

СТОХАСТИЧНО МОДЕЛИРАНЕ НА ВРЕМЕВИЯ РЕД ОТ ПОВЪРХНОСТНИТЕ ТЕМПЕРАТУРИ НА ГЛОБАЛНИЯ ОКЕАН ЗА ПЕРИОДА 1850–2014 г.

ЕВГЕНИ ВЛАДИМИРОВ¹, СТИЛИЯН ЕВТИМОВ²

¹ *Национален институт по метеорология и хидрология, София*
² *Катедра „Метеорология и геофизика“*

Евгени Владимиров, Стилиян Евтимов. СТОХАСТИЧНО МОДЕЛИРАНЕ НА ВРЕМЕВИЯ РЕД ОТ ПОВЪРХНОСТНИТЕ ТЕМПЕРАТУРИ НА ГЛОБАЛНИЯ ОКЕАН ЗА ПЕРИОДА 1850-2014 г.

Моделира се стохастично скаларният времеви ред от средните годишни повърхностни температури на глобалния океан за периода 1850–2014 г. Стриктно е аргументиран изборът на стохастичния модел $ARIMA(0,1,2) \times (0,0,1)_{18}$. Сезонността от осемнадесет години е интерпретирана като проявление на главния приливен цикъл. На базата на модела се дискутира декомпозицията на реда от океанските температури на стохастичен тренд и стационарна компонента и са оценени относителното значение на дългосрочния и непосредствения ефект от един шок и дисперсията на стохастичния тренд. Предложени са хипотезите, че локалните пикове в автокорелационната функция на диференцираните океански температури при 2, 36 и 11 години са съответно удвояване на годишния цикъл, проявление на климатичния цикъл на Brikner и слънчевия цикъл на Schwabe. Дългосрочната прогноза също е обсъдена.

Evgeni Vladimirov, Stilian Evtimov. STOCHASTIC MODELING OF THE TIME SERIES OF GLOBAL OCEAN SEA SURFACE TEMPERATURE FOR 1860-2014 PERIOD

A stochastic modeling of the scalar time series of the annual sea surface temperatures of the global ocean for the period 1850-2014 is carried out. The stochastic model $ARIMA(0,1,2) \times (0,0,1)_{18}$ is strictly argued. The seasonality of eighteen years is interpreted as a manifestation of the main tidal cycle. Based on the model, the decomposition of ocean temperature time series into a stochastic trend and a stationary component is discussed, as well as the importance of long-run and immediate effect of a shock together with variance of the stochastic trend are estimated. The hypothesizes that local peaks of the autocorrelation function of the differenced time series at lags 2, 36 and 11 years are a period doubling of the annual cycle, a manifestation of the Blikner's climate cycle and Schwabe's solar cycle respectively are hypnotized. The long term forecast are also debated.

Keywords: See surface temperature, global ocean, stochastic modeling, trends and cycles
PACS numbers: 92.70.Jn, 05.40.-a.

За контакт: Стилиян Евтимов, Катедра „Метеорология и геофизика.“, Физически факултет, Софийски университет „Св. Климент Охридски“, бул. Джеймс Баучър 5, 1164 София, тел.: +359 2 8161 413, E-mail: evtimov@phys.uni-sofia.bg

1. УВОД

Глобалният океан е една от основните компоненти на земната климатична система. Повърхностните температури на океана (SST) са важен параметър в климатичните модели. Данните за SST се използват за детектирането на климатичните изменения, тяхното предсказване и като гранични условия – за атмосферни реанализи или различни модели на глобалната циркулация.

Увеличеното покритие на инструменталните наблюдения през последния век и половина даде възможност за оценка на месечната повърхностна температура на глобалния океан и нейния ход през този период. Под температура на глобалния океан се разбира на практика средната по площ от измерванията в различни точки. Целта на предлаганата работа е построяването на стохастичен модел на скаларния времеви ред на годишната повърхностна температура на глобалния океан. Данните са от Британския отдел по изследване на климата за периода 1850–2014 г. Моделирането е проведено със съвременните постижения и техники за анализ на времеви редове.

Трябва да се отбележи, че стохастичното моделиране на повърхностни морски температури е с традиции. Общата концепция за климата като масивна Браунова частица е развита в [1] и на тази база се предлага едно просто стохастично уравнение. Конкретно за океана тези идеи се прилагат в [2]. В тази светлина настоящото изследване е продължение на тези традиции за натрупаните нови експериментални данни и с помощта на съвременните постижения на статистиката в изследването на нестационарни времеви редове. Не по-малко внимание е отделено и на физичната страна на получените формални статистически резултати.

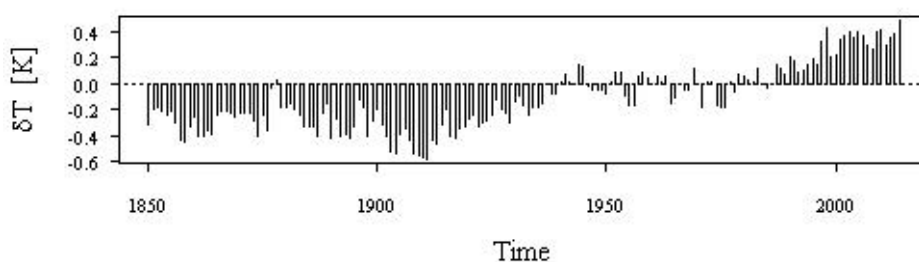
Съдържанието на настоящата работа е, както следва. Раздел 2 разглежда произхода и формирането на данните. Раздел 3 е посветен на провеждането на редица статистически тестове за установяване на природата на нестационарността на времевия ред от годишните повърхностни температурни аномалии на глобалния океан за периода 1850–2014 г. В раздел 4 се предлага един стохастичен модел, адекватно отчитащ особеностите в еволюцията на океанската температурна редица. Моделът се дискутира в раздел 5, където се разглеждат дългосрочната прогноза, физичните аспекти на получените формални резултати, а също се обосновават те и някои хипотези.

2. ДАННИТЕ

Повърхностната температура на океана в дадена точка се дефинира като интегралната температура на повърхностния слой с дълбочина 1–10

m. Под повърхностна температура на глобалния океан се разбира средната по площта на океаните. Средната годишна температура се получава, като се осредни средномесечната на 12-те календарни месеца.

Фиг. 1. представя времеви ред на средногодишните повърхностни температурни аномалии δT на глобалния океан за периода 1850–2014 г. Аномалиите са относно климатичната норма за периода 1961–1990 г.



Фиг. 1. Времеви ред на средногодишните повърхностни температурни аномалии на глобалния океан за периода 1850-2014 г.

