

УДК 004.588

ББК 74.5

Ч-751

В. З. Чокой
Иркутск, Россия

**ИНСТРУМЕНТЫ РЕГРЕССИОННОГО АНАЛИЗА И
ПРОГНОЗИРОВАНИЯ ПРОЦЕССОВ
В АВИАЦИОННО-ТЕХНИЧЕСКИХ СИСТЕМАХ**

Проведен обзор подходов к подбору для наблюдаемых процессов адекватных регрессий и их последующему использованию для решения практических задач интерполяции, экстраполяции и факторного анализа. Также рассмотрены вопросы формирования для наблюдаемых процессов прогнозных моделей на основе комбинации параболической регрессии, автокорреляционной функции и гармонических весов. Представлен обзор авторских средств компьютерного решения задач аппроксимации, интерполяции и экстраполяции, ориентированных на использование в образовательном процессе инженерных вузов и в эксплуатирующих организациях гражданской авиации.

Ключевые слова: наблюдаемый процесс, регрессионный анализ, парная и множественная регрессия, линеаризация нелинейных функций, метод наименьших квадратов, прогнозирование процессов, доверительные границы прогноза

V. Z. Chokoj
Irkutsk, Russia

**TOOLS FOR REGRESSION ANALYSIS AND FORECASTING
OF PROCESSES IN AIR-TECHNICAL SYSTEMS**

The paper reviews the approaches of adequate regressions selection for observed processes and their use for solving practical problems of interpolation, extrapolation of factor analysis. The issues of forming the predictive models for the observed processes on the base of combining the parabolic regression, autocorrelation function and harmonical weight are considered. The paper reviews the author's means of computer solutions of approximation, interpolation and extrapolation problems oriented toward educational process at engineering higher educational establishments and operating organizations of civil aviation.

Key words: observed process, regression analysis, double and multiple regression, linearization of nonlinear functions, least square method, forecasting of processes, confident limits of forecasting

Вопросы подбора адекватных математических моделей для наблюдаемых процессов актуальны для всех отраслей человеческой деятельности, и особенно – для управления. Остро проблема эффективного управления стоит там, где цена ошибок и риска весьма велика, например, в гражданской авиации. Представляется, что определяющим фактором эффективного управления в авиационной сфере является должное методическое, математическое, и программное обеспечение, позволяющие формировать и эффективно использовать корректные прогнозные модели. При этом важно, чтобы рекомендуемые решения и средства прогнозирования были доступны для широкого круга пользователей, не имеющих углубленной математической и компьютерной подготовки.

Традиционным подходом к моделированию процессов является регрессионный анализ, часто сводящийся к формированию уравнения линии, наилучшим образом аппроксимирующей облако точек из наблюдаемых значений. В большинстве случаев функция y и аргумент x – случайные величины, подверженные действию многочисленных факторов. Если на случайную величину x действуют факторы $z_1, z_2, \dots, \nu_1, \nu_2, \dots$, а на y действуют $z_0, z_1, \nu_1, \nu_3, \dots$, то наличие двух общих факторов z_2 и ν_1 позволяет говорить об определенной зависимости (регрессии) между x и y . Формальными задачами регрессионного обычно явля-

ются: установление формы зависимости (линейная, нелинейная, положительная, отрицательная); определение функции регрессии; оценка влияния на функцию отдельных факторов; решение задач экстраполяции и интерполяции. Ниже рассмотрены лишь задачи определения функции (уравнения), которая в последующем может быть использована для решения задач прогнозирования.

Для однофакторных (парных) линейных регрессий уравнение функции формируют по зависимости вида $y = a + b \cdot x$. Численные значения свободного члена a и коэффициента b определяют с использованием метода наименьших квадратов (МНК) по формулам:

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2};$$
$$a = \bar{y} - b \cdot \bar{x},$$

где n – число измерений функции и аргумента наблюдаемого процесса;

\bar{x} – среднее значение аргумента по всем измерениям;

\bar{y} – среднее значение функции по всем измерениям;

$\overline{y \cdot x}$ – среднее произведений значений функции и аргумента;

$\overline{x^2}$ – среднее значений квадратов аргумента;

\bar{x}^2 – квадрат среднего значения аргумента.

