https://doi.org/10.48368/bgs-2021.1.N5

Statistical methods for the analysis of climate variability

Tsvetelina Velichkova[1], *Natalya Kilifarska*[2]

[1] National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences, 3 "Acad. G.Bonchev", Sofia 1113, Bulgaria, tsvelichkova@geophys.bas.bg.
[2] Climate, Atmosphere and Water Research Institute, Bulgarian Academy of Sciences, 66, Blvd Tzarigradsko Chaussee, Sofia 1784, Bulgaria, natalya_kilifarska@yahoo.co.uk.

Key words: statistical methods, climate variability, geomagnetic field, surface temperature

Abstract
Climate variability is analyzed by applying linear and nonlinear statistical methods. Their use aims to compare and demonstrate the advantages and disadvantages of each technique in the analysis of climate series. The use of correlation analysis with a time lag confirmed the generally accepted view that carbon dioxide (CO_2) is the main initiator of global warming. The multiple linear regression showed that the selection of factor variables must be done very carefully, taking into account the potential dependences between them. Using nonlinear regression, it was found that the variations of the Earth's magnetic field is an alternative to CO_2 as an engine of modern global warming.

Статистически методи за анализ на измененията в климата

Цветелина Величкова[1], *Наталия Килифарска*[2]

[1] Национален институт по геофизика, геодезия и география, Българска академия на науките, ул. "Акад. Г. Бончев", бл. 3 София 1113, България, tsvelichkova@geophys.bas.bg.
[2] Институт за изследвания на климата, атмосферата и водите, Българска академия на науките, бул. "Цариградско шосе" 66, 1784 София, България, natalya_kilifarska@yahoo.co.uk.

Ключови думи: статистически методи, климатични изменения, геомагнитно поле, приземна температура

Резюме
Измененията в климата са анализирани чрез прилагането на линейни и нелинейни статистически методи, като тяхното използване има за цел да сравни и да демонстрира предимствата и слабостите на всеки един от тях при анализа на климатичните редове. Използването на корелационен анализ с отместване във времето потвърди общоприетото схващане, че въглеродният диоксид (CO_2) е основният инициатор на глобалното затопляне. Линейният многофакторен регресионен метод показа, че подбирането на факторните променливи трябва да става много внимателно, с отчитането на потенциалните зависимости между тях. С помощта на нелинейната регресия бе установено, че изменчивостта на магнитното поле на Земята е една алтернатива на CO_2 като двигател на съвременното затопляне на климата.

Увод

Климатът е комплексна система с много променливи. Някои се изменят линейно с времето, а други нелинейно в широк диапазон на пространствено-времеви скали. Поради тази причина, интерпретацията на климатични данни е голяма предизвикателство и се извършва с помощта на различни статистически методи.
За анализирането на климатичните редове най-често се използват линейни статистически методи, тъй като те са сравнително прости и лесни за прилагане. Съществуват определени изисквания, които трябва да бъдат спазени, за да бъдат коректно прилагани: 1) средната стойност и стандартното отклонение на времевите редове да не се изменят във времето; 2) връзките между въздействащите фактори и климатичния отговор са *линейни* и *стационарни* във времето.
Много изследвания показват, че динамиката на климата е нелинейна (Yiou et al., 1994; King, 1996; Miksovsky and Raidl, 2006; Wu et al., 2007; Halko and Tsilika, 2014). Нелинейността на климатичните редове и тяхната нестационарност често се свързва със случайния характер и респективно с непредсказуемостта на климата (IPCC, 2007). Тази гледна точка погрешно предполага, че

нелинейната изменчивост на климата не може да бъде разграничена от неговите случайни колебания. Ключов момент в прилагането на нелинейни методи за оценка на въздействието на факторите влияещи върху климата е правилното определяне на тяхната роля в глобалната променливост на климата. Също така, това зависи както от познаването на еволюцията им във времето, така и от функционалната зависимост между въздействащите фактори и климатичните изменения.

