

О НЕКОТОРЫХ ПРОБЛЕМАХ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ ВОССТАНОВЛЕНИЯ КЛИМАТА ЗА ПРОШЕДШЕЕ ТЫСЯЧЕЛЕНИЕ*

Е. В. ДМИТРИЕВ

Институт вычислительной математики РАН, Москва, Россия

e-mail: yegor@inm.ras.ru

On the basis of some well known reconstructions of past climate from proxy data, in particular MBH98-99 [Mann et al., 1998, 1999], it was concluded that the current global warming is unprecedented, at least during the last two millennia. On the other hand, it was shown [Storch et al., 2004], that the method MBH98 leads to the significant underestimation of the low-frequency variability. This time the results of [Storch et al., 2004] are interpreted as the proof of methodological inconsistency of the method MBH98. In this article it is demonstrated that this conclusion is premature.

Введение

Задача восстановления глобальных геофизических полей на несколько тысячелетий назад является одной из самых актуальных на сегодняшний день. Ее решение необходимо для объяснения текущих изменений климата, в частности, это должно помочь определению относительной величины антропогенного воздействия на современный климат. Одна из наиболее известных реконструкций палеоклимата по косвенным данным была опубликована в работах [1, 2]. В ней было показано, что низкочастотная изменчивость приземной температуры в Северном полушарии за последнее тысячелетие имеет практически горизонтальный линейный тренд, за исключением предыдущего столетия, где имеет место резкий скачок. Предложенный в [1] новый метод интерпретации данных косвенных измерений впоследствии стал называться методом Манна или MBH98, а соответствующая реконструкция — “хоккейной клюшкой”. Реконструкция Манна в этом смысле согласуется с некоторыми другими известными реконструкциями глобальной температуры, например [3, 4]. На этой основе в отчете IPCC 2001 года было сделано заключение о беспрецедентности текущего глобального потепления. Тем не менее значительная часть климатологов не захотела с этим согласиться, что отражено, в частности, в работах [5, 6].

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 04-05-64919).

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2005.

1. Некоторые наиболее часто используемые статистические методы интерпретации косвенных данных

Пусть неизвестное крупномасштабное геофизическое поле описывается случайным вектором $\mathbf{u} \in R^n$. Косвенные данные, которые будут использоваться в качестве предиктора, определим как случайный вектор $\mathbf{y} \in R^m$. Полагая, что физические процессы, описывающие текущее состояние атмосферы, определяют состояние источников косвенных измерений, будем считать, что связь между векторами \mathbf{u} и \mathbf{y} может быть задана в виде системы линейных уравнений

$$\mathbf{y} = G\mathbf{u} + \boldsymbol{\varepsilon}, \quad (1)$$

где G — линейный оператор, а $\boldsymbol{\varepsilon} \in R^m$ — случайный вектор ошибок измерений и линейного приближения. Положим также, что $E\mathbf{u} = E\boldsymbol{\varepsilon} = E\mathbf{u}\boldsymbol{\varepsilon}^T = 0$, где E — символ математического ожидания, а T — знак транспонирования. Решение обратной задачи (1), т. е. восстановление неизвестного вектора \mathbf{u} по известному \mathbf{y} , называют интерпретацией данных косвенных измерений. Поскольку, вообще говоря, матрица G не квадратная, а дисперсия вектора $\boldsymbol{\varepsilon}$ не нулевая, получение точного решения уравнения (1) невозможно и обратная задача является некорректной. В этом случае мы можем лишь надеяться получить решение, оптимальное в некотором смысле, например доставляющее минимум дисперсии нормы вектора ошибки.

Многомерная линейная регрессия (SMLR) решает, вообще говоря, прямую задачу. Однако ее можно применить и для решения обратной задачи (1), просто предположив, как это сделано в [7], что линейный оператор G неизвестен. SMLR дает наилучшую в среднеквадратичном смысле линейную оценку

$$\mathbf{u}_{smlr} = C_{uy}C_{yy}^{-1}\mathbf{y}.$$

При этом может быть вычислена ковариационная матрица ошибки решения

$$H_{smlr} = C_{uu} - C_{uy}C_{yy}^{-1}C_{yu}.$$

Ковариационные и взаимно-ковариационные матрицы определяются следующими соотношениями: $C_{uu} = E(\mathbf{u}\mathbf{u}^T)$, $C_{yy} = E(\mathbf{y}\mathbf{y}^T)$, $C_{uy} = E(\mathbf{u}\mathbf{y}^T)$.

