

Raport științific și tehnic pentru proiectul ParaSOL

etapa 3 (01/01/2024 – 20/06/2024)

-- Testare experimentală și validare funcțională --

Competiția:	Proiect experimental demonstrativ - PED2021
Nr. contract:	685PED din 21/06/2022
Cod proiect:	PN-III-P2-2.1-PED-2021-3625
Domeniul de cercetare:	Spațiu
Titlul :	Detecția obiectelor din Sistemul Solar și a resturilor spațiale folosind metode de calcul paralel
Acronim:	ParaSOL
Buget etapa 3:	144.908,00
Pagina web proiect:	https://planet.astro.ro/ParaSOL/
Instituția coordonatoare:	Institutul Astronomic al Academiei Române
Director de proiect:	Dr. Marcel M. Popescu
Partener 1 proiect (P1):	Universitatea din Craiova

1. Rezumatul etapei

Activitățile corespunzătoare ultimei etape a proiectului ParaSOL au inclus:

- ✓ *Act 3.1 - WP4.3 - Împachetarea serviciului de calcul pentru distribuție. Conectarea nodurilor de calcul la nivel central și distribuirea operațiilor*
- ✓ *Act 3.2 - WP5.2 Validarea folosind camerele mozaic de pe telescoapele mari*
- ✓ *Act 3.3 - WP5.3 - Identificarea secțiunilor de cod costistitoare din punct de vedere al timpului de procesare și optimizarea acestora*
- ✓ *Act 3.4 - WP5.4 - Caracterizarea performanțelor (rata de detecție, "fals pozitive", "fals negative", dimensiunea maximă a camerei pentru care softul poate rula, și cadența imaginilor pentru reducerea în timp real)*
- ✓ *Act 3.5 - WP1.5 - Diseminarea rezultatelor proiectului - etapa 3*

Pentru a putea realiza aceste activități am folosit sistemele de calcul parasol1 și parasol2 (instalate la Institutul Astronomic al Academiei Române), parasol3 parasol4 (instalate la Universitatea din Craiova). Sistemul de calcul parasol2 include un sistem RAID de stocare având capacitatea 60 TB (și încă 30TB instalați pentru paritatea sistemului) necesar pentru stocarea numărului mare de imagini generat în urma observațiilor efectuate. Pe mașina parasol1, pus în funcțiune în etapa anterioară, a fost rulează un sistem de control al versiunilor softurilor dezvoltate de tip git precum și documentația proiectului organizată pe pagini html de tip wiki. De asemenea pe mașina parasol1 este rulată interfața grafică a softului nostru, numită Webrella.

Aceasta este accesibilă online folosind un browser web (testată pe Firefox și Opera). Sitelele *parasol1* și *parasol2* pot fi accesate online.

2. Sumarul rezultatelor etapei și gradul de realizare al obiectivelor

1. Propus: Pachetul software Umbrella calificat ca nivel de maturitate tehnologică 4.

Realizat:

a). Pachetul software este disponibil folosind serverul de control al versiunilor (*git*) instalat pe sistemul de calcul *parasol1* (necesită autentificare):

<https://gitplanet.astro.ro/umbrella/>

Componentele pachetului software au fost testate sub sistemele de operare Ubuntu, ArchLinux și Windows.

b). Toate pipeline-urile rulează pe sistemul *parasol1* și pot fi accesate prin autentificare de către orice utilizator care folosește acest sistem. Modulele IPP și STU rulează pe sistemul de calcul *parasol2* și pot fi utilizate pe orice sistem de calcul ce folosește un sistem de operare Linux sau Windows și care permite framework-ul OpenCL.

c). Pentru vizualizarea și validarea rezultatelor rulării softului a fost realizată **interfața web Webrella**. Aceasta este accesibilă:

<https://gitplanet.astro.ro:9443/>

d). Toate modulele îndeplinesc criteriile de maturitate tehnologică de nivel 4.