Самият времеви ред е от базата данни HadSST3T на Climatic Research Unit of the University of East Anglia (<http://www.cru.uea.ac.uk/cru/data/temperature>) Ще коментираме накратко произхода и първичната обработката на наблюденията, върху които се базират изследваните данни, следвайки [3, 4].

Повърхностните морски температури се измерват по различни методики от началото на XX век, но за начало на т. нар "инструментален период" се смята 1850 г. По-ранните записи не са представителни в глобален мащаб. В момента най-богатата на информация за повърхностните морски температури е базата данни International Comprehensive Ocean-Atmosphere Data Set, ICOADS. Тя обединява в себе си United Kingdom Meteorological Office Marine Data Bank, US Merchant Marine Collection, World Ocean Database, Russian Marine Meteorological Data, Kobe Collection и др. Тази база данни включва измервания от кораби, океанографски станции, закотвени и плаващи буйове и изследователски плавателни съдове.

Съществуват три основни метода за измерване на повърхностните морски температури *in situ*: от кораби, от плаващи и от закотвени буйове. Повечето данни са от плаващи кораби, т.нар. Voluntary Observing Ships, които са снабдени със стандартизирана апаратура и съответните инструкции. Движението на плаващите буйове и вълните осигуряват представителна оценка на температурите за дълбочина до един метър. Закотвените буйове са в разнообразни варианти, което води до големи

разлики в грешката на измерванията.

Като първа стъпка наблюденията на повърхностните температури се подлагат на качествен контрол. Измерените стойности преминават набор от проверки. Изисква се наблюденията да са валидни относно момента на провеждането и местоположението, отхвърлят се, ако са във интервала ± 8 °C около климатичното средно, ако са под точката на замръзване на морската вода $-1,8$ °C, или са нереалистични в сравнение с измерванията в съседните точки.

Следваща стъпка е т.нар. gridding. По същество това е трансформиране на измерените температури до аномалии относно климата на избран базисен период във възлите на регулярна мрежа. След това за всяка клетка се пресмята претегленото средно от индивидуалните наблюдения и по редица характеристики за дадената клетка се определя най-представителното за нея измерване. Големината на клетките е $5^\circ \times 5^\circ$. Клетките с недостатъчно измервания остават „бели петна“. По-нататък получените стойности се трансформират в аномалии и се осредняват за всеки месец. Аномалиите представляват разликите между измерените стойности и средната температура за даден климатичен период. За базата данни HadSST3 климатичният период е 1961–1990 г. Накрая изчислените аномалии се оценяват за различните видове грешки, съставят се предположения за тяхното възникване и се разработват модели за тяхното описание.

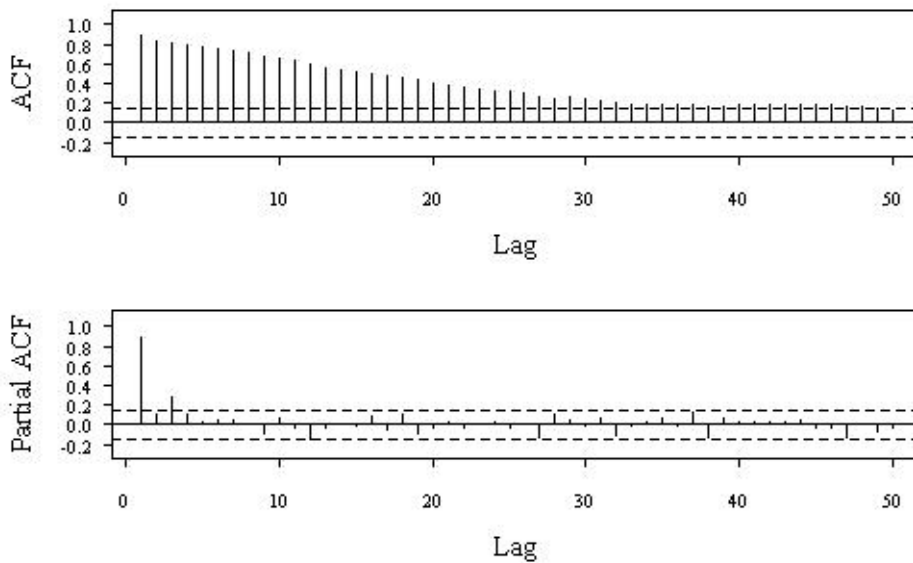
3. ПРЕДВАРИТЕЛНИ ТЕСТОВИ ПРОЦЕДУРИ

3.1. МОТИВАЦИЯ

От фиг. 1 визуално може да се направи изводът, че времевият ред от средните годишни повърхностни температурни аномалии на глобалния океан за периода 1850–2014 г. е нестационарен. Фиг. 2 представя автокорелационната и частичната автокорелационна функция на този ред. По хоризонталната ос са нанесени закъсненията на автокорелационната функция от 0 до 50 и на частичната автокорелационна функция от 1 до 50. Интервалът на дискретизация е една година. По вертикаланата ос на горния панел са разположени стойностите на автокорелационната функция, а на долния панел – на частичната автокорелационна функция. Двете, симетрични относно нулата, прекъснати хоризонтални линии маркират доверителните граници при предположението, че входният сигнал е бял шум. От фиг. 2 се вижда, че автокорелационната функция затихва бавно с увеличаване на закъснението. Поне до 40 години автокорелациите са над доверителната граница. Що се отнася до частичната автокорелационна функция, то тя има висок изолиран пик при закъснение единица, след което практически затихва под доверителните

граници. Както е известно, подобно поведение на извадките с крайна дължина индикира нестационарност [5].

И така, налице са основания за по-задълбочен анализ на нестационарността на изследвания времеви ред. От друга страна, адекватното стохастично моделиране също изисква диагностициране на природата на евентуалната нестационарност. Съвременният анализ на времевите редове разработва четири възможности: тренд-стационарност, стационарност след диференциране (unit root), структурни промени и дълга памет. Докато при тренд-стационарност оригиналната редица става стационарна след подходящо линейно детрендване, то при стационарност след диференциране това става след диференцирането на редицата. Под структурни промени се разбира стационарност около линеен тренд, но с една или няколко точки на структурни промени, в които параметрите на тренда претърпяват скок. Процесите с дълга памет се позиционират между стационарност и стационарност след диференциране. Това са времевите



Фиг. 2. Автокорелационна функция (горе) и частична автокорелационна функция (долу) на времеви ред от температурните аномалии

редове, които стационарират след прилагането на т.нар. филтър на дробното диференциране. Когато показателят на диференцирането е число между 0,5 и 1, оригиналната редица е нестационарна, защото нейната дисперсия расте с времето.

