

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра Автоматизированных систем управления

Расчетно-графическая работа

По дисциплине: Теория случайных процессов

Выполнил:

Проверил:

Новосибирск 20__

1. Цель работы

Изучить методологию анализа временного ряда. Построить модели на основе авторегрессии, скользящего среднего или экспоненциального сглаживания. Сделать прогноз по имеющимся данным и оценить качество полученной модели.

2. Описание исходных данных

В качестве исходных данных возьмем статистику выбросов загрязняющих атмосферу веществ в тоннах с 2001 года по 2014. Также имеются данные с 2015 по 2017 гг., их будем использовать для определения качества модели и прогноза.

| 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 33291 | 33930 | 34652 | 35751 | 35835 | 35510 | 35532 | 33952 | 32754 |
| 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | |
| 32353 | 32628 | 32469 | 32063 | 31228 | 31269 | 31617 | 32082 | |

Описание данных: наличие пропущенных значений, примечания.

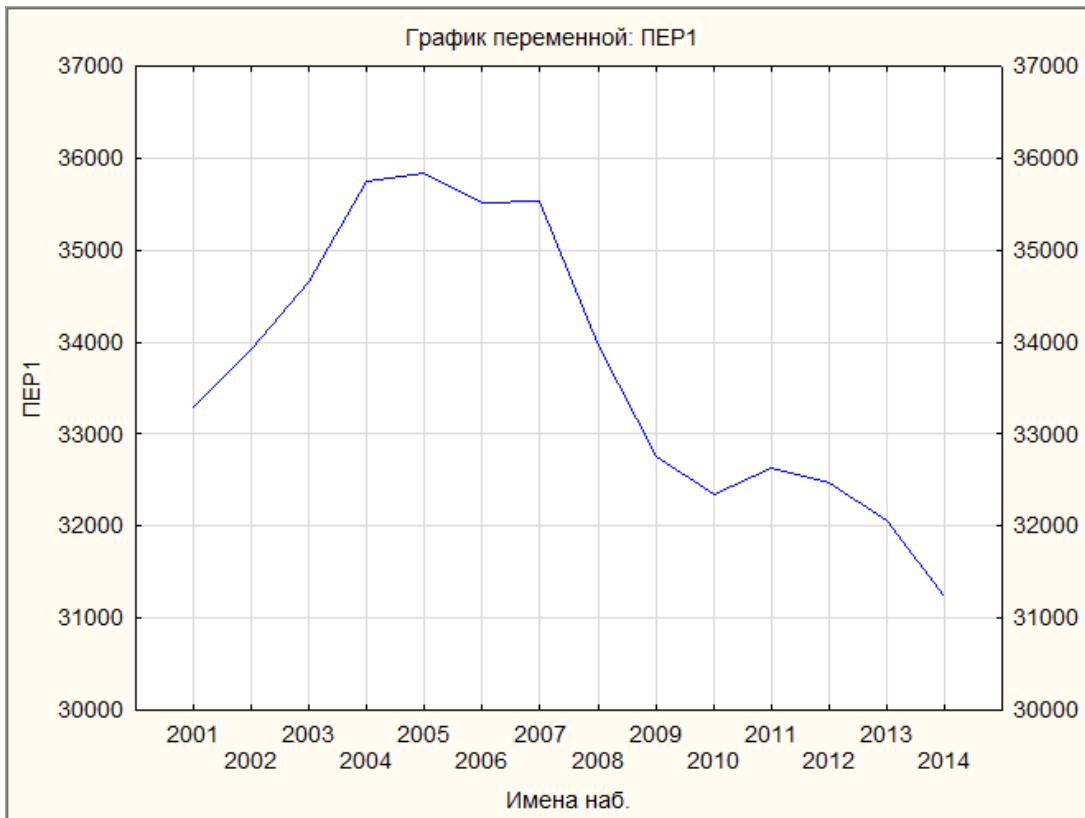


Рис 1. График исходных данных (2001 – 2014 гг.)

Визуальный анализ: необходимо определить компоненты ряда (тренд, сезонность), наличие выбросов, постоянство дисперсии. Если тенденции нет, то рассчитать среднее.

Графики автокорреляционной функции и частной автокорреляционной функции (проверка гипотез о наличии / отсутствии тренда, сезонной составляющей).

