

МАВКА: ПРОГРАМА СТАТИСТИЧНО ОПТИМАЛЬНОГО ВИЗНАЧЕННЯ ФЕНОМЕНОЛОГІЧНИХ ПАРАМЕТРІВ ЕКСТРЕМУМІВ. АЛГОРИТМ ПАРАБОЛІЧНИХ СПЛАЙНІВ ТА АНАЛІЗ ЗМІН БЛИСКУ НАПІВПРАВИЛЬНОЇ ЗОРІ Z UMa

К. Д. Андрич^{1,2} , І. Л. Андронов¹ , Л. Л. Чінарова^{3,1} 

¹Кафедра “Математика, фізика та астрономія”,
Одеський національний морський університет
e-mail: tt_ari@ukr.net

²Кафедра теоретичної фізики та астрономії,
Одеський національний університет ім. І. І. Мечникова
e-mail: katyaandrich@gmail.com

³Астрономічна обсерваторія,
Одеський національний університет ім. І. І. Мечникова
e-mail: lchinarova@gmail.com

(Отримано 17 грудня 2019 р.; в остаточному вигляді — 20 січня 2020 р.; прийнято до друку — 28 січня 2020 р.; опубліковано онлайн — 09 квітня 2020 р.)

Удосконалено програмне забезпечення МАВКА для апроксимації рядів спостережень біля екстремуму. Воно використано для аналізу змінності блиску пульсуючих та затемнюваних зір, але може бути корисним для аналізу сигналів будь-якої природи. Запропоновано новий алгоритм із використанням параболічного (квадратичного) сплайна. На відміну від традиційного визначення сплайна як кусково-заданої функції на фіксованих інтервалах, запропоновано сплайн із розділенням інтервалу на три підінтервали, але положення меж між підінтервалами є додатковими параметрами. Дефект сплайна дорівнює 1, тобто функція та її перша похідна неперервні, а друга може бути розривною на межах. Така функція є вдосконаленням “асимптотичної параболі” (Marsakova and Andronov 1996). Розглянуто залежність точності апроксимації фіксованого сигналу від розташування меж інтервалу. Порівняно оцінки точності параметрів методом найменших квадратів та “статистичного самотворення” (bootstrap).

Проаналізовано змінності напівправильної пульсуючої зорі Z UMa. Показано багатокомпонентну змінність об'єкта, зокрема, чотири періодичні коливання та суттєву змінність амплітуд і фаз окремих коливань.

Ключові слова: обробка даних, змінні зорі, екстремум, крива блиску.

DOI: <https://doi.org/10.30970/jps.24.1902>

I. ВСТУП

Досить часто з тих чи інших даних необхідно отримати феноменологічні параметри екстремуму, аби потім їх використовувати для подальшого фізичного аналізу процесу. І зазвичай найнеобхідніші параметри — це момент та величина цього екстремуму, іноді вся апроксимація з відповідними оцінками точності (стандартними похибками).

Ці параметри використовують переважно для дослідження змін періодів затемнюваних подвійних систем та пульсуючих зір [1–4]. Крейнер та ін. [5] опублікували 6-томну монографію, у якій зібрано 91 798 моментів мінімумів 1 138 затемнюваних систем. Чінарова та Андронов [6] визначили 6553 моменти екстремумів для 173 напівправильних зір та опублікували каталог цих та інших характеристик.

У подальшому такі копії публікують на сайтах [7–9] та додатково в окремих статтях (напр. у журналі “Open European Journal on Variable Stars” [10]) по мірі накопичення даних. В Американській асоціації спостерігачів змінних зір AAVSO використовують скорочення ToM (“time of minimum”). Утім для пульсуючих зір визначають здебільшого моменти максимумів та не завжди моменти мінімумів, тому

правильно було б розшифровувати скорочення ToM також і як (“time of maximum”).

Історично до одних із найперших методів відносять графічний метод Пікерінга, коли апроксимацію на міліметровці робили “від руки”, “на око”, а потім проводили на різних рівнях горизонтальні лінії, визначали середини хорд та проводили через них пряму до перетину з апроксимацією. Із розвитком обчислювальної техніки (починаючи з арифмометрів) стали застосовувати апроксимацію параболою (чи для асиметричних кривих — поліномом третього ступеня). Утім найбільш популярним методом для (теоретично) симетричних кривих став метод Кві та ван дер Вердена [11]. Оскільки відповідна тест-функція має багато локальних мінімумів, оцінка точності суттєво занижена, відповідно, вага такого значення неадекватно завищена.

Андронов [12] використовував не лише поліноми ступеня, що забезпечує найліпшу точність визначення моменту, а й поліноміальні сплайни. Апроксимація поліномами оптимального ступеня реалізовували різні автори (напр., [6, 12, 13]).

Сплайни зі змінним ступенем були запропоновані [13], які виправляли проблему “нефізичних” хвиль на апроксимаціях поліномами. Огляди наявних методів приведені [15–18].



Для зручності використання різних методів в одній програмі створено комп'ютерну програму MAVKA, у яку поступово додавали різноманітні функції [19–21].

У цій статті наведено порівняння двох базисних функцій: “асимптотична парабола” (запропонована [14]) та параболічний сплайн дефекту 1. Обидва методи поділяють наведений інтервал поблизу екстремуму на три частини. Кожен з підінтервалів апроксимується власною функцією, що дозволяє зменшити похибки результатів та краще описати спостереження, врахувати асиметричність чи інші особливості.

Проаналізовано вплив ширини інтервалу даних (лише екстремуму або й позаекстремальної частини) та асиметрії його меж на якість апроксимації та відповідні похибки визначених характеристик екстремуму. Для цього обидва методи були застосовані до набору згенерованих даних із нормальним розподілом шуму. Також методи проілюстровано застосуванням до реальних спостережень змінної пульсуючої зорі Z UMa із міжнародної бази даних AFOEV [22] із використанням методу статистичного самотворення (bootstrap)

та проаналізовано спостереження цієї зорі методами періодограмного та шкалограмного аналізу.

II. ОПИС МЕТОДІВ

A. Асимптотична парабола (Asymptotic Parabola, AP)

Цей метод був уперше запропонований у роботі В. І. Марсакової та І. Л. Андронова [14, 16]. Він показав свою ефективність під час визначення моментів екстремумів блиску змінних зір різних типів. Апроксимація складається з двох прямих ліній “асимптот”, які з'єднані параболою. При цьому сама функція та її перша похідна неперервні. Попередній досвід використання цього методу показав, що він є одним із найкращих для асиметричних максимумів пульсуючих зір.

Точки переходу від прямих на параболу C_4 та C_5 .

$$x_c(t) = \begin{cases} C_1 + C_2(-2v - D)D + C_3v, & t < C_4, \\ C_1 + C_2v^2 + C_3v, & C_4 \leq t \leq C_5, \\ C_1 + C_2(2v - D)D + C_3v, & t > C_5, \end{cases} \quad (1)$$

де $D = (C_5 - C_4)/2$, $v = t - (C_5 + C_4)/2$.

Із характерного вигляду функції зрозуміло, що дані, які будуть найефективнішими для такого наближення, охоплюватимуть не лише вузький інтервал біля екстремуму (який асимптотично збігатиметься з параболою), а й суттєву частину висхідної та низхідної гілок кривої блиску та не “суттєво закруглені” частини біля екстремумів протилежного виду (мінімумів, якщо визначаємо максимум та навпаки). Вибір початкової та кінцевої точок зазвичай відбувається за допомогою програм OO [23], MCV [24] і також реалізується в серії послідовно вдосконалених версій MAVKA [19–21].