3.2. ТЕСТ ЗА СТАЦИОНАРНОСТ НА DICKEY-FULLER

Ще подложим времеви ред за океанските температури на т.нар. тест на Dickey-Fuller [6]. Методиката на тестване следва [7]. В пълния си обем тестът протича на три стъпки.

Процедурата стартира с регресиране на диференцирания изходен ред върху линеен тенд, върху оригиналния ред при закъснение единица и върху краен брой ендогенни променливи. Ендогенните променливи всъщност са диференцираният изходен ред при закъснения от 1, 2 и т.н. Те се включват, за да се осигури некорелираност на регресионните остатъци. В нашия случай се оказва, че за тази цел са достатъчни две ендогенни променливи. Като начало се тества нулевата хипотеза, че редицата от аномалиите е стационарна след диференциране. Алтернативата е тренд-стационарност. Съгласно [8] при стойност на нестандартна t -статистика от 2,9177 съответната p -стойност е 0,1597 и нулевата хипотеза за стационарност след диференциране не може да се отхвърли дори при ниво на значимост 10%. При това положение се тества хипотезата за стационарност след диференциране без наклон на тренда, но с дрифт. Съгласно емпирично изчислените критични стойности в [9] при съответна статистика от 4,8435 тази хипотеза не може да се отхвърли при ниво на значимост 10%. В тази ситуация се тества хипотезата за стационарност след диференциране, но без тренд. Стойността на съответната статистика от 3,4967 показва, че и тази хипотеза не може да се отхвърли за ниво на значимост поне 10%.

На този етап излиза, че анализираният времеви ред е случайно блуждаене. Според следваната тестова методика обаче, трябва да се премине към тестова регресия без наклон на тренда. Отново само две от ендогенните променливи, за които стана въпрос по-горе, са достатъчни за некорелираност на регресионните остатъци. Сега първоначалната нулева хипотеза е стационарност след диференциране с дрифт. Алтернативата е стационарност около константа. Стойността на статистика е $-0,735$ и съгласно [8] съответната p -стойност е 0,8338. Нулевата хипотеза не може да бъде отхвърлена и трябва да се тества хипотезата за стационарност след диференциране при отсъствието на тренд. При стойност на съответната статистика от 0,6521 и тази хипотеза също не може да се отхвърли.

На този етап заключението е, че времевият ред от океанските температури е стационарен след диференциране без тренд и дрифт в генериращия данните процес. Това се потвърждава и от третата стъпка на следваната тестова методика, при която в тестовата регресия няма тренд. Отново две ендогенни променливи осигуряват некорелираност на остатъците. Нулевата хипотеза е стационарност след диференциране срещу алтернативата за стационарност. Стойността на статистика е $-1,065$ и съгласно [8] съответната p -стойност е 0,2585. Нулевата хипотеза не може

да се отхвърли.

И така, окончателното заключение е, че времеви ред от океанските температурни аномалии е стационарен след диференциране без тренд и дрифт. Остава обаче въпросът дали само еднократното диференциране е достатъчно за стационарирането на този ред.

За да изследваме този въпрос, ние повтаряме проведената вече тестова процедура, но не с оригиналния, а с диференцирания времеви ред. Оказва се, че остатъците на регресията са некорелирани още при една ендегенна променлива. Стойността на тестовата статистика още на първата стъпка от процедурата е $-14,1094$ и нулевата хипотеза за стационарност след диференциране се отхвърля при нива на значимост под 1%. Съответната p -стойност съгласно [8] е $1,093 \times 10^{-15}$. Според следваната методика този резултат е достатъчен, за да се направи изводът, че еднократно диференцираният времеви ред от океанските температурни аномалии е стационарен.

3.3. ТЕСТ ЗА СТАЦИОНАРНОСТ НА KWIATKOWSKI-PHILLIPS-SCHMIDT-SHIN

Според проведения тест на Dickey-Fuller времеви ред от океанските температурни аномалии е стационарен след диференциране. Както обаче е известно, при крайни извадки този тест в редица случаи е със сравнително малка мощност [7]. В [10] се предлага друг тип тестова процедура, при която нулевата хипотеза е стационарност срещу алтернативата стационарност след диференциране. Подобен подход е в съответствие с консервативната стратегия, че по-естествено е хипотезата, представляваща интерес, в случая стационарност след диференциране, да е алтернативата. Тестовият модел предполага, че изходният времеви ред е суперпозиция от тренд, случайно блуждаене и гаусов шум. Нулевата хипотеза е, че дисперсията на шоковете в случайното блуждаене е нула и следователно редът е тренд-стационарен. Алтернативата е ненулева дисперсия или стационарност след диференциране. Разглежданите варианти са два: стационарност около линеен тренд и стационарност спрямо константно средно ниво. Асимптотичните разпределения на съответните статистики са нетривиални и техните проценти са табулирани в [10].

Тестовата процедура изисква да се фиксира дължината на прозореца на Bartlett, използващ се за пресмятане на дисперсията на грешката за една стъпка. Оказва се, че дължината на този прозорец е 13 закъснения. Резултатите от самия тест са, както следва. В случая на нулева хипотеза за стационарност около константа стойността на съответната статистика е 1,0133 и тази хипотеза може да се отхвърли при нива на значимост от 1%. В случая на нулева хипотеза за стационарност около тренд статистиката е 0,1885. И тази хипотеза може да се отхвърли при ниво на значимост 2,5%.

И така, проведеният тест за стационарност отхвърля нулевите

хипотези за стационарност относно тренд с наклон или около константа. Отново обаче възниква въпросът достатъчно ли е само еднократно диференциране за стационарирането на времевия ред. За да изследваме този проблем, ние диференцираме оригиналния ред и отново прилагаме използваната вече тестова процедура за стационарност с дължина на прозореца на Bartlett 13 закъснения. За нулева хипотеза – стационарност около константа – стойността на съответната статистиката е 0,1799 и хипотезата не може да се отхвърли дори при ниво на значимост 10%. За нулева хипотеза – стационарност около тренд – статистиката е 0,0509 и тя също не може да се отхвърли при ниво на значимост 10%. Изводът е, че времевият ред от океанските температурни аномалии стационарира след еднократно диференциране.

3.4. ТЕСТ ЗА СТАЦИОНАРНОСТ СЛЕД ДИФЕРЕНЦИРАНЕ НА ZIVOT-ANDREWS

В т. 3.1 бе споменато, че по принцип съществува възможност един времеви ред да е стационарен около тренд с точки на резки структурни промени. Тук ние ще тестваме времевия ред от океанските температурни аномалии за тази възможност. За целта се използва т.нар. тест за стационарност след диференциране на Zivot-Andrews [11]. Той е изложен подробно в [7, 11]. Нулевата хипотеза е стационарност след диференциране. Алтернативата е стационарност около тренд с една точка на структурна промяна. Вариантите са три. При първия вариант скокът е само в константата на тренда, във втория случай – само в наклона, а в третия – скок както в константата, така и в наклона. Самата точка на структурна промяна не се задава предварително, а се определя в процеса на тестването, така че минимално да се фаворитизира нулевата хипотеза за случайно блуждаене с дрифт. Тестовите регресии включват и ендегенните променливи, коментирани в т. 3.2.