Для однофакторных (парных) нелинейных регрессий уравнение функции обычно формируют в классе типовых нелинейных зависимостей (степенной, экспоненциальной, показательной гиперболической, логарифмической и т. д.), которые, в первом приближении, умозрительно лучше аппроксимируют облако точек, полученных при наблюдениях. При этом применяют тот или иной способ линеаризации нелинейной функции, например, логарифмируют левую и правую части нелинейной зависимости, а затем для логарифмов вводят новые переменные. С учетом новых переменных получают линейное (линеаризованное) уравнение, коэффициенты которого рассчитывают рассмотренным выше способом. Для перехода к исходному нелинейному виду функции, после подстановки найденных численных значений a и b , выполняют обратную исходной

операцию – например, потенцирование, если при линеаризации выполнялось логарифмирование.

Для случаев, когда среди нелинейных зависимостей предпочтение отдается параболе, уравнение функции формируют по зависимости вида $y = a \cdot x^2 + b \cdot x + c$. Численные значения свободного члена a , коэффициентов b и c , как и в предшествующих случаях, определяют с использованием МНК, решая систему уравнений:

$$\begin{cases} (\sum_{i=1}^n x_i^4) \cdot a + (\sum_{i=1}^n x_i^3) \cdot b + (\sum_{i=1}^n x_i^2) \cdot c = \sum_{i=1}^n x_i^2 \cdot y_i \\ (\sum_{i=1}^n x_i^2) \cdot a + (\sum_{i=1}^n x_i) \cdot b + n \cdot c = \sum_{i=1}^n y_i \\ (\sum_{i=1}^n x_i^3) \cdot a + (\sum_{i=1}^n x_i^2) \cdot b + (\sum_{i=1}^n x_i) \cdot c = \sum_{i=1}^n x_i \cdot y_i \end{cases} \quad (1)$$

Для многофакторных (множественных) регрессий уравнение функции формируют на основе полиномов, свободный член и коэффициенты которых также определяют с помощью МНК.

Найденные тем или иным способом уравнения регрессий могут быть использованы в качестве прогнозных моделей лишь в том случае, если наблюдаемый процесс является временным массивом (то есть, одним из аргументов является время, или эквивалентные ему меры, например, кванты времени, этапы, шаги). Однако надо помнить, что вне диапазона экспериментальных значений регрессии корректны лишь в ближайших окрестностях правой границы диапазона, что обеспечивает возможность выполнения только краткосрочных прогнозов.

Прогнозные модели процессов. Модели процессов, пригодные для среднесрочных и долгосрочных прогнозов, чаще формируют не на основе «чистых» регрессий, а на базе других подходов, например, с использованием методов конечных разностей, комбинации регрессий с автокорреляционными функциями и гармоническими весами. Ниже представлен вариант прогнозной модели на основе последнего из упомянутых подходов – при допущении, что аргументом параболической парной регрессии являются равные кванты времени t .

Вначале с использованием системы (1) определяют значения свободного члена и коэффициенты уравнения параболы. Далее последовательно выполняются следующие процедуры.

1. Вычисление значений автокорреляционной функции наблюдаемого процесса, определение с помощью МНК отдельных фаз движения скользящего параболического тренда $Y_i = a_i \cdot t^2 + b_i \cdot t + c_i$ и проверка гипотезы о том, что отклонение от скользящего тренда исходного процесса является стационарным (процедуры проверки стационарности в данной статье не рассматриваются).