Хоча метод був запропонований для визначення екстремумів, його можна використовувати й для визначення середнього моменту та характерного часу переходу з однієї асимптоти на іншу, навіть за відсутності екстремуму. Наприклад, за діаграмою “O – C” для моментів екстремумів (чи відповідних фаз) за порівняно різкої зміни періоду [25].

B. Параболічний сплайн дефекту 1 (Parabolic Spline, PS)

Поліноміальні сплайни є кусково-заданими функціями, де в кожному інтервалі аргументу (часу, фази або ін.) задається поліномом із своїм набором коефіцієнтів. Їх часто використовують для апроксимації даних. Вони характеризуються “порядком” n , або найвищим ступенем полінома, та “дефектом” k , який визначається як найменша величина, для якої похідна порядку $(n - k)$ є неперервною в усіх точках, включаючи межі інтервалів [26–28].

Ця апроксимація також складається з трьох інтервалів, кожен з яких описується власною функцією (параболою). Точки переходу відповідають параметрам C_6 та C_7 .

Параболічний сплайн відрізняється від попереднього методу двома додатковими параметрами, які описують параболічні складники в бічних інтервалах:

$$x_c(t) = \begin{cases} C_1 + C_2t + C_3t^2 + C_4(C_6 - t)^2, & t < C_6, \\ C_1 + C_2t + C_3t^2, & C_6 \leq t \leq C_7, \\ C_1 + C_2t + C_3t^2 + C_5(t - C_7)^2, & t > C_7. \end{cases} \quad (2)$$

Метод дозволяє розширити ширину інтервалу апроксимації (і отже, збільшити кількість точок, які використовуються) та отримати точність, що відповідає результатам “асимптотичної параболі” на коротшому інтервалі.

Можливі ситуації, коли поділ інтервалу на три частини є неефективним. Наприклад, екстремум занадто гострий і центральна параболічна частина вироджується. Для такої ситуації у програмі MAVKA [21] реалізовано зменшення кількості інтервалів (а отже, й відповідних нелінійних параметрів). Отже, апроксимація “асимптотичною параболою” може складатися лише з двох асимптот, параболі та лише однієї асимптоти або взагалі лише параболі.

Так само “параболічний сплайн” за занадто малих бічних інтервалів, автоматично замінюється на звичайну параболу.

У межах реалізованого алгоритму лінійні параметри були визначені методом найменших квадратів, а нелінійні — методом диференціальних поправок. Ста-

тистичні похибки параметрів екстремуму виявлені за допомогою коваріаційної матриці похибок коефіцієнтів (напр. [30,16]).

III. ПОРІВНЯННЯ АСИМПТОТИЧНОЇ ПАРАБОЛИ ТА ПАРАБОЛІЧНОГО СПЛАЙНУ ДЕФЕКТУ 1 НА ШТУЧНИХ ДАНИХ ЗІ ЗМІННИМИ МЕЖАМИ

Ми застосували обидва методи до набору з 25 штучних інтервалів. Як сигнал обрано синусоїду з одиничною півамплітудою. Шум був згенерований з нормальним розподілом випадкової величини з теоретичним середньоквадратичним значенням 0.1. Ми змінювали межі інтервалів між екстремальними випадками, наведеними на рис. 1, для оцінки впливу ширини охоплення даних поза екстремумом на апроксимацію обома функціями.

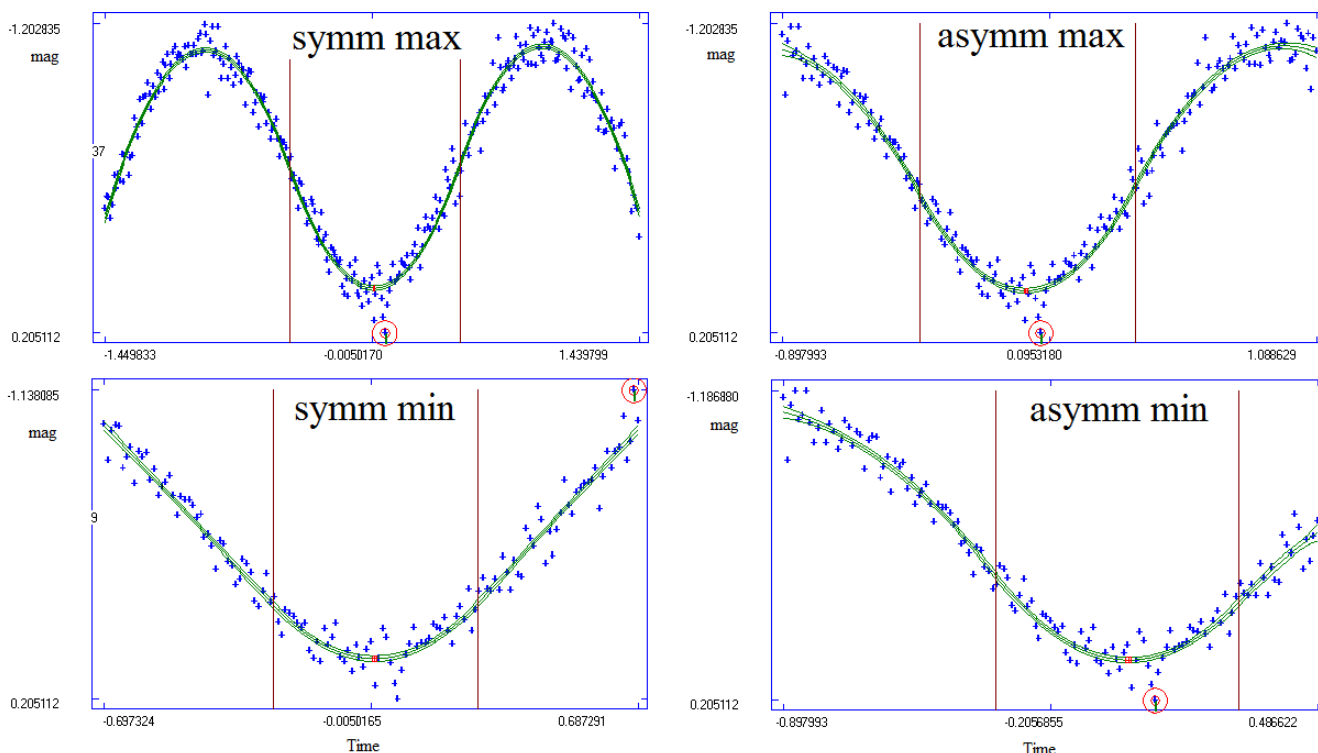


Рис. 1. Екстремальні за довжиною інтервали для тестування апроксимацій асимптотичною параболою (AP) та параболічним сплайном дефекту 1 (PS) — Згори — максимальна ширина інтервалу, внизу — найменша. Зліва — для інтервалів із симетричними (symm) даними відносно екстремуму, справа — із суттєво асиметричним (asymm). Апроксимації проведені параболічним сплайном, крім зліва внизу, де обидві апроксимації практично збігаються. Показано апроксимації $m_C(t)$ та “коридор похибок” $m_C(t) \pm \sigma[m_C(t)]$. Діаграми створені зі скріншотів програми MAVKA, яка показує інформацію про поточну точку (у центрі двох червоних кіл) та дозволяє послідовний перехід між точками або перехід на максимальну чи мінімальну точку. На результат апроксимації номер виділеної точки не впливає.