Резултатите от този тест за реда от океанските температури са, както следва. Оказва се, че в случая на алтернативата за структурна промяна само в константата на тренда две ендегенни променливи са достатъчни за некорелираност на остатъците. Годината на евентуалната структурната промяна е 1889. Стойността на съответната статистика е $-4,2797$ и нулевата хипотеза за стационарност след диференциране не може да бъде отхвърлена дори при ниво на значимост от 10%. В случая на структурна промяна само в наклона на тренда ендегенните променливи отново са две, а годината на структурната промяна е 1908. Стойността на тестовата статистиката е $-4,0678$. Нулевата хипотеза за стационарност след диференциране отново не може да бъде отхвърлена дори при ниво на значимост от 10%. Накрая, за алтернативата структурна промяна в наклона и константата на тренда при две ендегенни променливи и година на структурна промяна 1900 стойността на статистиката е $-4,3983$. И в този

случай нулевата хипотеза за стационарност след диференциране отново не може да бъде отхвърлена. Изводът е, че хипотезата за стационарност след диференциране на времевия ред от океанските температури не може да се отхвърли срещу алтернативите за стационарност с различен тип на структурна промяна в линейния тренд.

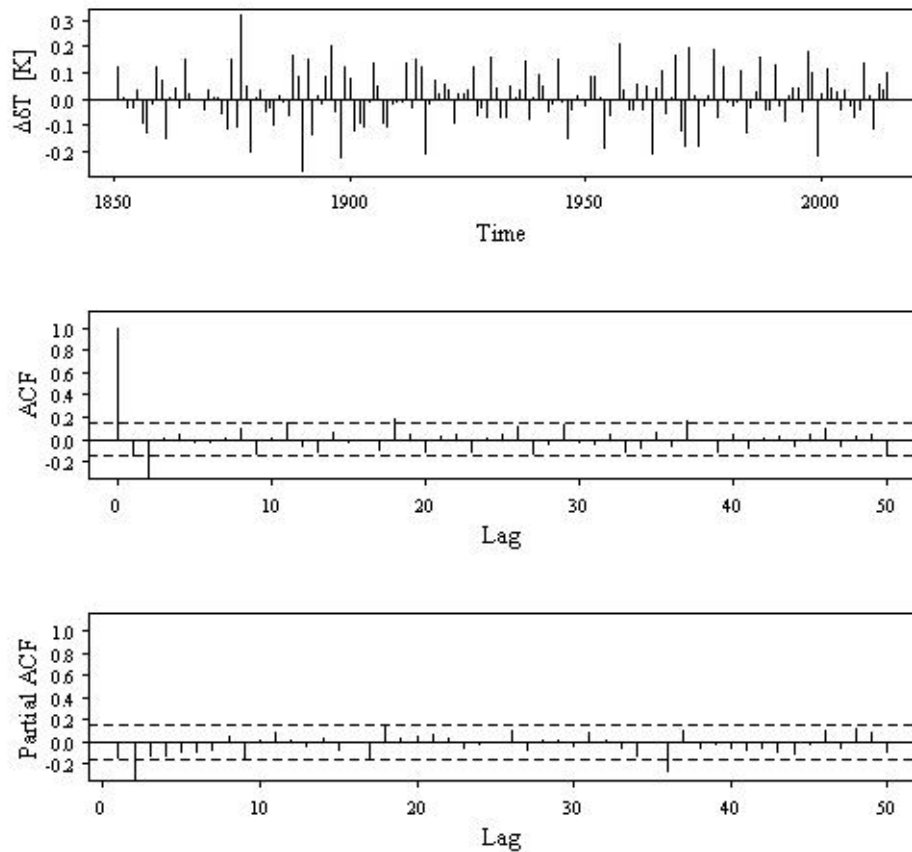
Остана да коментираме споменатата в т. 3.1 четвърта възможност за нестационарност, а именно, времевият ред да стационарира след филтрацията с филтъра на дробно диференциране. Според [12] указание за подобно поведение е отхвърлянето на нулевата хипотеза в тестовете за стационарност след диференциране, например на Dickey-Fuller, и едновременно с това отхвърляне на нулевата хипотеза в тестовете за стационарност, например теста на Kwiatkowski-Phillips-Schmidt-Shin. В нашия случай това обаче не е така. Както видяхме, тестът на Dickey-Fuller не отхвърля нулевата хипотеза за стационарност след диференциране. Можем да заключим, че в сравнение с възможността за стационарност след диференциране тази за дробно интегрируем процес е малко вероятна.

Да резюмираме. Проведените в този раздел тестови процедури дават основания за заключението, че нестационарността във времевия ред от океанските температурни аномалии е от типа стационарност след диференциране. Диференцирането на тази редица води до стационарност, като по всичко изглежда – не само без тренд с наклон, но и без дрифт.

4. СТОХАСТИЧНО МОДЕЛИРАНЕ

4.1. ИДЕНТИФИКАЦИЯ НА МОДЕЛА И ОЦЕНКА НА ПАРАМЕТРИТЕ

В раздел 3 бе установено, че след еднократно диференциране времевият ред за повърхностните температурни аномалии на глобалния океан става стационарен. Този резултат предполага моделиране по методиката на Box-Jenkins с подходящ Integrated autoregressive moving average (ARIMA) [13]. Процедурата стартира с предварителен анализ на диференцирания времеви ред, неговата автокорелационна функция и частична автокорелационна функция. Фиг. 3 представя диференцирания времеви ред $\Delta\delta T$ от средногодишните повърхностни температурни аномалии на глобалния океан за периода 1850–2014 г. и съответната автокорелационна функция и частична автокорелационна функция. На горния панел е представен диференцираният ред, в средата – автокорелационната функция, и на



Фиг. 3. Времеви ред (горе), автокорелационна функция (средата) и частична автокорелационна функция (долу) на диференцирания времеви ред от температурните аномалии

долния е частичната автокорелационна функция. Максималното закъснение е 50.