2. Вычисление гармонических коэффициентов и весов

$$m_{t+1} = \sum_{i=1}^t 1 / (n - 1)_i; d_{t+1} = (m_{t+1} + 1) / (n - 1), t = \overline{1, n - 1}.$$

3. Вычисление приращений процесса и средней приращений:

$$w_{t+1} = Y_{t+1} - Y_t, t = \overline{1, n - 1};$$

$$w_{cp} = \sum_{t=1}^n d_{t+1}^n \cdot w_{t+1};$$

$$S_w = \sqrt{\sum_{t=1}^{n-1} d_{t+1}^n \cdot (w_{t+1} - w_{cp})^2}.$$

4. Определение переменных $f'(l)$ и $f''(l)$ – точечных значений верхней и нижней доверительных границ прогноза на l -м шаге:

$$f'(l) = f''(l) \cdot \sum_{t=1}^{l+1} d_{n-t+1};$$

$$f''(l) = f'(l) \cdot S_w, l = \overline{0, n - 1}.$$

5. Вычисление прогнозных значений процесса и доверительных интервалов:

$$Y_{t+1}^* = Y_t^* + w_{cp}, t = n - 1 + l;$$

$$Y_{t+1}^* - f''(l) \leq Y_{t+1}^* \leq Y_{t+1}^* + f''(l), l = \overline{1, k + p}.$$

В выражениях пунктов 1...5 переменные обозначают: t – дискретное время (квант времени, шаг); l – глубина прогноза (число шагов); n – число шагов (замеров значений функции) по результатам наблюдений за процессом.

Функциональность и интерфейсные решения по инструментам регрессионного анализа и прогнозирования. В соответствии с рассмотренными выше подходами на факультете Эксплуатации летательных аппаратов Иркутского филиала МГТУ ГА сформирован пакет инструментов регрессионного анализа не-

прерывных процессов и прогнозирования квантованных процессов, характерных для систем типа авиапредприятий. Эти инструменты интегрированы в пакет Модельер 2.0 и для пользователей доступны через группу «Процессы в системах» головного меню (рис. 1).

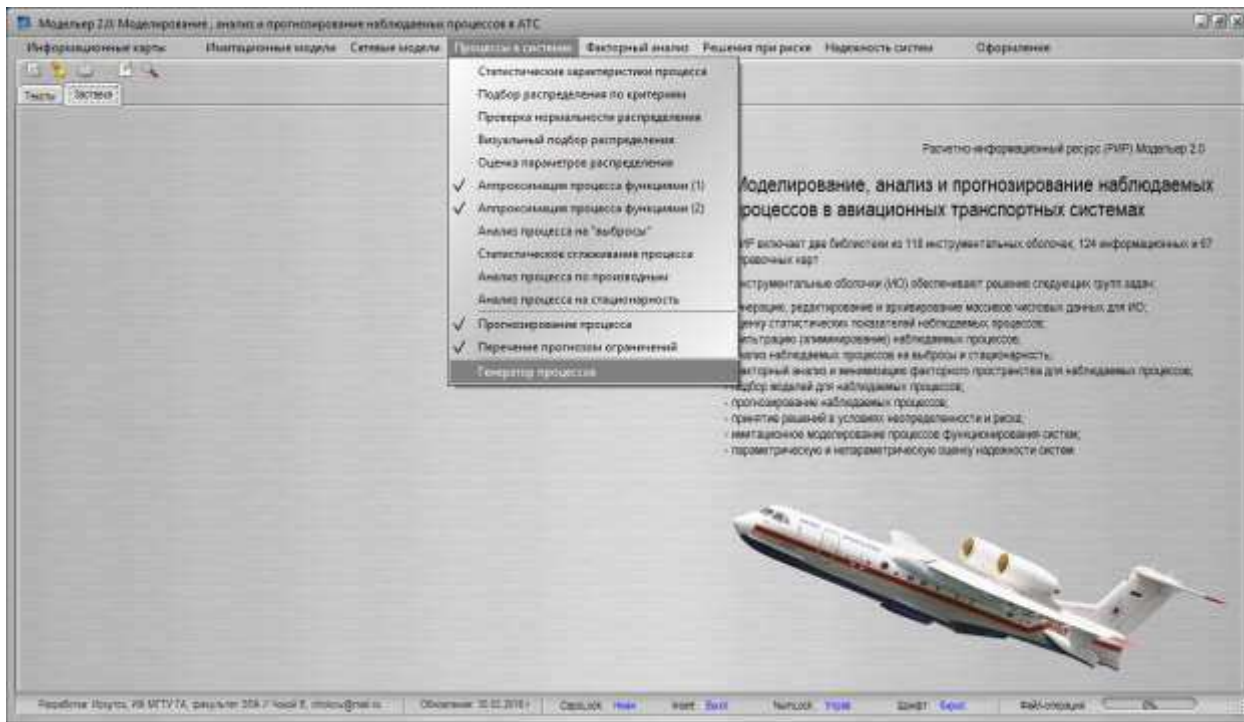


Рис. 1. Головная панель пакета Модельер 2.0

(раскрыта группа «Процессы в системах» головного меню)