За зменшення ширини інтервалів та “чистого” сигналу (без шуму) очікується, що всі апроксимації асимптотично сходяться до параболі (чи полінома більшого ступеня, найчастіше кубічної параболі).

Можливість апроксимації поліномом з автоматичним вибором ступеня полінома була реалізована різними мовами програмування [6, 13, 19, 21] й також у програмі MAVKA. Але кількість точок зменшується,

а шуми спостережень призводять до появи “нефізичних” хвиль на апроксимації, що нагадує ефект Гіббса для тригонометричних поліномів (напр. [35]). Це сут-

тєво погіршує якість апроксимації і може призвести до суттєвого зсуву визначеного моменту екстремуму.

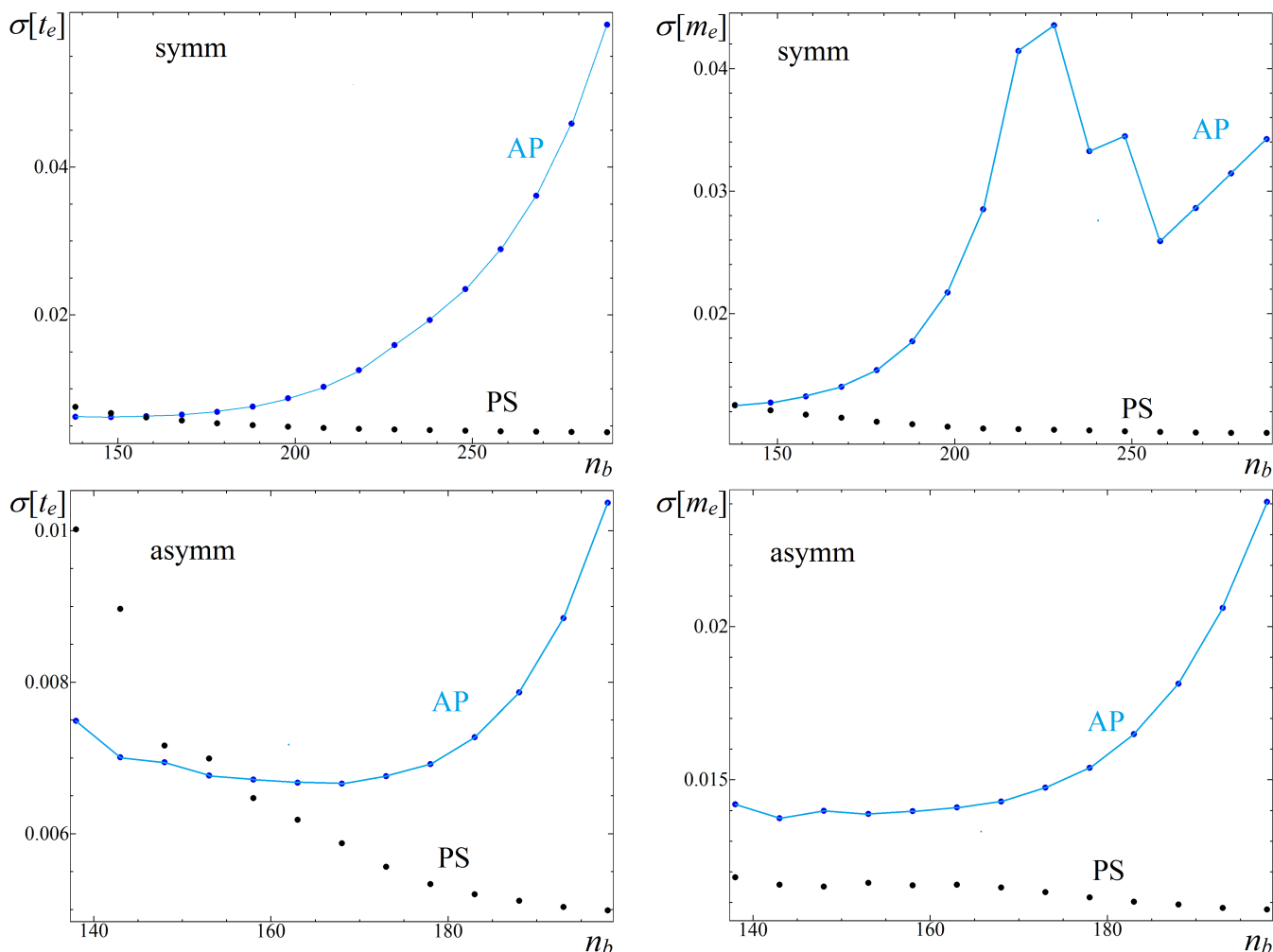


Рис. 2. Залежність оцінки точності визначення моменту екстремуму $\sigma[t_e]$ та відповідного значення зоряної величини $\sigma[m_e]$ від кількості точок для інтервалів, симетричних (symm) та асиметричних (asymm) відносно екстремуму, для асимптотичної параболи (AP) та параболічного сплайна (PS).

Тому для аналізу мінімальна ширина інтервалу була обмежена, як показано на рис. 1. Для меншої ширини розділення інтервалу апроксимації на 3 підінтервали стає формальним. Адже парабола може бути однаково представлена як кубічний сплайн за будь-якому діленню на підінтервали. Тест-функція стала як функція двох змінних (меж підінтервалів), отже матриця нормальних рівнянь у методі диференціальних поправок стає виродженою. У цьому випадку апроксимацію поводять квадратною параболою, а не сплайном.

Результати на кожному тестовому інтервалі для відповідного значення меж були усереднені за відхиленнями від ідеального (синусоїди без шуму) значення та оцінками точності для подальшого аналізу. Як видно на рис. 2, “асимптотична парабола” дуже чутлива до зміни меж апроксимації, на відміну від параболічного сплайна. Останній метод дозволяє проводити точну криву за наявності даних поза сусіднім ек-

стремумом. Однак, коли інтервал звужується, результати обидвох методів практично однакові і можуть відхилитися в межах випадково розподіленого шуму. Графіки у верхньому рядку рис. 2. відповідають симетричній зміні меж інтервалу, менша кількість точок відповідає вужчому інтервалу.

Для оцінки впливу асиметричності меж екстремуму ми зафіксували ліву межу на одному положенні, а праву змінювали покроково. Крайні випадки меж ширини інтервалів для цього аналізу наведені у правому стовпчику рис. 1. Як і в попередньому випадку, за занадто широкого інтервалу даних асимптотична парабола програє параболічному сплайну. Однак далі асимптотична парабола демонструє не гірші, а часом і кращі результати. Зі звуженням інтервалу правий підінтервал апроксимацій вироджується, спричинюючи помилку в роботі та навіть виродження матриці нормальних рівнянь для апроксимації параболічним сплайном.

Метод асимптотичної параболи стійкіший для таких змін. У цьому методі враховано випадки виродження якогось з підінтервалів. У таких ситуаціях апроксимація відбувається лише лінійною асимптотою та параболою або двома асимптотами.

Графіки в нижньому рядку рис. 2 демонструють залежності параметрів екстремуму для асиметричної зміни меж інтервалу, менша кількість точок відповідає вузькому інтервалу.