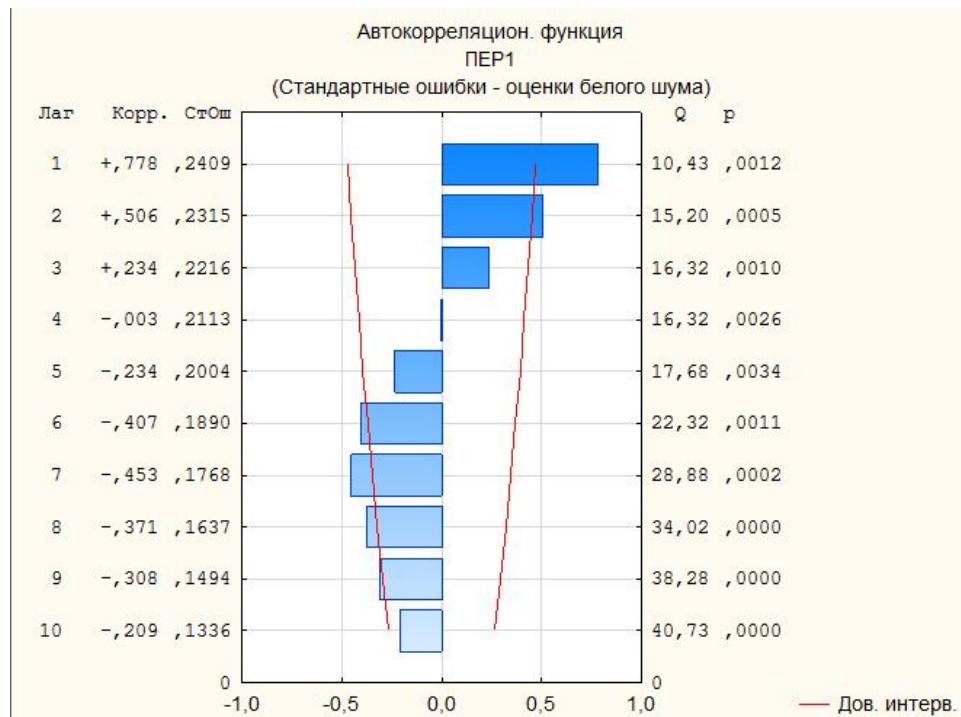


Рис 2. Автокорреляционная функция

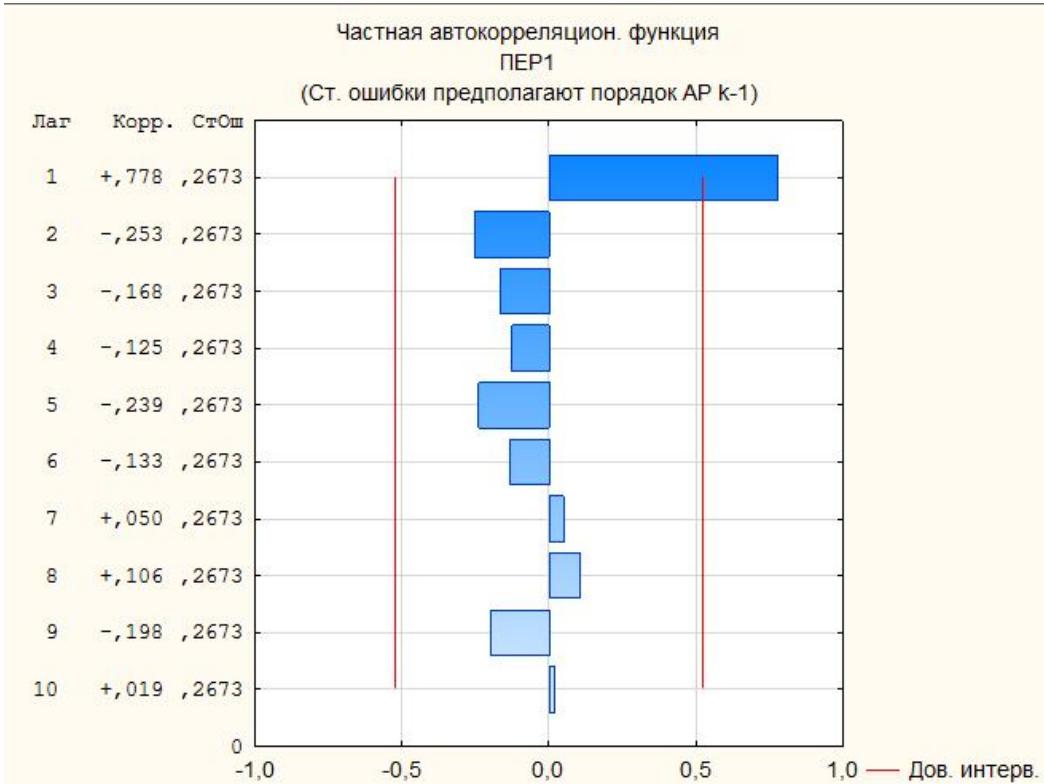


Рис 3. Частная автокорреляционная функция

В зависимости от результатов визуального анализа и вида автокорреляционной функции выбирается модель ряда.

3. Анализ данных и прогноз методом экспоненциального сглаживания.

Для сглаживания исходного ряда будем использовать методы сглаживания с трендом.

- **Экспоненциальное сглаживание с линейным трендом.**

