

SINTEZA LUCRARIII 2007-2010

Etapă 2007: Generarea de modele autoregresive și caracterizarea spectrală a seriilor de date.

Obiectivele etapei: Generarea de modele autoregresive de ordin 1 și ordin superior; Caracterizarea spectrală a acestora.

Scopul etapei: Cunoașterea caracteristicilor spectrale ale diverselor modele AR(p) în vederea aplicării lor la caracterizarea modului de organizare a genomului bacterian.

Activitățile prevăzute: Programare în MATLAB. Analiza spectrală a seriilor de date autoregresive; redactare articol pentru revista cotate ISI

Rezultate livrate pe etapă: Programe MATLAB pentru generarea de modele autoregresive și pentru efectuarea periodogramelor; Articol în revista cotate ISI.

SINTEZA ETAPEI 2007

Modelul autoregresiv AR(p), este un proces stohastic de ordin p dacă variabila aleatoare X_t este o serie staționară și pentru orice t este îndeplinită condiția:

$X_t - \varphi_1 X_{t-1} - \varphi_p X_{t-p} = Z_t$. Procesul stohastic $\{Z_t\}$ este un zgomot alb cu media nulă și varianța σ^2 iar $\varphi_1, \varphi_2, \dots, \varphi_p$ sunt constante reale. Constantele φ_p reprezintă factorii de interacțiune între termenii seriei. Astfel φ_1 poate fi privit ca măsura interacțiunii dintre un termen și vecinul de ordinul întâi, φ_p interacțiunea dintre un termen dat și vecinul de ordin p .

În vederea analizei corelației existente în genomurile bacteriene, am conceput următorii algoritmi de calcul:

- 1) PROGRAM MATLAB PENTRU GENERARE DE SPECTRE TEORETICE DE MODELE DE ORDIN SUPERIOR
Programul generează spectre ale proceselor AR. Parametrii relevanți sunt: constantele de interacțiune φ_k și dispersia σ . Programul poate genera mai multe spectre concomitent pentru a permite compararea lor.
- 2) PROGRAM MATLAB PENTRU COMPARAREA SERIILOR AUTOREGRESIVE GENERATE NUMERIC CU UN SPECTRU TEORETIC
Programul porneste de la o serie necorelată ("random") din care se generează o serie dorită de tip ARp. Peste spectrul seriei se suprapune un spectru teoretic, verificându-se coincidența lor.
- 3) PROGRAM MATLAB PENTRU STABILIREA PERIODICITĂȚILOR ÎN SPECTRUL MODELELOR AUTOREGRESIVE
Derivata funcției putere (spectrală) se anulează când spectrul atinge un maxim local care definește periodicitatea în spectru.

Studii realizate în cadrul etapei 2007

1. S-a efectuat un studiu cantitativ privind influența interacțiunilor la distanță mare între termenii seriei. Pentru aceasta am comparat un spectru AR(1) având $\varphi_1=0.99$ cu un spectru AR(p) caracterizat de $\varphi_1=0.9$ și $\varphi_n=0.09$ (unde $2 < n < 10$). S-a remarcat că la distanțe mari, spectrul AR(p) prezintă periodicități.
2. Am efectuat analiza datelor prin metoda Detrended Fluctuation Analysis (DFA) în scopul detectării corelațiilor de distanță lungă în serii de date nestaționare. Rezultatul analizei DFA se prezintă sub forma unei funcții de fluctuații $F(n)$ în funcție de mărimea n a intervalului analizat.

În reprezentare dublu logaritmică rezultă o dreaptă a cărei pantă reprezintă exponentul de corelare. Dacă dependența este liniară pe toată scala procesul prezintă corelare de distanță lungă. Dacă însă $\log F(n) = \log f(n)$ nu este o dreaptă, faptul indică o corelare limitată (corelare de distanță scurtă). Pentru a înțelege clar informația furnizată de DFA am analizat procese autoregresive de ordin 1 și 2 care sunt prin definiție controlate de corelare de distanță scurtă.

Analiza DFA pentru AR(1) arată că exponentul de corelare și domeniul de corelare reprezintă o pereche caracteristică de date pentru un tip dat de proces AR(1). Analiza modelului AR(2) arată că prezenta celei de a doua interacțiuni între termeni are ca rezultat creșterea exponentului de corelare comparativ cu AR(1). Concluzia generală este că analiza cantitativă a diagramelor DFA poate identifica un proces AR(1) respectiv AR(2).