Очікуваною є перевага апроксимації параболічним сплайном на широких інтервалах, оскільки кількість спостережень збільшується, а три параболи показують менше систематичне відхилення від сигналу, ніж асимптотична парабола. Для невеликих інтервалів поблизу екстремуму перевагу має асимптотична парабола.

Загальна рекомендація для розмітки інтервалів для визначення екстремумів: вибирати (за можливості) приблизно однакові за тривалістю ділянки кривої блиску. Для асимптотичної параболи брати якнайдовші відрізки висхідної та низхідної гілок кривої блиску, де зміна блиску близька до лінійної. З попереднього досвіду використання “асимптотичних парабол” відомо, що блиск на межах інтервалу приблизно вдвічі ближчий до сусіднього екстремуму, ніж до апроксимованого. Звісно, це залежить від конкретного виду кривої блиску. Для автоматизації процесу й було запропоновано вибір між різними функціями для фіксованого інтервалу даних.

Незважаючи на якісний характер рекомендації, для спостережень із порівняно низьким значенням відношення оцінки точності до амплітуди (що характерно не лише для фотографічних та візуальних спостережень зір з великою амплітудою, а й для величезної кількості малоамплітудних зір, відкритих з орбітальних обсерваторій) такий візуальний підхід (або різноманітні напівавтоматичні алгоритми) дає значення параметрів, які несуттєво відрізняються за невеликих змін інтервалу. Але статистична оцінка точності порівняно з дуже широкими чи, навпаки, вузькими інтервалами ліпша на десятки відсотків або навіть у рази.

IV. ЗАСТОСУВАННЯ ОБОХ МЕТОДІВ ДО СПОСТЕРЕЖЕНЬ Z UMa

Для подальшого аналізу використано 12578 спостережень Z UMa з міжнародної бази даних Французької асоціації спостережень змінних зір (AF0EV) [22], отриманих у діапазоні юліанських дат 2451630–2458026 (тривалість 17.5 років). Аналіз попередніх даних проведено [6].

Крива блиску пульсуючої зорі Z UMa була попередньо розмічена на інтервали поблизу екстремумів (як максимумів, так і мінімумів). Ми провели апроксимацію кожним методом окремо.

Ми аналізували точність моментів екстремуму, точність зоряної величини екстремуму та середньоквадратичне відхилення для обох методів відповідно до

номера інтервалу. Результати досить схожі, усі інтервали (згідно з розміткою) не містили позаекстремальної частини. Іноді відповідна точність дещо краща для апроксимації асимптотичною параболою.

Також ми порівняли обидва методи, аби алгоритм програми MAVKA самостійно визначив найкращий з них для кожного конкретного інтервалу. Вибір кращого методу відбувається за меншою оцінкою точності моменту екстремуму.

У результаті з 94 інтервалів асимптотична парабола виявилась кращою для 68 з них, а параболічний сплайн — для 26. Медіана відношення оцінки точності моменту екстремуму для PS та AP трохи менша за одиницю (0.86). Але для 34 інтервалів із 94 екстремум не попадає в центральний підінтервал апроксимації і тому відбракувався, хоча сама апроксимація була достатньо якісною. З урахуванням цих відбракованих даних медіана суттєво перевищує одиницю (2.36).

Це пов'язано з декількома причинами. По-перше, сплайн менш стійкий до шуму в спостереженнях, ніж асимптотична парабола, яка іноді може визначати не глобальний екстремум на інтервалі, а локальну флуктуацію шуму. По-друге, на цей момент апроксимація параболічним сплайном була реалізована за умови, що момент екстремуму має потрапляти в центральний підінтервал. Однак можуть бути ситуації, коли апроксимація проведена досить точно, але екстремум є на одному з бічних підінтервалів, що призводить до вильоту алгоритму під час спроби його розрахувати або до завеликих та некоректних похибок визначених параметрів.

Медіана відношення середньоквадратичних відхилень спостережень від апроксимацій PS та AP становить 0.986, що лише несуттєво менше від одиниці.

Асимптотична парабола є стійкішою до таких випадків й дозволяє коректно визначити характеристики екстремуму для будь-якого екстремуму. Тому ці методи не замінюють один одного на всіх інтервалах даних, а доповнюють, тобто для кожного окремого інтервалу потрібно вибрати оптимальний метод.

V. МЕТОД СТАТИСТИЧНОГО САМОТВОРЕННЯ ТА ПОРІВНЯННЯ МЕТОДІВ АПРОКСИМАЦІЇ НА СПОСТЕРЕЖЕННЯХ Z UMa

Щоб порівняти оцінки точності ми проаналізували метод “статистичного самотворення”. Іноді його називають калькою з англійської “бутстреп” (bootstrap). Сервіс перекладу Google дає “завантажувач”, що відповідає вживанню терміна в іншому комп’ютерному значенні — під час завантаження операційної системи.

Суть методу полягає в тому, щоб з наявного набору даних сформувати досить велику кількість вибірок, розмір кожної з яких збігається з вихідним інтервалом. Такі вибірки складаються з випадкових комбінацій початкових даних (“самотворення”), а вибір кожного елемента відбувається з можливістю повер-

нення, на відміну від “перемішування”, за якого всі елементи використовуються один раз за випадковим порядком.

Термін “bootstrap” запропонував Ефрон [30], його теорії та застосуванню присвячені монографії, напр. [31–33].

Отже, деякі елементи можуть траплятися декілька разів, а інших — немає взагалі. Для кожної отриманої вибірки визначають значення аналізованих статистичних характеристик з метою вивчити їх розкид, стійкість, розподіл. Цей метод є вдосконаленою модифікацією методу “складного ножа” (jackknife) [30], згідно з яким n нових вибірок даних отримують із початкової вибірки видаленням лише однієї точки, поспідовно змінюючи номер цієї видаленої точки.

Метод “bootstrap” є стандартним для оцінювання точності визначення моментів мінімумів затемнюваних подвійних систем на сайті Секції змінних зір та екзопланет Чеського астрономічного товариства [7] з використанням наближення для повного інтервалу затемнення [34]. Після встановлення моменту мінімуму для серії “самостворених” рядів даних визначаються перцентілі 0.025 та 0.975 як границі довірчого інтервалу для ймовірності 95%. На відміну від стандартного визначення точності як оцінки середньоквадратичного відхилення σ у [34] задають σ_- та σ_+ , що відповідають наведеним перцентілям.

Використання σ_- та/або σ_+ , замість стандартного середньоквадратичного відхилення σ , призводить до суттєвого зменшення статистичної ваги цієї величини $w = \text{const}/\sigma^2$. У припущенні неперервного нормального (гауссовського) розподілу [35, 36] $\sigma_- = \sigma_+ = 1.96\sigma$ для цієї ймовірності 95%. Тобто зменшення статистичної ваги у $1.96^2 \approx 3.84 \approx 4$ рази. Тому потрібно перетворювати два значення на одне, знаходячи середньоквадратичне значення $\sigma \approx ((\sigma_-^2 + \sigma_+^2)/2)^{1/2}/1.96$; значення, що відповідає середній вазі $\sigma \approx ((\sigma_-^2 + \sigma_+^2)/2)^{-1/2}/1.96$, або інтерперцентильному інтервалу $\sigma \approx (\sigma_- + \sigma_+)/2/1.96$. За незначного відхилення σ_-/σ_+ від одиниці всі три перелічені наближення практично не відрізняються. Суттєве відхилення показує негауссівський характер розподілу похибок, яке повинно бути проаналізовано додатково.

