

Метод “Гусеница”-SSA: прогноз временных рядов

Ссылка на полную версию работы находится на странице

<http://www.gistatgroup.com/gus/books.html>

6 Приложение: пример прогнозирования временного ряда методом “Гусеница”-SSA

Продemonстрируем методику применения метода “Гусеница”-SSA на примере прогноза временного ряда FORT (объемы месячных продаж крепленых вин в Австралии с января 1984 года по июнь 1994 года). Исходный ряд изображен на рис. 1. Естественно, любой прогноз временного ряда должен предваряться анализом ряда. В данном случае мы будем опираться на анализ рассматриваемого ряда, проведенный в Приложении В работы “Метод “Гусеница”-SSA: прогноз временных рядов”. Так как мы будем рассматривать как сам ряд длины $N = 174$, так и его подряд, состоящий из первых 120 точек, то там, где это необходимо, будем использовать обозначения FORT174 для всего ряда и FORT120 для укороченного ряда. Приведенные графики будут одновременно являться иллюстрацией понятий, возникающих при изложении теоретических аспектов метода.

Так как прогноз методом “Гусеница”-SSA применяется к рядам, управляемым (может быть, приближенно) линейными рекуррентными формулами, то начнем с исследования ряда FORT с этой точки зрения.

6.1 Описание временного ряда линейной рекуррентной формулой

В предыдущем разделе мы определили, что ряд может быть представлен в виде “сигнал+шум” и смогли приближенно отделить сигнал от шумовой компоненты, причем качество разделения довольно хорошее (w -корреляция между этими компонентами равна 0.004). При этом сигналу соответствует 11 собственных троек, что означает, что его траекторное подпространство должно хорошо аппроксимироваться подпространством размерности 11. В свою очередь, подпространство порождает линейную рекуррентную формулу размерности $83 = L - 1$ такую, что восстановленный ряд (в данном случае — сигнал) будет аппроксимироваться рядом, управляемой этой ЛРФ.

В таблице 1 представлена информация о первых 19 из 83 корней характеристического полинома данной ЛРФ. Корни упорядочены по модулю. Если в столбце “Тип” таблицы написано “компл.”, то это означает, что строка содержит информацию о двух комплексно-сопряженных корнях и указано значение модуля мнимой части корня. Первые 6 строк таблицы можно легко проинтерпретировать: 1-3 и 4-5 строки — соответствуют комплексно-сопряженным корням, порождающим гармоники с периодами примерно равными 6, 4, 2.4, 12 и 3. Модуль, больший единицы, указывает на возрастание амплитуды, меньший единицы — на убывание. Результаты согласуются с рис.8. Четвертая строка таблицы соответствует вещественному корню с модулем 0.997. Так как число главных корней должно быть равно как раз одиннадцати, то больше главных корней нет. Остальные корни являются побочными, по модулю меньшими единицы. Все корни изображены на рис. 2 на комплексной плоскости в координатах (Re, Im).

Проверим, насколько хорошо ряд описывается порожденной восстановленным подпространством ЛРФ (для дальнейших ссылок будем называть таким формулы *полными ЛРФ*). Максимальная ошибка глобальной аппроксимации равняется 132, что составляет меньше 10% от значений ряда. Напомним, что так как в качестве начальных значений ЛРФ были взяты первые 83 значений ряда, то ошибка аппроксимации считается, начиная с 84ой точки.

Удалим теперь все корни полинома кроме одиннадцати главных. Главные корни порождают новую (минимальную) ЛРФ, которая, если взять в качестве начальных точки 73ой по 83ю, то на

Таблица 1: Ряд FORT: характеристики корней характеристического полинома полной ЛРФ

N	Re	Im	Модуль	Частота	Период	Тип
1	0.497	0.871	1.003	1.053	5.969	компл.
2	-0.002	1.000	1.000	1.573	3.994	компл.
3	-0.870	0.489	0.998	2.630	2.389	компл.
4	0.997	0.000	0.997	0.000	Нет	вещ.
5	0.861	0.497	0.994	0.524	12.002	компл.
6	-0.478	0.866	0.989	2.075	3.028	компл.
7	-0.094	0.972	0.976	1.667	3.768	компл.
8	-0.391	0.894	0.975	1.983	3.168	компл.
9	0.796	0.563	0.975	0.615	10.212	компл.
10	0.401	0.888	0.975	1.147	5.480	компл.