CONCLUZII ETAPA 2007

- S-au construit programe MATLAB care generează spectrele teoretice de putere pentru orice tip de model autoregresiv precum și periodogramele pentru date numerice.
- Au fost analizate caracteristicile unui proces autoregresiv prin metoda Detrended Fluctuation Analysis, complementară analizei spectrale.
- S-a publicat o lucrare într-o revistă cotate ISI

Etapa 2008: Caracteristici de corelare ale modelelor autoregresive; Caracterizarea spectrală a seriilor de date privind secvențele de codare pentru organisme bacteriene model

Obiectivele etapei: Analiza caracteristicilor de corelare ale modelelor autoregresive și caracterizarea spectrală a seriilor de date cuprinzând lungimea secvențelor de codare pentru organisme bacteriene model.

Scopul etapei:

Analiza complementară (DFA) plus modelare AR1 a seriilor lungimii secvențelor de codare pentru *Bacillus subtilis* precum și pentru șapte linii celulare de *Escherichia coli*.

Activitățile prevăzute: Analiza prin metoda “Detrended fluctuation analysis” a seriilor de date; realizarea de programe în MATLAB; Modelarea seriilor de date cu AR1; redactarea unui articol pentru o revistă cotate ISI.

Rezultate livrate pe etapă:

- Program MATLAB pentru analiza structurii lungimii secvențelor de codare din genomul bacterian: extragerea tendinței, efectuarea și medierea spectrelor, modelarea cu procese autoregresive de ordinul 1
- Analiza cu programul elaborat mai sus, a genomului unor bacterii.
- Articol în revistă cotate ISI

SINTEZA ETAPEI 2008

1. În vederea analizei autoregresive a seriilor de date formate din lungimile secvențelor de codare în genomurile bacteriene, am conceput următorul PROGRAM MATLAB: Analiza și modelarea autoregresivă de ordinul 1 a genomului microbial (bacterii și archaeobacteria). Rezultatul calculului este exprimat în esență prin valoarea factorului de interacțiune ϕ_1 .

2. Analiza unor genomuri microbiene și modelarea lor cu procese autoregresive de ordinul 1

Seria lungimii secvențelor de codare în genom conține două componente asociate celor 2 lanțuri moleculare care stau la baza structurii de dublu helix a ADN-ului bacterian. Analiza AR1

a helixului plus respectiv minus releva deosebiri sesizate de marimea factorului ϕ_1 . De exemplu in cazul *B. Subtilis*, pentru helixul plus $\phi_1 = 0.52$ in timp ce pentru helixul minus $\phi_1 = 0.43$. In acelasi timp dispersia σ este mai mare in primul caz fata de al doilea.

3. Analiza neuniformitatii structurii de corelare:

Procedura implica segmentarea seriei de date integrale intr-un numar de serii de lungime mai mica dar cu o lungime rezonabila (in jur de 500 de date) pentru a acoperi mai mult de 2 ordine de marime. Seria integrala a fost impartita in 4 segmente, fiecare fiind analizat prin aplicarea transformatei Fourier si a modelarii segmentului respectiv cu un process AR1. In cazul model *B. Subtilis*, este remarcabil cum segmentele respective arata o interactiune extreme de slaba intre termenii seriei cu exceptia unei portiuni din interiorul genomului (al doilea sfert din lungimea seriei) unde interactiunea este puternica ($\phi_1=0.6$). Astfel interactiunea in sferturile de serie 1/4, 3/4 si 4/4 este practice inexistentă ($\phi_1<0.1$) in timp ce pentru al doilea sfert este mare ($\phi_1=0.6$). Analiza secventiala arata faptul ca structura seriei este nenuniforma. Pe de alta parte evaluarea intregii serii indica o interactiune apreciabila intre termenii seriei.

3. Analiza efectelor medierii spectrale:

Investigand efectul numarului de termeni mediatii in spectru, am observant ca in cazul medierii cu mai multi termeni palierul de la frecvente joase este foarte bine descris dar fitarea cu AR1 nu este suficient de adecvata in domeniul frecventelor intermediare si mari. O imbunatatire a descrierii se poate face cu un model de ordin superior (asa cum vom demonstra in etapa urmatoare a proiectului).