Отже, можна рекомендувати використовувати σ замість пар (σ_-, σ_+) .

Середньоквадратичне значення відхилення для неперервного нормального розподілу, якщо враховувати 95% середніх значень, становить $\sigma_{0.95} = 0.871\sigma$. Отож, за викидання 5% та під час обчислення σ потрібно враховувати цей коефіцієнт.

Для ілюстрації, з кривої блиску пульсуючої зорі Z UMa був виділений початковий ряд поблизу одного з 94 екстремумів (JD 2455999-2456172, $n = 351$). Спостереження отримано з міжнародної бази даних Французької асоціації спостережень змінних зір (AFOEV) [22].

Як критерій вибору меж інтервалу обрано частину від попереднього до наступного інтервалу, бо ширші інтервали, очевидно, будуть ліпше апроксимовані сплайном, ніж асимптотичною параболою, де висхідна та низхідна гілки апроксимовані прямими. Квадратичний сплайн враховує можливу кривизну кривої блиску в крайніх підінтервалах та тим самим дозволяє використовувати більші інтервали з більшою кількістю точок.

Квадратичний сплайн враховує можливу кривизну кривої блиску в крайніх підінтервалах та тим самим дозволяє використовувати більші інтервали з більшою кількістю точок.

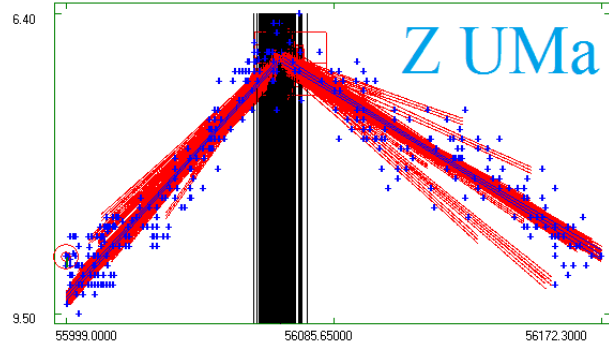


Рис. 3. Частина кривої блиску напівправильної пульсуючої змінної Z UMa біля одного з екстремумів. Апроксимації $(m_C(t))$ асимптотичною параболою показані разом із коридором похибок $(m_C(t) \pm \sigma[m_C(t)])$ для 200 рядів методу статистичного самотворення (bootstrap). Відповідні криві проведені від початку до кінця створеного ряду. Сині лінії всередині червоної смуги демонструють апроксимацію та коридор похибок для початкового набору даних.

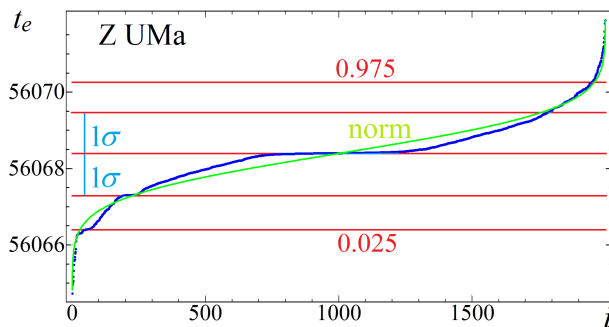


Рис. 4. Залежність оцінки моменту екстремуму t_e від номера реалізації після сортування за величиною. У методі статистичного самотворення, не враховуються перші та останні 2.5% чисел (показані верхньою та нижньою лініями). Внутрішніми горизонтальними лініями показано значення для початкового ряду та коридор похибок. Окремо наведена теоретична крива для нормального розподілу з відповідними вибірковими значеннями.

Створено 2000 рядів із використанням генератора випадкових чисел. До них були застосовані такі самі апроксимації, що й для початкового ряду. Апроксимації для початкового ряду з $\pm 1\sigma$ коридором похибок зображені на рис. 3. Для випадкових рядів показано накладання апроксимацій. Як видно, більшість кривих лежать у порівняно вузькому інтервалі. Але кілька кривих сильно відрізняються. Це є наслідком вибіркової надмірної концентрації даних в окремих точках та відповідного зменшення впливу інших точок.

Для ілюстрації, на рис. 3 наведено лише 200 апроксимацій із 2000. Основна їх частина є в порівняно вузькому діапазоні. Апроксимації проведені від початку до кінця згенерованого інтервалу, тому видно,

що апроксимації, які суттєво відрізняються від основної смуги, відповідають коротшим інтервалам.

На рис. 4 показана залежність оцінок моментів від номера (рейтингу за величиною) після сортування в порядку збільшення величини для 2000 “самостворених” рядів. Транспонована залежність $i/200$ від t_e називається “кумулятою”. Для даної вибірки по 2.5% з обох боків якраз збігаються із суттєвим надлишком кількості “викидів” порівняно з нормальним розподілом, який показано окремою лінією для вибіркового середнього та середньоквадратичного значення відхилення.

Отже, кількість таких “викидів” в обидва боки суттєво менша за 5%, тому “перцентильний” підхід до оцінки точності σ ефективний.

VI. ІНШІ МЕТОДИ АНАЛІЗУ СПОСТЕРЕЖЕНЬ ПУЛЬСУЮЧОЇ ЗОРІ Z UMa

У попередніх розділах були розглянуті методи аналізу даних в інтервалах навколо екстремумів, які не враховують більш віддалених точках.

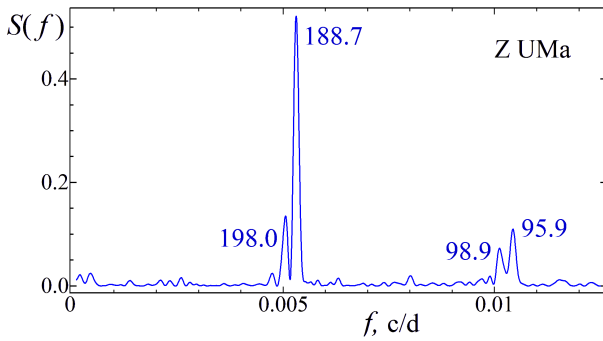


Рис. 5. Періодограма $S(f)$ для спостережень Z UMa з бази даних AFOEV. Статистично значущі піки позначаються значеннями відповідних періодів.

Утім, найбільш поширеним методом є періодограмний аналіз із застосуванням не “локальної”, а “глобальної апроксимації”. Тест-функція $S(f) = \sigma_C^2 / \sigma_O^2 = 1 - \sigma_{O-C}^2 / \sigma_O^2 = r^2$, де σ_O^2 , σ_C^2 та σ_{O-C}^2 є, відповідно, дисперсіями даних “O”, апроксимованих значень “C”, нев’язок (відхилень) “O – C”, а r — коефіцієнт кореляції між “O” та “C” (34). Для апроксимації методом найменших квадратів $m_C(t)$ використовується гармонійна функція

$$m_C(t) = C_1 - R \cdot \cos(2\pi f \cdot (t - T_0)), \quad (3)$$

де C_1 — середнє за період значення (яке загалом не збігається з вибірковою середньою), R — півамплітуда, $f = 1/P$ — частота, P — період, T_0 — початкова епоха максимуму блиску (тобто мінімуму зоряної величини m_C). На відміну від спрощених алгоритмів, принаймні 6 з яких їхні автори називають “перетворенням Фур’є” чи “дискретним перетворенням Фур’є”, у випадку спостережень, нерівномірних у часі, саме метод найменших квадратів є статистично оптимальним

та забезпечує незміщеність оцінок параметрів [36, 37, 16].

