9% ordinary matter (Ade *et al.* 2013). However, there are many open questions concerning DM, such as the particles DM consists of, interactions DM takes part in, and the distribution of DM in space, i.e. is it “warm”, “cold” or “hot”, self-interacting, repulsive or not, etc. These questions are directly connected with cosmological models of the Universe, as well as with each other. Cold dark matter (CDM) is supposed to be composed of weakly interacting massive particles (axions or WIMPs). Self-interacting dark matter (SIDM) is a particular kind of CDM with a nonzero impact distance for the interaction between DM particles (Kormendy & Freeman 2003, Rocha 2012). The cosmological N-body simulations within CDM and SIDM models (Schneider *et al.* 2010, Diemand *et al.* 2005, Stadel *et al.* 2009, Vogelsberger *et al.* 2012) have shown that the spatial distribution DM is strongly inhomogeneous, compactified in hierarchical structures surrounded by less dense continuously distributed DM, formed in “bottom-up” process. The lower estimated limit on the substructure masses depends on the particular kind of hypothetical DM particles, however, for these “primordial” clumps it is anyway below the typical planetary mass (Contini 2011). The mass density distribution in DM structures within the CDM model is often assumed to follow the cusped NFW profile of Navarro *et al.* (1996). In the velocity-dependent SIDM the density profile of a DM clump has a core and follows Burkert profile (Rocha 2012), or pseudo-isothermal sphere (Kormendy & Freeman 2003), contrarily to the cuspy coreless NFW profile typical for CDM clumps.

In Warm Dark Matter (WDM) models the substructure formation is a complex hybrid with the leading role is played by the “top-down” formation process (Knebe 2008), that is why the lower limit on the substructure mass is significantly higher than in CDM/SIDM ones: e.g.,  $10^6 - 10^{11} M_\odot$  (Hisano 2006). The most massive DM structures were predicted by CDM and WDM models both and they have been recognized already: these are galactic and galaxy cluster DM haloes. But should note that only the existence of the more massive members of this hierarchy of DM structures has been proven from the results of observations. The situation with substructures of stellar masses is not so clear, however, this information can be a crucial clue to understand is DM “warm” or “cold”, because in WDM cosmology there are no substructures of such low masses.

In this context, the role of Gravitational Lensing (GL) and particularly gravitational microlensing (GM) can hardly be overestimated. High amplification microlensing events (HAME) due to stellar-mass objects on the lightcurves of stars of extragalactic objects can give us possibility to reconstruct the density profile of a microlens. In the WDM cosmology, we may predict the non existence of DM gravitational microlensing, CDM and SIDM clumps have different density profiles distiguishable from each other as well as from the point mass microlens, as we had shown in Fedorova *et al.* (2016). The other interesting situation we can meet is a star or a star-mass black hole surrounded by a dense cloud of DM trapped by its gravity, or ultra compact mini halo (UCMH) proposed by Zhang (2011). In the simplest case such an object can be modeled up by a Schwarzschild or Chang-Refsdal lens with overcritical convergence.

Photometric signatures of DM substructure via gravitational microlensing have been widely discussed; see Mao *et al.* (2004), McKean *et al.* (2007), Oguri (2005), Shulga *et al.* (2014), Zhdanov *et al.* (2012) and references therein. In our previous works Fedorova *et al.* (2014), Fedorova *et al.* (2016) we also had considered the same models for DM clumps to understand are they distiguishable in principle from the photometrical and astrometrical observational data.

In this paper, we pay special attention to the question of classification of HAME lightcurves using the techniques of statistical machine learning (SML). We use several models of DM clumps to simulate the training and testing datasets of amplification curves of sources with different sizes microlensed by objects with various kinds of objects (star, cuspy/cored clump, UCMH) of different sizes. We focus our attention on the possibility to classify the microlens from HAME data and the percentage of misclassified cases in the sets worked out using the different SML techniques.