Nivelul de maturitate tehnologică TRL 4 corespunde validării în condiții de laborator a pachetului de date Umbrella și a componentelor sale (IPP, STU, Webrella). Pentru a face acest lucru am realizat o serie de programe largi de observații folosind telescoapele cu apertură între 25 cm și 2.54m și câmp de observație per imagine între 7×7 arcmin² până la $2^\circ \times 2^\circ$ respectiv 0.25m T025-BD4SB/IAAR, 0.8 m T80S – Javalambre South/Chile, 1.52 m TCS/IAC, 1.6 m KASI – Chile și 2.54 m INT/ ORM. În figura 1 este prezentată comparația între măsurătorile efectuate cu Umbrella/STU și coordonatele precise. Dispersia este mai mare pe INT din cauza problemelor mecanice ale telescopului (urmărire siderală imprecisă).

O altă componentă a validării pachetului de programe a fost detecția asteroizilor apropiați orbitei terestre necunoscuți la momentul observației. Menționăm aici ca exemple relevante (prin prisma faptului că sunt obiecte apropiate orbitei terestre și care la momentul descoperirii au avut viteze aparente până la 10 arcsec/min) următoarele descoperiri:

- ◆ 2023 DZ2 (MPEC 2023-F12)
- ◆ 2023 XC8 (MPEC 2023-X212)
- ◆ 2023 XN11 (MPEC 2023-X270)

◆ 2024 CW2 (MPEC 2024-C133)

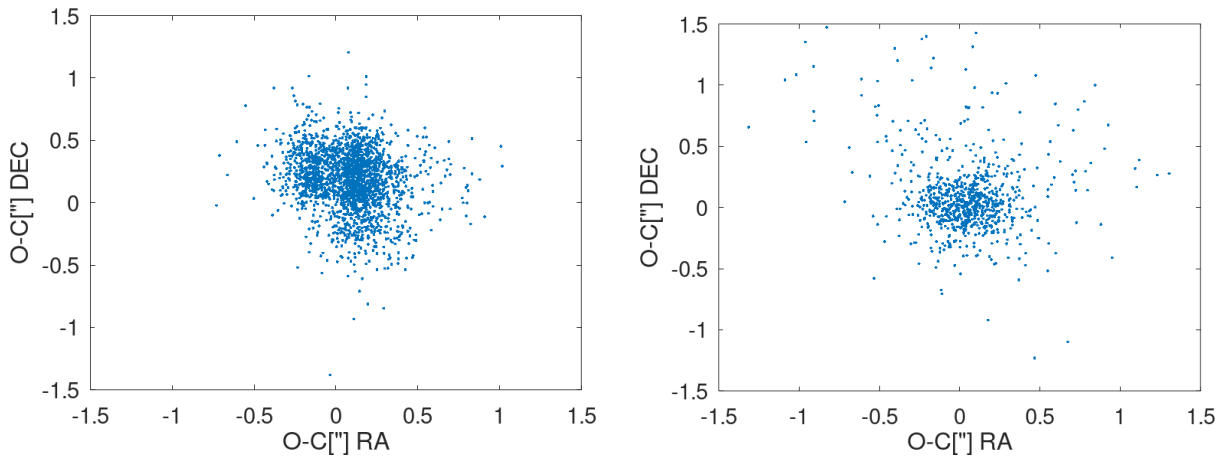


Fig. 1 Diferența dintre coordonatele măsurate de STU și coordonatele calculate din efemeridă (O – C) pentru obiectele cunoscute observate cu KASI (stânga) și INT (dreapta).