На рис. 5 показана періодограма в діапазоні частот 0.0001–0.0125 c/d (період от 406^d до 10000^d). Найбільше виділяються чотири піки. Періоди, які їм відповідають, наведені на рисунку. Можна зробити висновок про наявність двох близьких періодів 198.0^d та 188.7^d, а права пара піків відповідає їхнім гармонікам (подвоєним частотам). Подібна наявність двох близьких періодів (179.1^d ± 0.3^d та 162.4^d ± 0.8^d) спостерігалася в іншій системі RX Boo [38]. Огляди зір цього типу наведені, напр. [39–41].

Тому у програмі MCV ми обчислили параметри комбінованої моделі із $s = 4$ частотами, дві з яких є подвійними двох перших:

$$m_C(t) = C_1 - \sum_{j=1}^s R_j \cdot \cos(2\pi f_j \cdot (t - T_{0j})). \quad (4)$$

Періоди скореговано методом диференціальних поправок: $P_1 = 188.88(3)$, $P_2 = 197.89(4)$. У дужках наведено точність в одиницях останнього знака. Для гармонік $P_3 = P_1/2$, $P_4 = P_2/2$. Відповідні півамплітуди дорівнюють 0.585^m, 0.092^m, 0.186^m, 0.216^m із практично однаковою точністю трохи менше 0.005^m. Початкові епохи 2454008.6(3), 2453925.2(8), 2453893.4(8), 2453987.5(3). $C_1 = 7.898(3)$. Малі оцінки похибок пов’язані з великою загальною кількістю спостережень. Середньоквадратична точність апроксимації в моменти спостережень $\sigma[x_C] = 0.010^m$. Утім, $\sigma[O - C] = 0.377^m$, що значно більше від значень $\sigma_{O-C} = 0.12^m \dots 0.25^m$ (медіанне значення 0.19^m). Отже, індивідуальні цикли пульсацій, хоч і показують цикли биття, мають суттєвий аперіодичний складник змінності.

Для довгоперіодичних зір із моноперіодичними та порівняно стабільними кривими блиску оптимальною апроксимацією є тригонометричний поліном (кінцева сума ряду Фур’є) із статистично значущим ступенем [42,43]. Для виявлення повільних змін періоду та амплітуди часто використовують вейвлет-аналіз (wavelet analysis) [44–46], який іноді називають “спалах-аналіз”, та його модифікацію — “ковзаючі синуси” (Running Sine, RS) [47]. Комбінація тригонометричного полінома з локальним профілем затемнення [48–50] теж не ефективна для розглянутого типу коливань, як і визначення “характерних точок” як перетину постійного рівня [51–52].

У випадку Z UMa це не дуже ефективно, тому що “подвоєння” частоти пов’язано з “биттям” двох коливань із близькими частотами та їхніми гармоніками. Тому більш ефективним буде застосування “шкалограмного аналізу” методом “ковзаючих парабол” (“running parabola”, RP) [53]. Основна ідея методу полягає в дослідженні систематичних зсувів апроксимації від даних, що дозволяє визначити період та півамплітуду для гармонійного сигналу або їхні ефективні значення для випадку квазіперіодичних коливань (QPO). Із застосуванням “Λ”-шкалограмного аналізу [37], визначено $P_\Lambda = 190.7^d$ з ефективною пів-

амплітудою $R_A = 0.66^m$. Шкалограми з використанням різних взаємодоповнювальних функцій наведені на рис. 6. Очікувано цей метод не дозволяє розрізнити дві близькі частоти, тому період є проміжним і півамплітуда відповідає сумарним коливанням, а не окремій хвилі.

Ефективна ширина інтервалу згладжування $\Delta t = 60^d$ відповідає $\sigma[O - C] = 0.22^m$. Відповідна крива блиску зображена на рис. 7. Добре помітно, що амплітуда коливань за домінування швидкої змінності

суттєво занижена. Тому було обрано менше значення $\Delta t = 40^d$. Ця крива показує коливання значно краще.

Також показані локальні апроксимації алгоритмами “асимптотичної параболи” (AP) та “параболічного сплайна”, які ми реалізували у програмі MAVKA. Хоча візуальний вибір інтервалів проводився не лише за індивідуальними точками, а й з урахуванням “ковзаючих парабол”, використання інтервалів із змінною шириною поліпшує якість апроксимації.

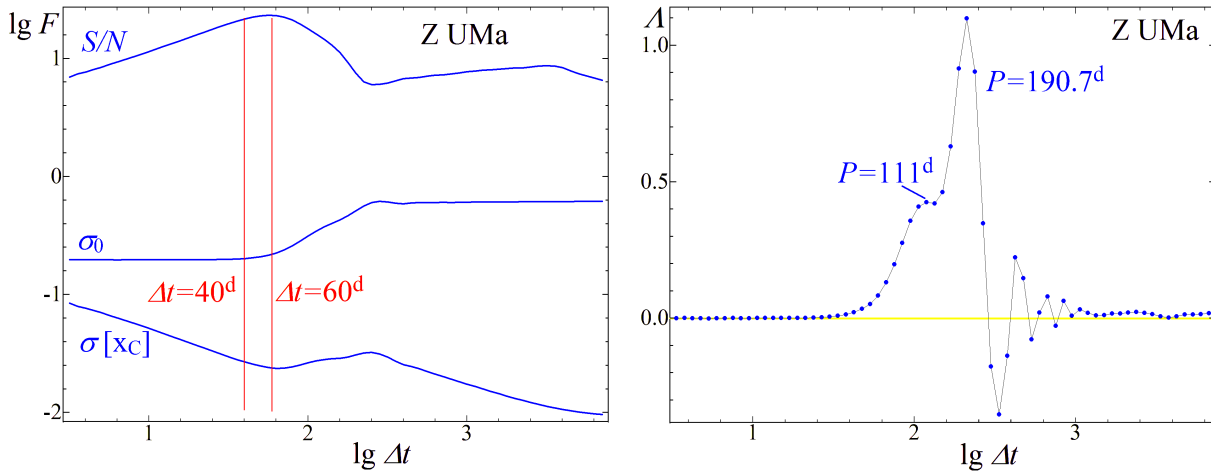


Рис. 6. Зліва: шкалограма “ковзаючою параболою” [53] для Z UMa. Вертикальні лінії відповідають статистично оптимальним значенням $\Delta t = 60^d$ і прийнятому значенню $\Delta t = 40^d$, що відповідає рідкому видимому подвоєнню частоти. Праворуч: “A” — шкалограма [15] та відповідні оцінки періоду.