Данни и методи

Проведените изследвания в настоящия труд се основават на месечни данни за относителното съдържание на O_3 на 70 hPa, приземната температура на 2 m над земната повърхност, взети от реанализа ERA 20 century, за периода 1900-2010 г. За анализа на въглеродния диоксид взехме неговите средно годишни стойности от обсерваторията Мауна Лоа в Хавай, както и исторически записи, получени от ледени ядки. Използвахме годишни стойности за галактичните космични лъчи, които взехме от (Usoskin et al., 2008) и редът от стойности беше продължен въз основа на калибрирани стойности от Московския неутронен монитор. Годишните данни за геомагнитното поле взехме от International Geomagnetic Reference Field model. В изследването използвахме още годишни стойности за броя на слънчевите петна, които взехме от WDC-SILCO, Кралската обсерватория на Белгия в Брюксел.

Като представител на линейната техника използвахме *крос-корелационният анализ* с отместване назад във времето на независимата променлива, тъй като някой от изследваните климатични параметри реагират със закъснение на приложеното въздействие. Корелационните коефициенти се изчисляват като отношението между крос-ковариацията и стандартните отклонения на двата реда от данни. Статистическата представителност на корелационните коефициенти се определя с помощта на двустранния t-тест на Student.

Също така, избрахме да използваме линейна многофакторна регресия, тъй като е един от най-често прилаганите статистически методи. Основната цел на множествената регресия е да се състави модел, с който да се анализира влиянието на две или повече независими променливи върху една зависима променлива. Този метод позволява да се оцени както индивидуалното, така и общото влияние на тези фактори. Независимите променливи, включени в модела, трябва да са количествено измерими, както и да не корелират помежду си, т.е. да не са обвързани помежду си с функционална зависимост. Адекватността на многофакторния регресионен модел се оценява посредством регресионния коефициент R, а също така и чрез F-критерия на Фишер.

За да илюстрираме нелинейната техника, приложихме нелинейна регресия. В отличие от корелационния анализ, целта на който е да определи подобието в изменчивостта на зависимата и независимите променливи, регресионният проблем има за цел да построи модел на предполагаемата зависимост между тях. Многофакторният и нелинейният регресионни коефициенти се изчисляват като отношението на описаната към неописаната от модела вариабилност на променливата Y, т.е. като квадратен корен от разликата между пълната вариабилност на зависимата променлива Y и погрешността на регресионния модел, нормирана на пълната вариабилност:

$$\boldsymbol{R} = \sqrt{\left(\frac{\sum\left(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}}\right)^2 - \sum\left(\boldsymbol{Y_i} - \widehat{\boldsymbol{Y_i}}\right)^2}{\sum\left(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}}\right)^2}\right)} \quad (\mathbf{1})$$

Първият член в числителя на формула (1), както и знаменателят, означават пълната изменчивост на зависимата променлива Y, а вторият член в числителя – моделната грешка. Поради тази причина, регресионният коефициент винаги е положително число. Този критерий дава оценка на адекватността на модела дори и в случаите, когато зависимата променлива не е „нормално“ разпределена.

Прието е, че квадратът на регресионния коефициент R^2, умножен на 100, отразява процента на изменчивост на моделираната променлива, който регресионният модел е в състояние да опише.

Изброените техники са предоставяни от професионалния статистически пакет STATISTICA 8, който предлага множество цифрови и графични критерии за определяне адекватността на статистическия модел.

Резултати получени с линейни статистически методи

За независими параметри избрахме двукратно изгладените по пет точки зимни стойности на озона (O_3) на 70 hPa, годишно усреднените стойности на въглеродния диоксид (CO_2), годишно и пространствено усреднените стойности на земно магнитно поле (F), изгладения ред с плъзгащ се 11- и 22-годишен прозорец на годишното средно на броят слънчеви петна (S_s), изгладения ред с 11- и 22-годишно плъзгащо се осредняване на средно-годишните стойности на галактичните космични лъчи (GCR). Приземната температура на въздуха (T_2m), измерена на 2 м. над земната повърхност през зимата (декември-март) е зависимата променлива. Изчислените корелационни коефициенти, заедно с времезакъснението в реакцията на приземната температура, са представени в Таблица 1.

Таблица 1. Линейни крос-корелационни коефициенти (r) на зимната глобална (пространствено усреднена) приземна температура (T_2m) с различни въздействащи фактори. Изчислените коефициенти са статистически значими при ниво на достоверност 2σ. На втория ред е показано времевото закъснение на отклика на температурата в години.