O altă componentă a validării pachetului de programe a fost detectia asteroizilor apropiați orbitei terestre necunoscuți la momentul observației. Menționăm aici ca exemple relevante (prin prisma faptului că sunt obiecte apropiate orbitei terestre și care la momentul descoperirii au avut viteze aparente de până la 10 arcsec/min) următoarele descoperiri:

- ◆ 2023 DZ2 (MPEC 2023-F12)
- ◆ 2023 XC8 (MPEC 2023-X212)
- ◆ 2023 XN11 (MPEC 2023-X270)
- ◆ 2024 CW2 (MPEC 2024-C133)

Documentația asociată pachetului de programe include: 1) Descrierea generală a metodelor IPP și a algoritmului STU este prezentată în secțiunea 2 din articolul publicat în *Astronomy & Astrophysics*, Popescu et al. 2023. În acest caz softul a fost utilizat în mediu operațional, în condiții de program larg de observații (all -sky survey).

<https://www.aanda.org/articles/aa/pdf/2023/08/aa46751-23.pdf>

2) Pe website-ul proiectului, în secțiunea Rezultate, se regăsesc toate documentele asociate prezentărilor la conferințe și seminarilor ținute de-a lungul proiectului.

3) Specificațiile tehnice sunt disponibile pe website-ul proiectului.

4) Proiectul include o documentație de tip html – wiki disponibilă intern pe serverul git al proiectului.

5) Software-ul dezvoltat include documentația funcțiilor și tipurilor de date interne, inclusiv a setărilor de configurare și metodelor folosite, într-un format XML ce poate fi exportat ca HTML sau PDF folosind software-ul de procesare a documentației Doxygen.

În concluzie, capabilitățile acestui pachet de programe au fost clar descrise și demonstrate în diferite aplicații (sesiuni de observații pentru detectarea obiectelor

noi, sau puțin cunoscute din apropierea Pământului), conform propunerii de proiect. 2) Sistemul funcționează ca un ansamblu în mediu real. 3) Sistemul a fost clar și riguros modelat și testat demonstrând satisfacerea cerințelor enunțate în propunerea de proiect. 4) Softul poate fi integrat pentru a procesa datele provenite de la programele largi de observații sau poate fi particularizat pentru aplicații specifice detecției de asteroizi din apropierea Pământului.

2. Propus: Document cu specificațiile tehnice ale acestui soft;

Realizat:

- Documentul cu specificațiile tehnice este disponibil pe website-ul proiectului.
https://planet.astro.ro/ParaSOL/wp-content/uploads/2022/12/SpecificatiiTehnice_Website.pdf
- Proiectul are atașat un sistem complet de documentație de tip wiki, disponibil pe serverul de git (necesită autentificare):
<https://gitplanet.astro.ro/parasol/parasol-issue-tracker/-/wikis/home>
- Manualul complet se regăsește pe serverul de git la adresa:
<https://planet.astro.ro/ParaSOL/wp-content/uploads/2024/06/CoreReferenceManual.pdf>

3. Propus: Raport de testare și validare

Realizat:

- Rapoartele de testare și validare sunt accesibile pe website-ul proiectului.
- Document cu rezultatele testării automate pe telescopul TCS, rezultatele testării manuale pe datele de la telescopul INT, rezultatele optimizării paramaterilor pe telescopul de 1.6 m KASI.
<https://planet.astro.ro/ParaSOL/wp-content/uploads/2024/06/RaportTestare.pdf>
 - Set de fișiere de tip .csv conținând rezultatele testării automate în cadrul programului larg de observații realizat cu INT și KASI.
<https://planet.astro.ro/ParaSOL/wp-content/uploads/2024/06/TestareAutomata.zip>

4. Propus: Al doilea articol acceptat spre publicare într-un jurnal cu factor ISI mai mare ca 1.5

Realizat:

1. Articol publicat: *Pre-perihelion monitoring of interstellar comet 2I/Borisov*;
Autori: Prodan, George P.; **Popescu, Marcel**; Licandro, Javier; Akhlaghi, Mohammad; de León, Julia; Tatsumi, Eri; Pastrav, Bogdan Adrian; Hibbert, Jacob M.; **Văduvescu, Ovidiu**; Simion, Nicolae Gabriel; Pallé, Enric; Narita, Norio; Fukui, Akihiko; Murgas, Felipe
Monthly Notices of the Royal Astronomical Society (Factor de impact 6.5),
Volume 529, Issue 4, pp.3521-3535, Aprilie 2024
<https://ui.adsabs.harvard.edu/abs/2024MNRAS.529.3521P/abstract>