Из охваченной в статье предметной области в данной группе головного меню представлены следующие инструменты:

- аппроксимация процесса парными регрессиями (рис. 2);
- аппроксимация процесса регрессией в виде степенного полинома (рис. 3);
- прогнозирование процесса комбинированным способом – параболическая регрессия + автокорреляционная функция + гармонические веса (рис. 4);
- моменты пересечения прогнозом линий-ограничений (рис. 5).

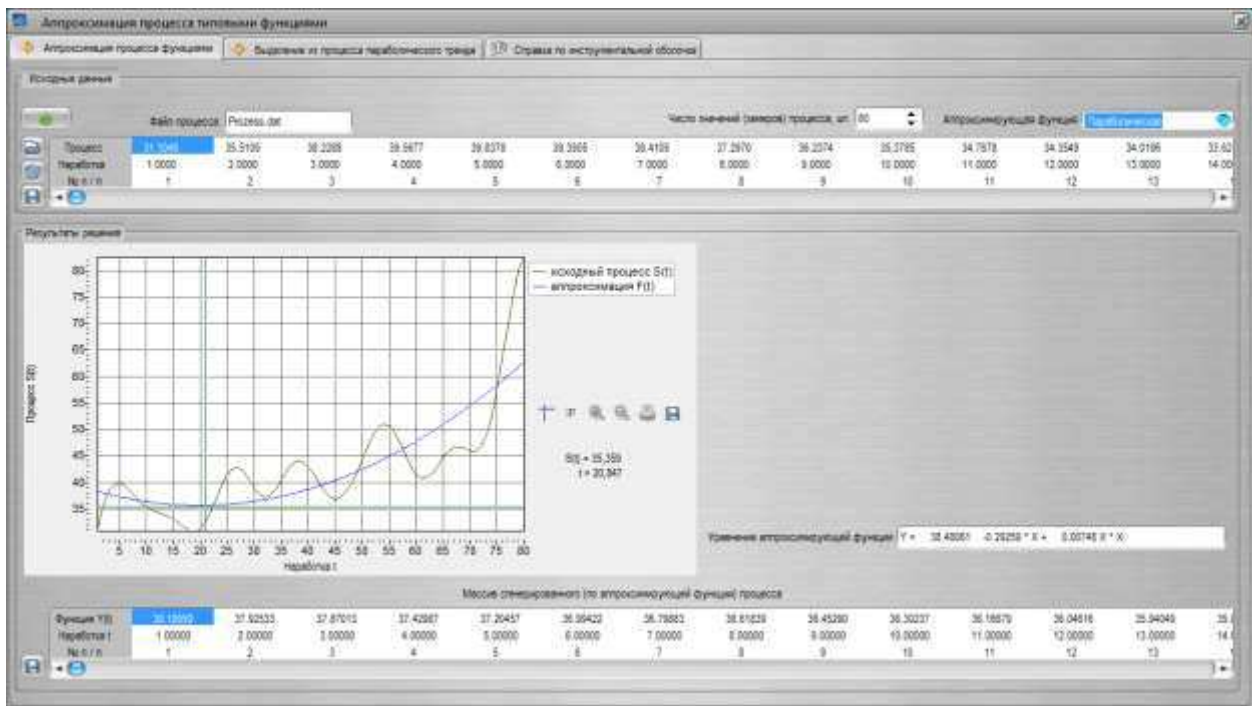


Рис. 2. Панель инструментов «Аппроксимация процесса типовыми функциями (парными регрессиями)»

На панелях всех перечисленных инструментов предусмотрена закладка «Справка по инструментальной оболочке», в которую загружается справочная информация, необходимая пользователям с недостаточным опытом работы (рис. 6). Справка содержит следующую информацию: проверка работоспособности инструмента, показания к применению инструмента, примеры решаемых задач, особенности подготовки исходных данных и интерпретации получаемых результатов.

