УДК 551.58

МНОГОФАКТОРНАЯ НЕЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ С ОБРАТНЫМИ СВЯЗЯМИ – ИМИТАЦИЯ РЕГИОНАЛЬНЫХ КЛИМАТОВ

Э. Н. Серга, канд. геогр. наук, доц.

Одесский государственный экологический университет,
ул. Львовская, 15, 65016, Одесса, Украина, Serga_ed@ukr.net

Предлагается усовершенствованная многофакторная нелинейная регрессионная модель с обратными связями с переопределёнными выражениями для коэффициентов системы уравнений. Данная модель апробирована как имитационная с целью установления связи между зонами интенсивных взаимодействий атмосферы и океана в акватории Северной Атлантики и однородными регионами на территории Восточной Европы. Главные компоненты энергетических взаимодействий составили исходные выборки влияющих факторов и откликов для модели. Значимые коэффициенты множественной корреляции, характеризующие степень адекватности модели, подтверждают возможность её практического применения для решения задач подобного характера.

Ключевые слова: многофакторная регрессионная модель, аппроксимирующие полиномы, влияющие факторы, коэффициенты уравнений, обратные связи.

1. ВВЕДЕНИЕ

Начиная с конца второй половины XX столетия вопросы исследования климата Земли приобрели особую остроту. Связано это, в первую очередь, с предполагаемым усилением парникового эффекта и, соответственно, повышением глобальной температуры воздуха. Свидетельством особого интереса к глобальному потеплению и его последствиям явилось принятие в июне 1992 года Рамочной Конвенции по изменению климата (Рио-де-Жанейро), которая в силу важности решаемых проблем была ратифицирована рядом государств, взявших на себя определённые обязательства, в том числе: повышать эффективность и интенсивность научных исследований по вопросам глобальных и региональных изменений климата, а так же вопросов оценки экологических и социально-экономических последствий глобального потепления. В 1997 году в Киото был принят Киотский протокол, главной задачей для каждой Стороны которого явилось недопущение превышения установленных объемов выбросов парниковых газов. На двадцать первой сессии конференции Сторон Рамочной Конвенции в Париже (декабрь 2015 г.) было принято Парижское соглашение, подписанное более чем 170 странами в апреле 2016 года, где в качестве общей цели обозначено сдерживание повышения средней температуры на Земле в пределах 2 ºС к 2100 году от уровня доиндустриальной эпохи, а также приложение усилий по сдерживанию этого показателя в пределах 1,5 ºС.

Гипотезы об изменении климата обусловили необходимость более детального рассмотрения механизмов взаимодействия компонентов климатической системы – атмосферы, океана, криосферы, деятельного слоя суши и биосферы, как с качественной, так и количественной стороны.

Процессы, протекающие внутри климатической системы, обладают разнообразными обратными связями и их вряд ли можно назвать линейными. Под воздействием этих процессов в климатической системе возбуждаются сложные собственные колебания разных временных масштабов. Для описания состояний климатической системы: прошлого, настоящего и будущего, с учетом внешних воздействий, необходимы сложные физико-математические модели [1].

Задачи, которые ставятся перед моделированием, описание самих моделей, их преимущества и недостатки, результаты практического применения, подробным образом рассмотрены в большом количестве научных трудов [например, 2 - 5]. Основной вывод, который признается разработчиками и пользователями существующих гидродинамических климатических моделей заключается в том, что на современном этапе развития науки нет идеальной модели, одинаково хорошо описывающей и предсказывающей климатические величины, а наилучшие показатели дает "средняя" по ансамблю модель. Для сравнительной оценки моделей создан ряд международных программ, таких как CMIP (Coupled Model Intercomparison Project), AMIP (Atmospheric Model Intercomparison Project), PMIP (Paleoclimate Model Intercomparison Project) и другие, позволяющих исследовать систематические ошибки, возникающие в результате работы отдельных моделей [2, 3].

Несмотря на предпочтительность применения гидродинамических моделей в исследовании климата, для количественного описания связей между влияющими факторами и откликами, а также изучения свойств климатической системы, целесообразно использовать аппарат математических моделей из разряда вероятностно-статистических [6]**.** В выборе такого аппарата предпочтение отдают регрессионным (например, в гидродинамико-статистическом прогнозировании, для статистической интерпретации с помощью множественной регрессии результатов интегрирования гидродинамических моделей атмосферы от ансамбля начальных состояний [7] или же в качестве одного из подходов к задаче "даунскейлинга" [4]) и нейросетевым моделям [8, 9].

2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЙ

В нашем случае, за основу взята многофакторная регрессионная модель [10]. В данной работе рассматриваются основные принципы построения системы нелинейных уравнений множественной регрессии с обратными связями вида

$$\hat{y}_l = a_0^{(l)} + \sum_{i=1}^{m} a_i^{(l)} x_i + \sum_{\substack{i=1 \\ (i \le j)}}^{m} \sum_{j=1}^{m} a_{ij}^{(l)} x_i x_j + \sum_{i=1}^{m} \sum_{\substack{j=1 \\ (i \le j \le s)}}^{m} \sum_{s=1}^{m} a_{ijs}^{(l)} x_i x_j x_s + \sum_{\substack{\nu=1 \\ (\nu \ne l)}}^{k} \alpha_\nu^{(l)} y_\nu, \left(l = \overline{1,k}\right). \qquad (1)$$

В уравнениях (1) $a_0^{(l)}, a_i^{(l)}, a_{ij}^{(l)}, a_{ijs}^{(l)}$ – коэффициенты регрессии при соответствующих степенях влияющих факторов x; m – количество влияющих факторов; k – количество откликов (уравнений в системе). Последний член равенства (1) характеризует вклад обратных связей. Особенности этих взаимосвязей определяют коэффициенты обратных связей α_ν.

Влияющие факторы x (далее факторы) в модели являются ортогональными, центрированными и нормированными. Параметры на выходе модели – центрированы и нормированы. На основании центральной предельной теоремы теории вероятности это дает основание принять гипотезу об их нормальном распределении.