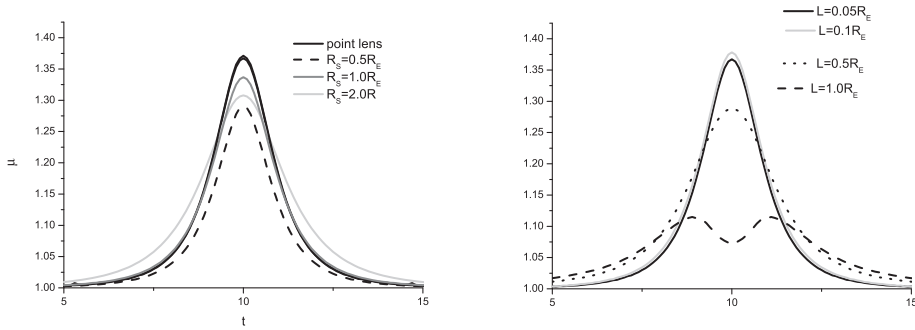
## 2. Microlens models and amplification curves

We use the microlensing equation in its normalized form; the distances will be expressed in units of an Einstein radius  $R_E^2 = \frac{4GM D_d s}{c^2 D_d D_s}$ , where  $D_s$  is the distance between the source and the observer,  $D_d$  the distance between the deflector and the observer, and  $D_{ds}$  the distance between the source and the deflector. Furthermore  $\mathbf{y} = \{y_1, y_2\}$  is the normalized angular source position,  $\mathbf{r} = \{r_1, r_2\}$  is the normalized angular image position,  $r = |\mathbf{r}|$ . We remind also that after normalization both these vectors are defined in the lens plane.

The normalized ray-tracing lens equation looks as follows:

$$\mathbf{y} = (1 - \sigma)\mathbf{r} - \mathbf{r} \frac{R_C^\alpha}{r^\alpha} F(r),$$

where  $F(r) = 1$  and  $\sigma = 0$  for star or BH (point) microlens;  $F(r) = \frac{r^\alpha}{R_C^\alpha}$  and  $\sigma = 0$  for cusped clump of the size  $R_C$  with density index  $\alpha$ ;  $F(r) = \frac{r^2}{r^2 + R_S^2}$  and  $\sigma = 0$  for



**Figure 1.** Left panel: Amplification curves for the source with size of  $0.05R_E$  at  $1.0R_E$  impact distance for different sizes of a SIDM clump. Right panel: Amplification curves for the source with  $1.0R_E$  impact distance microlensed by point lens for different sizes of a source.

cored SIDM clump of the size  $R_S$ , and  $F(r) = 1$  and  $\sigma = -2$  for the star or black hole surrounded by a trapped DM cloud (UCMH).

The ray-trace equation is supposed to be normalized to a corresponding length scale,  $\sigma = const$  stands for a convergence (optical depth) that can be due to, e.g., a background DM. To calculate the amplification curves the following formula was used:

$$\mu = \frac{\int B(\mathbf{r})rdr}{\int B(\mathbf{r})rdr}$$

where  $B(\mathbf{r})$  is a brightness distribution over the source. Below we show some amplification curves generated with different models corresponding to a straight line motion of the microlensing system with respect to the line-of-sight towards the remote point source (Fig. 1).

For the testing of the classification algorithms were created two sets of amplification curves with 300 curves for impact distances of a source in the range  $(0.0-3.0)R_E$ , source size in the range  $(0.05-1.0)R_E$ , and lens size  $(0.1-1.0)R_E$ .

### 3. SML classification of the simulate sets of amplification curves

Of course, the most correct way to distinguish the DM-induced HAME is to fit the concrete observational data. There exist many observational data compiled by EROS (Experience pour la Recherche d’Objets Sombres), OGLE (Optical Gravitational Lens Experiment) and other groups (Tisserand *et al.* 2007, Wyrzykowski *et al.* 2009), however, one should note that the fitting process is quite time-consuming every for one singular curve, and also, that there is some part of them which are not good enough to perform correct and statistically significant fitting. Using the SML techniques can be more effective to perform this task, paying the price of some percentage of misinterpretations. Here we test several SML algorithms to find one which gives the lowest percentage of outliers (badly interpreted inputs). To find a better way to classify the amplification HAME curves, we used the following SML techniques:

- linear regression;
- k nearest neighbours for  $k=3$  and  $k=5$  (non-parametric method to determine the parameters by means of averaging between its k nearest neighbours from the training set);
- random tree (non parametric supervised learning method to predict the parameters values usnig simple decision rules inferred from the input set features (Hastie 2009));
- random forest of k trees (uses a number of classifying trees on various subsets of data to improve the accuracy (Breiman 2001));

