О ВОССТАНОВЛЕНИИ ОТСУТСТВУЮЩИХ ДАННЫХ ВРЕМЕННЫХ РЯДОВ С СОХРАНЕНИЕМ ПОВЕДЕНИЯ ПОРОЖДАЮЩЕЙ СИСТЕМЫ

Л. В. Грахольская, С. А. Илларионов, А. О. Мантуров

Поволжский институт управления – филиал РАНХиГС, Саратов, Россия
E-mail: graholskayalv@yandex.ru, manturov1970@yandex.ru

В работе рассматривается задача реконструкции отсутствующих данных временного ряда, порождаемого динамической системой той или иной природы. Процедура реконструкции состоит во введении (импутации) данных, полученных путём интерполяции. При этом корректность процедуры реконструкции рассматривается в смысле оценки устойчивости поведения динамической системы. Как критерий корректности процедуры реконструкции предложено использование расчёта и сопоставления спектров ляпуновских характеристических показателей, проводимых для набора исходных данных, и для реконструированного набора. Путём постановки численного эксперимента показано, что использование сплайн-интерполяции сохраняет поведение порождающей системы в указанном выше смысле

ON RECOVERING MISSING TIME SERIES DATA WHILE PRESERVING THE BEHAVIOR OF THE SOURCE SYSTEM

L. V. Grakholskaya, S. A. Illarionov, A. O. Manturov

The paper considers the problem of missing data reconstructing of a time series generated by some dynamical system. The reconstruction procedure consists of data imputation that obtained by interpolation. So, the correctness of the reconstruction procedure is considered in the sense of evaluation the stability or instability of the behavior of the dynamic system. As a criterion for the correctness of the reconstruction procedure it is proposed to use the calculation and comparison of the Lyapunov Characteristic Exponents spectra carried out for the set of source data and for the reconstructed one. By numerical experiment it is shown that the use of spline interpolation preserves the behavior of the source system in the sense above

Задача реконструкции отсутствующих («потерянных») данных временного ряда наблюдаемой вида $\{x_i\}$, i=0,...,N, в котором отсутствуют данные $\{x_j, x_{j+l}\}$, 0<l<<N, j=0,...,N, является весьма распространённой, и встречается, в том числе, в задачах экономики, эконометрии и социологии [1,2]. Указанные данные могут отсутствовать по различным причинам (техническим - например, сбой регистрации или потери в результате действия помех, в результате действия «человеческого фактора» - непреднамеренного или преднамеренного внесения искажений в результаты регистрации, и т.д.). Как следствие, наборы данных с такими искажениями оказываются непригодными для применения (как минимум, корректного) целого ряда методов их анализа или обработки.

Создано большое количество методов реконструкции отсутствующих данных (imputation of missing data time series, см., например, [3]). Все эти методы имеют существенно различную вычислительную сложность, а практическая их реализация требует привлечения вычислительных мощностей различной ресурсоёмкости. Процедура импутации (imputation) с использованием указанных методов обычно ориентирована на сохранение статистических характеристик временного ряда, и эффективность, или же корректность результатов выполнения этой процедуры оценивается соответственно [2].

В то же время, во многих задачах анализа и обработки данных есть необходимость сохранения динамики, или же поведения во временной области, наблюдаемой системы (например, в задачах цифровой обработки сигналов [4], реконструкции модельных уравнений [5], и т.д.). Оценка эффективности импутации в данном случае на основе расчёта, например, известных статистик, не позволит оценить сохранение особенностей динамики временного ряда.

В связи с изложенным, целью настоящей работы является изучение возможности оценки корректности реконструкции отсутствующих данных временных рядов динамических систем, которая учитывала бы сохраняемость особенностей их поведения (устойчивость или неустойчивость) во временной области.

Будем рассматривать временные ряды, порождаемые динамическими системами с непрерывным временем [6,7]. Ограничим рассмотрение процедуры реконструкции методом сплайн-интерполяции, как относительно нетребовательным к вычислительным ресурсам и обеспечивающим гладкость восстанавливаемых данных до второй производной включительно [8,9].

В качестве источника данных (временных рядов) в работе использована динамическая система Рёсслера, представленная в форме системы обыкновенных нелинейных дифференциальных уравнений [10]. Выбор данной системы обусловлен тем, что её поведение хорошо изучено [11,12], и она демонстрирует широкое разнообразие динамических режимов при небольшом числе (коразмерность 3) управляющих параметров.

