УДК 681.514

DOI: https://doi.org/10.34680/2076-8052.2021.2(123).82-86

ПРОГНОЗИРОВАНИЕ НАДЕЖНОСТИ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ НА ОСНОВЕ МОДЕЛЕЙ АВТОРЕГРЕССИИ

А.А.Шерстнева

RELIABILITY FORECASTING ON THE BASE OF AUTOREGRESSION MODELS FOR INFOCOMMUNICATION SYSTEMS

A.A.Sherstneva

Сибирский государственный университет телекоммуникаций и информатики, Новосибирск, asherstneva@sibguti.ru

Рассматривается задача расчета показателей надежности инфокоммуникационных систем, основывающаяся, как правило, на статистических данных, сбор и обработку которых осуществляет система мониторинга. Для получения максимально приближенных к реальным практическим результатам расчётных показателей необходимо проводить большое число измерений. В этом смысле теория фильтрации находит широкое применение в разнообразных задачах оценивания. Популярность обусловлена возможностью эффективного решения технических вопросов и реализацией через программы математического моделирования. Статья направлена на прогнозирование параметров инфокоммуникационных систем на основе моделей авторегрессии. Целью статьи является применение методов регрессионного анализа и рассмотрение задачи оценки показателей по экспериментальным значениям для прогнозирования тренда данных. В результате предлагается решение задачи идентификации интересующего временного ряда переменных методом наименьших квадратов. Помимо теоретических выкладок, разработана программа, позволяющая автоматизировать процесс вычислений.

Ключевые слова: регрессионный анализ, метод наименьших квадратов, прогнозирование, изменение данных, интеллектуальный анализ данных, машинное обучение

Для цитирования: Шерстнева А.А. Прогнозирование надежности инфокоммуникационных систем на основе моделей авторегрессии // Вестник НовГУ. Сер.: Технические науки. 2021. №2(123). С.82-86. DOI: https://doi.org/10.34680/2076-8052.2021.2(123).82-86

The solution to the problems of calculating the reliability indicators of infocommunication systems is based, as a rule, on statistical data. Their collection and processing are carried out by the monitoring system. To obtain the most accurate results of indicators calculation, it is necessary to conduct a large number of measurements. The article aims to make a forecast the parameters based on autoregressive models. The article considers methods of regression analysis and solves a problem indicators estimation based on experimental values for data trend forecasting. As a result, a solution to the problem of identifying the time series of variables by least squares method is proposed.

Keywords: regression analysis, least squares estimation, forecasting, data trend, data mining, machine learning

For citation: Sherstneva A.A. Reliability forecasting on the base of autoregression models for infocommunication systems // Vestnik NovSU. Issue: Engineering Sciences. 2021. №2(123). P.82-86. DOI: https://doi.org/10.34680/2076-8052.2021.2(123).82-86

Введение

В процессе эксплуатации инфокоммуникационных систем возникают ситуации, когда необходимо быстро принимать решение о перенастройке системы для поддержания качества обслуживания на принятом уровне. Выбор верного решения в настоящем должен быть основан на достоверном прогнозе поведения системы в будущем. Поскольку значения параметров, от которых зависит поведение системы, являются совокупностью статистических данных, собираемых системой мониторинга, то задача анализа временных рядов, рассматриваемая в статье, актуальна на сегодняшний день. Однако желательно получение таких результатов анализа, которые были бы максимально приближены к реальным значениям показателей инфокоммуникационных систем. С этой целью строится авторегрессионная модель ряда на основе прошлого поведения системы. Для решения проблемы прогнозирования допускается, что полученная информация о системе актуальна и адекватно отображает поведение системы. В дальнейшем этот подход экстраполирует прошлое поведение на будущее.

Линия регрессии описывает, как меняется прогнозируемая зависимая переменная y (регрессионная переменная) при изменении независимых переменных x (предиктор). Другими словами, линия регрессии используется для прогнозирования значения y для данного значения x. Наиболее распространенной является линия регрессии наименьших квадратов, которая минимизирует вертикальное расстояние от точек данных до линии регрессии. Переменная растет линейно, т.е. в каждый равный период времени к значению переменной добавляется фиксированное приращение. Однако многие переменные, описывающие конкретные процессы в инфокоммуникационных системах, демонстрируют нелинейный рост. В этом случае целесообразно применять принцип наименьших квадратов, который обеспечивает способ эффективного выбора коэффициентов путем минимизации суммы квадратов ошибок [1,2].

