

ОСНОВНА ПОСЛЕДОВАТЕЛНОСТ НА ОБРАБОТКАТА И АНАЛИЗА НА ДАННИТЕ ОТ GPS ИЗМЕРВАНИЯ (АЛГОРИТЪМ)

Петър Ковачев

Национален институт по геофизика, геодезия и география – БАН
1113 София, България, ул. „Акад. Г. Бончев“, бл. 3
E-mail: pkov@abv.bg

Резюме: В [1] бяха разгледани някои теоретични аспекти и статистически инструменти, свързани с анализа и метрологичния контрол на часовниците на GPS приемниците. Въз основа на този анализ бяха създадени алгоритми и софтуер за тестване на GPS приемници. Отделни компоненти от този софтуер могат да бъдат използвани при обработката на временни редове в редица области от геодезията, геофизиката, военното дело, телекомуникациите, високоточната навигация и др.

Ключови думи: GPS, Временни редове, Геодезически параметри, Анализ.

1. Въведение

Геодезическите временни редове получени от високоточни геодезически GPS приемници съдържат геофизична информация. Под геодезически временни редове разбираме вариациите на пресметнатите *геодезически параметри* (координатите на точките и параметрите на часовника на приемника, координати на геоцентъра, UT1, координати на полюса) във времето.

Геодезическите временни редове съдържат геофизична информация. За да отделим геофизичната информация (*геофизичните параметри* – координатите на точките, тропосферно закъснение (Zenith Total Delay – ZTD), йоносферна концентрация на свободните електрони – TEC) от шума, трябва да използваме подходящи модели. Критерий за коректност на модела е отсъствието на тренд при анализа на остатъците. Ако остатъците имат случаен характер и спектър, характерен за белия шум, тогава целият систематичен компонент е отстранен от модела.

Преди да пристъпим към анализа на определени геофизични модели трябва да сме сигурни, че сме отстранили всички „инструментални“ систематични грешки на системата, нямащи отношение към моментното действително положение на измерваната точка.

За постигане на максимална точност и надеждност на резултатите при GPS измерванията от съществено значение е контролът на качеството на „суровите“ данни. Така наречените *некорелирани грешки* (причинени от отразени сигнали (multipath), от изменението във фазовия център на антените, от шумовете в приемниците, от местни заглушаващи сигнали и обструкции, пречещи на директното приемане на спътниковите сигнали) са една от основните причини за некоректното определяне на положението на точката, особено за GPS приложенията, работещи в реално време. Тези грешки не са общи за *референтната станция* и *подвижния приемник* и не могат да бъдат коригирани от DGPS/RTK метода.

Тъй като факторите, които влияят на точността на измерванията, са изключително много и не са постоянни във времето и пространството, а подходите за достигане на желаната точност (хардуерни и софтуерни) при различните видове и марки GPS приемници се различават, то се налага разглеждането на решението за конкретния приемник като „черна кутия“.

Обработката на GPS измервания формално може да се раздели на три етапа (фиг. 1). Всички модули от разработения софтуерен пакет се отнасят към етап **I. Предварителна обработка на данните** и етап **III. Анализ на параметрите на решението** (фиг. 1). Трите етапа са взаимосвързани.



Фигура 1. Основна схема при обработката на GPS измервания

2. Извличане на данните за анализ от конкретен файлов формат

Решенията (етап **II**), получени от високоточни геодезически приемници, могат да бъдат разделени в три различни категории в зависимост от периода между момента на измерването и времето, необходимо за обработка на данните.

Обработка на измерванията в реално време (Real Time (RT) processing)

Решенията се получават в рамките на 1 секунда или по-малко. Основните приложения на този метод са в навигацията и мониторинга на деформации, в случаите когато

координатите на измерваните точки могат бързо да се променят. Приложенията в реално време използват точно предсказаните орбити на спътниците и параметрите на часовниците им, както и диференциални поправки.

Обработка на измерванията в почти реално време (Near Real Time (NRT))

Решенията се получават в рамките на няколко часа. Положенията на точките се изчисляват последователно в работен порядък и ако има проблеми с изследвания феномен, те не могат да бъдат установени докато тече обработката на данните. Като пример за приложение на този метод е използването на GPS за метеорологичен мониторинг, където изискванията са финалният продукт да бъде получен в рамките на един (или няколко) часа. Основната разлика между RT и NRT е в използваните параметри на орбитата на спътниците, като NRT работи с изчислените им стойности, а не с предсказаните (пример: ултра-бързите орбитни елементи, изчислени от мрежата на IGS).

Постобработка

Това е най-точният и надежден метод без ограничения за времето на обработка с възможност за използване на прецизни ефемериди и последващи корекции.

Освен фирмения софтуер на производителите на GNSS приемници, съществуват няколко софтуерни пакета за високоточни GNSS приложения (Таблица 1):

BAHN/GPSOBS	European Space Agency (ESA)
Bernese Software	Astronomische Instituts Universitat Bern (AIUB), Switzerland
GAMIT/GLOBK	Massachusetts Institute of Technology (MIT), USA
GIPSY/OASIS	Jet Propulsion Laboratory (JPL), USA
CGPS22	Geological Survey of Canada, (GSC), Canada
DIROP	University of New Brunswick (UNB), Canada
EPOS P V3	GeoForschungs Zentrum (GFZ), Germany
GAS	University of Nottingham, Great Britain
GEODYN	Goddard Space Flight Center (NASA/GSFC), USA
GEOSAT	Norwegian Defense Research Establishment (NDRE), Norway
MSOP	National Aerospace Laboratory, Japan
OMNIS	Naval Surface Warfare Center (NSWC), USA
PAGE3	National Geodetic Survey (NGS), USA
TEXGAP/MSODP	University of Texas Center for Space Research (UTCSR), USA

Таблица 1. Софтуерни пакети за високоточни GNSS приложения

От тях най-разпространени са GIPSY/OASIS, GAMIT и Bernese, разработени в университети и научни институти. Те изпълняват следните функции:

- Обработка на всички измерени параметри, записани от геодезическите GNSS приемници;

- Перманентна обработка на всички точки от мрежата (часови, дневни и т.н. решения);
- Разрешаване на фазовата неопределеност за къси, средни и дълги бази;
- Моделиране на закъсненията на сигнала в йоносферата и тропосферата;
- Моделиране на орбитите на спътниците и параметрите на ротация на Земята;
- Геофизични модели на приливните явления в земната кора, както и на индиректното влияние на океанските приливи на положението на антената на приемника.