Исходные данные вводятся с клавиатуры в редактируемые одиночные поля и в таблицы, выбираются курсором из списков. Кроме этого табличные исходные данные могут загружаться из предварительно подготовленных файлов (активацией кнопки с пиктограммой папки).

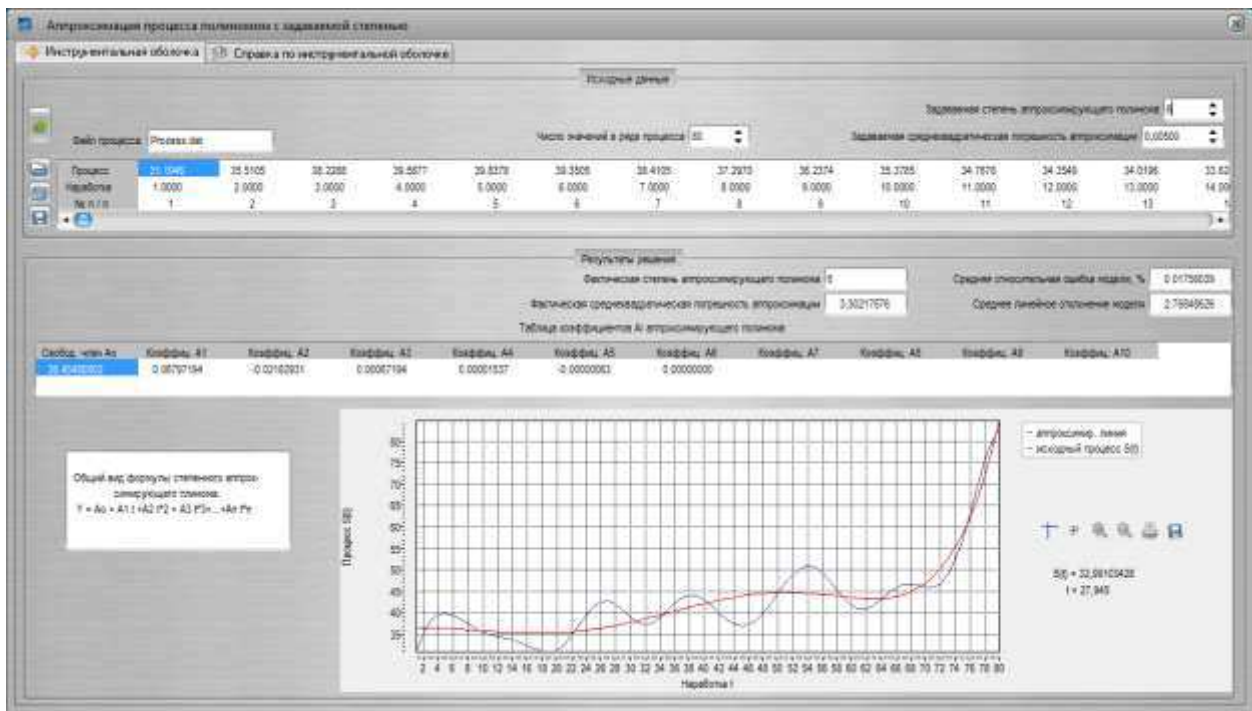


Рис. 3. Панель инструмента «Аппроксимация процесса степенным полиномом (парная регрессия)»