	2 * 5год. изгл. зим. O_3	Ср.год CO_2	Ср.год. F	11-год. изгл. S_S	22-год. изгл. S_S	11-год. изгл. GCR	22-год. изгл. GCR
Зим. T_2m	-0.57	**0.85**	**-0.72**	0.59	0.57	-0.48	-0.51
Време закъснение	00	**- 0**	**-1**	2-48	2-48	1-20	-1

Ако разгледаме таблицата, ще забележим, че корелацията между CO_2 и T_2m е най-силна r=0.85. Този резултат подкрепя общоприетото твърдение, че емисиите на CO_2 са основният двигател за наблюдаваното затопляне на климата (Cox et al., 2000; IPCC, 2007). Корелацията на приземната температура с изгладените слънчеви петна S_{s11}=0.59 и S_{s22}=0.57 е значително силна, но времезакъснението в реакцията на температурата е много голямо и трудно за обяснение. Зависимостта с измененията на GCR, надхвърлящи 22 години, е силна r=-0.51 и с отместване във времето от една година. Това предполага съществуването на механизъм, чрез който GCR могат да влияят върху приземната температура (Kilifarska, 2012). Корелационният метод дава доста добра свързаност между O_3 и T_2m r=-.057. Трябва да отбележим силната свързаност между геомагнитното поле F и приземната температура r= -0.72 с времезакъснение от 1 година. Този резултат предполага съществено влияние на измененията на геомагнитното поле върху климата на Земята.

Широка практика в научните изследвания е създаването на многофакторни регресионни модели, в които приноса на всеки от потенциалните фактори влияещи върху изменчивостта на анализираната променлива се определя вътрешно в модела като парциален регресионен коефициент. Частичните корелационни коефициенти, обаче, нямат самостоятелно значение. Най-често те се използват в етапа на формулирането на модела за изключването на факторите с несъществено влияние върху зависимата променлива.

Като илюстрация на възможностите на многофакторния регресионен анализ уточнихме регресионните коефициенти на уравнението:

$$T_{2m} = a_0 + a_1.smtO_3 + a_2.CO_2 + a_3.F + a_4.GCR_{22} \quad (2)$$

отчитащо съвместното влияние на четири фактора върху променливата T_2m. Със $smtO_3$ сме означили двукратно изгладените по пет точки зимни стойности на озона; CO_2 е редът с годишните стойности на въглероден диоксид; F е интензитетът на геомагнитното поле, а GCR_{22} – годишните стойности на галактическите космични лъчи, изглатени с 22-точков прозорец. Статистически значимият коефициент на многофакторната регресия има стойност R=0.86, а справката с коефициента на детерминираност (R^2) показва, че моделът е в състояние да обясни 76% от пълната изменчивост на приземната температура в изследвания период 1900-2010 г. Всички корелационни параметри са

статистически значими и според F- критерия на Фишер регресионният модел се отличава значимо от регресионния модел с нулеви параметри.
Сравнението на многофакторния регресионен коефициент с крос-корелационния коефициент на температурата T_2m с CO_2, както и с техния нелинеен регресионен коефициент (представен по-долу), показва незначително „подобрение“ на регресионния модел след включването на 3 допълнителни факторни променливи. Този резултат е силно озадачаващ и навежда на мисълта за съществуваща зависимост между факторните променливи. Действително, анализът на регресионните коефициенти между факторните променливи, както и коефициента им на толеранс (представени в Таблица 2) показва изключително силна зависимост между тях. Освен това, анализът на парциалните и полу-парциалните регресионни коефициенти показва, че те „обясняват“ много малък процент от изменчивостта както на приземната температура, така и на дисперсията на моделните грешки.

Таблица 2. Оценки на свързаността между факторните променливи участващи в многофакторния регресионен модел.

Redundancy of Independent Variables; Dependent Variable: wint.T_2m R-square column contains R-square of respective variable with all other independent variables				
	Toleran.	**R-square**	**Partial - Cor.**	**Semipart - Cor.**
smt. wintO3	0.161794	0.838206	0.310678	0.165753
ann.CO_2	0.034215	0.965785	0.516309	0.305733
ann F	0.015503	0.984497	0.200262	0.103657
22yr avr of GCR	0.059939	0.940061	-0.214122	-0.111164

Този експеримент показва, че подбирането на факторните променливи в регресионното уравнение трябва да става много внимателно, с отчитането на потенциалните зависимости между тях. Ако такава предварителна информация не съществува, то използването на разнообразни методи за определянето на адекватността на регресионния модел е абсолютно наложителна.
В следващия параграф ще направим оценка за ролята на различните въздействащи фактори в изменчивостта на климата, прилагайки нелинейна регресия.