Построим графики сглаженной функции, исходных данных и остатков.

Модели с разными параметрами: описание параметров, графики, анализ результатов.

Использовались коэффициенты сглаживания $\alpha = 0,9$ и $\gamma = 0,9$.

Синий – исходный график, красный – гладженный, зеленый – остатки (разница между исходным графиком и сглаженным)

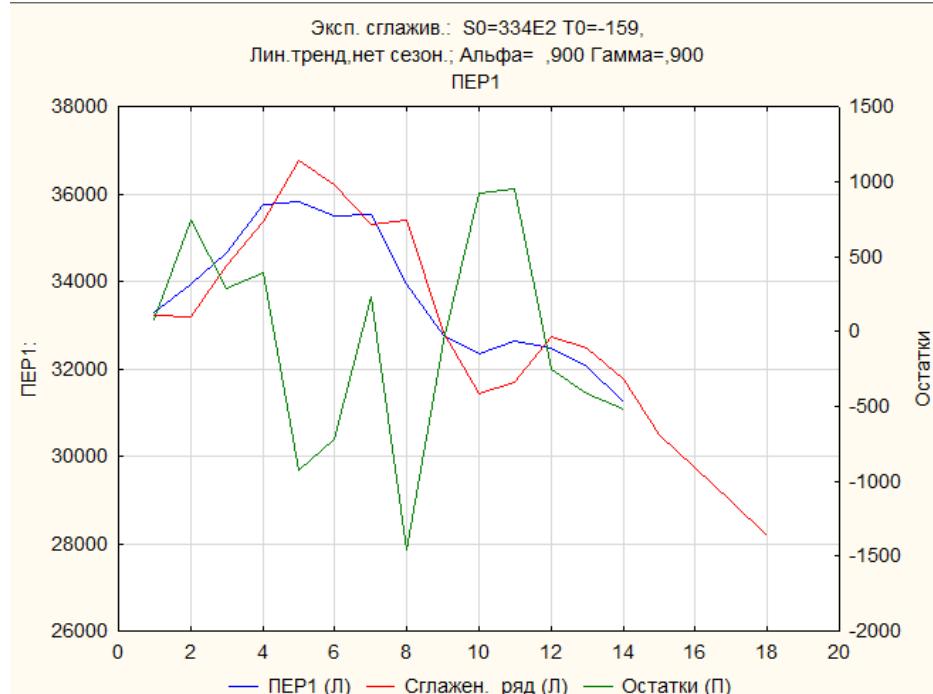


Рис 4. Экспоненциальное сглаживание с линейным трендом.