При моделировании инфокоммуникационная система представляется в виде графовой модели, включающей множество состояний с описанием созависимостей между ними. Соответственно, необходимы программные решения, которые позволят качественно улучшить прогнозирование системных параметров, повысить точность сложных вычислений и уменьшить время обработки массивов данных. В частности, при решении поставленной задачи методом наименьших квадратов необходимо решить систему уравнений в матричной форме из (*N*-*k*)-уравнений и *k*-неизвестных.

Постановка задачи

На рис.1 в виде графа состояний приведена модель функционирования сетевого элемента с точки зрения надежности. Здесь под сетевым элементом понимается любой элемент, который входит в состав инфокоммуникационной системы, будь то линия связи или функциональный блок. Состояние сетевого элемента контролируется системой мониторинга, которая включает как постоянный вид контроля, так и периодический. Система мониторинга позволяет получить разнообразные статистические данные и характеристики выборки [3,4].

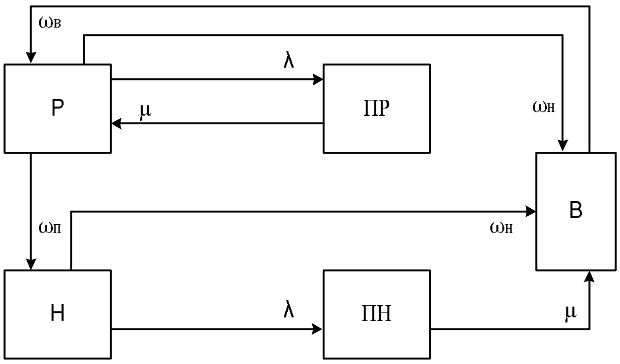

Рис.1. Граф состояний сетевого элемента

На рис.1 состояния обозначены символами Р, Н, В, которые соответствуют работоспособному, неработоспособному состояниям и состоянию восстановления, а также отдельно выделены состояния проверки сетевого элемента, который находится в работоспособном (ПР) и неработоспособном (ПН) состоянии.

Переход из одного состояния в другое осуществляется в результате обнаружения отказа элемента. Переходы обозначены через интенсивности отказов, обнаруживаемых системой контроля, периодического $(\omega_{\text{П}})$ и непрерывного $(\omega_{\text{Н}})$. В модели учитываются интенсивность проверки работоспособного состояния сетевого элемента (γ), интенсивность завершения периодического контроля (μ) и интенсивность восстановления $(\omega_{\text{В}})$. Для оценки, например, надежности сетевого элемента необходимо вычислить ключевые параметры надежности. В данном случае был использован матричный метод анализа вероятностных систем [3]. Получены расчетные формулы для коэффициента простоя и среднего числа проверок, приходящегося на одно восстановление:

$$K_{\text{П}} = \frac{\omega_{\text{П}}\omega_{\text{В}} + \omega_0(\omega_{\text{Н}} + \gamma)}{\omega_{\text{В}}\gamma + \omega_0(\omega_{\text{В}} + \omega_{\text{Н}} + \gamma)},$$

$$n_{\text{Н}} = \frac{\gamma(1+\omega_0)}{\omega_0(\omega_{\text{Н}} + \gamma)},$$

$$K_{\Gamma} = 1 - K_{\text{П}}.$$

Для удобства вычислений в полученных расчетных формулах принято: $\omega_0 = \omega_{\text{Н}} + \omega_{\text{П}}$.

По приведенным формулам можно вычислить текущие значения интересующих параметров, построить необходимые графики зависимости и сделать определенные выводы [4]. Но интерес представляет задача прогнозирования поведения системы при изменении факторов, влияющих на ее надежность, живучесть, продолжительность «жизненного цикла» и т.п. Изменение внешних и внутренних факторов системы приводит и к изменению параметров, включенных в расчетные формулы для приведенного на рис.1 примера. Используемые в модели и в расчетных формулах параметры перехода определяются по статистическим данным, которые собирает и обрабатывает система мониторинга в разные моменты времени. Таким образом, прогнозирование поведения системы в целом зависит от прогнозирования значений этих параметров. Решение задачи прогнозирования можно выполнить путем построения авторегрессионной модели ряда и дальше решить задачу идентификации авторегрессионного процесса, т.е. определения его параметров и оценки порядка с помощью возможностей программы математического моделирования Matlab.