На практике нам неизвестны точные значения вторых моментов. Вместо этого на основе соответствующих матриц наблюдений вычисляются их несмещенные оценки, использование которых при больших размерах предиктора может приводить к неустойчивому решению. Таким образом, SMLR часто применяется после некоторого оптимального сокращения размерности предиктора, которое осуществляется за счет фильтрации малоинформативных гармоник или непосредственно компонент предиктора. Для этого могут быть использованы разложение по эмпирическим ортогональным функциям (EOF) или сингулярным базисам взаимно-ковариационной матрицы (SVD), метод рекуррентной редукции (RR) или иные известные методы.

Вектор предиктора можно представить в виде разложения в конечный ряд

$$\mathbf{y} = \sum_{i=1}^m (\boldsymbol{\varphi}_i, \mathbf{y}) \boldsymbol{\varphi}_i \equiv \sum_{i=1}^m a_i \boldsymbol{\varphi}_i.$$

Разложение по базису собственных векторов матрицы C_{yy} обеспечивает максимальную скорость сходимости данного ряда и называется разложением по EOF. При этом дисперсии случайных величин a_i равняются соответствующим собственным числам λ_i . Регионы, в которых рассматриваемая геофизическая величина обладает наибольшей изменчивостью, оказывают наибольшее влияние на ее среднее значение, которое необходимо восстановить. Таким образом, можно предположить, что более изменчивые гармоники разложения по EOF будут наиболее информативными компонентами предиктора для регрессионной оценки. Определив пороговое значение λ_{\min} , которое определяет уровень шума, мы можем считать, что гармоники с изменчивостью $\lambda_i < \lambda_{\min}$ содержат преимущественно шум и должны быть исключены из предиктора. В частности, для случая одномерного предиктанта имеет место оценка

$$\hat{u}_{eof} = \sum_{i=1}^{\tilde{m}} \frac{(Eu a_i)}{\lambda_i} a_i,$$

где \tilde{m} — число оставшихся гармоник.

Разложение по векторам правого сингулярного базиса взаимно-ковариационной матрицы C_{uy} позволяет построить ортонормированные линейные комбинации, имеющие максимальные корреляции с предиктантом u . Для случая одномерного предиктанта имеется только один информативный сингулярный вектор правого сингулярного базиса взаимно-ковариационной матрицы C_{uy} . В этом случае окончательное выражение для оценки имеет довольно простой вид и может быть записано как

$$\hat{u}_{svd} = \frac{C_{uy} C_{yu}}{C_{uy} C_{yy} C_{yu}} C_{uy} \mathbf{y}.$$

Можно также непосредственно фильтровать компоненты предиктора. Для этого используют метод рекуррентной редукции. Для случая, когда линейный оператор G неизвестен и ни одна из компонент y_i вектора \mathbf{y} не является линейной комбинацией всех остальных, регрессионная оценка вектора \mathbf{u} и априорная оценка ковариационной матрицы ошибки могут быть получены за m шагов, начиная с $\mathbf{u}^{(0)} = E\mathbf{u}$ и $H^{(0)} = C_{uu}$, по следующим рекуррентным соотношениям:

$$C_{ki} = Ey_k y_i - \sum_{j=1}^{i-1} \frac{C_{ij}}{D_j} C_{kj}, \quad i = 1, 2, \dots, k-1,$$

$$\Delta y_k = y_k - \sum_{i=1}^{k-1} \frac{C_{ki}}{D_j} \Delta y_i, \quad D_k = Ey_k^2 - \sum_{i=1}^{k-1} \frac{C_{ki}^2}{D_j}, \quad \mathbf{P}_k = Eu y_k - \sum_{i=1}^{k-1} \frac{C_{ki}}{D_j} C_{u \Delta y_i},$$

$$\mathbf{u}^{(k)} = \mathbf{u}^{(k-1)} + \frac{\Delta y_k}{D_k} \mathbf{P}_k, \quad H^{(k)} = H^{(k-1)} - \frac{\mathbf{P}_k \mathbf{P}_k^T}{D_k}.$$

Максимальная скорость сходимости процесса достигается при выборе на каждом шаге рекуррентной редукции компоненты y_k , соответствующей наибольшему уменьшению нормы ковариационной матрицы ошибки $H^{(k)}$ на k -м шаге. Когда улучшение становится статистически не значимым, процесс останавливается.