- **Экспоненциальное сглаживание с экспоненциальным трендом**

Использовались коэффициенты $\alpha = 0,9$ и $\gamma = 0,9$

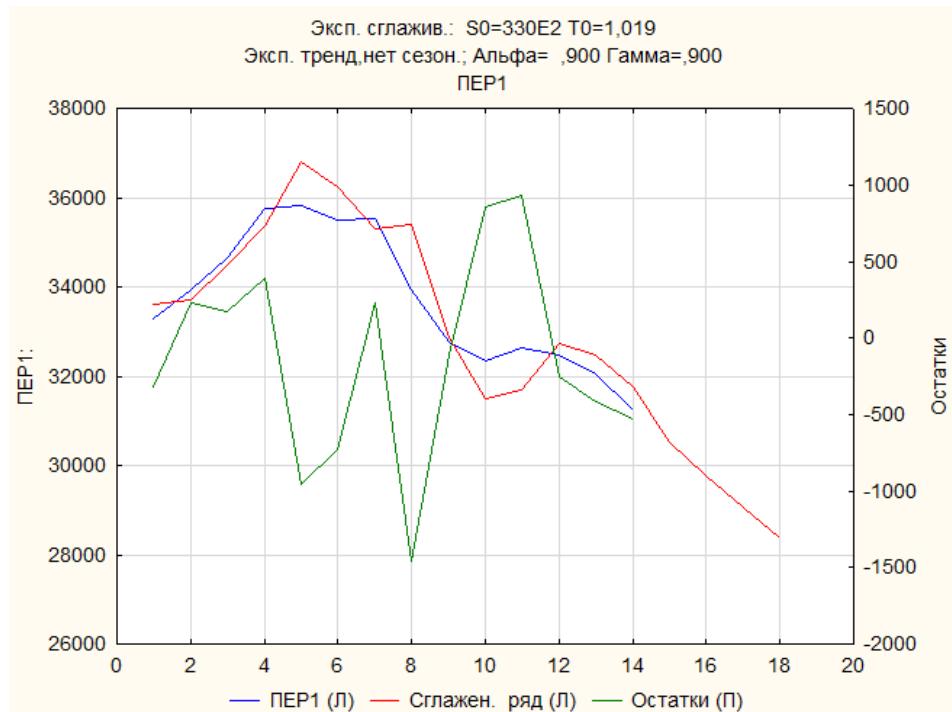


Рис 6. Экспоненциальное сглаживание с экспоненциальным трендом ($\alpha = 0,9$ и $\gamma = 0,9$)

Для лучшей модели (или для всех) – оценка точности, прогноз.

Приведем таблицы с данными сглаживания и ошибками.

| Итоговая ошибка | Эксп. сглажив.: S0=330E2 T0=1,019 (Таблица) Эксп. тренд,нет сезон.: Альфа= ,900 Гамма=,900 ПЕР1 | |
|--------------------------|---|--|
| | Ошибка | |
| Средн. ош. | -140,41810235 | |
| Ср. абсолют. ошибка | 540,70500439 | |
| Суммы квадратов | 6082256,68100924 | |
| Ср. квадрат | 434446,90578637 | |
| Средн. относ. ошибка | -0,40591637 | |
| Ср. абсолют. отн. ошибка | 1,60467493 | |

Рис 7. Итоговые ошибки сглаживания ряда.

В данной таблице:

- средняя абсолютная ошибка (MAE) равна 540, 7;
- средняя абсолютная ошибка в процентах (MAPE) – 1,6%;
- среднеквадратичная ошибка (MSE) – 434446,9.

| Набл. | Эксп. сглажив.: S0=330E2 T0=1,019 (Таблица) Эксп. тренд, нет сезон.; Альфа= ,900 Гамма=,900 ПЕР1 | | | |
|-------|---|-----------------|----------|--|
| | ПЕР1 | Сглажен. ряд | Остатки | |
| 1 | 33291,00 | 33608,98 | -317,98 | |
| 2 | 33930,00 | 33702,14 | 227,86 | |
| 3 | 34652,00 | 34481,01 | 170,99 | |
| 4 | 35751,00 | 35362,49 | 388,51 | |
| 5 | 35835,00 | 36786,85 | -951,85 | |
| 6 | 35510,00 | 36235,74 | -725,74 | |
| 7 | 35532,00 | 35303,01 | 228,99 | |
| 8 | 33952,00 | 35415,21 | -1463,21 | |
| 9 | 32754,00 | 32870,05 | -116,05 | |
| 10 | 32353,00 | 31495,01 | 857,99 | |
| 11 | 32628,00 | 31700,33 | 927,67 | |
| 12 | 32469,00 | 32721,31 | -252,31 | |
| 13 | 32063,00 | 32475,96 | -412,96 | |
| 14 | 31228,00 | 31755,76 | -527,76 | |
| 15 | | 30524,66 | | |
| 16 | | 29786,82 | | |
| 17 | | 29066,82 | | |
| 18 | | 28364,22 | | |

Рис 8. Данные сглаженного ряда и остатки.