Въпреки че тези софтуерни пакети използват едни и същи модели, между тях съществува съществена разлика в подхода на обработка на наблюденията.

За разлика от Bernese и GAMIT, GIPSY/OASIS не използва метода на двойното диференциране, а използвайки и филтър (*Square Root Information Filter* – SRIF), изчислява едновременно всички параметри на пълния модел, вместо да елиминира грешките от часовниците на спътниците и приемниците. Обикновено, този софтуер се използва за анализ на мрежи, като наблюденията от много станции се обработват едновременно. По този начин се определят:

- Параметрите на орбитите и часовниците на спътниците;
- Фазовата неопределеност;
- Координатите на станциите от мрежата и параметрите на часовниците им;
- Зависимите от положението на станциите закъснения на сигналите по трасето на разпространение през йоносферата и тропосферата.

За получаване на тропосферните параметри GIPSY-OASIS използва стохастичен модел, който разглежда неизвестното остатъчно закъснение на сигнала като вариращ във времето параметър, а шумът в данните се приема за random walk процес („Брауново движение“). Зенитното закъснение се моделира за всяка епоха като сума от предишното изчислено зенитно закъснение плюс шума от напълно случаен процес (process noise), ограничаващ измененията в закъснението на сигнала. За целта се използва Калманов филтър. Стохастичните модели се използват с допускането, че закъснението на сигнала варира в определени граници за къси интервали от време. По този начин закъснението на сигнала се определя по начин, подобен на този, по-който се определят параметрите на отклоненията на часовниците.

След обработката на измерванията се получават прецизните параметри на орбитите на спътниците и часовниците им за всяка епоха на наблюдение. Това дава възможност с използването на тези новополучени параметри да се изчислят точното положение на индивидуални станции, използвайки фазовите двучестотни измервания. Тази стратегия се нарича точно (прецизно) определяне на положението (*precise point position* – PPP) и често се използва заедно с GIPSY. Важно предимство на този метод е скоростта на получаването на решенията. Докато времето за изчисление нараства кубично с увеличаването на броя на станциите, когато те се обработват едновременно, то нараства линейно при прилагане на метода PPP.

Разработеният софтуерен пакет в момента включва три отделни от основната програма модула за извличане на параметрите на конкретното решение:

- Извличане на данните от конкретен GPS приемник (Trimble 4000SSE) при работа в диференциален режим [2]. Модулът е разработен на Visual Basic 6.0. Работи под Windows 98SE и Windows XP.
- Извличане на данните от GPS приемници, представени в UNAVCO COMPACT формат, получени с помощта на програмата TEQC на UNAVCO. Модулът е разработен на Visual Basic.NET. Модулът работи под Windows Vista и Windows 7.
- Извличане на данните от решенията на GPS приемници, получени с помощта на APPS (*Automatic Precise Positioning Service*) на Глобалната Диференциална GPS система (*Global Differential GPS - GDGPS*) на НАСА и JPL (*Jet Propulsion Laboratory*) и програмата RTG (*Real Time Gipsy*) – версия на програмата GIPSY/OASIS за работа в реално време. Модулът е разработен на Visual Basic.NET. Работи под Windows Vista и Windows 7.

Основният програмен пакет е разработен на Visual Basic.NET и включва следните модули.

3. Предварителна обработка на данните

Предварителната обработка на данните (*Data Inspection and Modification*) обикновено е необходима за постигане на максимална точност и надеждност на решението. Подпрограмите на този модул включват:

3.1. Редукция/редактиране на данните

- Конфигуриране на проект;
- Конфигуриране на сесия;
- Визуален контрол на данните;
- Привежда временните данни от входния файл в MJD;
- Обратна трансформация (MJD => година, месец, ден, час, минута секунда);
- Съединяване на прозволени брой файлове;
- Отделяне на данните от избран временен прозорец за допълнителен анализ.

Може значително да се намали броят на данните, като се извлече всяка десета стойност от данните за фазата, а данните за честотата съответно могат да се усреднят. Това повишава скоростта на обработка, но е недостатък, когато анализираме решението, използвайки дисперсиите със застъпване на данните или пълните дисперсии.

Могат да се съединят няколко последователни файла или да се раздели определен файл на части.

3.2. Конверсия фаза – честота

- Данните за фазата могат да бъдат конвертирани в данни за честотата чрез разделяне на първите разлики на τ :

$$y_i = (x_i - x_{i-1}) / \tau.$$

- Данните за честотата могат да бъдат конвертирани в данни за фазата чрез умножаването им с τ и добавяне към фазата:

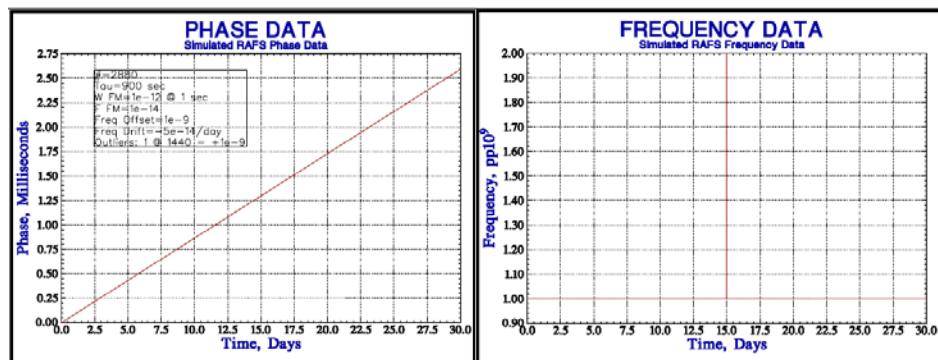
$$x_i = x_{i-1} + (y_{i-1}) \cdot \tau.$$

Конверсията *фаза* > *честота* не работи правилно, когато имаме две съседни еднакви

стойности на фазата. Тогава $y = 0$ и резултатът може да бъде приет като празнина в данните. Използването на много малка ненулева стойност (например, $1e-99$) решава този проблем. Но, две идентични съседни стойности на фазата могат да бъдат знак за проблеми в конкретното измерване или при квантуването на данните.