2 Articol publicat: *Hubble Asteroid Hunter. III. Physical properties of newly found asteroids*; García-Martín, Pablo; Kruk, Sandor; **Popescu, Marcel**; Merín, Bruno; Stapelfeldt, Karl R.; Evans, Robin W.; Carry, Benoit; Thomson, Ross; ***Astronomy & Astrophysics (Factor de impact 6.5)***, Volume 683, id.A122, 10 pp., Martie 2024
<https://www.aanda.org/articles/aa/pdf/2024/03/aa46771-23.pdf>

4 Articol publicat: *Photometry of the Didymos System across the DART Impact Apparition*; Moskovitz, Nicholas; ... Berteșteanu, Daniel N.; ... **Popescu, Marcel M.**; ...; ***The Planetary Science Journal (Factor de impact 3.4)***, Volume 5, Issue 2, id.35, 28 pp.; Februarie 2024
<https://iopscience.iop.org/article/10.3847/PSJ/ad0e74/pdf>

4. Articol în pregătire [estimat să fie trimis spre publicare în august 2024]: *Data-parallel methods for fast and deep detection of asteroids on the Umbrella platform*; Mălin Octavian Stănescu, Marcel M. Popescu, Ovidiu Văduvescu, Lucian Curelaru, Daniel Berteșteanu, Marian Predatu

3. Descrierea științifică și tehnică a activităților efectuate

WP4.3 - Împachetarea serviciului de calcul pentru distribuție. Conectarea nodurilor de calcul la nivel central și distribuirea operațiilor

Conectarea nodurilor de calcul la nivel central și distribuirea operațiilor : 1) Exportul datelor interne într-un format portabil; 2) Dezvoltarea unui serviciu de management al nodurilor și un serviciu de calcul pentru noduri; 3) Împachetarea serviciului de calcul pentru distribuție; 4) Dezvoltarea codurilor de lansare a calculelor și comunicarea între noduri.

În cadrul proiectului ParaSOL am implementat funcționalitate pentru a rula pipeline-urile de procesare a datelor astronomice prin rețea, astfel încât suita de software dezvoltată în cadrul proiectului să poată fi folosită în sisteme de calcul tip cluster. Script-ul din figura de mai sus rulează pe un calculator de control (1), pentru a apela prin rețea programele de configurare și rulare al un pipeline de procesare pe un calculator worker (2) folosind datele dintr-un nod de stocare aflat pe un alt calculator (3). Un astfel exemplu de rulare distribuită a fost realizat pentru un set de date de 6.4GB de pe telescopul T80S aflat în arhiva de date de pe calculatorul parasol2 a fost transferat automat pe calculatorul parasol, folosind o legătură de mare viteză (10Gbps) configurată între cele două calculatoare din cadrul nodului de calcul de la AIRA, într-un spațiu temporar dedicat și procesat folosind STU, toate aceste operații având loc printr-o comandă dată de calculatorul Portal2, care a funcționat ca nod de control pentru acest cluster în cadrul acestui test.

```
echo "Preparing for remote run"
ssh $REMOTE rm -r $TARGET
ssh $REMOTE mkdir $TARGET
ssh $REMOTE mkdir $TARGET/in
echo "Copying configuration file"
PIPERUNNER="/mnt/data-ssd/AstroTools/umbrella//v22/basichost-v0.6.1/BasicHost.exe"
REMOTE="parasol.astro.ro"
TARGET="/mnt/ramdisk/stu-test-retea/"
CONFIGTEMPLATE="/home/malin/config-T80S-auto.txt"
STORAGE="parasol2:/"
DATAPATH="/mnt/data-raid/photom/807_T80S/3_proc/ipp_science/20230523/EU03/*.fits"
```

Fig. 2 Scriptul (sus) folosit pentru a rula STU în rețea și fișierul de configurare (jos) necesar.