CONCLUZII ETAPA 2008

- Analiza fluctuatiilor detrendate demonstreaza ca seria secventelor de codare in genomul bacterian prezinta corelare la distanta scurta. Rezultatele au fost publicate intr-o revista cotata ISI.
- A fost realizat un program in MATLAB care modeleaza secventele de codare cu un proces autoregresiv de ordinul 1 (AR1).
- S-a investigat modelarea secventelor de codare la *B.subtilis*. Modelarea cu AR1 este adecvata doar intr-o prima aproximatie. Prin aceasta analiza s-a confirmat structura neomogena a genomului bacterian si prin urmare oportunitatea modelarii prin procese autotregresive.

Etapa 2009: Analiza secventelor de codare din genomul bacterian prin modele autoregresive de ordin superior

Obiectivele etapei:

Modelarea secventelor de codare din genomul bacterian prin modele autoregresive de ordin superior; Cercetari privind organizarea si modelarea genomului bacterian la diverse specii.

Scopul etapei:

1. Identificarea gradului optim al modelului autoregresiv care modeleaza seriile de date formate din lungimile secventelor de codare din genomul bacterian.
2. Determinarea, prin analiza parametrilor obtinuti, a diferentelor existente intre diferite specii bacteriene la nivel de codare a informatiei.

Activitati prevazute:

Programare MATLAB verificare serii de date provenite din genomul bacterian al unor cazuri model; Analiza spectrala si identificarea modelului autoregresiv optim pentru specii bacteriene selectate; redactare articol pentru revista cotate ISI; Participare la The 7th European Biophysics Congress, Italia; Participare la The XX-th International Symposium on Bioelectrochemistry and Bioenergetics, Sibiu

Rezultate livrate pe etapă:

- Program MATLAB destinat modelarii structurii lungimii secventelor de codare din genomul bacterian cu procese autoregresive de ordin superior.
- Analiza cu programul elaborat mai sus, a genomului unor bacterii.
- Au fost publicate 3 lucrari stiintifice: unul intr-o revista codata ISI si doua intr-o baza de date (arXiv) continand lucrari finalizate si nepublicate la data raportarii.

SINTEZA ETAPEI 2009

ADN-ul fiecărei specii bacteriene poate fi caracterizat fie investigând genomul ca întreg fie analizând separat cele două lanturi moleculare notate $I(+/-)$, $I(+)$ și $I(-)$. Datele au fost extrase de pe site-ul web al Laboratorului European de Biologie Moleculară (EMBL).

Prin analiza fluctuațiilor detrendate am demonstrat ca aceste serii prezinta corelare de distant scurta. Urmatorul pas consta din calculul periodogramelor si analiza acestora prin modele autoregresive. Succesiunea de operațiuni poate fi rezumata astfel:

1. Seria este detrendata prin extragerea unui polinom de gradul 10;
2. Se aplica transformata Fourier discreta si periodogramele se mediaza cu 21 termeni;
3. Se fiteaza spectrul mediat al seriilor $I(+/-)$, $I(+)$ și $I(-)$ cu un model $AR(p)$. Factorii de interațiune ϕ_k și dispersia σ depind de gradul de polinomului utilizat pentru procedura de detrendare, precum și de numărul de termeni spectrali utilizati pentru procedura de mediere.

Deși modelele de ordin superior fiteaza mai bine periodograma mediata, modelul autoregresiv de ordinul unu este de cele mai multe ori suficient pentru modelarea genomului bacterian.

În cazul unor bacterii si archaea, factorul de interațiune variază între 0.52 pentru *Bacillus subtilis* și aproape 0 pentru unele tulpini de *Haemophilus influenzae* și *Helicobacter pilori*. Metoda permite evaluarea diferentelor dintre specii si/sau dintre tulpinile aceleiasi specii. Exponentul de corelare are o valoare mica, ceea ce indica o corelare slaba între termenii seriei.

De asemenea am aratat ca exista o relatie între exponentul de corelare α si factorul de interațiune ϕ . Dependenta este liniara între α si ϕ pentru $I(+)$ si $I(-)$, în vreme ce pentru întreg genomul nu mai este liniara.