Актуальной является задача оценивания параметров авторегрессионного процесса и дисперсии. Для решения задачи определения порядка модели используется метод наименьших квадратов. Подход заключается в допущении того, что временной ряд является авторегрессионным процессом, и тогда порядок авторегрессии увеличивают до тех пор, пока сумма квадратов ошибок в схеме оценивания параметров авторегрессии не перестанет ощутимо уменьшаться.

Популярность теории фильтрации в задачах оценивания обусловлена возможностью эффективного решения технических вопросов и реализацией через программы математического моделирования. На базе теории фильтрации созданы системы навигации и управления космическими аппаратами и другими подвижными объектами. Предлагаемая теория также находит широкое применение и при расчете надежности инфокоммуникационных систем.

Теория

Согласно теории прогнозирования лучшим результатом нахождения минимума среднеквадратичной ошибки при линейном прогнозе случайной последовательности является расчет условного математического ожидания [1,2,5-7]:

$$\hat{x}_{t+\frac{\tau}{t}} = M\{x_{t+\tau} / x_0, x_1, \ldots, x_t\}.$$

С учетом приведенного выражения можно выполнить прогноз временного ряда для модели авторегрессии.

Авторегрессионная модель представляет собой:

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_P X_{t-P} + = Z_t.$$

Или в более удобной форме записи:

$$X_t = Z_t - \sum_{n-1}^{P} a_n X_{t-n},$$

где $a_0, a_1, a_2, \ldots, a_P$ — постоянные коэффициенты, а Z_t — белый шум.

Коэффициенты измеряют влияние каждого предиктора после учета влияния всех других предикторов в модели. Таким образом, коэффициенты измеряют предельные эффекты переменных предиктора. На практике, конечно, есть набор наблюдений, но значения коэффициентов $a_0, a_1, a_2, \ldots, a_P$ неизвестны. Они должны быть оценены на основе данных. Здесь при ссылке на оценочные значения используется обозначение $\widehat{a}_0, \widehat{a}_1, \widehat{a}_2, \ldots, \widehat{a}_P$.

С помощью функции *filter* программа математического моделирования Matlab позволяет создать авторегрессионную модель. Функция *filter* с входным вектором (*b*,*a*,*z*) фильтрует последовательность данных $Z[m]$, используя цифровой фильтр с коэффициентами $a_n (n = 0,1,2,\ldots,p)$ и $b_n (n = 0,1,2,\ldots,p)$.

Выходная последовательность X_m определяется выражением:

$$x[m] = \frac{1}{a_0}\left(\sum_{n=0}^{q} b_1 z[m-1] - \sum_{n=1}^{P} a_1 x[m-n] \right),$$

где p — порядок фильтра обратной связи, а q — порядок фильтра прямой связи.

Порядок фильтра прямой связи должен быть $q = 0$, чтобы создать авторегрессионую модель:

$$x[m] = \frac{1}{a_0}\left(b_0 z[m] - \sum_{n=1}^{P} a_n x[m-n] \right).$$

Это выражение показывает, что коэффициенты фильтра b_0 и a_0 должны быть равны единице для создания авторегрессионной модели порядка p.

$$x[m] = z[m] - \sum_{n=1}^{P} a_n x[m-n],$$

где $b = b_0 = 1$, $a = (1, a_1, \ldots, a_P)$ и последовательность $Z[m]$ — белый шум.

На рис.2 показана блок-схема функции *filter* Matlab с входной последовательностью $x[m]$ и выходной последовательностью $y[m]$.