Конверсията *честота* > *фаза* не може да бъде осъществена, когато съществуват празнини в данните. За да се запази непрекъснатостта на фазата, може да бъде използвана усреднена стойност на честотата, за да се премине липсващия интервал в данните.

Може да бъде използвана индивидуална стъпка вместо фиксирано τ за данни, неравномерно разположени във времето.



Фигура 2. Конверсия фаза – честота. Очевидна груба грешка, която трябва да се отстрани, за да се продължи анализът

3.3. Отстраняване на грубите грешки

Много е важно да използваме надежден метод за откриване и отстраняване на грубите грешки, базиран на робастните статистически методи [3].

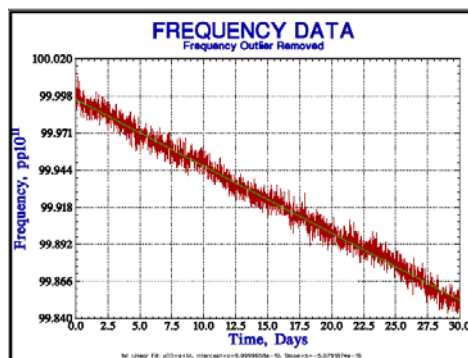
Грубите грешки се откриват много по-лесно в данните за честотата, отколкото за фазата.

Медианата е по-добър избор за определяне на централната стойност (грубите грешки влияят по-силно на средната стойност).

Отклонението от медианата е добър начин за откриване на груби грешки.

Използването на *абсолютното отклонение на медианата* MAD (Median absolute deviation) [4] е добър начин за откриване на груби грешки.

$$\text{MAD} = \text{Median} \{ |y(i) - m| / 0.6745 \},$$
където $m = \text{Median}\{y(i)\}$, а коефициентът 0.6745 прави MAD равен на стандартното отклонение за данни с нормално разпределение.



Фигура 3. Отстранени са грубите грешки. Вече може да се види дрейфът на честотата и шума в данните

Критерият $5 \cdot \text{MAD}$ обикновено е добър избор за отстраняване на грубите грешки.

3.4. Откриване на празнини в данните

В данните за фазата и честотата често се срещат празнини и груби грешки, които след отстраняването им също водят до празнини. (Например, нулева стойност се третира като празнина в данните, освен ако тя не се появява в началото или в края на данните за фазата.) За да се запазят истинските нулеви стойности в данните, на тяхно място се присвоява някаква минимална стойност (например: $1e-99$). Нулеви стойности в данните за честотата могат да се получат при конверсията на данните за фазата, когато две съседни точки имат равни стойности. Тази ситуация може да се реши, като нулата се замени с минимална стойност (например: $1e-99$).

За да се получат смислени резултати от анализа изследваните точки трябва да са равномерно разпределени. Някои от прилаганите дисперсии могат просто да прескочат липсващи точки.

3.5. Методи за запълване на празнините в данните (Gap Handling Methods)

- **Пропускане (Omit):** Използва се [*Edit function*] (или друг текстов редактор) за да се елиминира съществуващата празнина в данните. Този метод елиминира празнината, но също така изменя временната последователност в данните. Той е най-подходящ за отстраняването на няколко изолирани точки.

- **Изключване (Exclude):** Използва се [*Part function*] за да се изключат празнините в началото или края на данните. Той е най-подходящ за отстраняването на празнини намиращи се в началото или края на данните. Съкращава дължината на анализирания ред.
- **Прескачане (Skip):** Празнините могат да бъдат прескочени автоматично при използването на някои дисперсии (виж [1] и Таблица 2). Това е предпочитан метод, когато може да бъде използван.
- **Запълване (Fill):** Използва се [*Fill function*] за да се запълнят празнините с интерполирани стойности. Празнините автоматично се запълват при някои дисперсии (виж долната таблица). Този метод дава задоволителни резултати, ако празнините съставляват по-малко от 5 % от данните [5].

Дисперсия	Празнините се прескачат	Празнините се запълват
AVAR	X	
OVAR	X	
MVAR		X
TVAR		X
HVAR	X	
OHVAR	X	
TOTVAR		X
TOTMVAR	X	
TOTTVAR	X	
TOTHVAR	X	
Théo1	X	
TIE rms	X	
MTIE		X

Таблица 2. Таблицата показва използваните дисперсии (виж [1]) и начина на автоматично отстраняване на празнините в данните

3.6. Анализ на данните, съдържащи празнини

Методите за обработка на празнините в данните имат свои граници на приложение и за да се избягнат грешки, тези граници трябва добре да се разберат. Основният алгоритъм е следният:

Използва се [*Open function*], за да се прочетат данните за фазата; използва се [*Check function*], за да се отстранят грубите грешки; използва се [*Regularize function*], за да се допълнят липсващите точки; използва се [*Convert function*], за да се извърши конверсия до данни за честотата; след което така коригираните данни за честотата се подлагат на анализ.

За данните за фазата съдържащи празнини, но не и груби грешки за предпочитане е да се извърши директен анализ на стабилността. Във всички тези случаи е необходимо да се извърши допълнителна оценка на некоректните резултати. Понякога е по-благо-разумно просто да се анализират свободните от празнини данни.

Таблица 3 обобщава ограниченията при съществуващите методи за обработка на празнините в данните.