оставшейся части ряда она будет аппроксимировать ряд даже чуть лучше (максимальная ошибка равна 94). Если взять в качестве начальных первые 11 точек, то максимальная ошибка аппроксимации, сосчитанная на 12–174 точках, увеличится (что не удивительно) до 495. Таким образом, ряд достаточно хорошо аппроксимируется рядом, управляемым минимальной ЛРФ размерности 11 с корнями, изображенными на рис. 3. Поэтому, с точностью до констант, можно выписать аналитический вид этой аппроксимации. Для локальной аппроксимации соотношение результатов примерно такое же (абсолютные значения ошибок меньше).

Так как мы знаем точное значение периода, то можно еще подправить ЛРФ, изменив корни так, чтобы они в точности соответствовали периодам 6, 4, 2.4, 12 и 3 (будем называть эту формулу размерности 11 *исправленной минимальной ЛРФ*). Ошибки локальной аппроксимации при этом даже уменьшатся.

Выпишем вид ряда, управляемого исправленной минимальной ЛРФ:

$$f_n = C_1 0.997^n + C_2 0.994^n \sin(2\pi n/12 + \phi_2) + C_3 \sin(2\pi n/4 + \phi_3) + C_4 1.003^n \sin(2\pi n/6 + \phi_4) + C_5 0.998^n \sin(2\pi n/2.4 + \phi_5) + C_6 0.989^n \sin(2\pi n/3 + \phi_6). \quad (1)$$

Здесь коэффициенты C_i и ϕ_i определяются начальными данными. Мы выписали слагаемые в порядке убывания их вкладов в сумму на интервале от 1 до 174, а не в порядке убывания модулей порождающих их корней, который руководит упорядоченностью вкладов в будущем, построенном по соответствующей ЛРФ.

Таким образом, предварительно исследование показало, что ряд достаточно хорошо соответствует требуемой модели и можно приступать к прогнозированию.

6.2 Прогноз и доверительные интервалы

Устойчивый долгосрочный прогноз. Как было показано выше, ряд FORT оказался почти идеальным для прогноза: он состоит из сигнала, хорошо аппроксимируемого рядом, управляемым ЛРФ, и белого шума. На рис. 4 приведены первые 24 точки рекуррентного прогноза на 60 точек вперед по полной ЛРФ, построенной по подпространству, порожденному первыми одиннадцатью собственными тройками. Значения векторного варианта прогноза, а также рекуррентного прогноза вида (1) по исправленной минимальной ЛРФ почти совпадают с изображенным на рис. 4, поэтому мы их не приводим. Такое совпадение подтверждает устойчивость построенного прогноза.

На рис. 5 изображены бутстреп доверительные интервалы, соответствующие уровню доверия 0.95. При построении доверительных интервалов мы вначале убедились, что ряд остатков восстановления удовлетворяет необходимым условиям. А именно, можно считать, что ряд $\tilde{F}^{(2)} = F - \tilde{F}^{(1)}$, где $\tilde{F}^{(1)}$ — восстановленный по первым 11 собственным тройкам ряд, состоит из независимых реализацией нормально-распределенных случайных величин. Исследование остатков уже проводилось при анализе ряда. Гипотеза о независимости подтверждается, в то время как критерий Аббе для проверки гипотезы о нормальности дает пороговый уровень значимости 0.01. Некоторые отклонение

от нормальности могут немного исказить величину доверительных интервалов, однако качественно им все равно можно доверять. Заметим, что размер доверительных интервалов лишь незначительно растет в течении 5 лет.

Таким образом, для ряда FORT построен прогноз на 60 точек вперед, которому вполне можно доверять при условии, что структура ряда в дальнейшем не изменится.

Краткосрочный прогноз. Не всегда данные таковы, что по ним возможно построить прогноз на большое число точек вперед. Причинами этого могут быть следующие: 1) недостаточная длина ряда по сравнению с величиной шума и/или сложностью сигнала; 2) нелинейная структура сигнала, в том смысле, что он плохо аппроксимируется рядом, управляемым ЛРФ относительно небольшой размерности. Мы рассмотрим только первый случай.