Построение полиномов, составляющих систему (1), производится на основе метода вероятностной аппроксимации [11, 12], в котором определение коэффициентов аппроксимирующих полиномов осуществляется при условии минимума дисперсии ошибки аппроксимации $D_\varepsilon = \overline{\varepsilon_l^2} - \left(\overline{\varepsilon_l}\right)^2$ функции $y_l(x)$, где $\varepsilon_l(x) = y_l(x) - \hat{y}_l(x)$.

Учитывая особенности величин y_l и компонент вектора x, а также условие независимости моментов $z_0^{(l)}$, $z_p^{(l)}$, $z_{ps}^{(l)}$, $z_{psr}^{(l)}$, $r_{l\mu}$ ($p,s,r = \overline{1,m}$; $l,\mu = \overline{1,k\text{-}1}$; $\mu \ne l$; m – количество факторов в исходной выборке) от вида полинома (1) (они определяются значениями аппроксимирующей функции и законом распределения случайного вектора x), можно построить систему производящих функций [10]. Эта система позволяет получить оценки искомых коэффициентов системы уравнений (1) при условии, что найдены коэффициенты обратных связей $\alpha_\nu^{(l)}$.

Коэффициенты обратных связей можно определить, решив систему алгебраических уравнений: $\sum_{\substack{\nu=1 \\ (\nu \ne l)}}^{k} A_{\mu\nu}^{(l)} \alpha_\nu^{(l)} = B_\mu^{(l)}$, $\left(\mu = \overline{1,k}; \mu \ne l\right)$ [10].

Впервые такого рода модель была реализована при исследовании особенностей характеристик микроструктуры теплых приморских туманов и их связи с состоянием атмосферы [13].

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ И ИХ ОБСУЖДЕНИЕ

В первоначальном варианте модели предполагалось, что коэффициенты уравнений при значимых факторах и соответствующих их степенях, полученные до процедуры "просеивания", осуществляемой с помощью критерия Фишера, остаются неизменными и полностью формируют статистические связи с откликами, а незначимые факторы вносят незначительный вклад в общую дисперсию отклика и ими можно пренебречь.

Выражения для определения значений критерия Фишера имеют вид:

– для оценки общей значимости факторов

$$F = \frac{Д_x[\hat{y}_l]}{\sum_{i=1}^{m-t} \chi_i^t[\hat{y}_l]},$$

где $Д_x[\hat{y}_l]$ – дисперсия отклика, обусловленная влиянием всех факторов; $\chi_i^t[\hat{y}_l]$ – дисперсия i-го фактора (дисперсии ранжированы в порядке увеличения), $t = \overline{1, m-1}$;

– для оценки статистической значимости членов нелинейных уравнений регрессии, которые имеют вторую и третью степени

$$F = \frac{\chi_p^t[\hat{y}_l]}{\chi_i^{(t-1)}[\hat{y}_l]},\ t = 2{,}3;\ p,i = \overline{1,m}.$$

Однако выдвинутая гипотеза об общности значимых факторов для откликов, участвующих в модели, накладывает ограничение на её использование для более широкого круга физико-статистических задач, решаемых в области гидрометеорологии и климатологии. Так, например, при рассмотрении в качестве откликов характеристик региональных климатов больших территорий, влияющие на них факторы могут значительно различаться. В этом случае возникает ситуация, когда оказывается совершенно разною не только внешняя структура уравнений в системе, но и структура внутри каждого из них. При этом применение модели в первоначальном виде (без переоценки параметров системы уравнений с учетом исключения незначимых факторов, и незначимых степеней значимых и незначимых факторов) может привести к некорректным результатам, так как: во-первых, коэффициенты разных степеней значимых факторов и коэффициенты обратных связей, которые являются связующими звеньями в системе уравнений, содержат в себе значительное количество различных степеней незначимых факторов и незначимые степени значимых факторов; во-вторых, в результате вероятности возникновения ситуации несовпадения части из общей совокупности значимых факторов l-го и μ-го уравнений, появляются члены полиномов без участия обратных связей.

Как следствие, возникает необходимость рассмотреть максимальное количество возможных ситуаций, связанных с различными вариантами формирования выборок значимых факторов (общих l-го и μ-го уравнений и частных для каждого из них) раздельно для каждой степени и для каждого из уравнений системы. Разнообразие вариаций гипотетических ситуаций с общими значимыми факторами, в свою очередь, определяет необходимость преобразования производящих функций, и, соответственно, соотношений для расчета коэффициентов уравнений и обратных связей.

После выполнения этих процедур модель можно представить, как модифицированную версию многофакторной нелинейной регрессионной модели.

На первом этапе, при сохранении гипотезы о том, что все члены исходной выборки потенциально являются общими влияющими факторами для l-го и μ-го откликов, производится отбор значимых факторов с помощью соответствующих критериев (например, критерия Фишера). Оценивается степень нелинейности факторов (в результате, например, может оказаться, что p - ый фактор может быть значимым в 3-ей степени, но незначимым в первых двух) и значимость каждого элемента обратной связи. На втором этапе производится расчет коэффициентов системы уравнений, построение аппроксимирующих полиномов и определение модельных значений откликов.

Основные принципы необходимых преобразований и соотношения для расчёта оценок коэффициентов уравнения (1) в конечном виде для второго этапа построения модели изложены ниже (подробный вывод этих соотношений из-за громоздкости не приводится).

Для удобства введём обозначения для количества членов во множествах значимых факторов: q - первой степени, τ - второй, η - третьей, ρ - объединённого множества, состоящего из совпадающих факторов первой и третьей степени. Индексы $g,\xi,p,s,r,w,\gamma,\iota,\beta$ – обозначение номеров различных факторов в исходной выборке.