Данни за:	Дисперсия	Забележки
фазата	Всички	Няма ограничения
		Да се избягва конверсия честота – фаза
честотата	AVAR	Няма ограничения
	OVAR	Обикновено работи
	MVAR	Да се избягва при големи празнини
	TVAR	Да се избягва при големи празнини
	HVAR	Няма ограничения
	OHVAR	Да се избягва при големи празнини
	TOTVAR	Да се избягва при големи празнини
	MTOTVAR	Да се избягва при големи празнини
	TOTTVAR	Да се избягва при големи празнини
	TOTHVAR	Да се избягва при големи празнини

Таблица 3. Показва как различните дисперсии (виж [1]) работят с данни, съдържащи празнини

3.7. Отстраняване на отместването на честотата (frequency offset)

Отместването на честотата може да бъде определено от данните за фазата с един от двата метода:

- *Линейна апроксимация (Linear Fit):*
Този метод (оптимален за белия фазов шум, *white PM noise*) използва метода на най-малките квадрати (МНМК) за определяне на параметрите на уравнението

$$x(t) = a + bt,$$

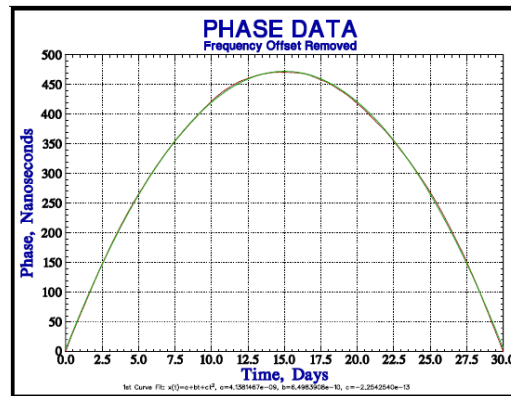
апроксимиращо данните, където $y(t) = b$ е наклонът на правата.

- *Методът на крайните точки (Endpoints):*
Методът просто използва разликата между първата и последната точка от данните, като наклонът е

$$y(t) = [x(\text{end}) - x(\text{start})] / (M - 1),$$

където M е броят на данните. Този метод (оптимален за белия честотен шум, *white FM noise*) можем да използваме като изберем двете крайни точки.

Отместването на честотата в честотната област обикновено се определя чрез усредняване (средната стойност) на данните.



Фигура 4. Изглед на данните след отстраняване на отместването на честотата

3.8. Отстраняване на дрейфа на честотата (frequency drift)

За отстраняване на дрейфа на честотата в данните за фазата, обикновено се използват три метода:

- *Апроксимация с полином от втора степен (Quadratic Fit)*: С помощта на МНМК се определят параметрите на полинома

$$x(t) = a + bt + ct^2,$$

където:

$$y(t) = x'(t) = b + 2ct,$$

а наклонът е равен на:

$$y'(t) = 2c.$$

Този метод е оптимален за белия фазов шум [6].

- *Вторите разлики (2nd Differences)*: Този метод използва средната стойност на вторите разлики на данните за фазата:

$$y(t) = [x(t+\tau) - x(t)]/\tau,$$

наклонът е

$$[y(t+\tau) - y(t)] / \tau = [x(t+2\tau) - 2x(t+\tau) + x(t)] / \tau^2.$$

Този метод е оптимален за случайно блуждаещия честотен шум (*random walk FM noise*) [7].

- *Три точки (3-Point)*: Методът използва три точки – от началото, средата и края на данните за фазата, наклонът е

$$4[x(end)-2x(mid)+x(start)]/(M\tau)^2,$$

където M е броят на данните.

За данните за честотата, операцията е еквивалентна на *метода на бисекцията** (*bisection method*) [8]. Методът на бисекцията е алгоритъм за намиране на корените на уравнение чрез разделяне на интервал от функцията на две, след което аналогично процесът се повтаря многократно за подинтервалите, за които функцията е равна на 0 (т.е. крайните стойности на съответния подинтервал имат различни знаци).

За анализ на дрейфа на честотата в честотната област обикновено се използват четири метода:

- *Линейна апроксимация (Linear Fit)*: Първият, по подразбиране, метод е линейната регресия с помощта на МНК на данните за честотата:

$$y(t) = a + bt,$$

където a е отместването (*intercept* – точката, в която правата пресича оста Y), b е наклонът $= y'(t)$. Този метод е оптимален за белия честотен шум.

- *Бисекция (Bisection)*: Вторият метод определя дрейфа чрез средните стойности на първата и втората половина от данните:

$$\text{наклонът} = 2 [y(\text{втора половина}) - y(\text{първа половина})] / (N\tau),$$

където N е броя на точките. Този метод е оптимален за белия честотен шум и за случайно блуждаещия честотен шум.

- *Логаритмична апроксимация (Log Fit)*: Третият метод използва логаритмичен модел:

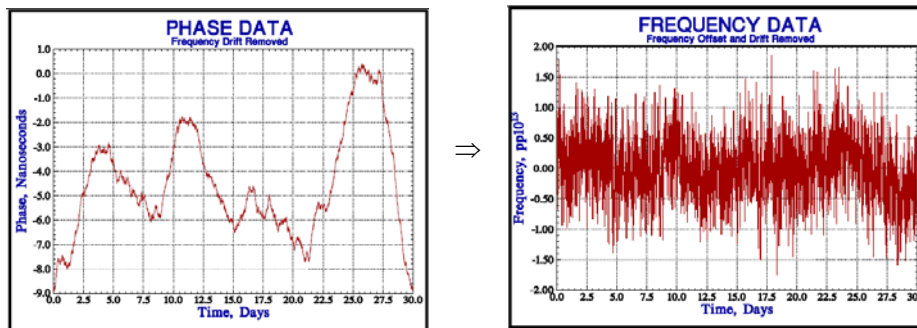
$$y(t) = a \cdot \ln(bt + 1),$$

където наклонът е $y'(t) = ab/(bt+1)$ [9] за честотните данни.

- *Дифузна апроксимация (Diffusion Fit)*: Четвъртият метод използва моделът:

$$y(t) = a + b(t + c)^{1/2},$$

където наклонът $= y'(t) = \frac{1}{2} \cdot b(t + c)^{-1/2}$.



Фигура 5. Изглед на данните след отстраняване на дрейфа на честотата: получаваме остатъците (на втората графика)

4. Анализ на данните

Не е достатъчно само да изчислим една от разгледаните по-горе дисперсии. Необходимо е също така да знаем към каква стойност се очаква да клони и колко близко се доближава до нея. Дисперсията на Алан е дефинирана така, че се очаква да има същата стойност като стандартната дисперсия за белия честотен шум.

Повечето от другите дисперсии, определящи характеристиките на източниците на сигнали за време и честота, са или дефинирани по същия начин, или са изместени оценки на дисперсията на Алан, за които трябва да бъдат добавени съответните корекции (в зависимост от типа на шума и дължината на извадката).