Укоротим временной ряд FORT до первых 120 точек так же, как это делалось при анализе ряда. Для того, чтобы было удобно ссылаться, назовем его FORT120, а исходный ряд — FORT174. Проведенный в Приложении В предыдущей работы анализ ряда показал, что к сигналу можно отнести первые 11 собственных троек сингулярного разложения траекторной матрицы ряда FORT120. Перемешивание двух гармоник для прогноза существенно только в том смысле, что включать/не включать собственные тройки с номерами 8-11 необходимо одновременно.

Также как и для ряда FORT174, построим прогноз на 60 точек вперед на основе первых 11 собственных троек. Структура порожденного соответствующими собственными векторами пространства для полного и укороченного рядов полностью аналогична. Разница в том, что ряд FORT120 короче, поэтому качество приближенной отделимости сигнала от шума должно быть хуже. Это подтверждает ошибка глобальной аппроксимации, максимальное значение которой равно 312 (вместо 132 для FORT174).

Для укороченного ряда мы имеем большое преимущество при оценке качества прогноза, состоящее в том, что мы можем сравнивать значения прогноза с реальными значениями ряда. На рис. 6 изображены значения рекуррентного прогноза (жирная сплошная линия) по сравнению с точками ряда FORT174. Видно, что на первых 24 точках прогноз качественно описывает будущее поведение, а затем начинаются довольно сильные расхождения. Приведем информацию о первых 19 из 59 корней характеристического полинома полной прогнозирующей ЛРФ для ряда FORT120 (таблица 2). Сравнение с таблицей 1 показывает, что поведение прогноза ряда FORT120 обусловлено большим модулем комплексно-сопряженных корней, соответствующих модулированной гармонике с периодом 6 (первая строка таблицы). На рис. 6 видно, что именно растущая по амплитуде 6-месячная гармоника в основном и отвечает за расхождение прогноза с реальными данными.

Таблица 2: Ряд FORT120: характеристики корней характеристического полинома полной ЛРФ

N	Re	Im	Модуль	Частота	Период	Тип
1	0.505	0.878	1.013	1.049	5.990	компл.
2	-0.885	0.480	1.007	2.644	2.376	компл.
3	0.000	1.000	1.000	1.571	4.001	компл.
4	0.997	0.000	0.997	0.000	нет	вещ.
5	0.862	0.496	0.994	0.522	12.033	компл.
6	-0.490	0.851	0.982	2.093	3.002	компл.
7	0.366	0.896	0.968	1.183	5.311	компл.
8	0.768	0.587	0.966	0.652	9.635	компл.
9	-0.128	0.957	0.966	1.704	3.688	компл.
10	-0.899	0.350	0.965	2.771	2.268	компл.

Для того, чтобы подтвердить меньшую устойчивость прогноза ряда FORT120 по сравнению с прогнозом ряда FORT174, построим доверительные интервалы. На рис. 7 видно, что размер бутстреп доверительных интервалов сильно увеличивается, даже захватывая область отрицательных значений.

Чтобы окончательно убедиться в том, что плохое качество долгосрочного прогноза является следствием недостаточной разделимости, а не того, что структура первых 120 точек ряда отлича-

ется от структуры последующих точек, применим исправленную минимальную ЛРФ, полученную при анализе ряда FORT174 и соответствующую формуле (1). Результат изображен на рис. 8 и показывает довольно хорошее качество прогноза (Жирная сплошная линия соединяет точки прогноза, а реальные значения помечены квадратиками).

Тестируя качество прогноза на укороченном ряде, мы имеем возможность количественно выразить качество прогноза, например, с помощью среднеквадратичного отклонения значений прогноза от реальных значений. Рекуррентный прогноза по полной прогнозирующей ЛРФ дает значение среднеквадратичной ошибки, равное 549, векторный вариант прогноза приводит к чуть меньшей ошибке 535, в то время как прогноз по исправленной минимальной ЛРФ уменьшает ошибку до 314.

Таким образом, исследованный пример подтверждает теоретические результаты о зависимости качества прогноза от точности разделимости.