Визуалният преглед на диференцирания ред води до извода, че той е стабилен по отношение на измененията на дисперсията с времето. От фиг. 3 се вижда също, че автокорелационната функция и частичната автокорелационна функция имат очевидни пикове над долната доверителна граница при закъснение 2 години, след което затихват в рамките на доверителната лента. Евентуални изключения са единичните локални пикове при закъснение 18 и 37 години за автокорелационната функция и при закъснение 18 и 36 години за частичната автокорелационна функция, които са на границите на значимост. Тези наблюдения водят до

извода, че подходящият стохастичен модел е сред множеството от ARIMA модели с еднократно диференциране, максимален ред на авторегресионния полином – две, максимален ред на полинома на пълзящото средно – също две, и евентуална сезонност с период 18 години, като стационарност след сезонно диференциране също не е изключена [13].

На този етап възниква въпросът как на практика да се определи “най-добрият” сред очертаното вече множество от модели. За целта в литературата се препоръчва това да е моделът, който минимализира някой от т.нар. информационни критерии [13]. Тези критерии, освен точността на апроксимирането на оригиналните данни, отчитат и броя на моделните параметри, а в някои случаи и дължината на времевия ред.

За оптимизиране на пресмятанията ние прилагаме стъпаловидния алгоритъм от [14]. Множеството на разглежданите сезонни ARIMA модели с дрефт ние определяме, фиксирайки: ред на диференцирането – единица, максимален ред на авторегресионния полином – пет, максимален ред на полинома на пълзящото – две, максимален ред на сезонния полином на пълзящото средно – две, максимален брой на моделните параметри – шест. Освен това ние оставяме открита и опцията за сезонно диференциране, като дали това се налага, или не, се определя от резултатите от теста за стационарност след сезонно диференциране на Osborn et al. [15].

Резултатите от съответните пресмятания показват, че и трите използвани информационни критерия – Akaike’s Information Criterion [16], Schwarz’s Information Criterion [17] и т.нар. AIC, Bias Corrected [7], недвусмислено фаворитизират модела $ARIMA(0,1,2) \times (0,0,1)_{18}$, т.е. ред две на полинома на пълзящото средно и ред едно на полинома на сезонното пълзящо средно. Пресмятанията показват още, че и в трите случая отчитането на сезонността подобрява моделите. Моделът $ARIMA(0,1,2) \times (0,0,1)_{18}$ е с по-ниски стойности на информационните критерии в сравнение с модела $ARIMA(0,1,2)$. Единствената разлика е, че докато Akaike’s Information Criterion дава предимство на $ARIMA(0,1,2) \times (0,0,1)_{18}$ с дрефт, то Schwarz’s Information Criterion и AIC, Bias Corrected поставят този модел на второ място, като фаворитизират същия модел, но без дрефт. В тази връзка ще отбележим следното. Докато Akaike’s Information Criterion не отчита дължината на редицата, то в Schwarz’s Information Criterion и AIC, Bias Corrected включват този параметър и с това са по-ефективни за по-малки извадки, както е в нашия случай [13]. Освен това разликата в стойностите на Akaike’s Information Criterion в полза на модела с дрефт е почти пренебрежима – под 0,07. И накрая, проведеният в раздел 3 тест за стационарност след диференциране на Dickey-Fuller доведе до извода за случайно блуждаене без дрефт. Затова ние ще предпочетем $ARIMA(0,1,2) \times (0,0,1)_{18}$ без дрефт.

Оценката на коефициентите на избрания модел по метода на максималното правдоподобие води до:

$$(1-L)\hat{y}_t = (1 + 0.27_{(0,08)}L^{18})(1 - 0.23_{(0,07)}L - 0.39_{(0,07)}L^2)\hat{\varepsilon}_t \text{ с } \hat{\sigma}_\varepsilon^2 = 0.007, \quad (1)$$

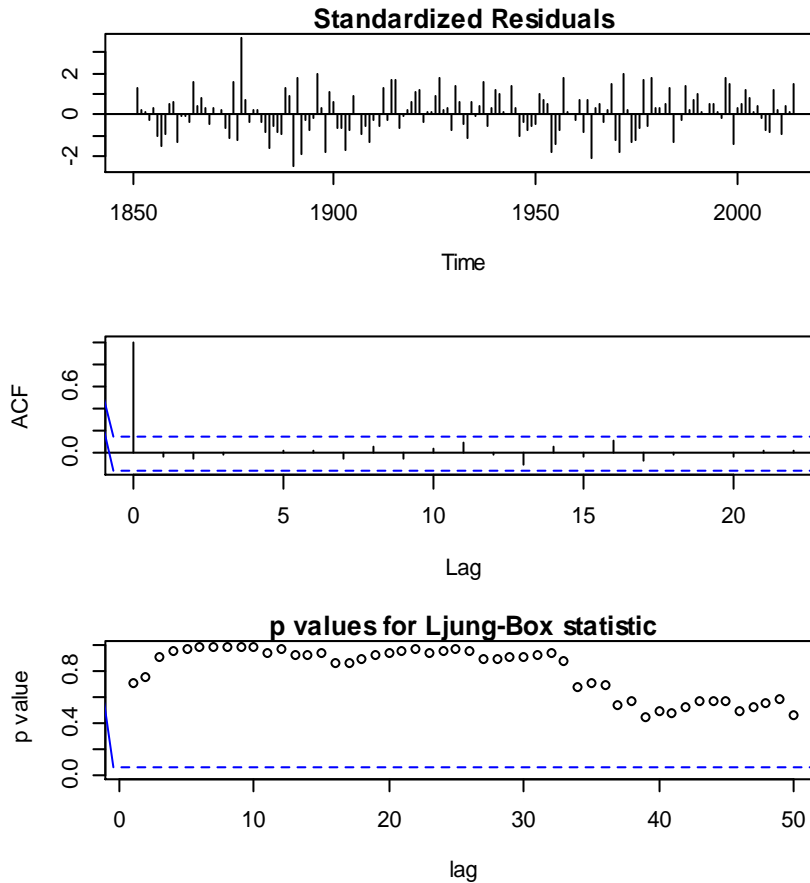
където t са годините, \hat{y}_t е редицата от океанските температурни аномалии, $\hat{\varepsilon}_t$ – белият шум, $\hat{\sigma}_\varepsilon^2$ – неговата дисперсия, $Ly_t = y_t - y_{t-1}$ е операторът на отместване с едно закъснение, а в по-малките скоби – са стандартните грешки.

4.2. ДИАГНОСТИКА НА МОДЕЛА

Стохастичните модели ARIMA предполагат нормално разпределение и некорелирани остатъци с постоянна по времето дисперсия. Тук ще проверим доколко изпълнени са тези предположения от нашия модел.