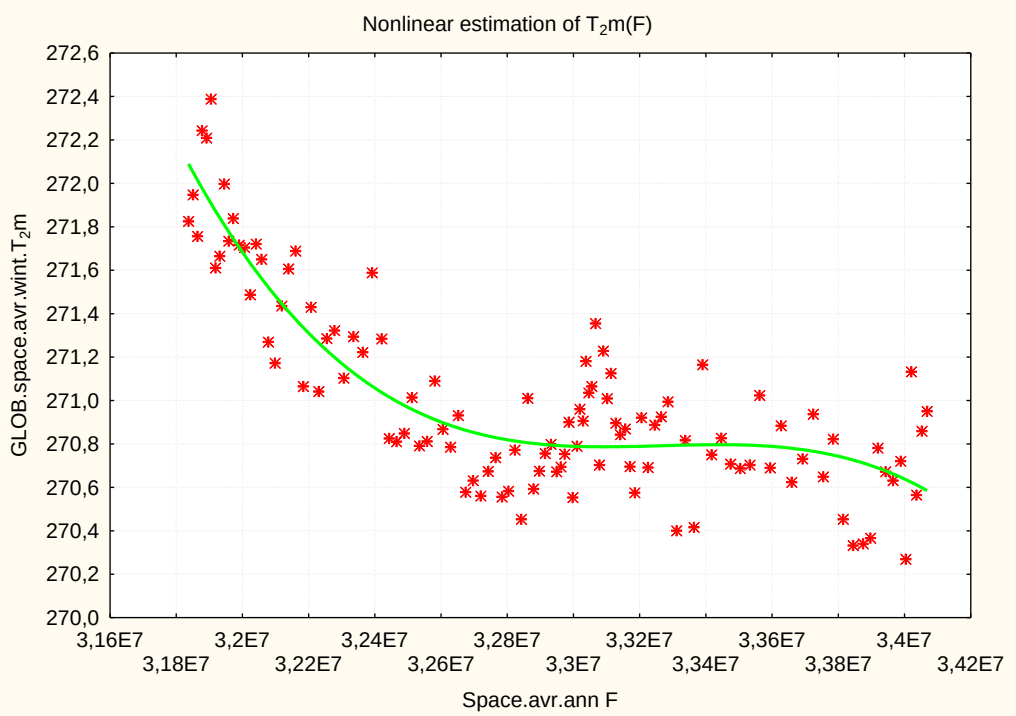
Резултати получени с нелинейни статистически методи

Пример за добър статистически модел, е моделът, чиито грешки (т.е. разликите между измерените и моделираните стойности на зависимата променлива, наричани още остатъци) са нормално разпределени и не корелират помежду си. Графичният анализ на наблюдаваните и моделните стойности, както и анализът на грешките, са много добър начин за откриване на отклонения от тези изисквания и респективно за точността на модела.
Така например, графиката на зависимостта между наблюдаваните и моделните стойности, в която всички точки трябва да са групирани около регресионната линия ни дава графична илюстрация за точността на модела. Големите флуктуации на наблюдаваните около моделните стойности, както всички нелинейни изкривявания, като например разпределението във формата на латинската буква ‘S’, е индикация за лошото качество на модела.
Графиката на съотношението между моделните и нормално разпределените грешки ни дава информация на „нормалността“ на разпределението на грешките. Така точките върху тази графика трябва да са близко разположени около регресионната линия. Друг критерий за доброто качество на модела е случайното разпределение на точките около нулевата линия (без наличието на структурни нееднородности) в графиката на зависимостта между моделните стойности и грешките (т.е. разликата между наблюдаваните и моделните стойности).