CONCLUZII ETAPA 2009

- S-a efectuat analiza fluctuațiilor detrendate si analiza autoregresiva pe periodogramele mediate ale seriilor date de lungimile secventelor coding pentru diferite specii de bacteria si archaea.
- Analiza fluctuațiilor detrendate (DFA) indica o corelare de distant scurta.
- Modelele autoregresive de ordin superior $AR(p)$ fiteaza mai bine periodogramele mediate, dar informatiile oferite sunt similare celor obtinute prin fitarea cu $AR(1)$.
- În analiza secventelor coding din genom, se impune separarea datelor pe cele doua helixuri. Dependenta dintre exponentul spectral de corelare si factorul de interațiune este liniara pentru cele doua helixuri (dominant si recesiv). Când investigam întreg genomul, dependenta nu mai este liniara, cele doua helixuri interfera si actioneaza unul asupra celuilalt similar unor zgomote.

ETAPA 2010:

Diseminarea rezultatelor obtinute in cadrul studiilor pentru proiectul ID32 - Organizarea secventelor de codare in genomul microbial; modelarea autoregresiva

Obiectivele etapei:

Diseminare si raport final

Scopul etapei

Diseminarea rezultatelor obtinute prin studiile realizate in cadrul proiectului ID 32 - Organizarea secventelor de codare in genomul microbial; modelarea autoregresiva.

Activitatile prevazute sunt:

Redactare de lucrari si comunicari; Participare la 1st International Conference on Analytical and Nanoanalytical Methods for Biomedical and Environmental Sciences, Brasov 18-20 iunie 2010 si Conferinta Nationala de Fizica, Iasi 2010.

Rezultate livrate pe etapa

1. 3 lucrari stiintifice publicate in revista indexate ISI.
2. O lucrare stiintifica publicata intr-o revista indexata BDI.
3. 2 comunicari la Conferinte Nationale si o comunicare la o Conferinta Internationala
4. O metoda de analiza: "Organizarea secventelor de codare in genomul microbial; modelarea autoregresiva"
5. Un software destinat analizei organizarii secventelor de codare din genom.

SINTEZA LUCRARILOR STIINTIFICE PUBLICATE:

- a) Vasile Morariu, Calin Vamos, Alexandru Pop, Stefan Soltuz, Luiza Buimaga-Iarinca, *Autoregressive modeling of the variability of an active galaxy*, **Rom. Journ. Phys**, 55 (7-8) 2010.

Studiul a fost realizat in vederea validarii modelului autoregresiv prin compararea rezultatelor obtinute prin aceasta metoda cu cele obtinute printr-o metoda complexa (Linear State Space Model).

Metoda de analiza propusa a fost testata pe o serie de date care cuantifica fluxul de raze X emis de galaxia Seyfert NGC5506. Variabilitatea fluxului de raze X a fost descrisa de König si Timmer. Rezultatele raportate de König si Timmer pentru procesul AR1 sunt continute in esenta in parametrul $\phi = 0.994$ si deviatia standard $\sigma = 0.722$.

Metoda prezentata de noi se constituie intr-o alternativa la LSSM fiind semnificativ mai simpla decat aceasta. Eroarea relativa a rezultatelor obtinute cu metoda propusa de noi este de 0.2% pentru ϕ si de 5% pentru σ si confirma valabilitatea metodei de modelare spectrala autoregresiva propusa.

- b) Vasile Morariu, Luiza Buimaga-Iarinca, *Autoregressive modelilg of coding sequence lengths in bacterial genome*, **Fluct. Noise Lett.**, 9(1) 47-59 (2010),

Fiecare specie de bacterii sau archaea se caracterizează printr-o serie de secvențe de codare n fiecare având o anumită lungime lk , unde $k = 1 \dots n$. In analiza noastră sunt investigate atat seria completă constând în succesiunea naturala de lungimi care se regasesc în genom, cat si seriile de secvente coding pe helixul dominant si respectiv pe cel recesiv. Datele au fost extrase de pe site-ul web al Laboratorului European de Biologie Moleculară (EMBL).

Aceste serii au fost analizate prin DFA, analiza spectrala si modelare cu modelul autoregresiv. Au fost identificati doi parametri: α - exponent de corelare și ϕ – factorul de interactiune pe care ii consideram adecvati descrierii cantitative a genomurilor bacteriene. Au fost luate in considerare

mai multe specii de bacterii sau/si tulpini diferite in cadrul aceleiasi specii. In plus am inclus o specie de archaea in vederea unui studiu comparativ intrucat archaea formeaza un domeniu distinct dar au o structură similară bacteriilor.