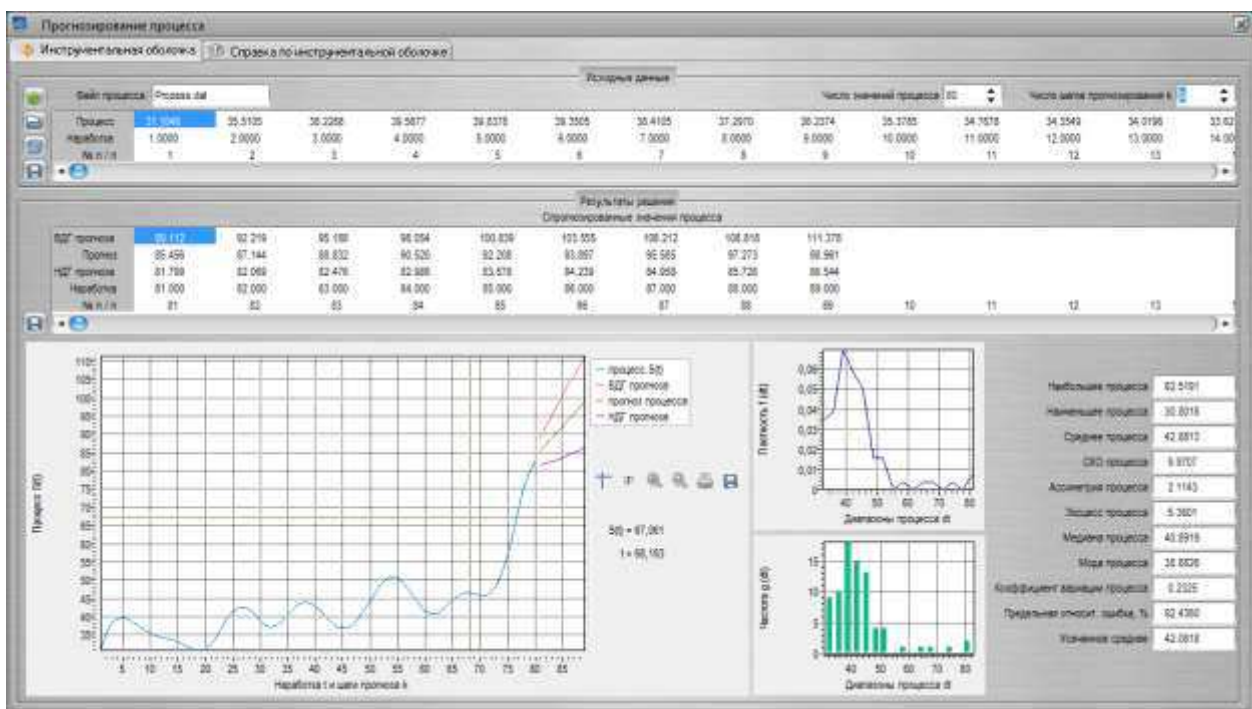


Рис. 4. Панель инструмента «Прогнозирование процесса»

Запуск инструмента на решение задачи осуществляется активацией курсором кнопки с пиктограммой стрелки. Результаты решения выводятся: в единичные не редактируемые поля, в не редактируемые таблицы и на графики. Таблицы исходных данных и результатов могут быть сохранены в файлах, для

чего активируется кнопка с пиктограммой дискеты. Дополнительно графические результаты могут быть: распечатаны на принтере (для чего активируется кнопка с пиктограммой принтера) и масштабированы (для чего активируются кнопки с пиктограммами лупы).