Използвахме нелинейната техника *non-linear estimation*, за да анализираме отново въздействието на гореспоменатите климатични фактори върху приземната температура (T_2m).
Високата корелация между магнитното поле на Земята и глобалната температура (виж Таблица 1) основателно възбужда интереса към възможността за „прогнозиране“ на изменчивостта на температурата въз основа на познанията ни за вариациите на магнитното поле. Приложихме кубична функция, за симулация на свързаността между двете променливи:

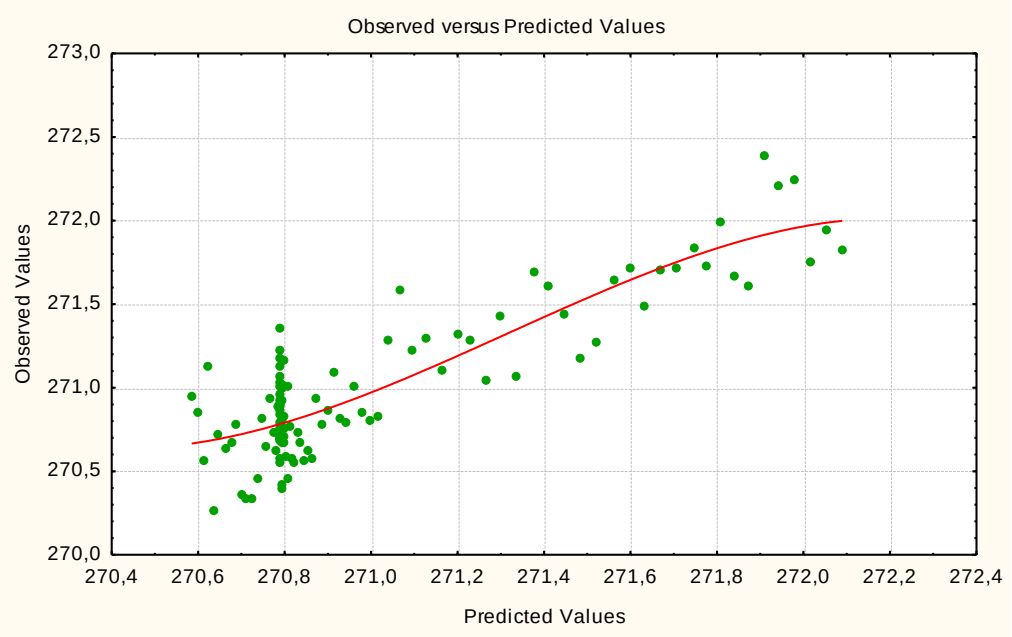
(3) $$T_2m = a_0 + a_1 * F + a_2 * (F)^2 + a_3 * (F)^3$$

където с F сме означили пространствено усреднения интензитет на геомагнитното поле, а коефициентите a_i се уточняват посредством регресионната процедура. Това доведе до съществено повишение на коефициента на регресия R= 0,87, в сравнение с линейния корелационен коефициент (r=-0,72, виж Таблица 1).
В допълнение към изложените критерии за достоверност (на които трябва да се подчинява всеки задоволителен статистически модел), сравнението на най-доброто (с минимизиране на грешката) графично изображение на данните (червените снежинки) с графиката на адаптираната функционална зависимост (зелената линия) бе използвано като допълнителна диагностика за точността на нашия модел (Фиг. 1).



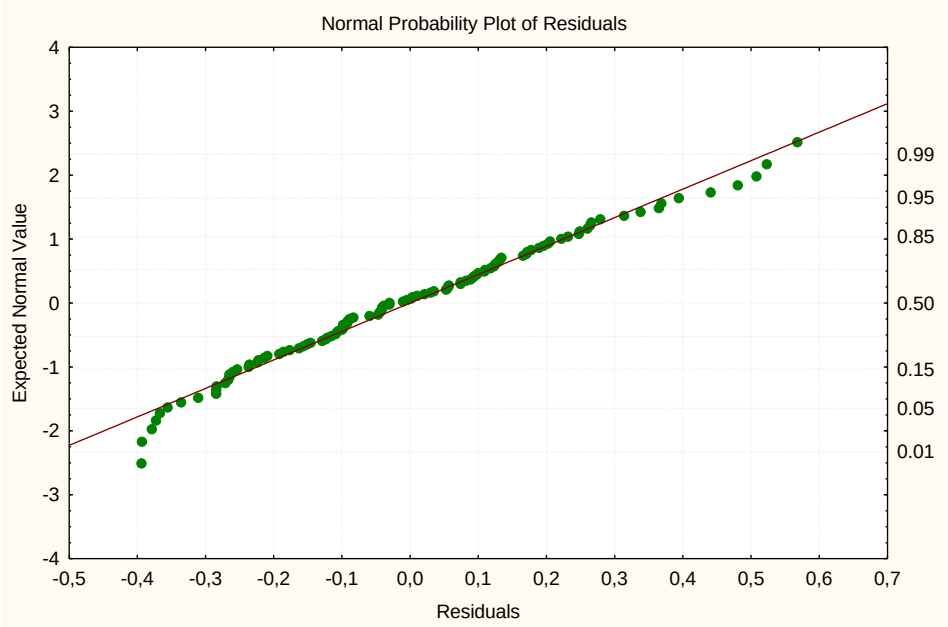
Фиг. 1 Графика на зависимостта между интензитета на геомагнитното поле (F) и приземната температура (пространствено усреднени) за периода 1900-2010г.