Analiza unui set numeros de periodograme mediate ne-a permis sa concluzionam ca organizarea CDS poate fi bine descrisă de modelul AR1, care implică doar interacțiunea între termenii vecini. Parametru ϕ al procesului autoregresiv pare să fie mai sensibil decat exponentul de corelare.

c) Vasile Morariu, Calin Vamos, Alexandru Pop, Stefan Soltuz, Luiza Buimaga-Iarinca, Oana Zainea, *Autoregressive modeling of biological phenomena*, **Biophys. Rev. Lett.** 5(3) 1-20 (2010),

Modelul autoregresiv de ordin 1 a fost aplicat in cazul mai multor fenomene naturale. Rezultatele obținute cuprind urmatoarele aspecte:

Lungimea secventelor de codare in cazul genomurilor microbiene.

Au fost analizate 17 tulpini a unor specii bacteriene (*Bacillus sp.*, *E.coli*, *Haemophilus Influenzae*, *Helicobacter pylori*).

- Interacțiunea dintre termenii seriei formate din lungimile secventelor de codare poate sa varieze între valori apropiate de zero si valori care descriu interacțiuni medii (in cazul *Bacillus subtilis* 0.52).

- Factorul de interacțiune este suficient de sensibil pentru a identifica in cadrul aceleiasi specii, pentru toate liniile speciei, similaritatile in modul de organizare al secventelor de codare. De asemenea, este evidenta diferenta între valorile factorilor de interacțiune rezultati din analiza unor specii diferite.

Mobilitatea scheletului atomic al proteinelor

Structura si dinamica unei proteine este determinata de coordonatele atomice si de factorul de temperatura (Debye-Waller) care este determinat prin difracție de raze X asupra proteinelor. Acesta constituie un parametru caracteristic al mobilitatii proteinelor. Datele analizate pentru un numar considerabil de proteine au fost obtinute din baza de date: Protein Data Bank (PDB), <http://www.rcsb.org>.

- Prin analiza spectrală am aratat ca seriile de date de acest tip nu sunt de tip 1/f. Reprezentarea in coordonate dublu-logaritmice a spectrelor mediate prezinta deviatii importante de la liniaritate.

- Aproximativ 40% din cazuri pot fi descrise printr-un model autoregresiv de ordin 1, in vreme ce restul sunt descrise in mod satisfactor de modele autoregresive de ordin superior.

Fenomenul de flickering al celulelor rosii

A fost analizat modul in care se produce fenomenul de fluctuatie membranara ("flickering") a eritrocitelor in plasma sangvina. Fluctuatiile membranare au fost inregistrate in regim secvential cu ajutorul unui montaj experimental care cuprinde: microscop inversat, camera CCD si calculator. Seriile de date formate din parametrii caracteristici pentru celula (aria, diametrul) au fost obținute prin prelucrarea cu un soft informatic a imaginilor secventiale inregistrate.

- Toate seriile analizate au avut un factor de interacțiune cuprins între 0.5 si 0.8 care indica o interacțiune medie.

- Fenomenul de flickering se modeleaza bine cu modele autoregresive de ordin 1.

- Fluctuatiile membranare ale eritrocitelor pot fi incadrate ca fenomen cu corelare pe distanta scurta.

Psihologie cognitiva

Au fost investigate serii de numere aleatoare generate de subiecti umani. Concluzia principala care se desprinde este ca mintea umana este incapabila sa genereze serii "random", algoritmul intrinsec fiind foarte bine descris de modele autoragresive. Mentionam ca acest tip de fenomen poate fi descris numai de modelele autoregresive de ordin superior. Modelele autoregresive de ordin 1 s-au dovedit insuficiente.

Luiza Buimaga-Iarinca, Silvia Neamtu, Mihaela Mic, Ioan Turcu, *Comparative analysis of Bacillus subtilis and Escherichia coli sakai genomes. The high order autoregressive method*, **Rom. J. Biophys**, **20(3)** 203-211 (2010).

Studiul reprezinta o analiza comparativa a doua dintre genomurile utilizate frecvent ca sisteme model: *Bacillus subtilis* si *Escherichia coli*. Am utilizat modele autoregresive de ordin superior (1-9) in vederea identificarii modelului optim care poate descrie corelarea la nivelul lungimilor secventelor de codare.

Am aratat ca modelul autoregresiv de ordin trei descrie cel mai bine aceasta corelare. Factorii de interactiune obtinuti sunt pozitivi in cazul ambelor genomuri, indicand o structura non-random a secventelor de codare. De asemenea am aratat ca cele doua genomuri bacteriene sunt diferit organizate.