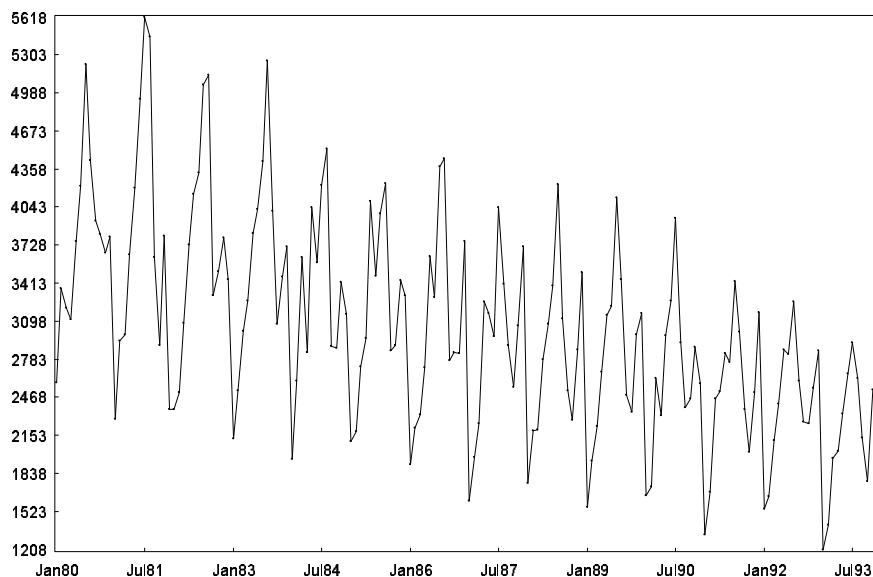


Рис. 1: FORT174: график исходного ряда месячных объемов продаж крепленых вин в Австралии

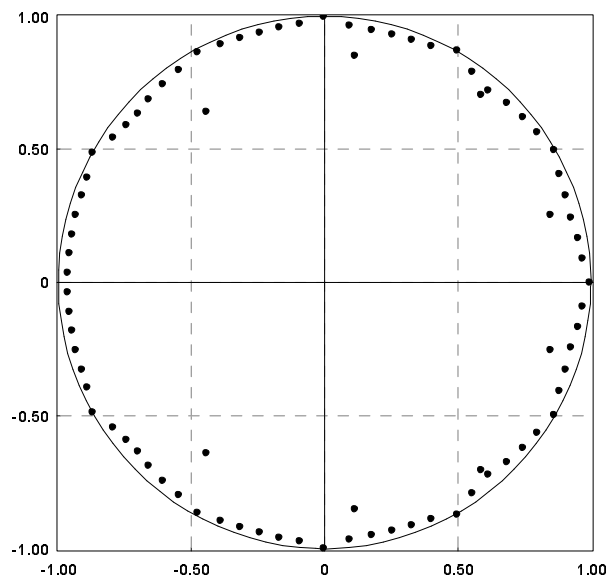


Рис. 2: FORT174: главные и побочные корни характеристического полинома полной ЛРФ

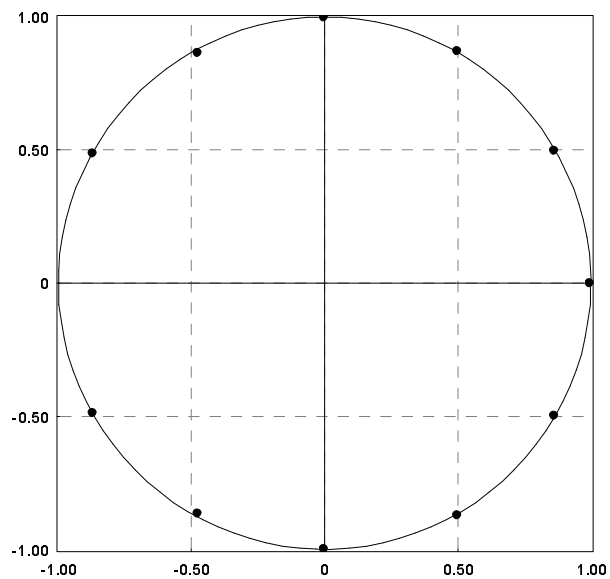


Рис. 3: FORT174: главные и побочные корни характеристического полинома полной ЛРФ