Фиг. 4 представя диагностиката на остатъците на предлагания модел. На горния панел е показан времевият ред на стандартизираните остатъци. Това е разликата между оригиналния ред и реда от прогнозите с една стъпка напред чрез анализирания модел, нормирани с дисперсията на грешката на едностъпковите прогнози [13]. Както се вижда, в поведението на стандартизираните остатъци не се забелязват по-съществени особености. Те са сравнително симетрично разпределени спрямо нулата, а дисперсията изглежда постоянна във времето. Изключение е годината 1877, за която съответният остатъкът е доста по-висок. оригиналният ред и реда от прогнозите с една стъпка напред чрез анализирания модел, нормирани с дисперсията на грешката на едностъпковите прогнози [13]. Както се вижда, в поведението на стандартизираните остатъци не се забелязват по-съществени особености. Те са сравнително симетрично разпределени спрямо нулата, а дисперсията изглежда постоянна във времето. Изключение е годината 1877, за която съответният остатъкът е доста по-висок.

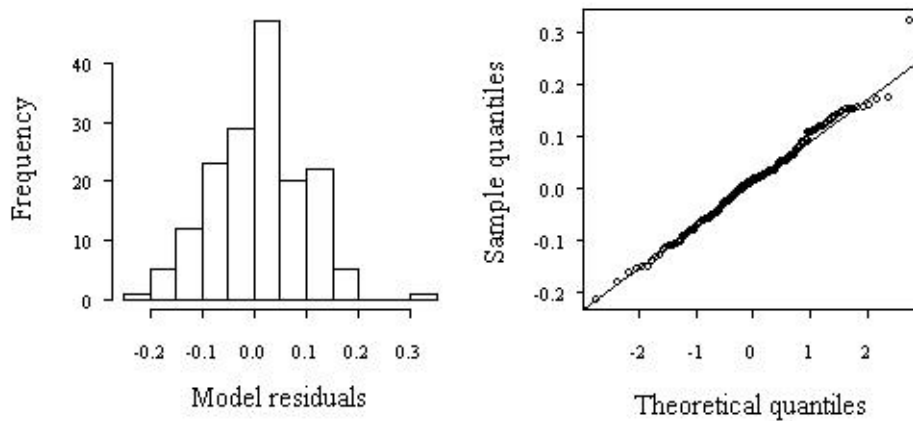
На панела в средата е представена автокорелационната функция на остатъците. Както се вижда, те са в рамките на доверителните граници, което показва некорелираност. На долния панел са дадени p -стойностите на т.нар. Ljung-Box-Pierce статистика. Този тест отчита големините на автокорелациите не поотделно за всяко закъснение, а като група [13]. Нулевата хипотеза е некорелирани остатъци. Както се вижда от фиг. 4, при тези високи p -стойности нулевата хипотеза не може да бъде отхвърлена. Остатъците са некорелирани поне до закъснения 50.



Фиг. 4. Диагностика на остатъците на модела $ARIMA(0,1,2) \times (0,0,1)_{18}$

Имат ли остатъците нормално разпределение? Фиг. 5 представя хистограмата и квантил-квантил диаграмата на разпределение на остатъците. По хоризонталната ос на квантил-квантил диаграмата са разположени теоретичните квантили на нормалното разпределение, а по вертикалната ос – квантилите на остатъците. Правата линия свързва първия и третия квантил. Плътното групиране на точките около правата линия е показател за нормално разпределение.

Както се вижда, картините на хистограмата и квантил-квантил диаграмата добре се съгласуват с предположението за нормално разпределение. Изключение е споменатата вече година 1877, която е изолирана в десния край на хистограмата и десния горен ъгъл на квантил-квантил диаграмата. Очевидно тази година е “особено” наблюдение.



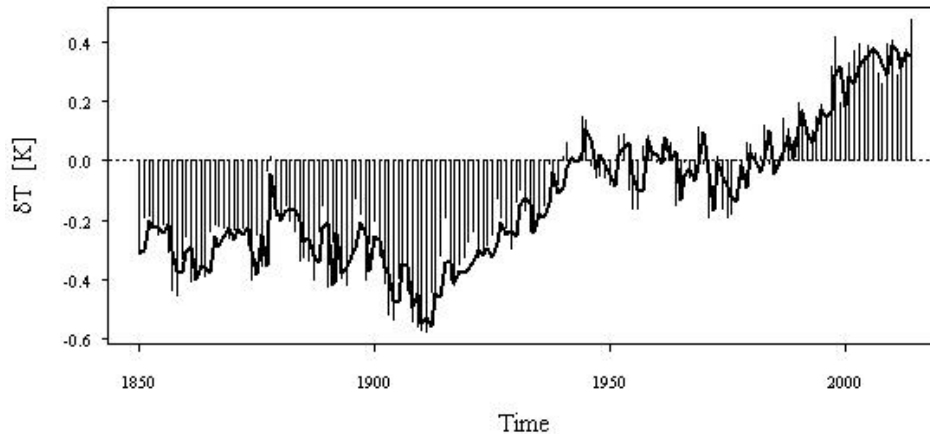
Фиг. 5. Хистограма (ляво) и квантил-квантил диаграма (дясно) на остатъците на модела $ARIMA(0,1,2) \times (0,0,1)_{18}$

Хистограмата и квантил-квантил диаграмата дават визуална оценка доколко една извадка е нормално разпределена. Да подложим остатъците и на по-строг статистически тест. За целта ние използваме теста на Jarque-Bera и теста на Shapiro-Wilk. И в двата случая нулевата хипотеза е нормално разпределение. Резултатите от тестовете са, както следва. Стойността на χ^2 -статистиката в теста на Jarque-Bera е 0,7991, а съответната p -стойност е 0,6706. Нулевата хипотеза не може да се отхвърли. Стойността на W -статистиката в теста на Shapiro-Wilk е 0,9903 и съответната p -стойност 0,3212. Хипотезата за нормално разпределение не се отхвърля. Окончателният извод е, че остатъците са нормално разпределени.

Да резюмираме. Предлаганият $ARIMA(0,1,2) \times (0,0,1)_{18}$ без дрефт минимализира информационните критерии, неговите остатъци са некорелирани и нормално разпределени. С това формалните изисквания за едно удовлетворително стохастично моделиране са изпълнени. Предлаганият модел е напълно приемлив от статистическа гледна точка.

4.3. ЕДНОСТЪПКОВА ПРОГНОЗА

Фиг. 6 представя времевия ред от средногодишните повърхностни температурни аномалии на глобалния океан за периода 1850–2014 г. и неговата апроксимация с прогнозата с една стъпка напред, направена с предлагания $ARIMA$ модел. С падащите линии е показана оригиналната



Фиг. 6. Времеви ред на и неговата апроксимация с модела $ARIMA(0,1,2) \times (0,0,1)_{18}$

редица δT , а с непрекъснатата линия – апроксимацията. Както се вижда, моделът сравнително добре апроксимира особеностите в хода на океанските аномалии. Това е прогноза вътре в извадката.