Коэффициенты $a_0, a_{gg}, a_{g\xi}$ не зависят от коэффициентов при первой и третьей степени переменных в уравнениях регрессии (1) (коэффициент a_0 зависит только от коэффициента a_{gg}). Коэффициенты первой и третьей степени, согласно производящим функциям [10], являются взаимозависимыми. В связи с этим, требуется получить выражения для расчёта коэффициентов обратных связей и коэффициентов уравнения (1) при возможных вариантах формирования выборок с номерами значимых факторов при вышеуказанных степенях.

Из свойств величин x_i, указанных выше, следует, что $M[x_i x_j]$, где M - операция математического ожидания, может принимать значения $\{0;1\}$, $M[x_i x_j x_p x_s]$ – $\{0;1;3\}$, $M[x_i x_j x_k x_p x_s x_r]$ – $\{0;1;3;15\}$.

При отсутствии общих значимых факторов первой (q) и третьей (η) степеней выражения для производящих функций принимают вид:

$$z_p^{(l)} = \sum_{i=1}^{q} a_i^{(l)} M[x_i x_p] + \sum_{\substack{\nu=1 \\ (\nu \neq l)}}^{k} \alpha_\nu^{(l)} M[y_\nu x_p],\ \left(p = \overline{1,q}\right),$$

$$z_{psr}^{(l)}=\sum_{i=1}^{\eta}\sum_{\substack{j=1\\(i\le j\le t)}}^{\eta}\sum_{t=1}^{\eta}a_{ijt}^{(l)}M[x_ix_jx_tx_px_sx_r]+$$
$$+\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}M[y_\nu x_px_sx_r],\ \left(p\le s\le r=\overline{1,\eta}\right).$$

При наличии общих значимых факторов первой и третьей степени, вид производящих функций для искомых коэффициентов при этих факторах зависит от количества последних.

Рассмотрим преобразование выражений для расчёта коэффициентов уравнения (1) при различных ситуациях. Соотношения для элементов матрицы $A^{(l)}$ и вектора $B^{(l)}$, необходимые для определения коэффициентов обратных связей, приводятся только для одного случая. Однако их можно получить, используя выведенные ниже конечные выражения для расчета коэффициентов уравнения (1).

I. Определение коэффициентов аппроксимирующих полиномов третьей степени в случае отсутствия совпадений в совокупностях значимых факторов первой и третьей степени в отдельно взятом уравнении системы (1).

В этом случае коэффициенты уравнения и коэффициенты обратных связей при соответствующих степенях определяются независимо друг от друга.

1. При условии $q\ge 1;\eta=1$. Для коэффициентов первой и третьей степени:

$$a_t^{(l)}=z_t^{(l)}-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_t^{(\nu)},\left(t=\overline{1,q}\right);$$

$$a_{ppp}^{(l)}=\frac{1}{15}\left[z_{ppp}^{(l)}-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{ppp}^{(\nu)}\right],\quad\left(p=\overline{1,\eta}\right).$$

Далее вид соотношений коэффициентов первой степени для условий 2 и 3 остаётся без изменений.

2. При условии $q\ge 1;\eta=2$.

Для коэффициентов третьей степени:

$$a_{ppp}^{(l)}=\frac{1}{12}\left[z_{ppp}^{(l)}-z_{pss}^{(l)}-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{ppp}^{(\nu)}+\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{pss}^{(\nu)}\right],$$
$$\left(p>s=\overline{1,\eta}\right);$$

$$a_{sss}^{(l)}=\frac{1}{12}\left[z_{sss}^{(l)}-z_{ssp}^{(l)}-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{sss}^{(\nu)}+\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{ssp}^{(\nu)}\right],$$
$$\left(s>p=\overline{1,\eta}\right);$$

$$a_{pss}^{(l)}=\frac{1}{12}\left[5z_{pss}^{(l)}-z_{ppp}^{(l)}-5\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{pss}^{(\nu)}+\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{ppp}^{(\nu)}\right],$$
$$\left(s>p=\overline{1,\eta}\right).$$

3. При условии $q\ge 1;\eta\ge 3$.

Для коэффициентов третьей степени:

$$a_{ppp}^{(l)}=\frac{1}{15}\left[z_{ppp}^{(l)}-\frac{1}{4}\left(5\sum_{j=1}^{p-1}z_{jjp}^{(l)}-\sum_{j=1}^{p-1}z_{ppp}^{(l)}+5\sum_{j=p+1}^{\eta}z_{ppj}^{(l)}-\right.\right.$$
$$\left.-\sum_{j=p+1}^{\eta}z_{jjj}^{(l)}\right)-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{ppp}^{(\nu)}+\frac{1}{4}\left(5\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=1}^{p-1}\alpha_\nu^{(l)}z_{jjp}^{(\nu)}-\right.$$
$$\left.\left.-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=1}^{p-1}\alpha_\nu^{(l)}z_{ppp}^{(\nu)}+5\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=p+1}^{\eta}\alpha_\nu^{(l)}z_{ppj}^{(\nu)}-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=p+1}^{\eta}\alpha_\nu^{(l)}z_{jjj}^{(\nu)}\right)\right],$$
$$\left(p=\overline{1,\eta}\right);$$

$$a_{pss}^{(l)}=\frac{1}{2}\left[z_{pss}^{(l)}-\frac{1}{5}z_{ppp}^{(l)}-\frac{1}{30}\left(5\sum_{j=1}^{p-1}z_{jjp}^{(l)}-\sum_{j=1}^{p-1}z_{ppp}^{(l)}+\right.\right.$$
$$\left.+5\sum_{j=p+1}^{\eta}z_{ppj}^{(l)}-\sum_{j=p+1}^{\eta}z_{jjj}^{(l)}\right)-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{pss}^{(\nu)}+$$
$$+\frac{1}{5}\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{ppp}^{(\nu)}+\frac{1}{30}\left(5\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=1}^{p-1}\alpha_\nu^{(l)}z_{jjp}^{(\nu)}-\right.$$
$$\left.\left.-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=1}^{p-1}\alpha_\nu^{(l)}z_{ppp}^{(\nu)}+5\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=p+1}^{\eta}\alpha_\nu^{(l)}z_{ppj}^{(\nu)}-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\sum_{j=p+1}^{\eta}\alpha_\nu^{(l)}z_{jjj}^{(\nu)}\right)\right],$$
$$\left(s>p=\overline{1,\eta}\right).$$

II. Определение коэффициентов аппроксимирующих полиномов третьей степени в случае наличия совпадений в совокупностях значимых факторов первой и третьей степени в отдельно взятом уравнении системы.