SINTEZA LUCRARILOR COMUNICATE:

- a) L. Buimaga-Iarinca, S. Neamtu, M. Mic, I. Turcu, *Characterization of bacterial genomes by autoregressive models*, **IC-ANMBES 2010** – Brasov, Romania, June 18-20 (2010)

Studiul prezinta modalitatea in care genomurile bacteriene, pentru care structura genomului e cea mai simpla, pot fi investigate prin metode matematice. Lungimile secventelor de codare din genom sunt transformate intr-o serie numerica care poate fi subiectul analizei statistice.

Corelarea informatiei in cadrul genomurilor este analizata prin aplicarea Transformatei Fourier si modelarea cu Modelul Autoregresiv. Datele au fost investigate atat cu modele autoregresive de ordin 1 cat si cu modele de ordine superioare. Am aratat ca parametrii rezultati in urma analizei, si anume factorii de interactiune (ϕ), pot sa surprinda asemanarile la nivel de corelare in cadrul genomului pentru toate tulpinile aceleiasi familii bacteriene. De asemenea valorile acestor parametrii sunt semnificativ diferite pentru familii bacteriene diferite.

- b) Silvia Neamtu, Luiza Buimaga-Iarinca, Mihaela Mic, Ioan Turcu, *High order autoregressive methods used for coding sequence analysis in B. subtilis and E. coli chromosomes*, **Conferința Națională de Fizică**, Iasi, 23-25 Septembrie (2010)

In cadrul acestui studiu am investigat aplicabilitatea modelelor autoregresive superioare. Modelele au fost testate pe doua serii de date obtinute din succesiunea lungimilor secventelor de codare in cadrul a doua genomuri model (*B. subtilis* si *E. coli*). Am aratat ca desi modelul autoregresiv de ordin 3 este cel care descrie cel mai bine modul de corelare a informatiei in aceste doua genomuri, modelul autoregresiv de ordin 1 este la randul sau suficient pentru a oferi o descriere rezonabila a organizarii informatiei in aceste genomuri.

- c) Luiza Buimaga-Iarinca, Silvia Neamtu, Mihaela Mic, Ioan Turcu, *The influence of noise and trend in autoregressive modeling*, **Conferința Națională de Fizică**, Iasi, 23-25 Septembrie (2010)

Studiul este axat pe investigarea utilitatii modelului autoregresiv de ordin 1 pe serii apartinand fenomenelor naturale. Acestea in mod obisnuit contin zgomote si trenduri, iar modelul autoregresiv functioneaza numai pentru serii stationare. Mentionam ca o serie este stationara numai daca valoarea medie, deviatia standard si functia de corelare sunt invariante in timp.

In vederea testarii aplicabilitatii metodei am generat serii artificiale cu caracteristici autoregresive peste care am suprapus zgomote si trenduri (polinomial si sinusoidal) cu intensitati cu variabilitatea controlata. Aceste serii au fost investigate cu modelul autoregresiv de ordin 1 pentru a vedea in ce masura zgomotele si trendurile aplicate seriilor initiale afecteaza valoarea parametrului caracteristic (factorul de interactiune).

Am aratat ca zgomotul alb cu intensitati mici (10^{-2} , 10^{-1}) nu are efect asupra modelului AR1. Cand suprapunem zgomot alb cu acelasi ordin de marime ca si datele din serie, modelul AR1 devine un model mult mai complex. Zgomotele mai mari (10^1 si 10^2) acopera cu totul seria, valorile lui ϕ ramanand aproape de zero. Trendurile, cel polinomial si cel sinusoidal nu au efect remarcabil asupra modelului. Cand suprapunem aceste tipuri de trenduri pe serii zgomotoase valoarea factorului de interactiune pentru valori ale lui ϕ mici ($\phi \in [0.3, 0.6]$) este foarte mica, aproape de zero, iar valorile pentru factorii de interactiune mari ($\phi \in [0.7, 0.9]$) sunt reduse puternic. De asemenea, se observa ca pentru valori ale lui ϕ intre 0.6 si 0.8 analiza este perturbata cel mai puternic. La adaugarea unui zgomot alb cu intensitate de acelasi ordin de marime ca al datelor din seria initiala, observam ca valoarea initiala a factorului de interactiune este redusa cu aproape 50 procente, ceea ce sugereaza ca modelul este puternic afectat de zgomot. Pe de alta parte, existenta trendurilor nu afecteaza semnificativ valorile lui ϕ .