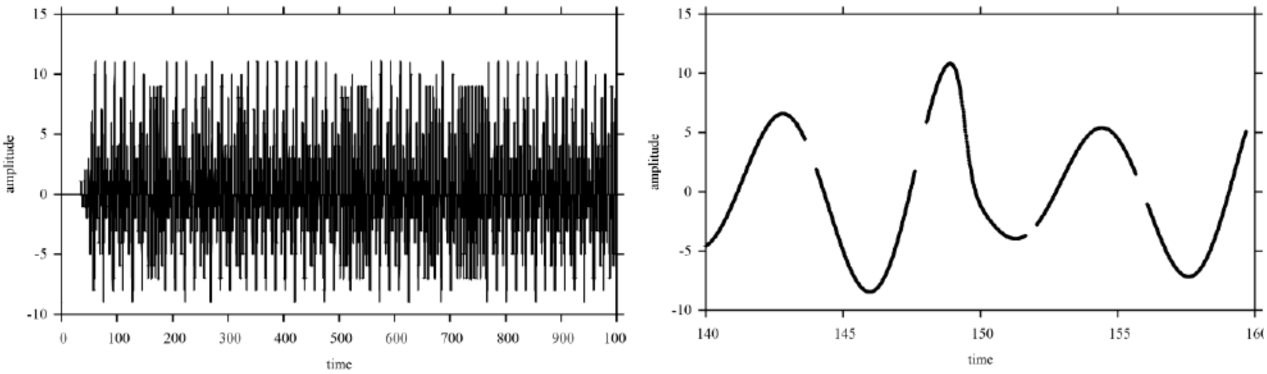
В качестве временного ряда использовалась последовательность отсчётов $x(t)$, порождаемая численным решением модельных уравнений системы [10]. Численное интегрирование выполнялось на временах эволюции $0<T<1000.0$ с шагом $h=0.01$ с использованием метода Рунге-Кутты 4го порядка. Пример полученного тестового временного ряда приведён на рис.1а.

Далее, для имитации потери данных из тестового временного ряда формировались экспериментальные реализации различной длительности T, из которых с заданной периодичностью (400 отсчётов) удалялись подпоследовательности отсчётов той или иной длительности l, $0< l<200$ (рис.1б). При формировании экспериментальных реализаций в их последовательность также не включался переходной процесс, т.е. последовательность отсчётов, соответствующих времени установления стационарной динамики системы.

Восстановление отсутствующих данных выполнялось при помощи вычисления и встраивания соответствующих интерполированных значений. Интерполированные значения вычислялись как отсчёты кубического сплайна, при этом число удалённых и восстановленных отсчётов совпадало.

Оценка корректности восстановления отсутствующих данных выполнялась путём расчёта спектра ляпуновских характеристических показателей

(ЛХП) [7] по экспериментальным реализациям.



а *б*

Рис. 1. Реализация временного ряда x(t), порождаемого численной моделью системы Рёсслера, параметры модели a=0.2, b=0.2, c=5.7 (а); фрагмент временного ряда с имитацией утраты части данных (б)

Для расчёта ЛХП был использован метод, предложенный в [13] и реализованный численно в составе пакета *Nolds* для языка Python [14]. Восстановление данных в настоящей работе считалось корректным, если для исходной и восстановленной реализаций совпадали сигнатуры спектров ЛХП, а модули самих показателей отличались несущественно. Корректность использованного метода расчёта ЛХП была подтверждена экспериментально, и также была определена оптимальная длина экспериментальной реализации T_o=300.0, рис. 2.

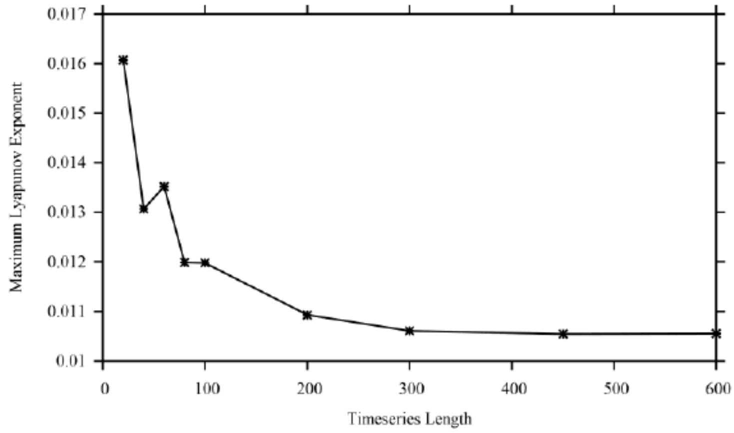


Рис. 2. Оценки наибольшего ЛХП λ_{max} в зависимости от длины T используемой для расчёта реализации