Метод МВН98 направлен на решение обратной задачи. На первом этапе вычисляется наилучшая в среднеквадратичном смысле оценка неизвестного линейного оператора G . Затем методом наименьших квадратов находится решение $\hat{\mathbf{u}}_{\text{Mann}}$ уравнения (1). В работе

[2] рассматривается случай, когда предиктант является скалярной величиной и окончательные оценки решения и ковариационной матрицы ошибки могут быть записаны в виде простых выражений [8]:

$$\hat{\mathbf{u}}_{\text{Mann}} = \frac{C_{uu}}{C_{uy}C_{yu}}C_{uy}\mathbf{y}, \quad H_{\text{Mann}} = \frac{C_{uu}^2 C_{uy}C_{yy}C_{yu}}{(C_{uy}C_{yu})^2} - C_{uu}.$$

Сравнение ошибок решения по методу Манна и остальным рассматриваемым в этой работе методам не может выявить существенного преимущества какого-либо из них при использовании конечных выборок для оценки статистических моментов. С одной стороны, метод Манна основан на несмещенной оценке, и при этом решение относительно устойчиво к увеличению числа малоинформативных компонент предиктора. Таким образом, мы можем надеяться на более точное воспроизведение низкочастотной составляющей точного решения. С другой стороны, использование оценки линейного оператора прямой задачи по недостаточно большой статистике может повлечь за собой существенные погрешности решения обратной задачи.

2. Численные эксперименты

Проведем тестирование вышеизложенных методов применительно к задаче восстановления средней температуры Северного полушария (СТСП), т. е. посмотрим, как будет меняться точность решения при различной точности косвенных измерений. Для этого удобно использовать так называемые псевдокосвенные данные, которые получаются добавлением шума различного уровня к точным данным. Тот факт, что инструментальные измерения доступны только за относительно небольшой период времени, существенно затрудняет валидацию методов. Разумной альтернативой является использование данных, полученных на основе моделей общей циркуляции атмосферы и океана. В данной работе были использованы модельные данные (ERIK) [9] среднегодовой температуры Северного полушария за период с 991 по 1991 год, полученные с помощью совместной модели ECHO-G [10]. Конечно, данные ERIK не могут претендовать на хорошее воспроизведение палеоклимата, но тем не менее позволяют на качественном уровне воспроизвести Средневековое потепление, Малый ледниковый период, в частности минимумы Маундера и Дальтона, а также современное потепление климата. Для построения псевдокосвенных данных были использованы реальные координаты косвенных индикаторов палеоклимата в Северном полушарии в соответствии с работой [2], всего 324 точки, и выделены 93 ближайших к ним узлов сетки, на которой считалась модель. Численные эксперименты, проводимые в данной работе, являются продолжением исследований, опубликованных в статье [5].

В реальности косвенные измерения содержат значительные ошибки. В частности, характерные корреляции дендрохронологий с приземной температурой лежат в диапазоне 0.5 ... 0.7. Таким образом, характерный уровень шума составляет 60% от общей дисперсии. В работе [5] данные косвенных измерений загроуляются белым шумом с уровнем от 50 до 80%. При этом метод MBH98 значимо недооценивает низкочастотную изменчивость точного сигнала. Что же происходит при использовании других известных методов? Реконструкции модельной средней температуры Северного полушария, произведенные различными методами по псевдокосвенным данным с различным уровнем белого шума, показаны на рис. 1, а, который соответствует результатам, представленным в работе [5], однако можно сделать вывод, что ранее используемые методы значительно больше недооценивают точное решение. С другой стороны, мы также можем видеть, что, несмотря на

большую среднеквадратичную ошибку, корреляции всех реконструкций с точным решением остаются весьма высокими. Следовательно, решение можно существенно улучшить, изменив соответствующим образом линейный оператор, с помощью которого осуществляется оценка. Это и происходит, если отказаться от удаления линейного тренда на периоде калибровки (рис. 1, б). При этом реконструкция по методу МВН98 опять же оказывается точнее остальных.