Да резюмираме. В този раздел построихме прост стохастичен модел на океанските температурни аномалии за периода 1850–2014 г. Оказа се, че моделът е от тип $ARIMA(0,1,2) \times (0,0,1)_{18}$. Прави впечатление, че моделът е с стационарност след диференциране, сезонност 18 години и само 4 параметъра. Диагностиката показва, че моделът е напълно приемлив от формално статистическа гледна точка.

5. ДИСКУСИЯ И ЗАКЛЮЧЕНИЕ

5.1. ДЪЛГОСРОЧНА ПРОГНОЗА

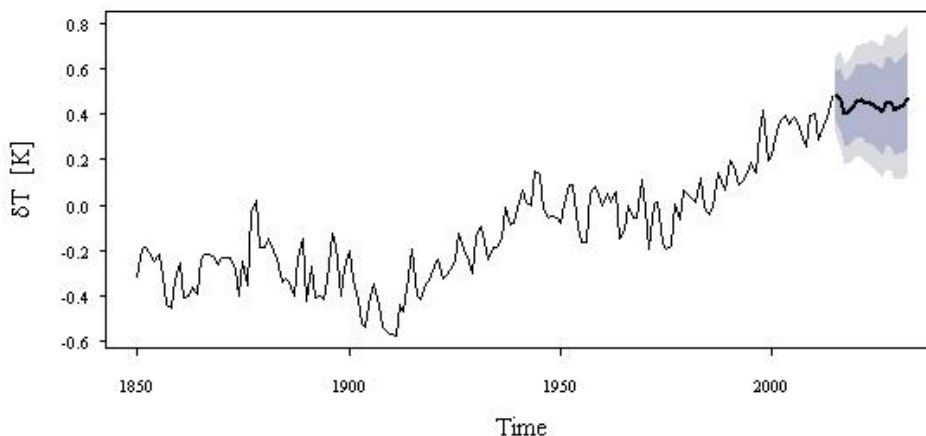
По принцип дългосрочната прогноза на стационарните след диференциране процеси е обречена на неуспех, защото грешката на прогнозата расте неограничено с времето. Тук ще демонстрираме една сравнително дългосрочна прогноза с нашия модел, но с ограничен хоризонт въвн от извадката.

Ние преоценяваме коефициентите на модела $ARIMA(0,1,2) \times (0,0,1)_{18}$, използвайки данните само за периода 1850–2000 г. Новият модел е

$$(1-L)\hat{y}_t = (1 + 0.29_{(0,09)}L^{18})(1 - 0.24_{(0,07)}L - 0.39_{(0,07)}L^2)\hat{\varepsilon}_t, \hat{\sigma}_{\varepsilon}^2 = 0.007. (2)$$

Сравнявайки с модела (1), виждаме, че уравнението е практически същото.

Разликите са в сезонния коефициент с 0,02 и в първия коефициент на



Фиг. 7. Времеви ред и прогноза за 20 години напред с модела $ARIMA(0,1,2) \times (0,0,1)_{18}$

пълзящото средно с 0,01, което е значително по-малко от стандартните грешки. С това уравнение ние прогнозираме 20 години напред, т.е. от 2001 до 2020 г. Получените резултати са представени на фиг. 7. Прогнозата е показана с черната непрекъсната линия. В по-тъмния отенък на сивото са 80-процентните доверителни нива на прогностичните интервали, в по-светлия нюанс – 90-процентните нива. Прави впечатление, че действителните температурни аномалии попадат в границата на 80-процентната доверителна лента.

Можем да спекулираме и че в следващите 2–3 години ни очаква повишаване на глобалната океанска температура.

5.2. СТОХАСТИЧЕН ТРЕНД И ЦИКЛИЧНОСТ

Бе установено, че времевият ред от средногодишните повърхностни температурни аномалии на глобалния океан за периода 1850–2014 г. се моделира статистически с един прост $ARIMA(0,1,2) \times (0,0,1)_{18}$ модел. Две неща правят веднага впечатление: имаме стационарност след еднократно диференциране и сезонност от 18 години. Какво означава това?

Съгласно [18] тренд-стационарните модели се характеризират със следните общи свойства. Времевият ред се завърща трайно към тренда, в частност, при стационарност – към средното ниво. Дългосрочната прогноза е сходяща към тренда. Случайните смущения, наречени шокове, затихват с времето и ефектът от тях е краткотраен. Ако редът обаче е стационарен след диференциране, нещата са противоположни [18]. Редът не се завърща

трайно към тренда или средното ниво. Дисперсията на прогностичната грешка е разходяща, защото нараства с времето. Ефектът от шоковете е дълготраен.

Нашият модел е стационарен след диференциране и е без дрифт. Това означава, че доколкото аномалиите нямат трайна тенденция да се завръщат към нулата, то температурата на глобалния океан не би трябвало трайно да се завръща към климатичната норма от периода 1960–1990 г. Дългосрочната прогноза на температурата на глобалния океан е най-малкото проблематична, защото дисперсията на прогнозата расте с времето. Ефектът на случайните шокове върху глобалния океан е дълготраен. Очевидно тези изводи са противоположни на концепцията, развита в [2], където предлаганият стохастичен модел е стационарен.

Да обсъдим по-подробно въпросът за шоковете. Известно е, че ако един ARIMA модел се представи във вида $(1-L)\hat{y}_t = \psi(L)\hat{\varepsilon}_t$, то числото $\psi(1)$ има смисъл на значението на дългосрочния ефект от един шок спрямо непосредствения ефект от него [19]. В този смисъл $\psi(1)$ е естествена мярка за значението на дългосрочния ефект от един шок. Ако $\psi(1)=1$, дългосрочният ефект е равен на непосредствения. При $\psi(1)>1$ дългосрочният ефект е по-голям от непосредствения, ако $\psi(1)<1$ – по-малък. В нашия случай полиномът $\psi(L)$ е дясната част на уравнението (1), защото в модела няма авторегресионна част. Като положим формално $L \rightarrow 1$, получаваме $\psi(1)=0,48$. Дългосрочният ефект на един шок върху времевия ред на температурата на глобалния океан е по-малък от непосредствения.

Известно е също, че всеки ARIMA модел може да се представи като сума от случайно блуждаене и стационарен процес [19]. Случайното блуждаене се нарича стохастичен тренд. Този тренд акумулира шоковете с дългосрочен ефект. Стационарната част на сумата акумулира шоковете с краткосрочен ефект върху нивото на времевия ред. Самата декомпозиция не е еднозначна, но при доста общи предположения, независимо каква конкретно е тя, дисперсията на стохастичния тренд е една и съща, равна на $\psi(1)^2 \hat{\sigma}_\varepsilon^2$ [19]. Замествайки със стойностите от нашия модел, оценката за стандартното отклонение на стохастичния тренд е 0,04 К.