На рис. 2, с увеличением длины реализации можно наблюдать монотонную сходимость рассчитанного наибольшего ЛХП λ_{max} к некоторому предельному значению, характеризующему устойчивость поведения порождающей системы [10], что согласуется с известными оценками для поведения $\lambda_{max} = \lambda_{max}(T)$ [15-17]. Таким образом, вычисление λ_{max} оказывается чувствительным к длине используемой для расчёта реализации. Следовательно, получаемые стационарные (мало изменяющиеся) значения λ_{max}, рассчитанные для различных длин реализации, больших чем T_o, можно использовать как индикатор стационарности динамики, или отсутствия потерь отсчётов временного ряда.

На рис. 3 и 4 приведены результаты численных экспериментов по оценке корректности процедуры восстановления отсутствующих данных для динамической системы Рёсслера.

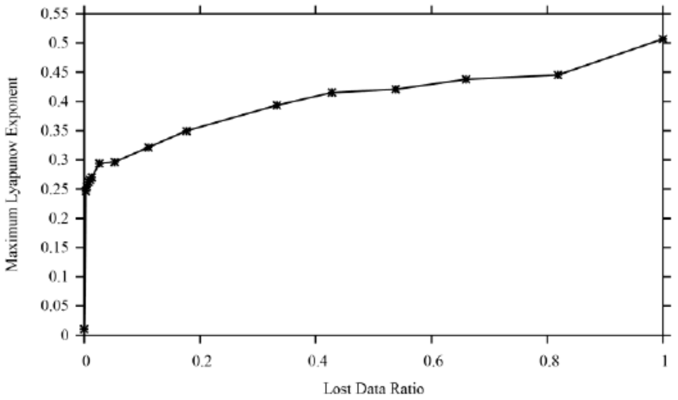


Рис. 3. Оценка наибольшего ЛХП λ_{max} в зависимости от количества отсутствующих (удалённых) данных в реализации

При последовательном уменьшении количества отсутствующих данных (измерялось как отношение числа удалённых отсчётов к числу оставшихся, Lost Data Ratio, α) от 1.0 (потеря данных ~ 46.4%) до 0.013 (~ 1.4%) рассчитанное значение λ_{max} менялось в пределах 0.5 - 0.25 (рис.3), что существенно превышает точные известные значения λ_{max} ~ 0.01. Практически, это соответствует случаю «коротких» временных рядов на рис. 2.

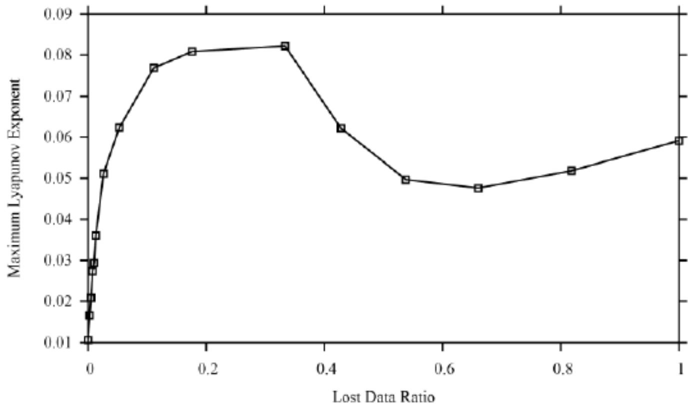


Рис. 4. Оценка наибольшего ЛХП λ_{max} в зависимости от количества восстановленных данных

Результаты расчёта λ_{max} для временных рядов, отсутствующие данные в которых были восстановлены при помощи кубической сплайн-интерполяции, приведены на рис.4. Как можно видеть, при таких же потерях данных полученные оценки λ_{max} оказываются существенно ближе к точному значению (λ_{max} ~ 0.015 - 0.08), что, очевидно, можно рассматривать как приемлемую точность восстановления исходного временного ряда. Обращает внимание немонотонность поведения зависимости $\lambda_{max} = \lambda_{max}(\alpha)$. Такое поведение, очевидно, возможно объяснить регулярностью чередования подпоследовательностей удалённых отсчётов во времени.