Drept urmare, la analiza unor serii de date apartinand unor fenomene reale, primul lucru pe care il recomandam este indepartarea trendului. O simpla fitare polinomiala poate arata daca exista trenduri, periodice sau aperiodice, acestea fiind indepartate printr-o operatie de diferentiere. Zgomotul alb poate fi indepartat prin proceduri conventionale de filtrare. Seriile rezultante, libere de zgomote si trenduri, sunt cele care pot fi subiectul analizei autoregresive.

METODE PROPUSE IN CADRUL PROIECTULUI

Metoda Modelarii autoregresive ca substitut la Metoda LSSM a fost validata in cadrul INCDTIM sub numele "Organizarea secventelor de codare in genomul microbial; modelare autoregresiva".

PRODUSE REZULTATE IN CADRUL PROIECTULUI

"Software destinat analizei organizarii secventelor de codare din genom".

Produsul este rezultatul intregului studiu realizat in cadrul acestui proiect de cercetare, fiind format din mai multe module. Programul permite analiza autoregresiva a seriilor de date cu modelul autoregresiv de ordin 1 (Rezultatul a fost prezentat in Raportul de faza din 2008, programul calculeaza factorii de interactiune pentru AR1) precum si analiza autoregresiva cu modele de ordin superior (Rezultatul a fost prezentat in Raportul de faza din 2009, programul calculeaza factorii de interactiune ai modelelor AR de ordin superior). In plus fata de aceste module programul mai are doua module auxiliare. Primul dintre acestea permite generarea de serii matematice artificiale cu proprietati autoregresive cunoscute (setate de catre utilizator). Cel

de-al doilea permite modelarea si generarea unor serii artificiale cu modele autoregressive de ordin superior).

CONCLUZII FAZA 2010

- Rezultatele obtinute in cadrul acestui proiect de cercetare au fost valorificate prin publicarea in reviste de specialitate si prin comunicarea acestora in cadrul unor conferinte stiintifice.
- Programele informatice dezvoltate in intervalul 2007-2010 constituie un produs finit, functional, inregistrat si validat in cadrul INCDTIM.

CONCLUZII GENERALE

Proiectul de cercetare pe care il finalizam se incadreaza in domeniul cercetarii fundamentale, abordand o directie de investigare extrem de actuala, cu caracter interdisciplinar axata pe intelegerea organizarii genomului. Principalele rezultate constau in elaborarea unor metode matematice de analiza a secventelor de codare din genom si dezvoltarea unor produse de tip software capabile sa puna in practica aceste metode.

Drept indicatorii minimali de performanta din planul de realizare sunt prevazuti 3 lucrari stiintifice publicate in reviste indexate ISI si 2 lucrari stiintifice publicate in reviste BDI.

In intervalul 2007 – 2010 in care s-a derulat proiectul au fost publicate un numar de 8 lucrari stiintifice publicate in reviste indexate ISI si 3 lucrari stiintifice publicate in reviste BDI. O alta lucrare stiintifica este in proces de evaluare. Rezultatele obtinute au fost de asemenea comunicate printr-un numar de 9 participari la Conferinte nationale si internationale de profil.

Lucrari publicate in reviste indexate ISI

1. Vasile V. Morariu, Luiza Buimaga-Iarinca, Călin Vamoș And Ștefan M. Șoltuz, *Detrended fluctuation analysis of autoregressive processes*, **Fluct. Noise Lett.**, vol 7 No 3, L249-L255 DOI No: [10.1142/S0219477507003908](https://doi.org/10.1142/S0219477507003908) (2007)
2. Oana Zainea, V. Morariu, *The length of coding sequences in a bacterial genome: evidence for short-range correlation*, **Fluct. Noise Lett.**, vol 7 No 4, L501-L506 DOI No: [10.1142/S0219477507004173](https://doi.org/10.1142/S0219477507004173) (2007)
3. V. Morariu, *Distribution and correlation of the coding sequence lengths in bacterial genome* **Rev. Chim.** **59**, 11, 1201-1204 (2008)
4. L. Buimaga-Iarinca, V. Morariu, *Short range correlation of the erythrocyte membrane fluctuations*, **J. Phys.: Conf. Ser.** **182**, 012005 (2009)
5. V. Morariu, L. Buimaga-Iarinca, *Autoregressive modeling of coding sequence lengths in bacterial genome*, **Fluct. Noise Lett.**, 9(1), L47-L59 (2010).
6. V. Morariu, C. Vamos, A. Pop, S. Soltuz, L. Buimaga-Iarinca, *Autoregressive Modeling of the Variability of an Active Galaxy*, **Rom. Journ. Phys** **55**, (7-8), (2010)
7. Vasile V. Morariu, Luiza Buimaga-Iarinca, *Autoregressive modeling of coding sequence lengths in bacterial genome*, **Fluct. Noise Lett.**, **9**, **1** 47-59 DOI: [10.1142/S021947751000006X](https://doi.org/10.1142/S021947751000006X) (2010)
8. V. Morariu, C. Vamos, A. Pop, S. Soltuz, L. Buimaga-Iarinca, O. Zainea, *Autoregressive modeling of biological phenomena*, **Biophys. Rev. Lett.** **5**(3), 1-20, (2010)