Ситуация 1. Совпадение только одного в совокупностях значимых факторов первой и третьей степени, т. е. общей выборке принадлежит только один элемент ($\rho=1$). Обозначим его через w. Тогда для коэффициентов первой и третьей степени имеем:

$$a_w^{(l)}=\frac{1}{2}\left[5z_w^{(l)}-z_{www}^{(l)}-5\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_w^{(\nu)}+\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{www}^{(\nu)}\right];$$

$$a_{www}^{(l)}=\frac{1}{6}\left[z_{www}^{(l)}-3z_w^{(l)}-\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_{www}^{(\nu)}+3\sum_{\substack{\nu=1\\(\nu\ne l)}}^{k}\alpha_\nu^{(l)}z_w^{(\nu)}\right].$$

Ситуация 2. Совпадение двух в совокупностях значимых факторов первой и третьей степени, т.е. выборке, состоящей из общих факторов принадлежит два элемента ($\rho = 2$). Тогда для коэффициентов первой и третьей степени имеем:

$$a_w^{(l)} = \frac{1}{2}\left\{6z_w^{(l)} - \sum_{j=1}^{w-1} z_{jjw}^{(l)} - \sum_{j=w}^{\rho} z_{wjj}^{(l)} - 6\sum_{\substack{\nu=1\\(\nu\neq l)}}^{k} \alpha_\nu^{(l)} z_w^{(\nu)} + \right.$$

$$\left. + \sum_{\substack{\nu=1\\(\nu\neq l)}}^{k}\sum_{j=1}^{w-1} \alpha_\nu^{(l)} z_{jjw}^{(\nu)} + \sum_{\substack{\nu=1\\(\nu\neq l)}}^{k}\sum_{j=w}^{\rho} \alpha_\nu^{(l)} z_{wjj}^{(\nu)}\right\}, \ \left(w = \overline{1,\rho}\right);$$

$$a_{www}^{(l)} = \frac{1}{6}\left[z_{www}^{(l)} - 3z_w^{(l)} - \sum_{\substack{\nu=1\\(\nu\neq l)}}^{k} \alpha_\nu^{(l)} z_{www}^{(\nu)} + 3\sum_{\substack{\nu=1\\(\nu\neq l)}}^{k} \alpha_\nu^{(l)} z_w^{(\nu)})\right],$$

$$\left(w = \overline{1,\rho}\right);$$

$$a_{wss}^{(l)} = \frac{1}{2}\left[z_{w\gamma\gamma}^{(l)} - z_w^{(l)} - \sum_{\substack{\nu=1\\(\nu\neq l)}}^{k} \alpha_\nu^{(l)} z_{w\gamma\gamma}^{(\nu)} + \sum_{\substack{\nu=1\\(\nu\neq l)}}^{k} \alpha_\nu^{(l)} z_w^{(\nu)})\right],$$

$$\left(w < \gamma = \overline{1,\rho}\right).$$

Ситуация 3. Совпадение трёх и более в совокупностях значимых факторов первой и третьей степени, т.е. общей выборке принадлежит более двух элементов ($\rho \ge 3$).Тогда для коэффициентов первой и третьей степени имеем:

$$a_w^{(l)} = \frac{1}{2}\{(\rho+4)z_w^{(l)} - \sum_{j=1}^{w-1} z_{jjw}^{(l)} - \sum_{j=w}^{\rho} z_{wjj}^{(l)} -$$

$$- \sum_{\substack{\nu=1\\(\nu\neq l)}}^{k} \alpha_\nu^{(l)}[(\rho+4)z_w^{(\nu)} - \sum_{j=1}^{w-1} z_{jjw}^{(\nu)} - \sum_{j=w}^{\rho} z_{wjj}^{(\nu)}]\}, \ \left(w = \overline{1,\rho}\right);$$

$$a_{w\gamma\beta}^{(l)} = z_{w\gamma\beta}^{(l)} - \sum_{\substack{\nu=1\\(\nu\neq l)}}^{k} \alpha_\nu^{(l)} z_{w\gamma\beta}^{(\nu)}, \left(w < \gamma < \beta = \overline{1,\rho}\right).$$

Определение коэффициентов $a_{www}^{(l)}$, $a_{wss}^{(l)}$ – аналогично Ситуации 2.

III. Выше представленные выражения для расчета коэффициентов системы уравнений (1) относятся только для различных случаев совпадения значимых факторов l-го уравнения со значимыми факторами ν-го и μ-го уравнений. Если же значимые факторы l-го уравнения не совпадают с ν-м и μ-м, то из выражений для определения коэффициентов исключаются члены, определяющие влияние обратных связей. Используя соответствующую систему производящих функций, можно получить конечные формулы для определения коэффициентов системы полиномов третьей степени при несовпадающих в разных уравнениях значимых факторов. В таком случае, коэффициенты обратных связей рассчитываются только для совпадающих в уравнениях факторов. В качестве примера приводится общий вид выражения для расчёта элементов матрицы $A^{(l)}$ для Ситуации 3, при условии $\rho \ge 3$; $(q-\rho) \ge 1$; $(\eta - \rho) \ge 3$; $\tau \ge 2$

$$A_{\mu\nu}^{(l)} = r_{\mu\nu} - \sum_{\iota=1}^{q-\rho} z_\iota^{(\nu)} z_\iota^{(\mu)} - \frac{1}{2}\sum_{g=1}^{\tau} z_{gg}^{(\nu)} z_{gg}^{(\mu)} -$$