Така, правилно приложени и коригирани, всяка от тези дисперсии клони към една и съща очаквана стойност, чиято статистическата надеждност (*confidence*) нараства с увеличаването на големината на извадката.

Неопределеността на резултатите може да бъде оценена и използвана за определяне на доверителни интервали, като се използва χ^2 -разпределението за дисперсията.

Това изисква познаването на еквивалентния брой на степени на свобода (*equivalent number of degrees of freedom, EDF*), които също са функция на вида на шума и дължината на извадката.

4.1. Разпознаването на вида на шума

Разпознаването на вида на шума е важно не само за разбирането на физическите процеси, протичащи в осцилатора и влияещи на неговата стабилност, но също така, за да бъдат приложени корекциите към изместените оценки и да бъде определена доверителната вероятност.

Често е необходимо да определим доминиращия шумов процес в осцилатора чрез спектралната плътност на относителната нестабилност на честотата:

$$S_y(f) = h_\alpha f^\alpha,$$

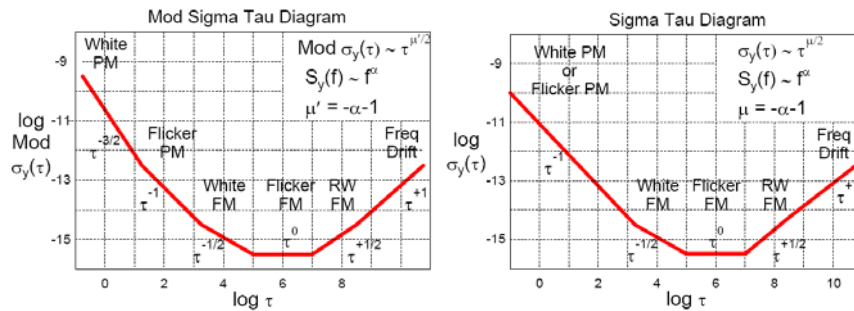
където ($\alpha = 2 \div -4$), за да направим анализ на относителната нестабилност на честотата. Например познаването на вида на шума е необходимо, за да определим (еквивалентния брой на степените на свобода, а оттам – и доверителния интервал и вертикалните скали на грешките (*error bars*)). Освен това, от съществена важност е да познаваме типа на шума, за да коригираме систематичните отклонения в по-новите разработени Пълни (Total) и Thêol дисперсии. Въпреки че типът на шума може да бъде предварително известен или определен „ръчно“, желателно е да имаме някакъв аналитичен метод за неговото разпознаване, който да бъде приложен автоматично като част от алгоритмите за анализ на стабилността на генераторите за всяка точка от данните особено ако трябва да бъдат въведени корекции и/или вертикални скали на грешките.

Видът на шума може да е известен предварително (*a priori*) или може да бъде определен от предварителния анализ на Сигма – тау диаграмите ($\sigma_y^2(\tau) - \tau$), представени в логаритмичен мащаб.

4.2. Сигма – тау диаграми

Сигма–тау диаграмите ($\sigma_y^2(\tau) - \tau$) (фиг. 5), представени в логаритмичен мащаб, показват зависимостта на стабилността на честотата от времето на усредняване и са добро средство за определяне на видовете шумове в осцилатора, които имат различни наклони на тези графики.

За предпочитане е вида на шума да бъде определян автоматично за всеки период на усредняване в границите на изследвания интервал от данните с едновременно определяне на корекциите към изместените оценки и да бъде направена оценка на грешките.



Фигура 6. Сигма – тау диаграми ($\sigma_y^2(\tau) - \tau$)

За да определим автоматично видът на шума, трябва да пресметнем допълнителни дисперсии (например, *модифицираната дисперсия на Алан*), чиито отношения с желаната дисперсия (например, *дисперсия на Алан*) са зависими от различните видове шум и са предварително известни [10].

За тази цел могат да бъдат използвани отношенията (коефициентите) **B1**, **B2**, **B3** и **R(n)** на Барнес [10].

B1 е отношението между стандартната дисперсия и дисперсията на Алан с мъртво време $r = T/\tau$, където T е периодът между измерванията, τ е времето на усредняване, а μ е степенният показател на τ за съответния шумов процес:

$$B1(N, r, \mu) = \sigma^2(N, T, \tau) / \sigma^2(2, T, \tau).$$

B2 е отношението между дисперсията на Алан с мъртво време $r = T/\tau$ към дисперсията на Алан без мъртво време ($r = 1$):

$$B2(r, \mu) = \sigma^2(2, T, \tau) / \sigma^2(2, \tau, \tau).$$

B3 е отношението между стандартната дисперсия с мъртво време $r = T/\tau$ с кратно $M = \tau/\tau_0$ на основното време на усредняване τ_0 към стандартната дисперсия със същото мъртво време и време на усредняване τ :

$$B3(N, M, r, \mu) = \sigma^2(N, M, T, \tau) / \sigma^2(N, T, \tau).$$

R(n) е отношението между модифицираната и класическата дисперсия на Алан:

$$R(n) = \text{Mod } \sigma_y^2(\tau) / \sigma_y^2(\tau).$$

Доминиращият шум може да бъде определен чрез сравняване на получения емпиричен коефициент $B1$ с теоретично очаквания за конкретния шум (при същия коефициент на усредняване). Този метод, макар и несъвършен, е доста ефективен в повечето случаи. Основните недостатъци са: 1) невъзможност да бъдат разграничени белия и трептящият фазови шумове (*white and flicker PM noise*); 2) ограничената му точност за големи интервали на усредняване, когато съществуват малко на брой точки за анализ.

Първият недостатък може да бъде преодолян чрез допълнителен $\mathbf{R}(\mathbf{n})$ тест. Тази техника може да бъде приложена и за останалите усъвършенствани дисперсии. Вторият недостатък може да бъде избегнат чрез предварително определяне на вида на шума с възможно най-голям коефициент на усредняване на данните. Друго допълнително ограничение е, че отношението $R(n)$ не е съдържателно при коефициент на усредняване равен на 1.