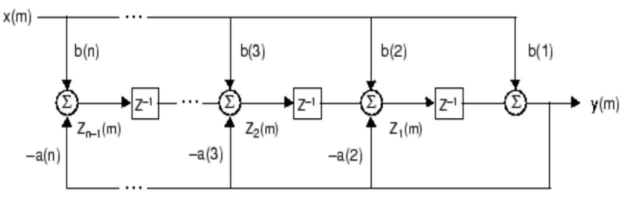

Рис.2. Схема функции «фильтр» в программе математического моделирования

Выражение авторегрессионной модели может быть представлено в виде [5-9]:

$$\sum_t \left(X_t - (\underbrace{-a_1}_{c_1} X_{t-1} \underbrace{-a_2}_{c_2} X_{t-2} - \ldots \underbrace{-a_k}_{c_k} X_{t-k}) \right)^2 = \sum_t Z_t^2.$$

С учетом того, что $-a_i = c_i$, выражение, приведенное выше, можно интерпретировать как сумму ошибок в зависимости от параметра c_i.

$$q(c) = \sum_t (X_t - (c_1 X_{t-1} c_2 X_{t-2} .. c_k X_{t-k}))^2 = \sum_t Z_t^2.$$

Это выражение приводит к методу МНК с параметрами $c_1, \ldots, c_k$ и матричной записи, состоящей из (*N*-*k*)-уравнений и *k*-неизвестных:

$$\underbrace{\begin{bmatrix} X_{k+1} \\ X_{k+2} \\ \vdots \\ X_N \end{bmatrix}}_{X} - \underbrace{\begin{bmatrix} X_{k1} & X_{k-1} & \cdots & X_1 \\ X_{k+1} & X_k & \cdots & X_2 \\ \vdots & & \ddots & \\ X_{N-1} & X_{N-2} & \cdots & X_{N-k} \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}}_{c} = \underbrace{\begin{bmatrix} z_{k+1} \\ z_{k+2} \\ \vdots \\ z_N \end{bmatrix}}_{Z} \quad N \geq 2 \cdot k.$$

Минимальное количество наблюдений — $n = 2k$ для решения системы уравнений, которая привела бы к интерполяции экспериментальных данных. Поэтому для оценки методом наименьших квадратов требуется $n > 2k$ наблюдений. Параметр c_i, который минимизирует сумму квадратов ошибок, необходимо вычислить $2k$ раз.

Применяя метод МНК, получаем решение системы уравнений:

$$\hat{c} = (H^T H)^{-1} H^T X$$
$$\Rightarrow \hat{a} = -\hat{c}.$$

МНК-оценка фильтра с коэффициентами $\hat{a} = (\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_k)$ и дисперсией σ_Z^2 определяется с помощью матрицы проекции P.

Матрица проекции, также известная как матрица влияния, отображает вектор значений отклика (зависимых переменных) к вектору прогнозируемых значений. Применяя матрицу влияния и ее свойства [1, 2], выражение для определения оценки дисперсии принимает вид:

$$P = H(H^T H)^{-1} H^T,$$

$$\hat{\sigma}_Z^2 = \frac{X^T (1-P) X}{N - 2 \cdot k}.$$

Практическая реализация

Для генерации авторегрессионного процесса необходимо отфильтровать белый шум с помощью рекурсивного цифрового фильтра.

Авторегрессионный процесс может быть сгенерирован функцией *filter*. Последовательность нормально распределенных случайных величин является реализацией дискретного белого шума. Коэффициенты фильтра принимают значения: a_1 = 0,5, a_2 = 0,3, a_3 = 0,1, a_4 = 0,7, a_5 = 0,3. При заданных условиях и нулевом значении коэффициентов a_0 и b_0 можно определить три входных вектора функции *filter*. Таким образом, *a* и *b* представляют собой векторы коэффициентов, а *z* — вектор с последовательностью белого шума. Получаем последовательность авторегрессионного процесса x с определенными коэффициентами фильтра и последовательностью белого шума. Порядок авторегрессионного процесса равен $p = 5$, так как коэффициенты фильтра $a_0,\ldots,a_5$.

Следующей задачей является оценка параметров a_i, при $i = 0,1,2,3,4,5$ и дисперсии σ_Z^2 авторегрессионного процесса. Программная реализация в среде Matlab осуществляется в соответствии с теоретическими выкладками для нахождения матрицы коэффициентов и матрицы проекции.