$$- \sum_{\substack{g=1\\(g<\xi)}}^{\tau}\sum_{\xi=1}^{\tau} z_{g\xi}^{(\nu)} z_{g\xi}^{(\mu)} - \frac{1}{15}\sum_{p=1}^{\eta-\rho}\left\{z_{ppp}^{(\nu)} - \frac{1}{4}\left(\sum_{j=1}^{p-1}[5z_{jjp}^{(\nu)} - z_{ppp}^{(\nu)}] + \right.\right.$$

$$\left.\left. + \sum_{j=p+1}^{\eta-\rho}[5z_{ppj}^{(\nu)} - z_{jjj}^{(\nu)}]\right)\right\} z_{ppp}^{(\mu)} -$$

$$- \frac{1}{2}\sum_{\substack{p=1\\(p<s)}}^{\eta-\rho}\sum_{s=1}^{\eta-\rho}\left\{z_{pss}^{(\nu)} - \frac{1}{5}z_{ppp}^{(\nu)} - \frac{1}{30}\left(\sum_{j=1}^{p-1}[5z_{jjp}^{(\nu)} - z_{ppp}^{(\nu)}] + \right.\right.$$

$$\left.\left. + \sum_{j=p+1}^{\eta-\rho}[5z_{ppj}^{(\nu)} - z_{jjj}^{(\nu)}]\right)\right\} z_{pss}^{(\mu)} -$$

$$- \frac{1}{2}\sum_{\substack{p=1\\(p<s)}}^{\eta-\rho}\sum_{s=1}^{\eta-\rho}\left\{z_{pps}^{(\nu)} - \frac{1}{5}z_{sss}^{(\nu)} - \frac{1}{30}\left(\sum_{j=1}^{s-1}[5z_{jjs}^{(\nu)} - z_{sss}^{(\nu)}] + \right.\right.$$

$$\left.\left. + \sum_{j=s+1}^{\eta-\rho}[5z_{ssj}^{(\nu)} - z_{jjj}^{(\nu)}]\right)\right\} z_{pps}^{(\mu)} - \sum_{p=1}^{\eta-\rho}\sum_{\substack{s=1\\(p<s<r)}}^{\eta-\rho}\sum_{r=1}^{\eta-\rho} z_{psr}^{(\nu)} z_{psr}^{(\mu)} -$$

$$- \frac{1}{2}\sum_{w=1}^{\rho}\left\{\left[(\rho+4)z_w^{(\nu)} - \sum_{j=1}^{w-1} z_{jjw}^{(\nu)} - \sum_{j=w}^{\rho} z_{wjj}^{(\nu)}\right] z_w^{(\mu)} + \right.$$

$$\left. + \frac{z_{www}^{(\nu)} z_{www}^{(\mu)}}{3} - z_w^{(\nu)} z_{www}^{(\mu)}\right\} - \frac{1}{2}\sum_{\substack{w=1\\(w<\gamma)}}^{\rho}\sum_{\gamma=1}^{\rho}[(z_{w\gamma\gamma}^{(\nu)} - z_w^{(\nu)})z_{w\gamma\gamma}^{(\mu)} +$$

$$+ (z_{ww\gamma}^{(\nu)} - z_\gamma^{(\nu)}) z_{ww\gamma}^{(\mu)}] - \sum_{w=1}^{\rho}\sum_{\substack{\gamma=1\\(w<\gamma<\beta)}}^{\rho}\sum_{\beta=1}^{\rho} z_{w\gamma\beta}^{(\nu)} z_{w\gamma\beta}^{(\mu)},$$

$$\left(\mu, \nu = \overline{1,k}; \nu, \mu \neq l\right).$$

Вектор свободных членов $B_\mu^{(l)}$ для l-го уравнения можно получить из выражения для матрицы $A_{\mu\nu}^{(l)}$ если заменить индекс ν на индекс l, при условии, что $\left(\mu = \overline{1,k}; \mu \neq l\right)$.

Таким образом, оптимальная усовершенствованная регрессионная модель будет иметь вид системы уравнений

$$\hat{y}_l = a_0^{(l)} + \sum_{i=1}^{\zeta} a_i^{(l)} x_i + \sum_{\substack{i=1 \\ (i \le j)}}^{\xi} \sum_{j=1}^{\eta} a_{ij}^{(l)} x_i x_j +$$

$$+ \sum_{i=1}^{\gamma} \sum_{\substack{j=1 \\ (i \le j \le s)}}^{\beta} \sum_{s=1}^{w} a_{ijs}^{(l)} x_i x_j x_s + \sum_{\substack{\nu=1 \\ (\nu \ne l)}}^{k} \alpha_\nu^{(l)} \hat{y}_\nu, \left(l = \overline{1,k}\right),$$

где $\{\zeta,\xi,\eta,\gamma,\beta,w\} \in R_x$, R_x – исходное множество влияющих факторов, $\hat{y}_\nu$ – модельное значение ν-го значимого отклика, k – количество значимых откликов для l-го уравнения (отклика).

4. КРАТКОЕ ИЗЛОЖЕНИЕ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА

Эксперимент, демонстрирующий адекватность представленной нелинейной регрессионной модели, был проведен с использованием гидрометеорологических данных, представленных в массиве ЭРА-40 в период с 1957 по 2003 годы, с ноября по февраль, заданных в узлах регулярной сетки точек 2,5º × 2,5º [14]. Его цель заключалась в установлении особенностей связей взаимодействий процессов, определяющих гидрометеорологические характеристики на указанных ниже поверхностях территории Северной Атлантики (район влияющих факторов: 30 - 90º с.ш. и 70º з.д. - 20º в.д.), с аналогичными процессами над подстилающей поверхностью территории Восточной Европы (район откликов: 30 - 65º с.ш. и 25 - 55º в.д.).