STU este distribuit în condiții normale folosind updater-ul dezvoltat în cadrul acestui proiect, disponibil atât pentru Windows cât și pentru Linux, având inclusiv pachete asamblate pentru distribuțiile Debian și ArchLinux. De asemenea, pentru distribuția rapidă și eficientă pe sistemele de calcul ParaSOL a mai multor variante de software (lucru necesar pentru a avea reproductibilitate a rezultatelor în timp precum și pentru a permite experimente cu versiuni preliminare în timp ce procesarea de rutină folosește versiuni stabile și testate), am dezvoltat un sistem de instalare, numit bundletool, urmărind convențiile de instalare a software-ului astronomic folosite în cadrul acestui proiect.

În schimb, în cazul sistemelor de tip cluster aflate într-un mediu de producție, pentru a asigura un mediu controlat de execuție pentru toate sistemele de procesare a datelor, suplimentar metodei obișnuite, este necesară distribuția unui pachet ce conține toate dependențele necesare cu o versiune bine controlată. În acest sens am dezvoltat o imagine pentru containere de tip systemd-nspawn, folosind la bază o distribuție ArchLinux. Ca atare, această imagine este ușor transferabilă, atât prin copiere directă cu rsync, cât și ca arhivă tarball. Folosind această infrastructură, software-ul dezvoltat în cadrul acestui proiect va putea fi folosit cu ușurință în producție pe scară largă odată ce maturitatea necesară este atinsă.

WP5.2 Validarea folosind camerele mozaic de pe telescoapele mari

Validarea softului în cadrul unor programe largi de observații s-a făcut folosind telescoapele de 2.54m INT și 1.6 m KASI. Cantitatea de date generată de aceste telescoape a fost suficientă pentru validarea propusă. Respectiv, au fost făcute observații în următoarele nopți:

1) 2.54 m INT: 20230227, 20230228, 20230301, 20230302, 20230303, 20230410, 20230411, 20231018, 20231029, 20231030, 20231031, 20231101, 20231102, 20240318, 20240319, 20240507, 20240508, 20240509, 20240528, 20240529, 20240530, 20240531. În total au fost observate 535 de câmpuri unice, o suprafață de cer acoperită de 133 grade^2 . Camera mozaic este prezentată în Fig. 6.

2) 1.6 m KASI: 20230216, 20230217, 20230218, 20231022, 20231023, 20231205, 20231207, 20231208, 20231210, 20240111, 20240210. În total au fost observate 133 de câmpuri unice care acoperă o suprafață de cer de 532 grade^2 . Camera mozaic include 4 x 8 CCD-uri acoperind o suprafață de cer de 4 grade^2 .

Datorită faptului că am obținut timp de observație pe aceste două telescoape (pe bază de competiție pentru timp de telescop) și cantitatea de date ne-a permis să testăm o multitudine de cazuri în mediu operațional (program larg de observații), nu am mai folosit imagini din arhive pentru alte telescoape.