В таблице показана МНК-оценка коэффициентов при допущении, что реальный порядок модели p известен.

МНК-оценка параметров авторегрессионного процесса

Коэффициент	Расчетное значение	Отклонение
$\hat{a}_1$	0,4749	0,0251
$\hat{a}_2$	0,2884	0,0116
$\hat{a}_3$	0,0910	0,0090
$\hat{a}_4$	0,6725	0,0275
$\hat{a}_5$	0,2883	0,0117
	Σ	0,0847

Все оценочные коэффициенты МНК-исследования несколько меньше реальных коэффициентов, но тем не менее максимально приближены к реальным. Дисперсия оценивается согласно приведенному в теоретической части выражению:

$$\sigma_Z^2 = 0{,}8935.$$

По сравнению со встроенной функцией в программе математического моделирования дисперсия *var(z)* = 0,8902. Полученное число максимально приближенно к реальному значению.

Заключение

Подход к прогнозированию, основанный на анализе временных рядов, строится на предпосылке о том, что механизмы поведения системы не меняются со временем с тем, чтобы будущее поведение системы определялось прошлым процессом. В этом заключается и основной недостаток подхода. Параметры моделей неизбежно меняются со временем, вносятся сезонные коррективы, меняется тренд данных. Поэтому данный подход особенно результативен в случае краткосрочного прогнозирования на небольшие отрезки времени. Поиск наиболее подходящих оценок коэффициентов можно рассматривать как обучение модели. Метод регрессионного анализа является видом машинного обучения, использование которого востребовано и актуально в настоящее время.

Для нахождения коэффициентов авторегресионного процесса были написаны программы в среде математического моделирования Matlab. Оценка порядка модели достигается использованием метода наименьших квадратов. Для нахождения порядка модели также была написана программа по заданным параметрам.

При построении моделей коротких временных рядов статистические показатели адекватности модели, такие как коэффициент детерминации, дисперсии параметров и прогнозов, а также процедуры проверки гипотез относительно параметров и построения доверительных интервалов, становятся практически не работоспособными, поскольку на коротких выборках невозможно выявить статистические закономерности. Поэтому для выбора наиболее адекватной (наилучшей) модели временного ряда среди возможных моделей необходимо привлекать экспертную информацию. По существу, в этом случае только экспертные суждения являются той доступной информацией, на основе которой можно выбрать наилучшую модель.

Решением проблемы становится комбинация статистических прогнозов и экспертных систем, предназначенных для решения задач идентификации моделей. Это позволяет значительно улучшить технологию сбора и анализировать уже структурированную информацию. Сочетание методологии мнения эксперта и объективных методов позволяет улучшить надежность прогнозов в условиях недостатка наблюдаемых данных. С помощью экспертных систем становится доступной информация о тенденциях и темпах изменения изучаемого показателя, возможных нижних и верхних границах значений временного ряда на периоде упреждения прогноза.

Чтобы улучшить описанную в статье методику, необходимо построение комплексной модели, которая будет описывать не только внешнюю, но и внутреннюю структуру системы, отражающую внутренние зависимости параметров и влияние внешних факторов. Безусловно, это означает существенно более сложные модели, их построение требует серьезной статистической информации и большого количества наблюдений, иными словами, глубокого механизма изучения процесса. Оценивание таких параметров будет рассмотрено в дальнейшем.

Очевидно, что прогнозы строятся на основе актуальной и достоверной информации в настоящем. Причем для улучшения прогнозирования информация обновляется постоянно, а в современных системах даже в режиме реального времени это позволяет существенно снизить указанные недостатки и корректировать возможности модели. Более того, при прогнозировании поведения реальных процессов пользуются, как правило, не только одним методом, а комплексом доступных инструментов, математических и имитационных.