При формировании начальных выборок в качестве гидрометеорологических характеристик использовались:

а) на уровне подстилающей поверхности – её среднемесячные значения температуры, температуры воздуха в слое 2 м, разности указанных температур, потока скрытого тепла, зональной составляющей скорости ветра (для выборок факторов и откликов);

б) на уровнях 850 гПа и 700 гПа – среднемесячные значения температуры воздуха, геопотенциала, массовой доли водяного пара, зональной составляющая скорости ветра, относительного вихря скорости (для выборок факторов).

Из совокупности этих данных в каждой точке пространства Северной Атлантики для каждой из трёх поверхностей были сформированы пять 45-мерных векторов гидрометеорологических характеристик (в каждом узле 15 векторов), для Восточной Европы – 5 векторов (только для поверхности).

Для выявления зон активных взаимодействий атмосферы и океана в районе Северной Атлантики и определения регионов, оказывающих существенное влияние на формирование особенностей климатических режимов в районе откликов, к влияющим факторам, представляющим характеристики тепло- и влагообмена в приповерхностном слое, а также тепло-, влагосодержание и циркуляционные свойства воздуха на уровнях 850 гПа и 700 гПа, перед включением их в модель были применены методы кластерного и компонентного анализа.

С помощью компонентного анализа из совокупностей исходных данных в каждом узле сетки путем ортогонального преобразования были выделены три главных компоненты, исчерпывающие >80% их суммарной дисперсии. Векторы соответствующих главных компонент в узлах выделенного пространства Северной Атлантики, составили исходную выборку для кластерного анализа. В общей сумме получилось 9 выборок для каждого месяца. В каждой выборке 925 45-мерных вектора. В отдельные выборки были выделены координаты собственных векторов матрицы корреляций, необходимые для анализа взаимодействий в системе атмосфера-океан. При этом предполагалось, что значение координаты собственного вектора соответствует весовому коэффициенту гидрометеорологической характеристики в крупномасштабном процессе, выражаемом главной компонентой.

Метод кластерного анализа УИМКД (Универсальный итерационный метод кластеризации данных) [15], применённый к каждой из указанных выборок, позволил получить в полях соответствующих главных компонент однородные регионы. Аналогичные процедуры были использованы и при подготовке выборок откликов. Репрезентативные векторы главных компонент в однородных регионах составили далее исходные выборки для проверки работоспособности описанной регрессионной модели. Определение значимости факторов в уравнениях системы (1) осуществлялось с помощью критерия Фишера (процедура просеивания).

Конечным этапом формирования структуры математической модели откликов в различных регионах Восточной Европы на процессы энергетического взаимодействия атмосферы и океана в Северной Атлантике явилась оценка её адекватности. Для сравнительного анализа были рассчитаны коэффициенты множественной корреляции, характеризующие меру адекватности моделей: исходной (коэффициенты системы уравнений, полученные до процедуры просеивания, остаются без изменений, из системы уравнений устраняются только незначимые факторы, отклики и соответствующие им коэффициенты) и преобразованной (с перерасчётом коэффициентов системы уравнений для значимых факторов и обратных связей).

В табл. 1 представлены вариации коэффициента множественной корреляции фактических и модельных значений откликов. Пример приводится для влияющих факторов в декабре (репрезентативные векторы однородных зон первых трёх главных компонент для уровней: подстилающая поверхность, 850 и 700 гПа) и откликов в январе и феврале (репрезентативные векторы однородных зон первой главной компоненты для уровня: подстилающая поверхность), то есть с последействием 1-2 месяца.

В данной таблице так же указаны варианты "прямого" воздействия влияющих факторов, т.е. без учёта последнего члена системы уравнений (1) (без учёта обратных связей) и с наличием его. Такой подход расширяет возможности анализа результатов и позволяет оценить степень влияния значимых факторов на l-ый отклик через μ-ый отклик. Значимость проверялась с помощью критерия Стьюдента: $t_{кр}(\alpha,\nu) = 2{,}02$, уровень значимости $\alpha = 0.05$, число степеней свободы $\nu = 42$, минимальный значимый коэффициент корреляции $R_{\min} = 0.29$.

5. ВЫВОДЫ

Численные эксперименты, в том числе и представленные в работах [16, 17], показали, что рассмотренная многофакторная нелинейная регрессионная модель обладает рядом достоинств. Во-первых, нелинейные члены, присутствующие в системе уравнений, являются значимыми. Во многих случаях учет их повышает достоверность модели, что находит отражение в усилении тесноты корреляционных связей между модельными и фактическими значениями. Во-вторых, так как большинство гидрометеорологических характеристик, как правило, статистически связаны между собой, и, соответственно, могут иметь общие влияющие факторы, то систему уравнений в модели логичнее представлять в виде аппроксимирующих полиномов с обратными связями. Обратные связи, к тому же, позволяют выделить ту часть общего фактора для l - го и μ - го откликов, которая влияет на l - ый отклик через взаимодействие с μ - ым откликом. Это дает дополнительные возможности исследователю для анализа особенностей работы модели, в частности то, каким образом представлена статистическая связь отклика со значимым фактором в l - ом уравнении. В-третьих, одним из требований, предъявляемых к факторам в модели, является их ортогональность, что позволяет использовать в исходной выборке большое количество факторов без ухудшения информативности модели и, естественно, должно повышать степень объективности моделирования. В-четвёртых, принципы метода вероятностной аппроксимации при построении модели позволяют использовать в алгоритмах расчёта её параметров выборки переменных различных объёмов.