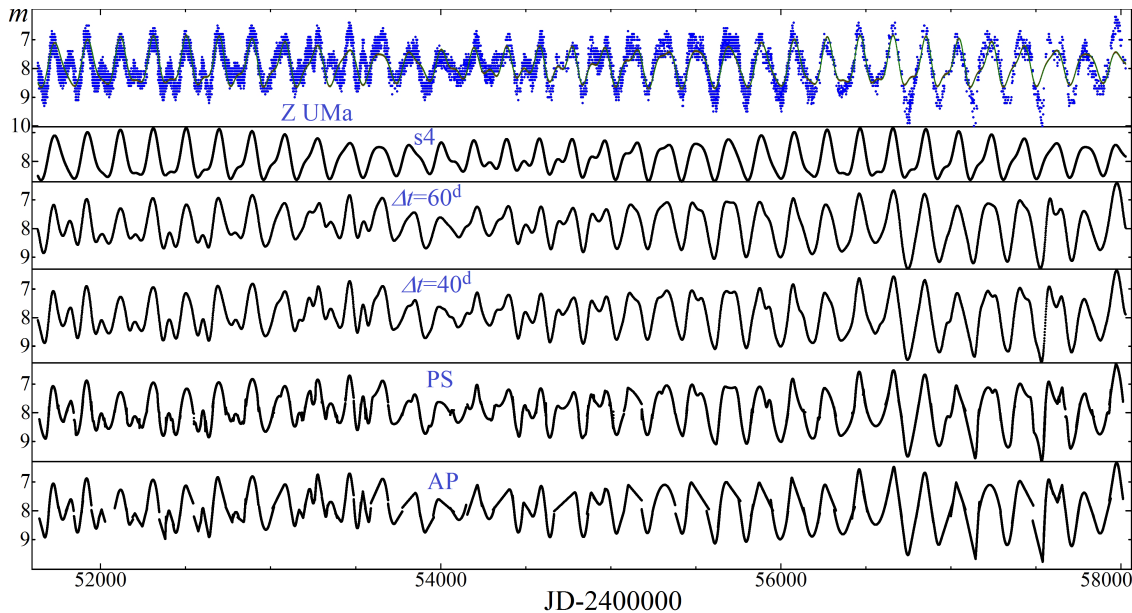


Рис. 7. Спостереження Z UMa та їх наближення: s4 - наближення з 4 синусоїдальними компонентами; $\Delta t = 60^d$ та $\Delta t = 40^d$ — апроксимація методом “ковзаючих парабол” із відповідною шириною фільтра; PS — апроксимація параболічним сплайном по розмічених інтервалах поблизу екстремумів; AP — апроксимація “асимптотичною параболою” тих самих інтервалів.

ВИСНОВКИ

Використання взаємодоповнювальних методів апроксимації дозволяє аналізувати сигнали та виявляти різні компоненти періодичної, квазіперіодичної та аперіодичної змінності. Розглянуто різні методи, ефективні для різних типів змінності.

У створеній програмі МАВКА реалізовано застосування 11 видів апроксимуючих функцій, загалом 21 різна функція, якщо враховувати різні ступені поліномів та симетричних поліномів. Вдосконалено програму МАВКА та тестування нових методів (включаючи описані в цій статті). Надалі сайт програми плануємо за адресою <http://uavso.org.ua/mavka>.

Використання кусково-заданих функцій виграє в часі обчислень порівняно з аналітичними функціями, які потребують обчислення через ряди.

Використання для кусково-заданих функцій інтервалів із межами, що є додатковими параметрами, на відміну від класичних сплайнів, дозволяє зменшити загальну кількість параметрів та поліпшити якість апроксимації

Реалізовано можливість автоматичного вибору статистично оптимальної апроксимації з обраних користувачем.

Реалізовано можливість автоматичного перебору інтервалів даних із розмічених користувачем.

Досліджено ефективність різних апроксимацій залежно від ширини інтервалу та розташування його

меж. Дано рекомендації для розмітки інтервалів. Показана ефективність локальних наближень із інтерактивно визначеною шириною інтервалу.

Досліджено статистичні властивості параметрів моделі та апроксимацій методом статистичного самотворення (bootstrap) для тестових та реальних даних. Для додаткової функції ваги порівняння розподілу оцінок коефіцієнтів із нормальним розподілом проведено [54].

Для реальних спостережень із великими сезонними перервами в них, рекомендації можуть бути враховані лише частково.

Досліджено змінність напівправильної пульсуючої подвійної зорі Z UMa. Показано наявність багатокomпонентної змінності об'єкта, зокрема чотирьох періодичних коливань та суттєвої змінності амплітуд і фаз окремих коливань.

Розроблені методи забезпечують статистично оптимальну апроксимацію, що відповідає найкращій точності визначення модельного параметра.

Подяки Дослідження виконано як частина міжнародної програми “Між-Довготна Астрономія” (Inter-Longitude Astronomy [55]) та проектів “Українська віртуальна обсерваторія” [56] та “Астроінформатика” [57]. Дякуємо Французькій асоціації спостерігачів змінних зір (AFOEV) [22] за спостереження Z UMa та В. І. Марсакової, Д. Є. Твардовському, Pavol Dubovsky і рецензентам за обговорення.

-
- [1] В. П. Цесевич, *Переменные звёзды и способы их исследования* (Наука, Москва, 1971).
- [2] *Нестационарные звёзды и способы их исследования. Том 3. Затменные переменные звёзды*, под ред. В. П. Цесевича (Наука, Москва, 1971).
- [3] *Методы исследования переменных звёзд*, под ред. В. Б. Никонова (Наука, Москва, 1971).
- [4] *Звёзды и звёздные системы*, под ред. Д. Я. Мартынова (Наука, Москва, 1981).
- [5] J. Kreiner, Chun-Hwey Kim, Il-Seong Nha, *An Atlas of O-C Diagrams of Eclipsing Binary Stars* (Wydawnictwo Naukowe Akademii Pedagogicznej, Krakow, 2001).
- [6] L. L. Chinarova, I. L. Andronov, *Odessa Astron. Publ.* **13**, 116 (2000).
- [7] BRNO O-C Gate, <http://var2.astro.cz/ocgate/>.
- [8] BAV, <https://www.bav-astro.eu/index.php/veroeffentlichungen/lichtencknecker-database/lkdb-b-r>.
- [9] AAVSO, <http://aavso.org/>.
- [10] OEJV, <http://var.astro.cz/oejv/>.
- [11] K. K. Kwee, H. van Woerden, *Bull. Astron. Inst. Netherlands* **12**, 327 (1956).
- [12] I. L. Andronov, *Publ. Astron. Inst. Czechoslov. Acad. Sci.* **70**, 161 (1987).
- [13] V. V. Breus, *Odessa Astron. Publ.* **20**, 32 (2007).
- [14] V. I. Marsakova, I. L. Andronov, *Odessa Astron. Publ.* **9**, 127 (1996).
- [15] I. L. Andronov, *ASP Conf. Ser.* **335**, 37 (2005).
- [16] I. L. Andronov, V. I. Marsakova, *Astrophys.* **49**, 370, (2006); <https://doi.org/10.1007/s10511-006-0037-8>.
- [17] I. L. Andronov, M. G. Tkachenko, L. L. Chinarova, *Astrophys.* **60**, 57 (2017); <https://doi.org/10.1007/s10511-017-9462-0>.
- [18] Z. Mikulášek, *Astron. Astrophys.* **584**, A8 (2015); <https://doi.org/10.1051/0004-6361/201425244>.
- [19] K. D. Andrych, I. L. Andronov, L. L. Chinarova, V. I. Marsakova, *Odessa Astron. Publ.* **28**, 158 (2015).
- [20] K. D. Andrych, I. L. Andronov, L. L. Chinarova, *Odessa Astron. Publ.* **30**, 57 (2017); <https://doi.org/10.18524/1810-4215.2017.30.118521>.
- [21] K. D. Andrych, I. L. Andronov, *Open Eur. J. Variable Stars* **197**, 65 (2019).
- [22] AFOEV, <http://cdsarc.u-strasbg.fr/afoev/>.
- [23] I. L. Andronov, *Odessa Astron. Publ.* **14**, 255 (2001).
- [24] I. L. Andronov, A. V. Baklanov, *Astron. School Rep.* **5**, 264 (2004); <https://doi.org/10.18372/2411-6602.05.1264>.
- [25] I. L. Andronov, L. L. Chinarova, preprint arXiv:1910.08799 (2019); *Astron. Astrophys. Trans.* (in press).
- [26] Н. П. Корнейчук, *Сплайны в теории приближения* (Наука, Москва, 1984).
- [27] Дж. Форсайт, М. Малькольм, К. Моулер, *Машинные методы математических вычислений* (Мир, Москва, 1980).
- [28] Б. П. Довгий, А. В. Ловейкін, Є. С. Вакал, Ю. Є. Вакал, *Сплайн-функції та їх застосування* (Видавничо-