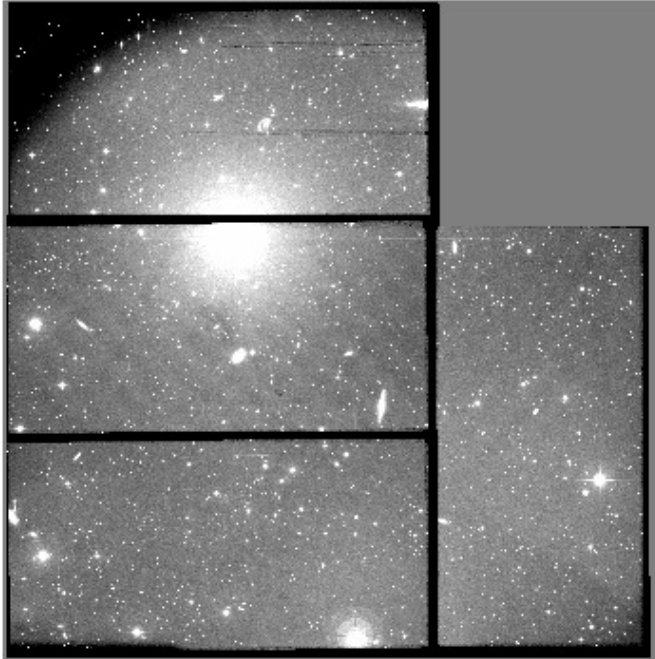


Fig. 3 Cele patru CCD-uri ale camerei mozaic WFC de pe INT. Datorită rezoluției mici a unui CCD de pe INT și vitezei ridicate de căutare a obiectelor (tipic 15"/min), toate cele 4 CCD-uri de pe INT sunt asamblate într-un singur macrotile.

Un exemplu de diferențe între pozițiile observate/ măsurate și cele calculate pentru obiecte cunoscute folosind efemerida acestora este prezentat în Fig. 1 din acest raport.

WP5.3 - Identificarea secțiunilor de cod costistitoare din punct de vedere al timpului de procesare și optimizarea acestora

Realizarea unei serii de teste extinse pentru a determina secțiunile din cod unde softul funcționează lent și cu performanțe reduse. Implementarea codului pentru optimizarea acestora. Cunoașterea impactului modificărilor prin rularea unui set extins de teste. Figura de mai jos prezintă performanțele rulării pachetului Umbrella pe diferite tipuri de imagini.

Our typical runs on, with an AMD Radeon RX 6800 XT:

- WFC on INT: 4×9 Mpx, $0.33'' \text{ px}^{-1}$, 12×1 min cadence $10'' \text{ min}^{-1}$ search cone
- Acquisition time: 12 min
- Runtime at full granularity: 26 s per CCD, with 2 s for actual ST scan
- T80S: 1×80 Mpx, $0.55'' \text{ px}^{-1}$, $20 \times \sim 1.5$ min cadence, $15'' \text{ min}^{-1}$ search cone
- Acquisition time: 30 min
- Runtime: 7 min per CCD, with 2.5 min for actual ST scan

Fig. 4 Performanțele algoritmului STU

WP5.4 - Caracterizarea performanțelor (rata de detecție, "fals pozitive", "fals negative", dimensiunea maximă a camerei pentru care softul poate rula, și cadența imaginilor pentru reducerea în timp real)

O analiză completă a performanțelor algoritmului STU a fost realizată folosind imaginile de la KASI. Pentru a face acest lucru s-a efectuat o variație a tuturor parametrilor. Rezultatele sunt prezentate în Anexa 3 a raportului de testare.

<https://planet.astro.ro/ParaSOL/wp-content/uploads/2024/06/RaportTestare.pdf>

MedianThreshold	1.25	1.5	1.59	1.67	1.75	2	2.5			
Positive detection	272	272	272	271	270	267	256			
Negative detector	2122	1938	1805	1671	1511	968	507			
positive / negative	13	14	15	16	18	28	50			
MinThreshold	1.5	2	2.1	2.25	2.5	2.58	2.75	3	3.25	3.5
Positive detection	281	281	277	277	277	276	273	272	270	269
Negative detector	3497	3184	3107	2965	2669	2561	2319	1938	1587	1284
positive / negative	8	9	9	9	10	11	12	14	17	21

Fig. 5 Secțiune din raportul de testare care arată raportul dintre detecțiile pozitive și numărul de artefacte relativ la pragurile de detecție.

Director de proiect
CS II Dr. Marcel Popescu