Таблица 1 – Коэффициенты множественной корреляции между модельными и фактическими значениями откликов на территории Восточной Европы на влияющие факторы в акватории Северной Атлантики

Месяц фактора	Месяц отклика	Номер кластера отклика	С учетом вклада обратных связей		Без учета вклада обратных связей	
			без перерасчета коэффициентов	с перерасчетом коэффициентов	без перерасчета коэффициентов	с перерасчетом коэффициентов
декабрь	январь	1	0.69	0.83	0.63	0.76
		2	0.44	0.72	0.40	0.60
		3	0.60	0.80	0.56	0.77
		4	0.73	0.83	0.67	0.79
	февраль	1	0.57	0.66	0.43	0.54
		2	0.39	0.63	0.23	0.53
		3	0.28	0.64	0.14	0.25
		4	0.40	0.68	0.37	0.50

Данная модель была применена в качестве имитационной для определения степени влияния зон экстремальных взаимодействий указанных ранее гидрометеорологических характеристик (выраженных в главных компонентах) регионов Северной Атлантики на аналогичные процессы районов Восточной Европы.

Коэффициенты множественной корреляции между фактическими и модельными значениями откликов, полученные по результатам апробации предлагаемой многофакторной модели нелинейной регрессии, показали не только более тесную связь, но и то, что в некоторых случаях незначимая связь становилась значимой.

СПИСОК ЛИТЕРАТУРЫ

1. Randall D. A., Wood R. A., Bony S., Colman R., Fichefet T., Fyfe J., Kattsov V., Pitman A., Shukla J., Srinivasan J., Stouffer R.J., Sumi A., Taylor K. Climate models and their evaluation. In: *Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007. (Eds: Solomon S., Qin D., Manning M. et al.)
2. Дымников В. П., Лыкосов В. Н., Володин Е. М. Моделирование климата и его изменений: Современные проблемы // Вестн. РАН. 2012. Т. 82. С. 227–236.
3. Жуликов С. Е. Математическое моделирование краткосрочного прогноза погоды // Вестн. ТГУ. 2009. Т. 14, вып. 5. С. 1021–1026.
4. Зарипов Р. Б. Обзор современных методов повышения детализации метеорологических полей // Динамика окруж. Среды и глобал. Изменения климата. 2010. №1. С. 1–11.
5. Катцов В. М., Мелешко В. П. Современные приоритеты фундаментальных исследований климата // Тр. ГГО им. А. И. Воейкова. 2008. Вып. 557. С. 3–19.
6. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. 471 с.
7. Мелешко В. П., Гаврилина В. М., Мирвис В. М., Матюгин В. А., Пичугин Ю. А., Вавулин С. В. Гидродинамико-статистический долгосрочный прогноз метеорологических полей по модели ГГО. 2. Результаты оперативных испытаний и перспективы улучшения прогностической схемы // Метеорология и гидрология. 2002. № 10. С. 5–17.
8. Maqsood I., Khan M. R., Abraham A. An ensemble of neural networks for weather forecasting . Neural Computing & Applications, 2004, vol.13, no 2, pp. 112–122.
9. Taylor, J. W., Buizza R. Neural Network Load Forecasting with Weather Ensemble Predictions. *IEEE Trans. on Power Systems*, 2002, vol. 17 (3). pp. 626-632.
10. Школьный Е. П. Многофакторная регрессионная модель физико-статистического метода прогноза погоды // Труды УкрНИГМИ. 1976. Вып. 134. С. 3–24.
11. Майборода Л. А., Школьный Е. П. Атмосфера и управление движением летательных аппаратов. СПб: ВИТИ, 2010. 572 с.
12. Пономарев В. М., Майборода Л. А. Об одной задаче многомерной регрессии // Изв. АН СССР, Техн. Кибернетика. 1971. № 3. С. 9–18.
13. Домрачёв А. Е. Метеорологические условия формирования микроструктуры тёплых туманов Одессы: Дис…канд. географ. наук: 11.00.09. Одесса, 1989. 177 с.
14. Служба данных ECMWF ERA-40. http://www.ecmwf.int/products/data.
15. Серга Э. Н. Универсальный итерационный метод кластеризации данных // Укр. гідрометеорол. ж. 2013. №12. http://uhmj.odeku.edu.ua/uk/category/2013-uk/12-uk/
16. Shkol'nyi E. P., Serga E. N. Influence of processes in the ocean-atmosphere system in North Atlantic on the intraannual variation on climatic characteristics on the territory of Ukraine . *Physical Oceanography*, 2009, vol. 19, no. 4, pp.240–253.
17. Серга Е. М. Результати чисельних експериментів по статистичної моделі динаміки клімату України // Міжвідомчий науковий зб. України: Метеорологія, кліматологія та гідрологія. 2004. Вип. 48. С. 23–32.

REFERENCES

1. Randall D. A., Wood R. A., Bony S., Colman R., Fichefet T., Fyfe J., Kattsov V., Pitman A., Shukla J., Srinivasan J., Stouffer R.J., Sumi A., Taylor K. Climate models and their evaluation. In: *Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007. (Eds: Solomon S., Qin D., Manning M. et al.)
2. Dymnikov V. P., Lykosov V. N., Volodin E. M. Simulation of climate and climate change: Modern problems. *Vestnik Rossiyskoy akademii nauk – Bulletin of the Russian Academy of Sciences*, 2012, vol. 82. pp. 227–236. (In Russian)
3. Zhulikov S. E. Mathematical modeling of short-term weather forecast. *Vestnik Tambovskogo universiteta – Bulletin of Tambov University*, 2009, vol. 14, issue. 5, pp. 1021–1026. (In Russian)
4. Zaripov R. B. A review of modern methods for increasing detailing of meteorological fields. *Dinamika okruzhayushchey sredy i globalnyye izmeneniya klimata – The dynamics of the environment and global climate change*, 2010, no 1. pp. 1–11. (In Russian).
5. Katsov V. M., Meleshko V. P. Current priorities of the fundamental climate research. *Trudy GGO – Proceedings of the Main Geophysical Observatory*, 2008, vol. 557, pp. 3–19. (In Russian)
6. Ayvazyan S. A., Enyukov I. S., Meshalkin L. D. *Prikladnaya statistika. Osnovy modelirovaniya i pervichnaya*

obrabotka dannykh [Applied statistics. Bases of modelling and initial data processing.]. Moscow: Finansy i statistika, 1983. 471 p.