Освен с отношението $\mathbf{B1}$, доминиращият шум на осцилатора може да се определи с автокорелационната функция (ACF) (с лаг (интервал, стъпка) равен на 1). Вторият метод е за предпочитане и се използва когато броят на данните е най-малко 30.

ACF с лаг (стъпка) k се дефинира като:

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2}$$

където z_t е временен ред, μ е средноаритметичното на z , σ_z^2 е дисперсията, а E е математическо очакване.

Автокорелационната функция обикновено се изчислява по формулата:

$$r_k = \frac{\frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})}{\frac{1}{N} \sum_{t=1}^N (z_t - \bar{z})^2},$$

където \bar{z} е средноаритметичното, а N е броят на данните [12].

Автокорелацията със стъпка 1 е ефективен метод за определяне на шумовете в осцилатора (W PM, F PM, W FM, F FM, RW FM, FW FM или RR FM), изхождайки от модела, описващ *спектралната плътност* $S_y(f)$ (*power law spectral densities*) на зависещата от времето променлива $y(t)$ с първите членове от степенния ред:

$$S_y(f) \propto h_\alpha f^\alpha,$$

където: f е съответната честота на Фурие или честотната лента (*sideband frequency*) в Hz, h_α е коефициент, показващ нивото на този тип шум, а степента α характеризира различните видове шумове. В разглеждания по-долу алгоритъм

$$h_\alpha f^\alpha = (2 \sin \pi f)^{-2\delta} \quad [13].$$

При $\delta < 1/2$ процесът се стабилизира (т.е. параметрите получават стойности, оставащи неизменни във времето) и стойността на автокорелацията става $\rho_1 = \delta / (1-\delta)$ [14].

Следователно типът на шума може да се определи по формулата $\delta = r_1 / (1+r_1)$. При честотни данни белият фазов шум има стойност $\rho_1 = -1/2$, фликер шумът (фазовият) има стойност $\rho_1 = -1/3$, а белият честотен шум има стойност $\rho_1 = 0$.

За по-разходящите шумове се образуват първите разлики, докато се постигне стабилизиране на параметрите, което се определя от критерия $\delta < 0.25$. Методът за идентифициране на шума използва $p = -\text{round}(2\delta) - 2d$, където $\text{round}(2\delta)$ е 2δ , закръглено към най-близкото цяло число, а d показва броя на образуваните разлики, за да стане $\delta < 0.25$. Ако z е извадка от честотните данни $y(t)$ с период на усредняване τ , то $\alpha = p$; ако z е извадка от фазовите данни $x(t)$ с период на усредняване τ , то $\alpha = p + 2$, където α е степента в израза f^α и така шумът се определя за този период на усредняване. Този метод дава отлични резултати при идентифицирането на всички често срещани шумове от експоненциален тип, както и в по-комплицираните случаи при наличие на смесени шумове.

4.3. Алгоритъм за идентифициране на шума

Алгоритъмът за идентифициране на шума е сравнително прост. Входните данни за изчисление са векторът z_1, \dots, z_N на фазовите или честотните данни, минималният брой разлики, които ще се образуват, d_{\min} (по подразбиране = 0), и максималният брой разлики d_{\max} . Данните, които се получават, са p , оценка за параметъра α на преобладаващия шум и (по желание) стойността за d . Входните данни трябва да бъдат приведени към периода на усредняване τ , за който ще се изчислява оценката.

Ето защо, преди да се приложи алгоритъмът, може да се наложи фазовите данни да бъдат редуцирани, а честотните – да се мащабират с помощта на подходящ коефициент.

За d_{\max} трябва да се избере стойност 2 или 3, ако се използва съответният дисперсионен анализ (Алан или Адамар). α приема стойностите $p + 2$ или p за фазови и честотни данни съответно, като резултатът може да се закръгли на цяло число, въпреки че дробната част може да се използва за оценка на смесените шумове.

Алгоритъмът е бърз и изисква изчисляването само на една стойност за автокорелацията, както и първите разлики за няколко епохи от време. Алгоритъмът е независим от стойностите на дисперсията.

Алгоритъм за идентифициране на шума

Done = False, $d = 0$

While Not Done

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z_i$$

$$r_i = \frac{\sum_{i=1}^{N-1} (z_i - \bar{z})(z_{i+1} - \bar{z})}{\sum_{i=1}^N (z_i - \bar{z})^2}$$

```


$$\delta = \frac{r_1}{1+r_1}$$

If  $d \geq d_{\min}$  And ( $\delta < 0.25$  Or  $d \geq d_{\max}$ )
     $p = -2(\delta + d)$ 
    Done = True
Else
     $z_1 = z_2 - z_1, \dots, z_{N-1} = z_N - z_{N-1}$ 
     $N = N - 1$ 
     $d = d + 1$ 
End If
End While

```

Бележка: p може да се закръгли до най-близкото цяло число.

Noise Type	α	Phase Data* x(t)	d=0 ACF of Phase Data	Lag 1 Autocorrelation, r_1^{\dagger}					
				d=0		d=1		d=2	
				x(t)	y(t)	x(t)	y(t)	x(t)	y(t)
W PM	2			0	-1/2	-1/2	-2/3	-2/3	-3/4
F PM	1			≈0.7	-1/3	-1/3	-3/5	-3/5	-5/7
W FM	0			≈1	0	0	-1/2	-1/2	-2/3
F FM	-1			≈1	≈0.7	≈0.7	-1/3	-1/3	-3/5
RW FM	-2			≈1	≈1	≈1	0	0	-1/2

Фигура 7. Автокорелация с лаг 1: представяне и резултати

Стойностите в сивите полета се използват за идентификация на шума ID (Identification) за конкретния шум и тип на данните.

Методът, използващ автокорелация с лаг 1, дава добри резултати, като надеждно идентифицира чистите шумове за стойности от $\alpha = 2$ до -4 при брой на елементите в извадката 30 или повече, и в повечето случаи идентифицира преобладаващия тип на шума при смесени шумове, ако той е с поне 10 % по-силен от останалите. При наличие на смесица от близки по характеристики шумове, дробната част на стойностите може да е индикатор за съотношението помежду им.

Преди извършването на анализа данните трябва да бъдат подложени на предварителна обработка, за да бъдат отстранени грубите грешки, прекъсванията и систематичните грешки.