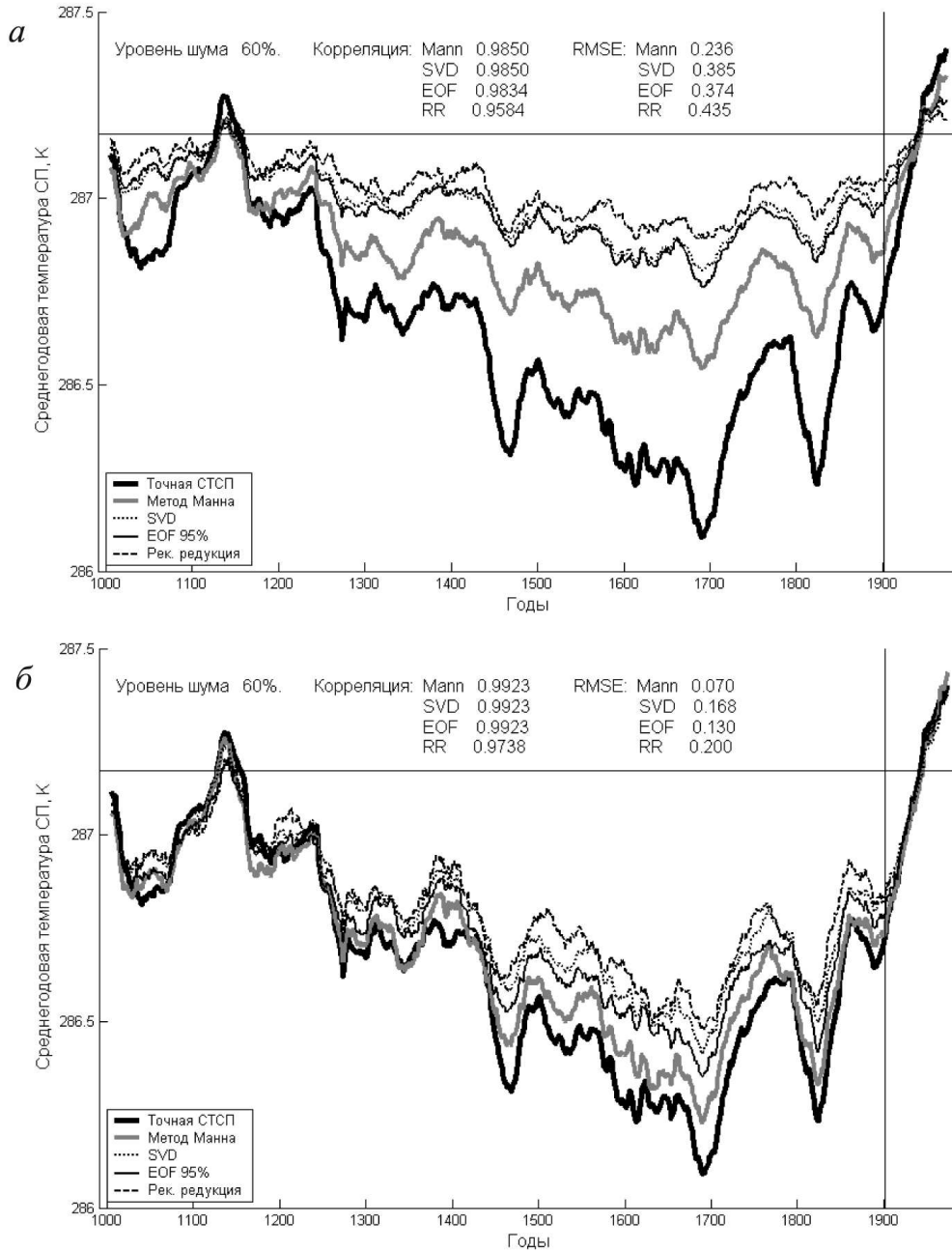


Рис. 1. Среднегодовая приземная температура в Северном полушарии с 991 по 1991 год, сглаженная 31-летним скользящим средним, восстановленная по методам МВН98, регрессии с EOF и SVD фильтрациями предиктора и рекуррентной редукции по зашумленным 60% белым шумом псевдокосвенным данным: а — с удалением линейного тренда на этапе калибровки, б — без удаления.