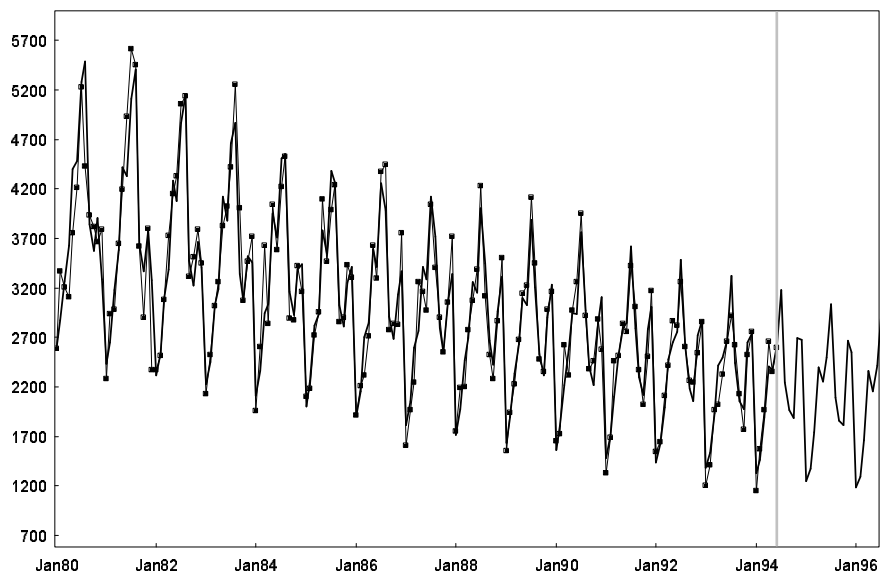


Рис. 4: FORT174: Рекуррентный прогноз на 24 точки вперед

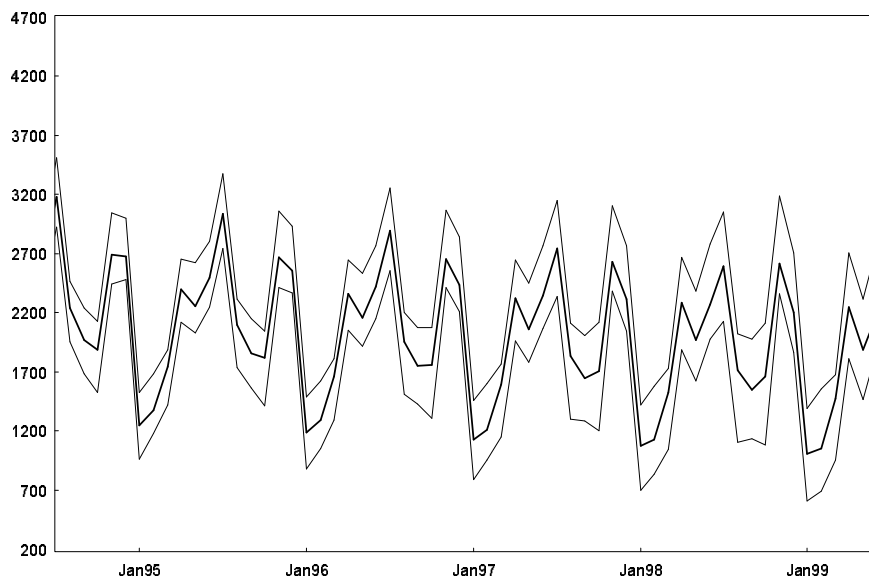


Рис. 5: FORT174: Рекуррентный прогноз на 60 точек и бутстреп доверительные интервалы