N	%	α
32	16	- 0.4 до 1.4
64	6	- 0.4 до 0.7
128	1	- 0.3 до 0.5
256	0	- 0.2 до 0.3
512	0	- 0.2 до 0.2
1024	0	- 0.1 до 0.2

Таблица 4. Процентът на грешно определените стойности спрямо N (броят данни в извадката)

4.4. Еквивалентният брой на степените на свобода

След като сме определили типа на шума, еквивалентният брой на степени на свобода (EDF) за различните статистики се получават от таблици 5, 6, 7, 8 и 9.

AVAR EDF

EDF за дисперсията на Алан с пълно застъпване на данните (*fully overlapping Allan variance AVAR*) се определя по следните приблизителни формули (Таблица 5), за всеки от различните видове шумове

Тип на шума	AVAR EDF , където N е броят на данните за фазата, $m = \tau / \tau_0$	
W PM	$\frac{(N+1)(N-2m)}{2(N-m)}$	
F PM	$\exp\left[\ln\left(\frac{(N-1)}{2m}\right)\ln\left(\frac{(2m+1)(N-1)}{4}\right)\right]^{1/2}$	
W FM	$\left[\frac{3(N-1)}{2m} - \frac{2(N-2)}{N}\right] \frac{4m^2}{4m^2+5}$	
F FM	$\frac{2(N-2)^2}{2.3N-4.9}$	За $m = 1$
	$\frac{5N^2}{4m(N+3m)}$	За $m > 1$
RW FM	$\left[\frac{N-2}{m}\right] \left[\frac{(N-1)^2 - 3m(N-1) + 4m^2}{(N-3)^2}\right]$	

Таблица 5. AVAR EDF за всеки от различните видове шумове

MVAR EDF

EDF за модифицираната дисперсия на Алан (*modified Allan variance (MVAR)*) се определя по същия начин като дисперсията на Адамар с пълно застъпване на данните (**HVAR EDF**), като аргументите се променят както следва: MVAR и TVAR EDF за N, m и $\alpha = \text{MVAR EDF}$ за $N+1, m$ и $\alpha - 2$ (валидна е за $-2 \leq \alpha \leq 2$).

TOTVAR EDF

EDF за пълната дисперсия (TOTVAR) се получава по формулата: $b(T/\tau) - c$, където T е дължината на записаните данни, τ е периодът на усредняване, a, b и c са коефициенти, зависещи от типа на шума, както е показано на таблица: (Таблица 6).

Тип на шума	TOTMVAR EDF коефициенти	
	b	c
W FM	1.50	0
F FM	1.17	0.22
RW FM	0.93	0.36

Таблица 6. TOTVAR EDF за всеки от различните видове шумове

TOTMVAR EDF

EDF за пълната модифицирана дисперсия (*total modified variance TOTMVAR*) се получава по същата формула: $b(T/\tau) - c$, където T е дължината на записаните данни, τ е периодът на усредняване, a, b и c са коефициенти, зависещи от типа на шума, както е показано на Таблица 7:

Тип на шума	TOTMVAR EDF коефициенти	
	b	c
W PM	1.90	2.10
F PM	1.20	1.40
W FM	1.10	1.20
F FM	0.85	0.50
RW FM	0.75	0.31

Таблица 7. TOTMVAR EDF за всеки от различните видове шумове

HVAR EDF

EDF за дисперсията на Адамар с пълно застъпване на данните (*fully overlapping Hadamard variance (HVAR)*) може да се определи по алгоритъм разработен от Гринхол [15], базиращ се на обобщената автоковариационна функция. HVAR EDF се определя или чрез събиране (за малки m и малък брой на периодите), или в ограничена форма за големи m , където $1/\text{EDF} = (1/p)(a_0 - a_1/p)$, с коефициенти, както следва: (Таблица 8)

Тип на шума	HVAR EDF коефициенти	
	$a0$	$a1$
W FM	7/9	1/2
F FM	1.00	0.62
RW FM	31/30	17/28
FW FM	1.06	0.53
RR FM	1.30	0.54

Таблица 8. HVAR EDF за всеки от различните видове шумове

4.5. Доверителни интервали

Разглежданите дисперсии имат разпределение:

$$\chi^2 = \frac{EDF \cdot s^2}{\sigma^2},$$

където χ^2 е разпределението хи-квадрат, s^2 е извадковата (емпиричната) дисперсия, σ^2 е съответната “истинска” дисперсия (от която се интересуваме, но можем да определим само с известно приближение), а EDF е *еквивалентният брой на степените на свобода* (не обезателно цяло число). EDF зависи от броя на анализирани точки и от вида на шума. По много причини χ^2 е дефинирана така, че EDF участва изрично в горното уравнение.

За разглежданите дисперсии могат да бъдат определени едностранни или двустранни доверителни интервали (*confidence intervals*) и вертикални скали на грешките (*error bars*) при съответната доверителната вероятност (*confidence factor*), базиращи се на χ^2 -разпределението.

Стандартната процедура е да се избере доверителната вероятност p , да се пресметне съответстващата χ^2 стойност, да се определи EDF в зависимост от съответната дисперсия, типа на шума и броя на точките в извадката и тогава да се поставят статистически граници на дисперсията.

За двустранен тест:

$$\sigma_{\min}^2 = s^2 \frac{EDF}{\chi^2(p, EDF)} \text{ и } \sigma_{\max}^2 = s^2 \frac{EDF}{\chi^2(1-p, EDF)}$$

Пресмятането (оценката) на доверителния интервал на *дисперсията на Алан* зависи не само от типа на обработката (без застъпване на данните, със застъпване, модифицирана или временна) и номиналната ѝ стойност, но и от коефициента на усредняване и броя на данните, а така също и от избраната статистическа доверителна вероятност и от типа на шума.

Прости доверителните интервали (Simple Confidence Intervals)

Най-простият метод за определяне на доверителни интервали не взима под внимание типа на шума и се визуализира на графиката с вертикални скали на грешките

(за $\pm 1\sigma$ (68%)) с размери $\pm \sigma_y(\tau) / \sqrt{N}$, където N е броят на данните, използвани за изчисляването на дисперсията на Алан.