7. Meleshko V. P., Gavrilina V. M., Mirvis V. M., Matyugin V. A., Pichugin Yu. A., Vavulin S. V. Statistical hydrodynamic long-range forecast of meteorological fields with the MGO model. 2. Operational test results and prospects of improving the prognostic scheme. *Meteorologiya i gidrologiya – Meteorology and hydrology*, 2002, no. 10, pp. 5–17. (In Russian)
8. Maqsood I., Khan M. R., Abraham A. An ensemble of neural networks for weather forecasting . Neural Computing & Applications, 2004, vol.13, no 2, pp. 112–122.
9. Taylor, J. W., Buizza R. Neural Network Load Forecasting with Weather Ensemble Predictions. *IEEE Trans. on Power Systems*, 2002, vol. 17 (3). pp. 626-632.
10. Shkol'nyy E. P. Multi-factor regression model of physical and statistical method of weather forecasting. *Trudy Ukrainskogo nauchno-issledovatelskogo gidrometeorologicheskogo institute – Proc. of the Ukrainian Scientific Research Hydrometeorological Institute*, 1976, no. 134, pp. 3–24. (In Russian).
11. Mayboroda L. A., Shkol'nyy E. P. *Atmosfera i upravlenie dvizheniem letatel'nykh apparatov* [Atmosphere and traffic control of flying machines]. Sankt-Petersburg: VITI, 2010. 572 p.
12. Domrachyov A. E. *Meteorological conditions of formation of microstructure of warm fogs in Odessa: Diss…Cand. Geogr. Sci.: 11.00.09*. Odessa, 1989. 177 p. (In Russian)
13. *Sluzhba dannykh ECMWF ERA-40* [Data Service ESMWF ERA-40]. http://www.ecmwf.int/products/data.
14. Serga E. N. The universal iterative method of clusterization of data. *Ukraïns'kij gidrometeorologichnij zhurnal – Ukr.hydrometeor.j.* 2013, no 12. http://uhmj.odeku.edu.ua/uk/category/2013-uk/12-uk/ (In Russian).
15. Shkol'nyi E. P., Serga E. N. Influence of processes in the ocean-atmosphere system in North Atlantic on the intraannual variation on climatic characteristics on the territory of Ukraine . *Physical Oceanography*, 2009, vol. 19, no. 4, pp.240–253.
16. Serga E. M. The results of numerical experiments on a statistical model of the dynamics of climate in Ukraine. Mizhvidomchiy naukovyy zb. Ukrayiny: Meteorolohia, klimatolohia ta hidrolohiya – Interdepartmental Scientific Collection of Articles of Ukraine: Meteorology, Climatology and Hydrology, 2004, no. 48. pp. 23-32. (In Ukraine)

MULTIFACTOR NON-LINEAR REGRESSION MODEL WITH FEEDBACKS – REGIONAL CLIMATE SIMULATION

E. N. Serga, Cand. Geogr. Sci., Assoc. Prof.

Odessa State Environmental University,
15, Lvivska St., 65016 Odessa, Ukraine, Serga_ed@ukr.net

The article analyzes the improved multifactor non-linear regression model with feedbacks and overdetermined expressions for coefficients of system of equations. It studies probable situations occurring when sampling of influencing factors takes place after selection process is complete. It also shows change of type of generating functions and expressions for determining the elements of matrixes $A_{\mu\nu}^{(l)}$ and vectors of free terms $B_{\mu}^{(l)}$ required in order to get coefficients of feedbacks. The presented system of equations of non-linear regression model takes into account the influence of feedbacks between responses as independent term of equations. General significance of influencing factors and their powers is defined using Fisher's test.

To determine connection between zones of intensive interaction of hydrometeorological characteristics of the North Atlantic and uniform regions of Eastern Europe the model was tested as a simulation one. To reveal zones of active interaction between the atmosphere and the ocean in the North Atlantic region and to determine the regions causing a significant impact on formation of peculiarities of climatic regimes in the regions of response, methods of cluster and component analysis were applied to influencing factors serving as characteristics of heat and moisture exchange in the near-surface layer and also characteristics of heat, moisture content and circulation properties of air at the 850 hPa and 700 hPa levels before including the latter in the model.

Main components of interactions of hydrometeorological characteristics were included into initial sets of samples of the model's influencing factors and responses. Significant multiple correlation coefficients which characterize the degree of adequacy of the model prove the possibility of its practical use when solving similar problems.

Keywords: multifactor regression model, polynomial approximant, influencing factors, coefficients of equations, feedbacks.

БАГАТОФАКТОРНА НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ З ОБЕРНЕНИМИ ЗВ'ЯЗКАМИ – ІМІТАЦІЯ РЕГІОНАЛЬНИХ КЛІМАТІВ

Е. М. Сєрга, канд. геогр. наук, доц.

Одеський державний екологічний університет,
вул. Львівська, 15, 65016, Одеса, Україна, Serga_ed@ukr.net

Пропонується удосконалена багатофакторна нелінійна регресійна модель з оберненими зв’язками з перевизначеними виразами для коефіцієнтів системи рівнянь. Дана модель апробована як імітаційна з метою встановлення зв’язку між зонами інтенсивних взаємодій атмосфери й океану в акваторії Північної Атлантики і однорідними регіонами на території Східної Європи. Головні компоненти енергетичних взаємодій склали вихідні вибірки впливаючих факторів і відгуків для моделі. Значні коефіцієнти множинної кореляції, що характеризують ступінь адекватності моделі, підтверджують можливість її практичного застосування для вирішення задач подібного характеру.

Ключові слова: багатофакторна регресійна модель, апроксимуючі поліноми, впливаючі фактори, коефіцієнти рівнянь, обернені зв'язки.

Дата першого подання: **11. 04. 2017**
Дата надходження остаточної версії: **03. 05. 2017**
Дата публікації статті: **29. 06. 2017**