На рисунке 8 представлен прогноз данных на 4 года вперед (15 – 18 наблюдения). Исходная таблица данных содержит данные, которые мы не использовали в анализе. Сделаем сравнение имеющихся данных и прогноза, чтобы оценить качество выбранной модели.

Подсчитаем ошибки за период (3 года), который не учитывался в анализе, по следующим данным:

| Исходные данные (не учитывали в анализе, период 2015-2017гг.) | Прогноз, полученный программой | Остатки |
|---|--------------------------------|---------|
| 31269 | 30524,66 | 744,34 |
| 31617 | 29786,82 | 1830,18 |
| 32082 | 29066,82 | 3015,18 |

- MAPE = 5,855795414%
- MAE = 1863,232985
- MSE = 4331637,623

Наиболее показательной является ошибка MAPE. Она показывает величину ошибки прогноза. Точность прогноза определяется как 100% - MAPE.

Величина MAPE подсчитанная программой получилась 1,6 %. Имея данные, не использованные в прогнозе, мы получили MAPE, равную 5,855795414%. Значит, точность прогноза составляет не 98,396%, а несколько ниже – 94,1443%. Будем считать, что такая точность прогноза является хорошей, и предскажем значения на 2018 и 2019 год. Получим следующие данные о выбросах загрязняющих веществ в атмосферу: 2018 год - 28364,2185 тонн, 2019 год - 27678,6018 тонн.

4. Анализ данных и прогноз методом авторегрессии.

Модели ARMA с разными параметрами: анализ, коррелограммы, прогноз, оценка точности.

Подберем коэффициенты р для построения модели и сделаем прогноз на три года вперед.

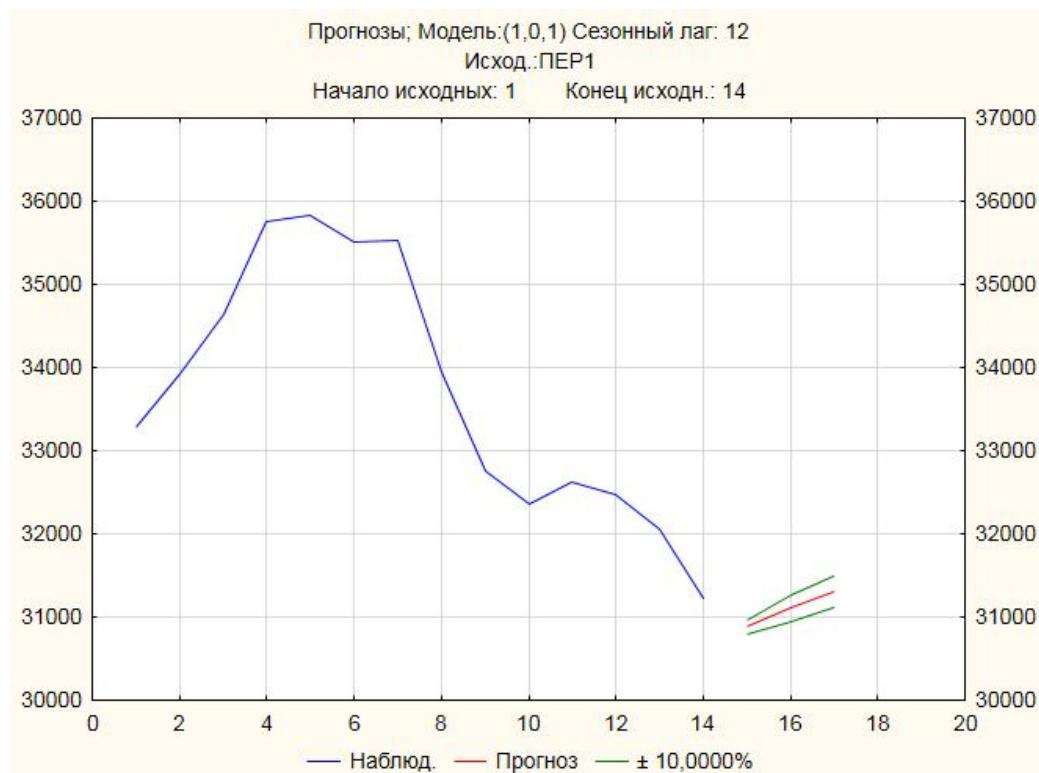


Рис 9. График исходных данных (синего цвета), прогноз (красного цвета) и погрешность прогноза (зеленого цвета).