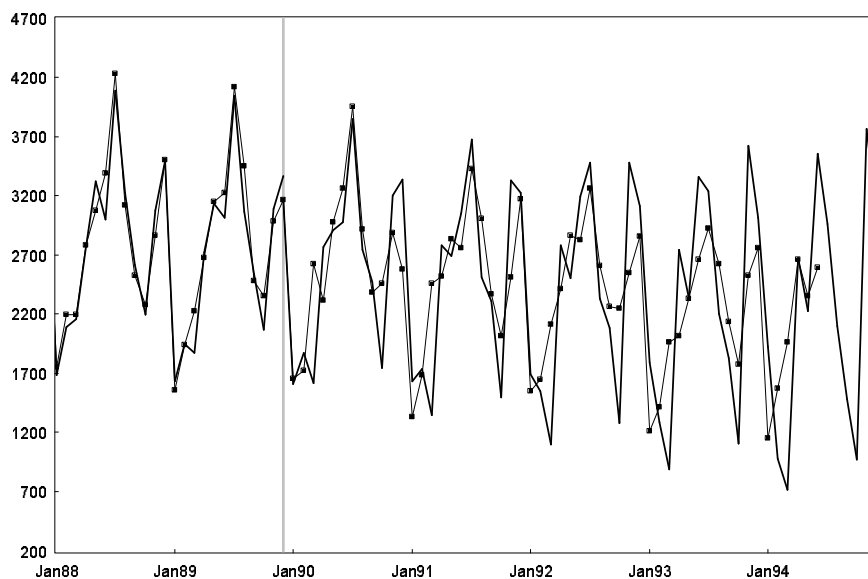


Рис. 6: FORT120: Рекуррентный прогноз по полной ЛРФ на 60 точек на фоне реальных значений

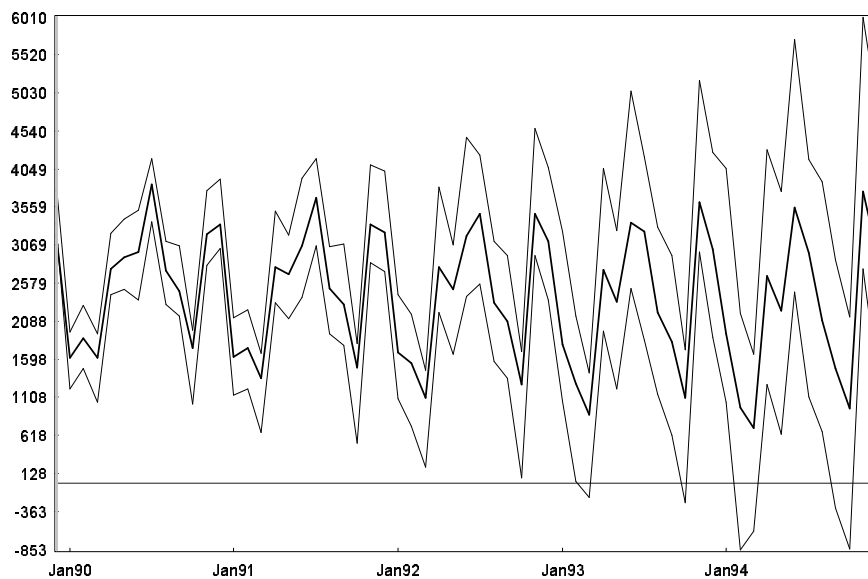


Рис. 7: FORT120: Рекуррентный прогноз на 60 точек и бутстреп доверительные интервалы

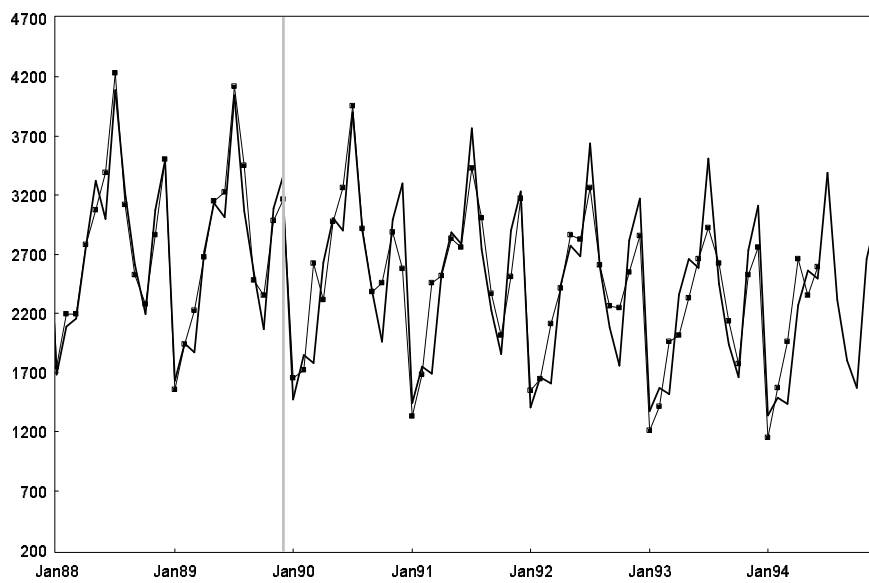


Рис. 8: FORT120: Рекуррентный прогноз по исправленной минимальной ЛРФ, построенной по ряду FORT174, на 60 точек вперед на фоне реальных значений