От фигурата се забелязва, че усилването на F, води до понижаване на T_2m. Това е резултат, който съответства на заключенията направени от редица автори (Valet & Meynadier 1993; Worm 1997; Knudsen & Riisager 2009; Kitaba et al. 2013).
Доброто съответствие между прогнозните и наблюдаваните стойности показва възможността на кубичната функция да моделира ковариацията на геомагнитното поле и приземната температура (Фиг. 2).

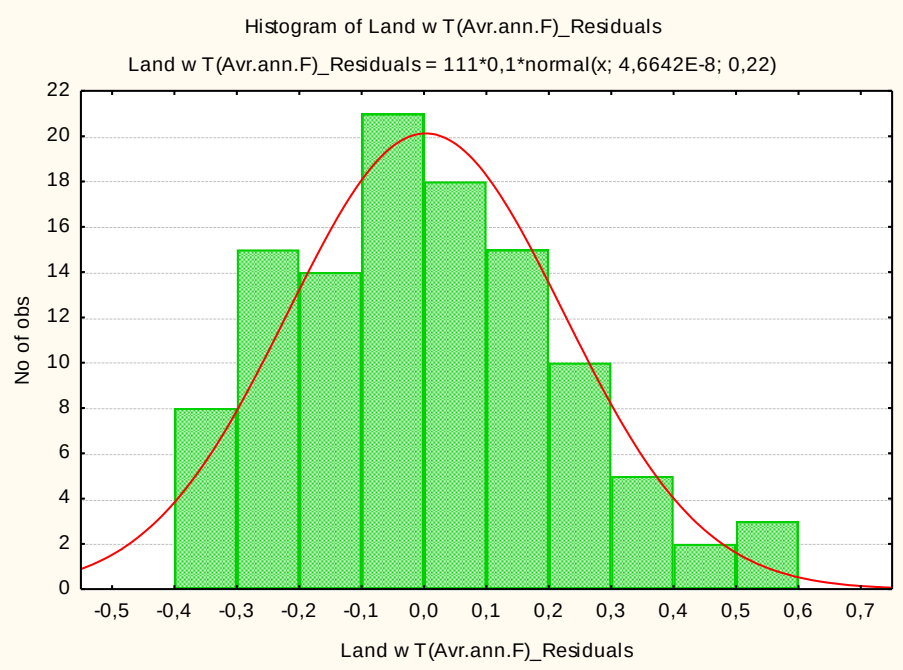


Фиг. 2 Диаграма на разсейване между прогнозните спрямо наблюдаваните стойности.

Други две доказателства за високата точност на модела дават графиките на разпределението на остатъците (Фиг. 3) и тяхната хистограма (Фиг. 4), които потвърждават, че грешките между моделните и измерените стойности са нормално разпределени.