Приведем таблицу данных, содержащую прогноз и погрешность прогноза.

| Набл. N | Прогнозы; Модель:(1,0,1) Сезонный лаг: 12 (Таблица.sta) | | | | |
|---------|---|----------------------|-----------------------|-----------|--|
| | Исход.:ПЕР1 | | | | |
| | Начало исходных: 1 Конец исходн.: 14 | | | | |
| Набл. N | Прогноз | Нижний ± 10,0000% | Верхний ± 10,0000% | Ст. ошиб. | |
| 15 | 30881,88 | 30796,13 | 30967,63 | 666,812 | |
| 16 | 31107,31 | 30952,38 | 31262,25 | 1204,818 | |
| 17 | 31308,37 | 31115,37 | 31501,37 | 1500,853 | |

Рис 10. Таблица данных со значениями прогноза и погрешности прогноза.

Построим автокорреляционную функцию, чтобы проверить адекватность модели.

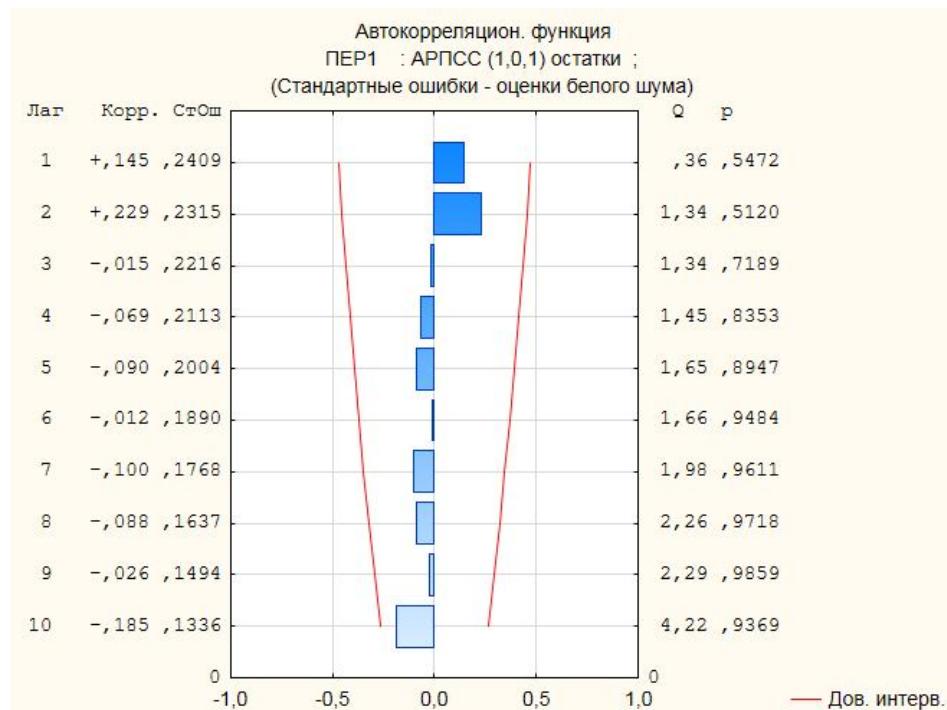


Рис 11. Автокорреляционная функция остатков.

Как видно из графика, остатки распределены хаотично и не выходят за пределы доверительного интервала. Значит, полученная модель адекватна.

Подсчитаем ошибки за период (3 года), который не учитывался в анализе, по следующим данным:

| Исходные данные (не учитывали в анализе, период 2015-2017гг.) | Прогноз, полученный программой | Остатки |
|---|--------------------------------|---------|
| 31269 | 30881,88 | 387,12 |
| 31617 | 31107,31 | 509,69 |
| 32082 | 31308,37 | 773,63 |

- MAPE = 1,753841%
- MAE = 556,8133
- MSE = 336049,7

Величина MAPE подсчитанная программой получилась 1,6 %. Имея данные, не использованные в прогнозе, мы получили MAPE, равную 1,753841%. Значит, точность прогноза составляет не 98,396%, а немного ниже – 98,246159%. Будем считать, что такая точность прогноза является очень хорошей, и предскажем значение на 2018 год, которое будет равно 31847 тонн.

5. Вывод

Сравнительный анализ моделей, описание результатов.