Фактор на шума K_n (K_n Noise Factor)

По-точно определяне на доверителните интервали може да се извърши, като се вземе предвид типът на шума, който може да се определи с $B1$ (вече споменатото отношение между стандартната (N извадкова) дисперсия и дисперсията на Алан). След което с получения от Таблица 9 **Фактор на шума K_n** се умножава доверителният интервал, за да се получи окончателната стойност.

Шум	K_n
Random Walk FM	0.75
Flicker FM	0.77
White FM	0.87
Flicker PM	0.99
White PM	0.99

Таблица 9. Фактор на шума K_n

Извадки със застъпване на данните (*Overlapping Samples*)

По-добра статистическа надеждност може да се получи, като използваме максимално наличните данни, формирайки всички възможни разлики за зададения период на усредняване. Въпреки, че тези извадки с пълно застъпване на данните не са статистически независими, те помагат да се подобри надеждността на пресметнатата дисперсия.

χ^2 -критерият може да се използва за изчисляването на едностранни и двустранни доверителни интервали и вертикални скали на грешките с избрана доверителната вероятност p .

Този метод, базиран на определянето на степените на свобода на съответните шумове на осцилатора, се прилага за пресмятането на доверителния интервал при използването на дисперсията на Алан със застъпване на данните.

Доверителните интервали за модифицираната дисперсия на Алан и за временната дисперсия също се определят с χ^2 -разпределението, използвайки броя на MVAR (*modified Allan variance*) степените на свобода за съответния тип шум, коефициента на усредняване и броя на данните.

4.6. Брой на точките, участващи в анализа (# Analysis Points)

Броят на точките, участващи в анализа се използва за определяне на доверителните интервали. Без празноти в данните (*gaps*), той е даден в Таблица 10, където: N е броят на фазовите данни, $N = M+1$; M е броят на данните за честотата; $m = AF$ е коефициентът на усредняване (*averaging factor*) [13, 14].

Дисперсия	Брой на точките
Normal Allan	$M/m - 1$
Overlapping Allan	$M-2m+1 = N-2m$
Modified & Time	$M-3m+2 = N-3m+1$
Total	$M-m = N-m-1$
Mod & Time Total	$M-3m+2 = N-3m+1$
Hadamard	$M/m - 2$
Overlapping Hadamard	$M-3m+1 = N-3m$

Таблица 10. Брой на точките, участващи в анализа

5. Заключение

Процедурите за анализ и метрологичен контрол на часовниците на GPS приемници [1] могат да се извършват по следния алгоритъм:

1. Отваря се файлът с данните за фазата.
2. Извършва се визуален контрол на данните.
3. Извършва се конверсия фаза - честота
4. Отстраняват се грубите грешки.
5. Отстраняват се отместването и дрейфът на честотата.
6. Разглеждат се остатъците.
7. При необходимост се запълват празнините.
8. Извършва се статистически анализ с описаните в публикацията методи.
9. Резултатите от анализа се представят в графичен и текстов формат.

При анализа на други параметри, представени като временен ред може да се следва същия алгоритъм или да се изберат подходящите модули от програмата. Описаната методика може да намери приложение в GPS системите за мониторинг и контрол, което ще бъде описано в следваща публикация.

Литература

- [1] Ковачев, П., Методи за анализ и метрологичен контрол на часовниците на GPS приемници, Годишник на секция "Информатика" Съюз на учените в България, Том 3, 2010,67-89, ISSN 1313-6852
- [2] Kovachev, P., (2008), Specialized software for analysis of differential GPS data, Bulgarian Academy of Sciences, Geodesy 21, Sofia 2008, 137-146, ISSN 0324-1114
- [3] Gernot M.R. Winkler, "Introduction to Robust Statistics and Data Filtering", *Tutorial at 1993 IEEE Freq. Contrl. Symp., Sessions 3D and 4D*, June 1, 1993

- [4] Barnett, V. and T. Lewis, *Outliers in Statistical Data*, 3rd Edition, John Wiley & Sons, Chichester, 1994, ISBN 0-471-93094-6
- [5] Barnes, J.A., "Tables of Bias Functions, B1 and B2, for Variances Based on Finite Samples of Processes with Power Law Spectral Densities", *NBS Technical Note 375*, January 1969
- [6] Sullivan, D.B., D.W Allan, D.A. Howe, F.L.Walls (Editors), Characterization of Clocks and Oscillators, *NIST Technical Note 1337*, U.S. Department of Commerce, National Institute of Standards and Technology, March 1990
- [7] Weiss, M., Allan D., Howe D., Confidence on the Second Difference Estimation of Frequency Drift, *Proc. 1992 IEEE Freq. Contrl. Symp.*, pp. 300-305, June 1992
- [8] Weiss M., Hackman C., "Confidence on the Three-Point Estimator of Frequency Drift", *Proc. 24th Annu. PTTI Meeting*, Dec. 1992, pp. 451-460
- [9] MIL-O-55310B-Military specification for crystal oscillators. U.S. government
- [10] Howe, D., R.Beard, C. Greenhall, F. Vernotte, W. Riley, "A Total Estimator of the Hadamard Function Used for GPS Operations", *Proceedings of the 32nd Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting*, November 2000, pp. 255-268
- [11] Box, G., G. Jenkins, *Time Series Analysis, Forecasting and Control*, Chapter 2, Holden-Day, Oakland, California, 1976, ISBN 0-8162-1104-3
- [12] Riley, W.J., C.A. Greenhall, Power Law Noise Identification Using the Lag 1 Autocorrelation, *Proc. 18th European Frequency and Time Forum*, April 2004
- [13] Brockwell, P., R. Davis, *Time Series: Theory and Methods*, 2nd Edition, Eq. (13.2.9), Springer-Verlag, New York, 1991
- [14] Howe, D., R. Beard, C. Greenhall, F. Vernotte and W. Riley, "A Total Estimator of the Hadamard Function Used for GPS Operations", *Proceedings of the 32nd Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting*, pp. 255-268, November 2
- [15] Greenhall C.A., "Recipes for Degrees of Freedom of Frequency Stability Estimators", *IEEE Trans. Instrum. Meas.*, Vol. 40, No. 6, pp. 994-999, December 1991