Проводя подобные тесты статистических методов, исследователи часто ограничиваются белым шумом по времени и пространству, чего может быть совсем недостаточно. Например, если мы говорим о дендрохронологиях, то не стоит забывать, что скорость роста деревьев в значительной степени определяется не только температурой, но и другими геофизическими параметрами, такими как состав и влагосодержание почвы, уровень солнечной радиации, которые могут быть существенным образом скоррелированы по пространству. Таким образом, для построения псевдокосвенных данных более корректно добавление красного по пространству шума. Напомним, что для стационарного случайного

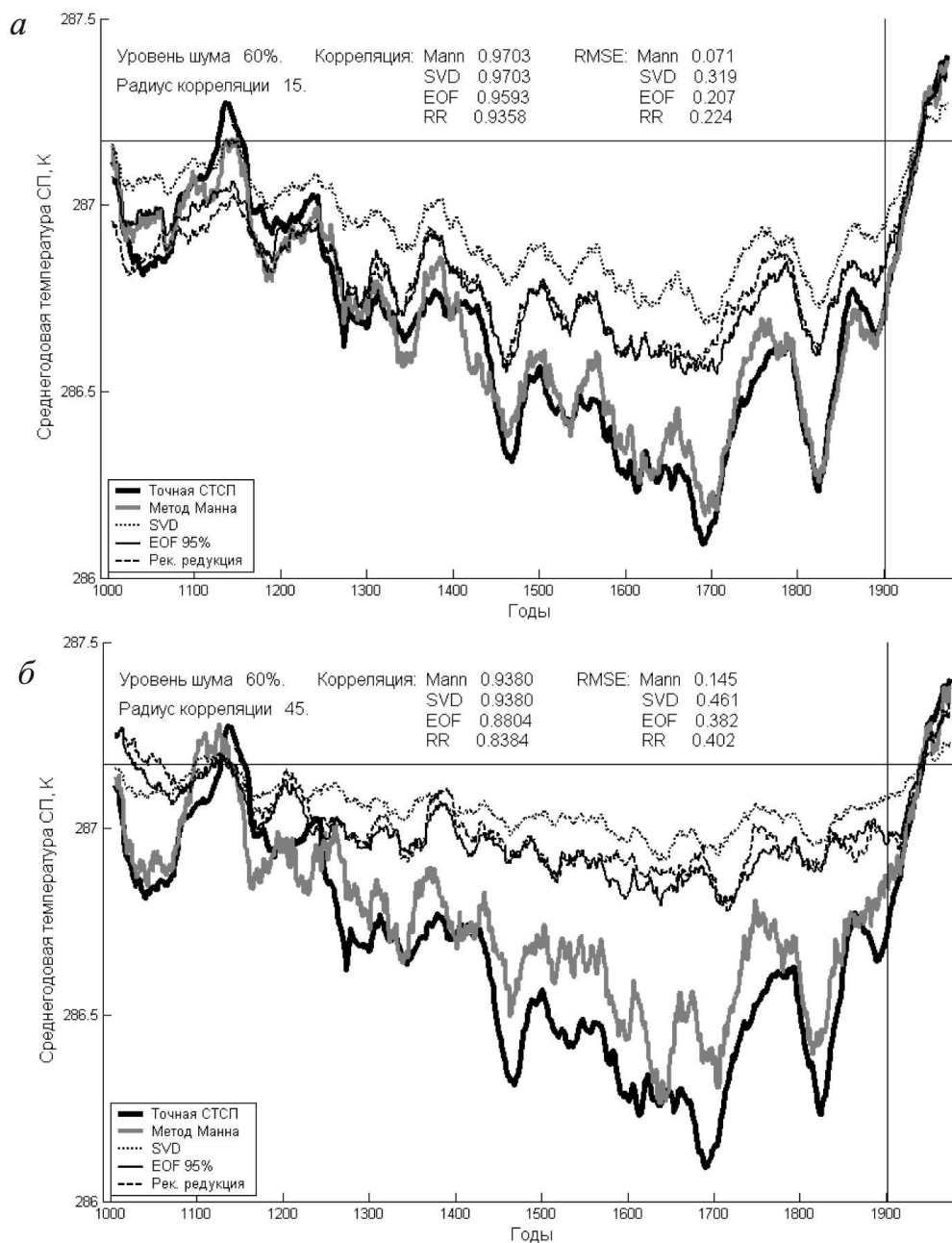


Рис. 2. То же, что на рис. 1, однако загроуление псевдокосвенных данных производится 60 % красным шумом, имеющим различный пространственный радиус корреляции: *а* — 15 град. широты, *б* — 45 град. широты.

процесса, который имеет корреляционную функцию $B(r)$, где $r > 0$ — расстояние между двумя произвольными сечениями, можно определить величину

$$R = \frac{1}{B(0)} \int_0^{\infty} B(r) dr,$$

называемую радиусом корреляции. Используя ковариационную функцию вида

$$B(r) = e^{-r/R},$$

можно преобразовать стандартный белый гауссовский шум в красный с произвольным радиусом корреляции.