1. Домбровский В.В., Пашинская Т.Ю. Оптимальные стратегии прогнозирующего управления системами со случайными параметрами, описываемыми многомерной регрессионной моделью с марковским переключением режимов // Вестник Томского гос. ун-та. Управление, вычислительная техника и информатика. 2019. №48. С.4-12. DOI: https://doi.org/10.17223/19988605/48/1
2. Домбровский В.В., Пашинская Т.Ю. Прогнозирующее управление системами с марковскими скачками и авторегрессионным мультипликативным шумом с марковским переключением режимов // Вестник Томского гос. ун-та. Управление, вычислительная техника и информатика. 2018. №44. С.4-9. DOI: https://doi.org/10.17223/19988605/44/1

3. Шерстнева А.А. Оценка параметров одномерного и двумерного распределения случайных величин // Вестник НовГУ. 2020. №5 (121). С.63-67. DOI: https://doi.org/10.34680/2076-8052.2020.5(121).63-67
4. Шерстнева О.Г., Шерстнева А.А. Анализ сети связи с учетом показателей надежности // Вестник РГРТУ. 2020. №73. С.52-58. DOI: https://doi.org/10.21667/1995-4565-2020-73-52-58
5. Bergmeir C., Hyndman R.J., Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction // Computational Statistics and Data Analysis. 2018. Vol.120. Issue C. P.70-83. DOI: https://doi.org/10.1016/j.csda.2017.11.003
6. Wickramasuriya S.L., Athanasopoulos G., Hyndman R.J. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization // J American Statistical Association. 2019. Vol.114(526). P.804–819. DOI: https://doi.org/10.1080/01621459.2018.1448825
7. Harrell F.E. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. New York, USA: Springer, 2015. 582 p.
8. Паклин Н.Б., Орешков В.И. Бизнес-аналитика. От данных к знаниям. Учебное пособие. 2-е изд., СПб: Питер, 2013. 704 с.
9. Вилков А.П., Родионова Т.Е. Использование систем одновременных уравнений для получения моделей описания технических объектов // Современные проблемы проектирования, производства и эксплуатации радиотехнических систем. 2016. №10. С.175-177.

References

1. Dombrovskiy V.V., Ob"edko T.Yu. Optimal'nye strategii prognoziruyushchego upravleniya sistemami so sluchaynymi parametrami, opisyvaemymi mnogomernoy regressionnoy model'yu s markovskim pereklyucheniem rezhimov [Optimal Predictive Control Strategies for Systems with Random Parameters Described by a Multivariate Regression Model with Markov Mode Switching]. Tomsk State University Journal of Control and Computer Science, 2019, № 48, pp. 4-12.
2. Dombrovskiy V.V., Pashinskaya T.Yu. Prognoziruyushchee upravlenie sistemami s markovskimi skachkami i avtoregressionnym mul'tiplikativnym shumom s markovskim pereklyucheniem rezhimov [Predictive control of systems with Markov jumps and autoregressive multiplicative noise with Markov mode switching]. Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitelnaja tehnika i informatika [Tomsk State University Journal of Control and Computer Science], 2018, no. 44, pp. 4-9.
3. Sherstneva A.A. Otsenka parametrov odnomernogo i dvumernogo raspredeleniya sluchaynykh velichin [Parameters estimation of one-dimensional and two-dimensional distribution of random variables]. Vestnik NovSU, 2020, no. 5 (121), pp. 63-67.
4. Sherstneva O.G., Sherstneva A.A. Analiz seti svyazi s uchetom pokazateley nadezhnosti [Analysis of the communication network with reliability indicators]. Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta, 2020, no. 73, pp. 52-58.
5. Bergmeir C., Hyndman R.J., Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics and Data Analysis, 2018, no. 120, pp. 70–83.
6. Wickramasuriya S.L., Athanasopoulos G. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. J American Statistical Association, 2019, no. 114(526). pp. 804–819.
7. Harrell F.E. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. New York, USA, Springer, 2015. 568 p.
8. Paklin N.B., Oreshkov V.I. Biznes-analitika. Ot dannykh k znaniyam [Business analytics. From data to knowledge]. Uchebnoe posobie. Saint Petersburg, 2013. 704 p.
9. Vilkov A.P., Rodionova T.E. Ispol'zovanie sistem odnovremennykh uravneniy dlya polucheniya modeley opisaniya tekhnicheskikh ob"ektov [Using systems of simultaneous equations to obtain models for describing technical objects]. Sovremennye problemy proektirovaniya, proizvodstva i ekspluatatsii radiotekhnicheskikh sistem. UGTU, 2016, no. 10, pp. 175-177.