**Table 1.** Correlations for parameters of HAME amplification curves classified with various SML methods.

Method	Impact distance, $R_E$	Lens size, $R_E$	Source size, $R_E$	$\sigma$
Linear regression	0.62	0.74	0	0.953
k-NN, k=3	0.87	0.954	0.988	0.986
k-NN, k=5	0.85	0.943	0.98	0.987
Random tree	0.74	0.788	0.986	0.989
Random forest	0.83	0.953	0.991	0.995
Decision table	0.82	0.941	0.967	1
Multilayer perceptron	0.84	0.975	0.988	0.997

- multilayer perceptron (artificial neural network model (Collobert & Bengio 2004), mapping the input data set to output through multiple conditional layers of nodes connected with each other).

To perform the training/testing processes, we have used the publicly available software Weka 3.8 provided by the University of Waikato: <http://www.cs.waikato.ac.nz/ml/weka/>.

The speediest and the most accurate appears to be the Random Forest technique, giving only 0.9% of outliers determining the lens size. This looks quite promising for further tasks of the determining the lens type and size from an observational data, however one should note that the situation with this data can be more complicated due to errors (which also have to be taken into account) or incomplete data on particular lightcurve. Also, we stress that we need to include in our consideration more complex situations (such as, for instance, binaries), different brightness profiles of the source as well as foreground field (for the extragalactic cases).

## References

- Ade, P. A. R., *et al.* 2013, *A&A*, 1303, 5062.
- Collobert, R., Bengio, S., 2004. Proc. Int'l Conf. on Machine Learning (ICML).
- Contini, E., De Lucia, G., & Borgani, S. 2012, *MNRAS*, 420, 2978.
- Diemand, J., Moore, B., & Stadel, J. 2005, *Nature*, 433, 389-391.
- Fedorova, E., Del Popolo, A., Zhdanov, V. I., *et al.* 2014, Renc. Moriond 49th. Cosmology, eds: E.Auge, J.Dumarchez and J. Tran Thanh Van *al.* Proc., 407.
- Fedorova, E., Sliusar, V. M., Zhdanov, V. I., *et al.* 2016, *MNRAS*, 457, 4147.
- Hastie, T., Tibshirani, R., Friedman, J., *Elements of Statistical Learning*, Springer, 2009.
- Hisano, J., Inoue, K. T., & Takahashi, T. 2006, *Phys. Lett. B.*, 643, 141.
- Knebe, A., Arnold, B., Power, C., & Gibson, B. K. 2008, *MNRAS*, 386(2), 1029.
- Kormendy, J. & Freeman, K. C. 2003, IAU Symp.220 Eds: S.D.Ryder *et al.* San Francisco: Astron. Soc. Pacific., 377; Astro-Ph/0407321.
- Mao, S., Jing, Y., Ostriker, J. P., & Weller, J. 2004, *ApJ*, 604, L5.
- McKean, J. P., *et al.* 2007, *MNRAS*, 378(1), 109.
- Navarro, J., Frenk, C., & White, S. 1996, *ApJ*, 462, 563.
- Oguri, M. 2005, *MNRAS*, 361(1), L38.
- Rocha, M. E., *et al.* 2012, *MNRAS*, 430(1), 81.
- Shulga, V., *et al.* Dark energy and dark matter in the Universe. V.2 , 2014, Akademperiodics, 357p.
- Stadel, J., Potter, D., Moore, B., *et al.* 2009, *MNRAS*, 398, L21.
- Tisserand, P., Le Guillou, L., Afonso, C., *et al.* 2007, *A&A*, 469, 387.
- Wyrzykowski, L., Kozłowski, S., Skowron, J., *et al.*, 2009, *MNRAS*, 397, 1228.
- Vogelsberger, M., *et al.* 2012, *MNRAS*, 423, 3740.
- Zhdanov, V., Alexandrov, A., Fedorova, E., *et al.* 2012, *ISRN A&A*, 2012, ID 906951.
- Zhang, D. 2011, *MNRAS*, 418(3), 1850.