Lucrari publicate in reviste BDI

9. V. Morariu, *A limiting rule for the variability of coding sequence length in microbial genomes* [arXiv:0805.1289v1](https://arxiv.org/abs/0805.1289v1) [q-bio.GN]
10. V. V. Morariu *Microbial genome as a fluctuating system: Distribution and correlation of coding sequence lengths*, [arXiv:0805.4315v1](https://arxiv.org/abs/0805.4315v1) [q-bio.GN]
11. Luiza Buimaga-Iarinca, Silvia Neamtu, Mihaela Mic, I. Turcu, *Comparative analysis of bacillus subtilis and escherichia coli sakai genomes. The high order autoregressive method*, **Rom. J. Biophys.** 20(3) 203-211 (2010)

Lucrari trimise spre publicare

12. V. Morariu, E. Horvath, G. Matte, I. Turcu, *A limiting rule for the variability of coding sequence length in microbial genomes*, **Fluct. Noise Lett.** (2010) (in evaluate la referenti)

Comunicari la Conferinte

1. C. Morari, D. Bogdan and I. Turcu, *Towards rapid DNA sequencing - ab-initio study of nucleotide sandwiched between Au(111) plates*, the **7th European Biophysics Congress** Genoa, Italy, July 11-15 (2009)
2. C. Morari, D. Bogdan and I. Turcu, *Electronic structure of DNA's base pairs in-between noble metal plate: "ab initio" study*, **XXth International Symposium on Bioelectrochemistry and Bioenergetics**, Sibiu, Romania, 10-14 May (2009)
3. Diana Bogdan, I. Turcu and C. Morari, *Ab-initio study of DNA nucleotides squeezed between Au(111) electrodes: applications to DNA sequencing technology*, **Conferința Națională de Biofizică CNB-2009**, Cluj-Napoca, 1-3 octombrie (2009)
4. L. Buimaga-Iarinca, V. Morariu, *Short range correlation of the erythrocyte membrane fluctuations*, **PIM 2009**, Cluj-Napoca, Romania, 24-26 September (2009).
5. Luiza Buimaga-Iarinca, *Noise and trend influences in autoregressive modeling*, **Exploratory workshop: New tendencies in computational biophysics**, West University, Timisoara, Romania, 4-6 June (2009)
6. L. Buimaga-Iarinca, V. Morariu, *Short range correlation of the erythrocyte membrane fluctuations*, **Conferința Națională de Biofizică CNB-2009**, Cluj-Napoca, 1-3 octombrie (2009)
7. L. Buimaga-Iarinca, S. Neamtu, M. Mic, I. Turcu, *Characterization of bacterial genomes by autoregressive models*, **First International Conference on Analytical and Nanoanalytical Methods for Biomedical and Environmental Sciences IC-ANMBES – 2010**, Brasov, Romania, June 18-20 (2010)
8. Luiza Buimaga-Iarinca, Silvia Neamtu, Mihaela Mic, Ioan Turcu, *The influences of noise and trend in autoregressive modeling*, **Conferința Națională de Fizică**, Iasi, 23-25 Septembrie (2010)
9. Silvia Neamtu, Luiza Buimaga-Iarinca, Mihaela Mic, Ioan Turcu, *High order autoregressive models used for bacterial genome analysis*, **Conferința Națională de Fizică**, Iasi, 23-25 Septembrie (2010)