На рис. 2 изображены реконструкции климата различными методами по псевдокосвенным данным с характерным уровнем красного шума 60 %. Радиус корреляции измеряется в градусах широты, а расстояние между точками определяется с учетом сферичности земной поверхности. Приведенный пример демонстрирует, что точность реконструкций падает с увеличением радиуса пространственной корреляции шума. Основной вклад в ошибку для всех методов дает недооценка низкочастотной изменчивости, что соответствует работе [5]. С другой стороны, очевидно, что реконструкция по методу Манна значительно лучше воспроизводит низкочастотную изменчивость по сравнению с другими рассматриваемыми в этой работе методами.

Отметим также, что временные зависимости, существующие в косвенных данных, также могут существенно влиять на решение. В частности, реконструкции по кольцам деревьев связаны с добавлением в предиктор низкочастотного сигнала — зависимости скорости роста деревьев от их возраста. При решении некорректных задач иногда даже небольшие изменения низкочастотной изменчивости предиктора могут вызвать существенные изменения в решении. Ошибка решения также зависит от точности оценки соответствующих первых и вторых моментов. С увеличением размерности задачи возрастают и требования к оценкам, поэтому ошибка решения может существенно определяться длиной периода калибровки и возможно, что в ближайшее время количество данных инструментальных измерений скоро станет таким, что проблема недооценки решения исчезнет сама собой.

Выводы

Проведено тестирование некоторых статистических методов, которые наиболее часто используются для восстановления глобальных геофизических полей, составляющих климат прошлого, по данным косвенных измерений. Численные эксперименты показали, что при характерном для косвенных данных уровне шума 60 % метод МВН98 значительно лучше восстанавливает низкочастотную изменчивость точного решения, чем другие ранее используемые в палеоклиматологии методы. Метод МВН98 значительно менее чувствителен к увеличению радиуса пространственной корреляции шума, которое влечет за собой существенный рост ошибки для метода линейной регрессии с различными фильтрациями предиктора. Данная работа является продолжением исследований, опубликованных в статье [5], и ее результаты говорят о преждевременности известных утверждений о методологической несостоятельности метода МВН98 для интерпретации косвенных данных.

Список литературы

- [1] MANN M.E., BRADLEY R.S., HUGHEST M.K. Global-scale temperature patterns and climate forcing over the past six centuries // *Nature*. 1998. Vol. 392. P. 779–787.
- [2] MANN M.E., BRADLEY R.S., HUGHEST M.K. Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations // *Geophysical Research Letters*. 1999. Vol. 26, N 6. P. 759–762.
- [3] JONES P.D., BRIFFA K.R., BARNETT T.P., TETT S. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control run temperatures // *The Holocene*. 1998. Vol. 8, N 4. P. 455–471.
- [4] BRIFFA K.R. Annual climate variability in the Holocene: interpreting the message of ancient trees // *Quaternary Sci. Reviews*. 2000. Vol. 19. P. 87–105.
- [5] STORCH H., ZORITA E., JONES J. ET AL. Reconstructing past climate from noisy data // *Science*. 2004. Vol. 306. P. 679–682.
- [6] MCKITRICK R., MCINTYRE S. Corrections to the Mann et. al. Proxy data base and northern hemispheric average temperature series // *Energy and Environment*. 2003. Vol. 14, N 6. P. 751–771.
- [7] ЧАВРО А.И., ДЫМНИКОВ В.П. Методы математической статистики в задачах физики атмосферы. М.: ИВМ РАН, 2000. 210 с.
- [8] ДМИТРИЕВ Е.В. Анализ методов обработки и оценок погрешностей интерпретации косвенных измерений геофизических величин // *Информационно-измерительные и управляющие системы*. 2004. Т. 2, № 5. С. 12–22.
- [9] ZORITA E., GONZALEZ-ROUCO F., LEGUTKE S. Testing the Mann et al. (1998) Approach to paleoclimate reconstructions in the context of a 1000-Yr control simulation with the ECHO-G coupled climate model // *J. Climate*. 2003. Vol. 16, N 9. P. 1378–1390.
- [10] LEGUTKE S., VOSS R. The Hamburg atmosphere ocean coupled circulation model // *Deutsches Klimarechenzentrum Rep*. 1999. N 18.

Поступила в редакцию 2 июня 2005 г.