- поліграфічний центр “Київський університет”, Київ, 2016).
- [29] I. L. Andronov, *Astron. Soc. Pacific Conf. Ser.* **292**, 391 (2003).
- [30] B. Efron, *Ann. Statist.* **7**, 1 (1979).
- [31] B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall/CRC, 1993).
- [32] M. Hollander, D. A. Wolfe, *Nonparametric Statistical Methods*, 2nd ed. (Wiley, New York, 1999).
- [33] J. Shao, D. Tu, *The Jackknife and Bootstrap*, 2nd ed. (Springer, 1996).
- [34] L. Brát, Z. Mikulášek, O. Pejcha, http://var2.astro.cz/library/1350745528_ebfit.pdf.
- [35] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. (Wiley, New York, 2003).
- [36] I. L. Andronov, *Odessa Astron. Publ.* **7**, 49 (1994).
- [37] I. L. Andronov, *ASP Conf. Ser.* **292**, 391 (2003).
- [38] I. L. Andronov, L. S. Kudashkina, *Astron. Nachr.* **309**, 323 (1988); <https://doi.org/10.1002/asna.211309050>
- [39] L. S. Kudashkina, *Astrophys.* **62**, 556 (2019); <https://doi.org/10.1007/s10511-019-09604-4>.
- [40] Л. С. Кудашкина, *Кинем. физ. небес. тел* **19**, 193 (2003).
- [41] I. L. Andronov, V. I. Marsakova, L. S. Kudashkina, L. L. Chinarova, *Adv. Astron. Space Phys.* **4**, (2014); <https://doi.org/10.17721/2227-1481.4.3-8>.
- [42] L. S. Kudashkina, I. L. Andronov, *Odessa Astron. Publ.* **9**, 108 (1996).
- [43] I. L. Andronov, L. S. Kudashkina, *Odessa Astron. Publ.* **30**, 93 (2017).
- [44] I. L. Andronov, *Kinem. Phys. Celest. Bodies* **14**, 490 (1998).
- [45] L. L. Chinarova, *Odessa Astron. Publ.* **23**, 25 (2010).
- [46] L. S. Kudashkina, I. L. Andronov, *Odessa Astron. Publ.* **23**, 6 (2010).
- [47] I. L. Andronov, L. L. Chinarova, *Częstochowski Kalendarz Astron.* 2014, **10**, 171 (2013); arXiv:1308.1129 (2013).
- [48] I. L. Andronov, *Astrophys.* **55**, 536 (2012); <https://doi.org/10.1007/s10511-012-9259-0>.
- [49] I. L. Andronov *et al.*, *J. Astron. Space Sci.* **32**, 127 (2015); <https://doi.org/10.5140/JASS.2015.32.2.127>.
- [50] M. G. Tkachenko *et al.*, *J. Phys. Stud.* **20**, 4902 (2016); <https://doi.org/10.30970/jps.20.4902>.
- [51] I. L. Andronov, L. L. Chinarova, W. Han, Y. Kim, J.-N. Yoon, *Astron. Astrophys.* **486**, 855 (2008); <https://doi.org/10.1051/0004-6361:20079056>.
- [52] I. L. Andronov, K. D. Andrych, *Odessa Astron. Publ.* **27**, 38 (2014).
- [53] I. L. Andronov, *Astron. Astrophys. Supp. Ser.* **125**, 207 (1997); <https://doi.org/10.1051/aas:1997217>.
- [54] I. L. Andronov *et al.*, *ASP Conf. Ser.* **511**, 43 (2017).
- [55] I. L. Andronov, V. P. Kulynska, *Ann. Astron. Novae* **1**, 167 (2020), http://astronomianova.org/pdf/AAN1_20.20.pdf; preprint arXiv:1912.13096 (2019).
- [56] I. B. Vavilova *et al.*, *Kinem. Phys. Celest. Bodies* **28**, 85 (2012); <https://doi.org/10.3103/S0884591312020067>.
- [57] I. B. Vavilova *et al.*, *Proc. Int. Astron. Union* **12**(S325), 361 (2017); <https://doi.org/10.1017/S1743921317001661>.

MAVKA: PROGRAM OF STATISTICALLY OPTIMAL DETERMINATION OF PHENOMENOLOGICAL PARAMETERS OF EXTREMA. PARABOLIC SPLINE ALGORITHM AND ANALYSIS OF VARIABILITY OF THE SEMI-REGULAR STAR Z UMa

K. D. Andrych^{1,2}, I. L. Andronov¹, L. L. Chinarova^{3,1}

¹*Department of Mathematics, Physics and Astronomy, Odesa National Maritime University, Ukraine, e-mail: tt_ari@ukr.net*

²*Department of Theoretical Physics and Astronomy, Odesa I. I. Mechnikov National University, Ukraine, e-mail: katyaandrich@gmail.com*

³*Astronomical Observatory, Odesa I. I. Mechnikov National University, Ukraine, e-mail: lchinarova@gmail.com*

Advanced MAVKA software for the approximation of extrema observations is used to analyze the variability of the brightness of pulsating and eclipsing stars, but may be useful in analyzing signals of any nature. A new algorithm using a parabolic (quadratic) spline is proposed. In contrast to the traditional definition of a spline as a piecewise-defined function at fixed intervals, a spline is proposed to be divided into three intervals, but the positions of the boundaries between the intervals are additional parameters. The spline defect is 1, that is, the function and its first derivative are continuous and the second derivative can be discontinuous at the boundaries. Such a function is an enhancement of the “asymptotic parabola” (Marsakova and Andronov 1996). The dependence of the fixed signal approximation accuracy on the location of the boundaries of the interval is considered. The parameter accuracy estimates using the least squares method and the bootstrap are compared. It is recommended to use the difference between the 0.975 and 0.025 percentiles (divided by $2 \cdot 1.96$) as the accuracy estimate of a given parameter in the bootstrap method.

The variability of the semi-regular pulsating star Z UMa is analyzed. The presence of multicomponent variability of an object, including four periodic oscillations (188.88(3), 197.89(4) days and halves of both) and significant variability of the amplitudes and phases of individual oscillations is shown. The approximation using the parabolic spline is only slightly better than the asymptotic parabola for our sampling of the complete interval. It is expectedly better for larger subintervals. The use of different complementary methods allows us to get a statistically optimal phenomenological approximation.