Полученные результаты подтверждают возможность оценки эффективности реконструкции отсутствующих данных в реализациях, порождаемых динамическими системами, путём расчёта ляпуновских характеристических показателей. Приведённые выше результаты были получены для динамической системы, и, безусловно, их распространение на более широкий класс систем требует отдельного изучения. На основе полученных результатов может также быть разработан итерационный алгоритм реконструкции отсутствующих данных временных рядов; в таком алгоритме применение какой-либо процедуры импутации данных будет оцениваться расчётом ЛХП, и по результатам расчёта будет выполняться уточнение процедуры импутации на следующем шаге.

СПИСОК ЛИТЕРАТУРЫ

1. *Андерсон Т.* Статистический анализ временных рядов : пер. с англ. / ред. пер. Ю. К. Беляев. М. : Мир, 1976. 755 с.
2. *Литтл Р. Дж. А., Рубин Д. Б.* Статистический анализ данных с пропусками: пер. с англ. - М. : Финансы и статистика, 1990. 336 с.
3. *Ahn H., Sun K., and Kim K.P.* Comparison of Missing Data Imputation Methods in Time Series Forecasting // Comput. Mater. Contin. 2022. Vol. 70. № 1. Pp. 767-779.
4. *Гольденберг Л. М. и др.* Цифровая обработка сигналов: Справочник / Л. М. Гольденберг, Б. Д. Матюшкин, М. Н. Поляк. М. : Радио и связь, 1985. 312 с.
5. *Bezruchko B., Smirnov D.* Nonlinear Dynamical Models from Chaotic Time Series: Methods and Applications // Handbook of Time Series Analysis B. Schelter, M. Winterhalder, J. Timmer. 2006 Ch. 8. P. 193-223.
6. *Рабинович М. И., Трубецков Д. И.* Введение в теорию колебаний и волн / 3-е изд. Ижевск : Науч.-издат. центр «Регулярная и хаотическая динамика», 2000. 560 с.
7. *Кузнецов С. П.* Динамический хаос. М. : ГИЗ ФИЗМАТЛИТ, 2006. 356 с.
8. *Самарский А. А., Гулин А. В.* Численные методы : Учеб. пособие для вузов. М. : Наука. Гл. ред. физ-мат. лит., 1989. 432 с.
9. *Wahba G.* Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics // Society for Industrial and Applied Mathematics, 1990. 181 p.
10. *Rössler O. E.* An equation for continuous chaos // Phys. Lett. A. Vol. 57. № 5. 1976. pp. 397-398.
11. *Castro V., Monti M., Pardo W.B., Walkenstein J. A., Rosa E. Jr.* Characterization of the Rössler system in parameter space // Int. J. Bifur. Chaos. 2007. Vol. 17. № 3. pp. 965-973.
12. *Genesio R, Innocenti G., Gualdani F.* A global qualitative view of bifurcations and dynamics in the Rössler system // Phys. Lett. A. 2008. Vol. 372. Pp. 1799-1809.
13. *Eckmann J. P., Kamphorst S. O., Ruelle D., and Ciliberto S.* Liapunov exponents from time series // Physical Review A. 1986.Vol. 34. № 6. Pp. 4971-4979.
14. *Schölzel C.* Nonlinear measures for dynamical systems (Version 0.5.2). Zenodo. [Electronic resource]. URL: http://doi.org/10.5281/zenodo.3814723 (access date: 25.09.2024).
15. *Eckmann J. P., Ruelle D.* Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems // Physica D: Nonlinear Phenomena. 1992. Vol. 56. № 2-3. Pp. 185-187.
16. *Fischer C., Náprstek J.* Strategies for computation of Lyapunov exponents estimates from discrete data. In: Jan Chleboun and Pavel Kůs and Petr Přikryl and Miroslav Rozložník and Karel

Segeth and Jakub Šístek and Tomáš Vejchodský (eds.) // Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Hejnice. 2018. Institute of Mathematics CAS, Prague. 2019. Pp. 55-62.

17. *Sprott J. C.* Numerical Calculation of Largest Lyapunov Exponent. [Electronic resource]. URL: https: https://sprott.physics.wisc.edu/chaos/lyapexp.htm (access date: 23.08.2024).