Ние предлагаме следната физична интерпретация на приведените по-горе формално-статистически резултати. Шоковете в повърхностната температура на океана се дължат на флуктуациите в атмосферата, термохалинната циркулация и криосферата. Поради високия топлинен капацитет на водата естествените кандидати за промените в стохастичния тренд са флуктуациите в термохалинната циркулация и криосферата. Тези флуктуации са с дългосрочен ефект. Атмосферните флуктуации би трябвало да се интерпретират като шокове с краткосрочен ефект. С оглед на направената по-горе оценка, ефектът от тях преобладава, но е

краткосрочен.

Накрая да се спрем и на сезонността. По-новите спътникови наблюдения показват, че дължащата се на приливния цикъл от 18,61 години дисипация в дълбоките слоеве на океана чрез произтичащото смесване води до изменения в общата океанска циркулация. Двата периода от 18 и 18,61 години практически съвпадат. Можем да направим извода, че сезонността в нашия модел при закъснение от 18 години е отражение на влиянието на главния приливен цикъл. Доколкото главният приливен цикъл се дължи на влиянието на Луната, то можем да кажем, че сезонността в нашия модел отразява влиянието на Луната върху глобалния океан.

Да се върнем към фиг. 3. Там са дадени автокорелационната и частичната автокорелационна функция на еднократно диференцирания времеви ред. В тази връзка бе коментирано, че над границите на значимост са автокорелациите при закъснение 2 години. Единични локални пикове над границите на значимост имаме и в автокорелационната функция при закъснение 18 и 37 години, а също и в частичната автокорелационна функци при закъснение 18 и 36 години.

Корелацията при закъснение 2 години можем да интерпретираме като удвоен годишен цикъл [20]. Удвояването на периода е основна бифуркация в теорията на нелинейните динамични системи. Доколкото глобалният океан е нелинеен, то можем да интерпретираме високите корелации при закъснение 2 като удвояване на годишния цикъл от нелинейния глобален океан.

Интересно е, че пикът при 36 години може да се свърже с като че ли забравения 35 ($34,8 \pm 7$)-годишен климатичен цикъл на Brikner, един от първите изследователи на периодичните колебания на климата от края на XIX век. Остава да завършим с пика на 11 години, който е точно на границата на значимост. Една очевидна е неговата интерпретация е като отклик на слънчевия цикъл на Schwabe.

6. ЗАКЛЮЧЕНИЕ

Стохастичното моделиране на времевия ред от средните годишни повърхностни температурни аномалии на глобалния океан за периода 1850–2014 г. доведе до един $ARIMA(0,1,2) \times (0,0,1)_{18}$ модел. В рамките на предложения модел могат да се направят редица изводи.

Температурата на глобалния океан няма да се завръща трайно към климатичната норма за периода 1960–1990 г. Дългосрочната прогноза на температурата на глобалния океан е най-малкото проблематична, защото дисперсията на грешката на прогнозата расте с времето.

Ефектът на случайните шокове върху глобалния океан е дълготраен. Дългосрочният ефект на един шок върху температурата на глобалния океан

е по-малък от непосредствения. Оценката на стандартното отклонение на стохастичния тренд на температурата на глобалния океан е 0,04 К. Правдоподобна е хипотезата, че промените в стохастичния тренд на повърхностната температура на океана се дължат на шоковете, свързани с флукуациите в термохалинната циркулация и криосферата. Ефектът от атмосферните флукуации преобладава, но е краткотраен.

Правдоподобна е и хипотезата, че сезонността от 18 години в модела се обуславя от главния приливен цикъл, т.е. от влиянието на Луната. Хипотезите, че локалните пикове в автокорелационната функция на диференцираните океански температури при 2, 36 и 11 години са отражение съответно на удвоения годишен цикъл, климатичния цикъл на Brikner и на слънчевия цикъл на Schwabe също имат своето основание.

ЛИТЕРАТУРА

- [1] Hasselmann, K. *Tellus*, 1976, **28**, 473.
- [2] Frankignoul, C., K. Hasselmann. *Tellus*, 1977, **29**, 284.
- [3] Kennedy, J.J., N.A. Rayner, R.O. Smith, M. Saunby, D.E. Parker. *J. Geophys. Res.*, 2011, **116**, D14103, doi:10.1029/2010JD015218.
- [4] Kennedy, J.J., N.A. Rayner, R.O. Smith, M. Saunby, D.E. Parker. *J. Geophys. Res.*, 2011, **116**, D14104, doi:10.1029/2010JD015220.
- [5] Brockwell, P.J., R.A. Davis. *Introduction to Time Series and Forecasting*, 2 ed., Springer-Verlag New York, Inc., 2002.
- [6] Dickey, D. A., W. A. Fuller. *Journal of the American Statistical Association*, 1979, **75**, 427.
- [7] Pfaff, B. *Analysis of Integrated and Cointegrated Time Series with R*. Springer Science+Business Media, LLC, 2008.
- [8] MacKinnon, J.G. *Journal of Applied Econometrics*, 1996, **11**, 601.
- [9] Dickey, D. A., W. A. Fuller. *Econometrica*, 1981, **49**, 1057.
- [10] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, Y. Shin. *Journal of Econometrics*, 1992, **54**, 159.
- [11] Zivot, E., Donald W.K. Andrews. *Journal of Business & Economic Statistics*, 1992, **10**, 251.
- [12] Kirchgässner, G., J. Wolters, U. Hassler. *Introduction to Modern Time Series Analysis*, 2 ed., Springer-Verlag Berlin Heidelberg, 2013.
- [13] Shumway, R. H., D.S. Stoffer. *Time Series Analysis and Its Applications With R Examples*, 2 ed., Springer Science+Business Media, LLC, 2006.
- [14] Hyndman, R.J., Y. Khandakar. *Journal of Statistical Software*, 2008, **26**(3).
- [15] Osborn, D.R., A.P.L. Chui, J. Smith, C.R. Birchenhall. *Oxford Bulletin of Economics and Statistics*, 1988, **50**, 361.
- [16] Akaike, H. *Journal of Econometrics*, 1981, **16**, 3.
- [17] Schwarz, H. *The Annals of Statistics*, 1978, **6**, 461.
- [18] Hamilton, J. D. *Time Series Analysis*. Princeton, NJ, 1994.
- [19] Cochrane, J, H. *Time Series for Macroeconomics and Finance*, U Chicago, 1997.
- [20] Stephenson, D.B., V. Pavan, R. Bojariu. *Int. J. Climatol.*, 2000, **20**, 1.