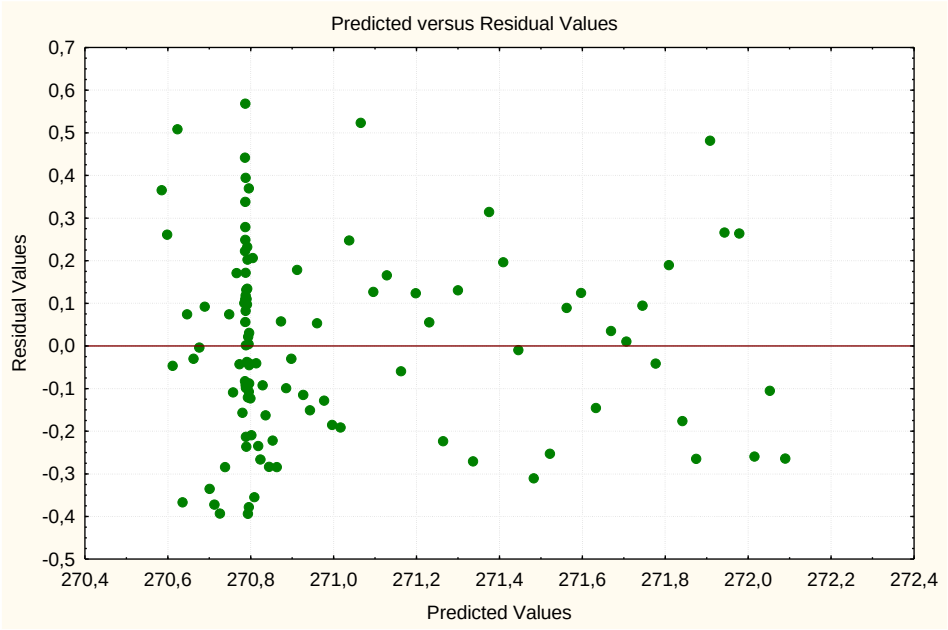


Фиг. 3 Нормално разпределение на разликите между реалните и моделните стойности на T2m.

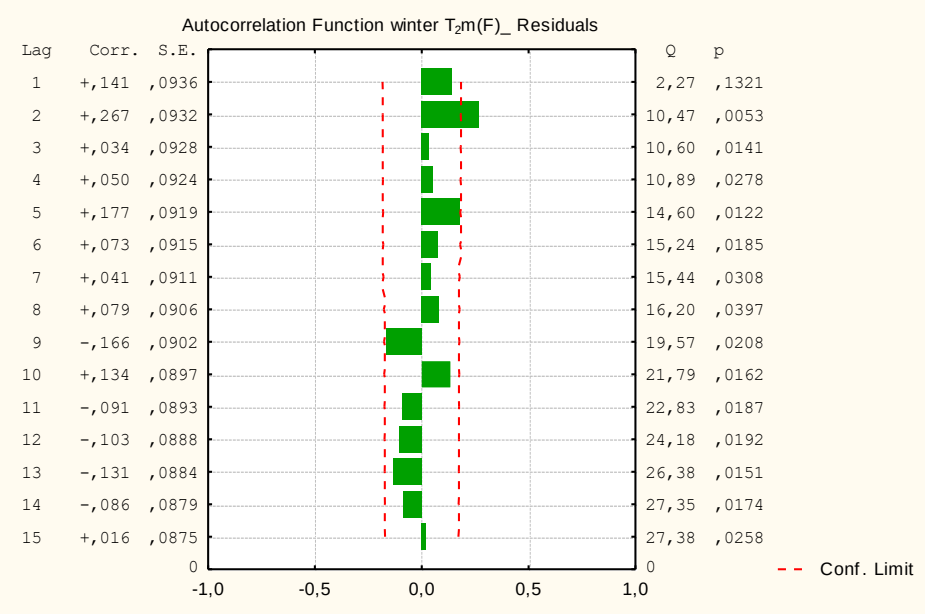


Фиг. 4 Хистограма на разпределението на остатъците.

В добавка, грешките от интерполирането на ковариацията между температурата и геомагнитното поле посредством функционалната зависимост (3) са случайно разпределени спрямо нулевата линия, което е един от споменатите критерии за един качествен статистически модел (Фиг. 5).
На Фиг. 6 сме показали и друга диагностика за качеството на модела $T_2m=f(F)$, а именно автокорелацията на остатъците. Тя показва, че грешките, допуснати при моделирането на функционалната зависимост между геомагнитното поле и температурата, не са свързани помежду си , т.е. са случайно разпределени. Това гарантира, че зависимостта на T_2m от F се описва добре от избрания модел с кубична функция.



Фиг. 5 Диаграма на разсейване на прогнозните стойности спрямо остатъците.



Фиг. 6 Автокорелация на остатъците на T2m=f(F).