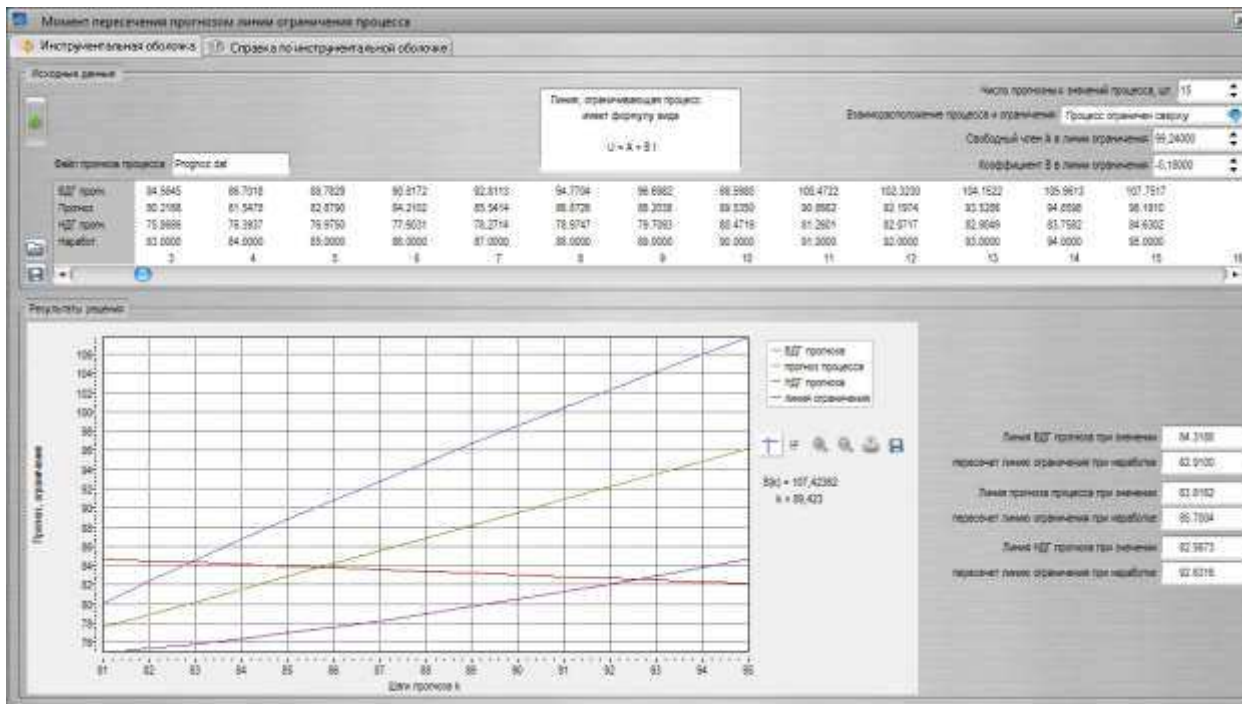


Рис. 5. Панель инструмента «Моменты пересечения прогнозом линии ограничения процесса»



Рис. 6. Пример закладки «Справка по инструментальной оболочке»

Для съема информации с графика о промежуточных значениях аргумента и функции используется визирка, активируемая кнопкой с пиктограммой перекрестия. Цвет визирки может быть подобран при активации кнопки с пиктограммой палитры. Текущие значения функции и аргумента, снимаемые визиркой с графика, выводятся в поля под кнопками графиков.

Для информационного обеспечения работы пользователей в пакет включены информационные карты, доступ к которым возможен через группу «Информационные карты» головного меню. Эти карты содержат иллюстрированные материалы как теоретического, так и практического характера и предназначены для самостоятельной проработки актуальной предметной области перед использованием инструментальных оболочек. Данные материалы также могут быть использованы и при проведении лекций.

Пакет Модельер 2.0 представляет собой автономное полнофункциональное windows-приложение, функционирующее на типовых IBM-подобных ЭВМ с операционной системой Windows-xx. Для инсталляции пакета на жестком диске достаточно 1,8 Гб памяти.

Библиографический список

1. *Головченко В. Б.* Комбинирование моделей неопределенности. Новосибирск: Наука, 2002. 122 с.
2. *Головченко В. Б.* Прогнозирование временных рядов по разнородной информации. Новосибирск: Наука. Сибирская издательская фирма РАН, 1999. 88 с.
3. *Дрейпер Н. Смит Г.* Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. М.: Статистика, 1973. 342 с.
4. *Лейфер Л. А.* Индивидуальное прогнозирование изменения технического состояния. Авторегрессионная модель / Л. А. Лейфер и др. // Надежность и контроль качества, 1987, № 12. С. 27–31.

References

1. Golovchenko V.B. (2002). Combining the uncertainty models. Novosibirsk: Science, 2002. 122 p.

2. Golovchenko V.B. (1999). Forecasting of temporal series by miscellaneous information. Novosibirsk: Science. Siberian publishing company of the RAS, 1999. 88 p.
3. Draper N., Smith G. (1973). Applied regression analysis / N. Draper, G. Smith. M.: Statistics, 1973. 342 p.
4. Lafer L.A. (1987). Individual forecasting of changes of the technical condition. Autoregression model / L.A.Lafer et al. // Reliability and quality control, 1987, № 12. P. 27-31.