Обобщение на получените нелинейни регресионни коефициенти, е представено в Таблица 3. В най-долния ред на таблицата е представено процентното съотношение на изменчивостта на приземната температура описвано от различните нелинейни модели. Сравнението на приноса на разглежданите фактори в изменчивостта на приземната температура показва, че вариациите с времето на пространствено усредненото геомагнитно поле е равностойна алтернатива на увеличението на антропогенните газове в следствие от развитието на човешката цивилизация – поне за разглеждания период 1900-2010 г. Другите два фактора с потенциално голям принос в съвременното затопляне на климата са дългопериодичните стойности на галактическите космически лъчи и озона на 70hPa.

Таблица 3. Нелинейни регресионни коефициенти (R) на зимната глобална (пространствено усреднена) приземна температура (T2m) с различни въздействащи фактори.

	O_3 + 2 * 5год. изгл. зим. O_3	**Ср.год CO_2**	**Ср.год. F**	**11-год. + 22-год. изгл. S_S**	**11-год. + 22-год. изгл. GCR**
R	0.66	**0.87**	**0.87**	0.34	0.64
R^2	44%	**76%**	**76%**	12%	41%

Заключение

Целта на днешния доклад бе да обърне внимание на преимуществата и недостатъците на линейните и нелинейни статистически методи при анализа на климатичните промени.
Корелационният анализ с отместване назад във времето на независимата променлива потвърди общоприетото схващане, че въглеродният диоксид има водещ принос за глобалното затопляне на Земята. Също така, анализът разкри значително влияние на времевите изменения на земното магнитно поле върху вариациите на приземната температура, в разглеждания 110-годишен период. Експериментът с линейната многофакторна регресия демонстрира, че подбирането на факторните променливи в регресионното уравнение трябва да става много внимателно, с отчитането на потенциалните зависимости между тях. Ако такава предварителна информация не съществува, то

използването на разнообразни методи за определянето на адекватността на регресионния модел е абсолютно наложителна.
Нелинейният статистически анализ показа, че вариациите на геомагнитното поле са в състояние да „обяснят“ същия процент от изменчивостта на приземната температура за периода 1900-2010 г., както и CO_2 – а именно 76%. Крайната оценка за качеството на избрания от нас модел, се основава на редица критерии за диагностика, които потвърдиха неговата точност.

Благодарности:

Авторите изказват своята благодарност на Фонд научни изследвания - договор No. ДН 14/1 11.12.2017 и Министерство на образованието и науката – Национална програма “ Млади учени и постдокторанти”, DCM # 577 / 17.08.2018, с чиято помощ бе осъществено изследването.

Литература

Cox, P., Betts, R., Jones, C. et al. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408, 184–187 (2000). https://doi.org/10.1038/35041539

IPCC, 2007, (Intergovernmental Panel on Climate Change), Climate Change 2007: The Physical Science Basis, edited by Solomon S., et al., Cambridge Univ. Press, New York, 966pp.

Kenny, D.A., 1979, Correlation and causality, John Wiley & Sons Inc, New York.

Kilifarska, N., 2012, Mechanism of lower stratospheric ozone influence on climate. Corporate Governance An International Review. 6. 279-290.

Kitaba, I., Hyodo, M., Katoh, S., Dettman, D.L., Sato, H., 2013, Midlatitude cooling caused by geomagnetic field minimum during polarity reversal, PNAS, 110, 1215–1220, https://doi.org/10.1073/pnas.1213389110.

Knudsen, M.F., Riisager, P., 2009, Is there a link between Earth’s magnetic field and low-latitude precipitation?, Geology 37, 71–74. https://doi.org/10.1130/G25238A.1.

Usoskin, I.G., Mursula, K., Solanki, S.K., Schuessler, M., Kovaltsov, G.A., 2002, A physical reconstruction of cosmic ray intensity since 1610, J. Geophys. Res., 107(A11), doi:10.1029/2002JA009343.

Valet, J., Meynadier, L., 1993, Geomagnetic field intensity and reversals during the past four million years, Nature, 366, 234. https://doi.org/10.1038/366234a0.

Worm, H.-U., 1997, A link between geomagnetic reversals and events and glaciations, Earth and Planetary Science Letters, 147, 55–67, https://doi.org/10.1016/S0012-821X(97)00008-3.

Yiou, P., et al., 1994, Nonlinear variability of the climatic system from singular and power spectra of late Quaternary records, Clim. Dyn., 9, 371–389.