Гауссовские процессы для прогнозирования временных рядов.

Семёновых Кирилл ФРКТ

Структура:

- регрессия на основе гауссовских процессов
- приближённый Байесовский вывод
- модель ARIMA прогнозирования временных рядов
- прогнозирование временного ряда

Гауссовский процесс.

- Гауссовский процесс — случайный процесс, для которого распределение любого конечномерного среза является случайным вектором с многомерным нормальным распределением.
- Для того, что бы полностью задать гауссовский процесс, нужно задать его среднее значение и ковариационную функцию.

Задача регрессии:

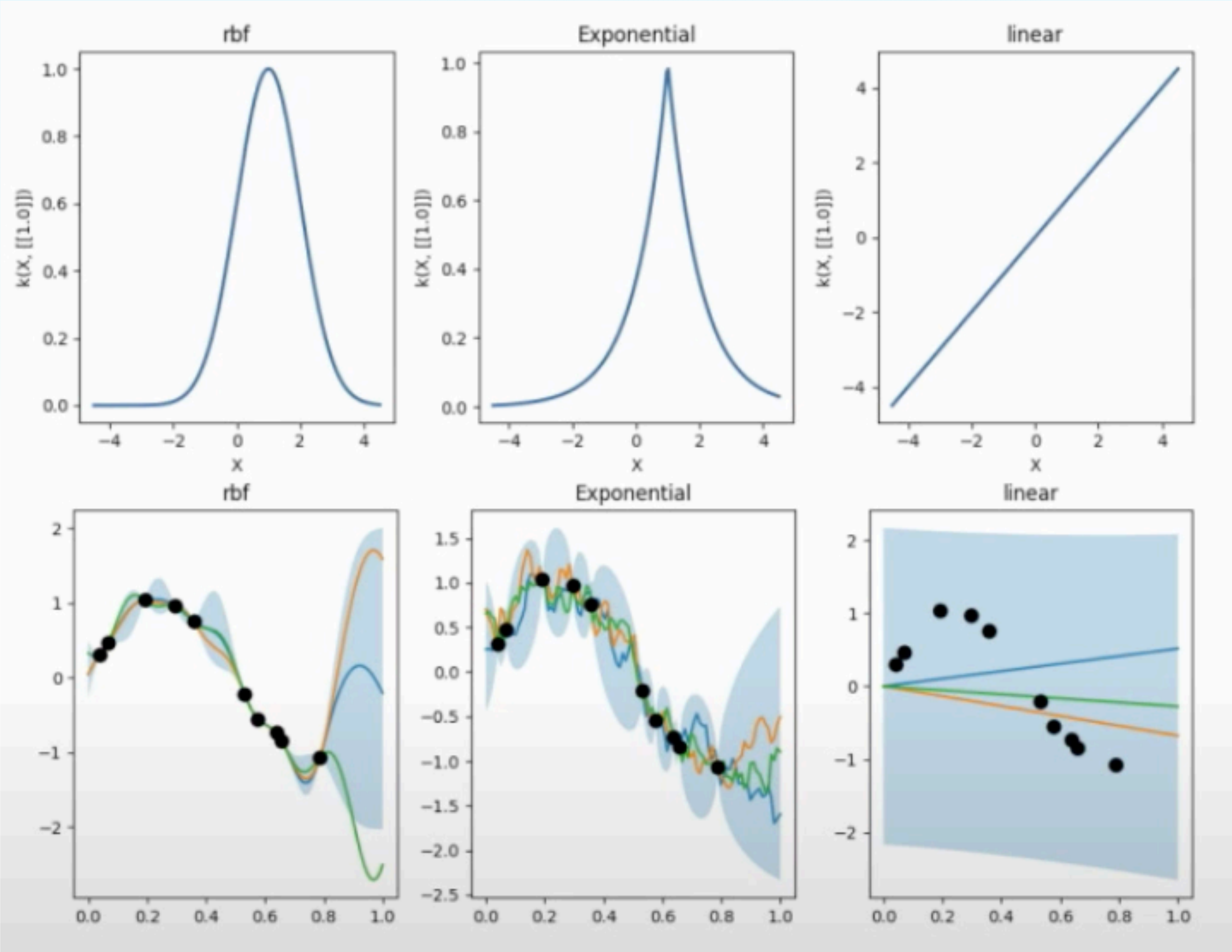
- есть набор данных D = (X, y)
- Мы хотим построить регрессионную модель f(x) неизвестной функции y(x), такую что f(x) ≈ y(x) и получить оценку неопределенности модели σ^2(x) для прогноза y(x).
- Для оценки качества аппроксимации используется независимая тестовая выборка D_t = (X_t , y_t).

Регрессия на основе гауссовских процессов:

- дана выборка D = (X, y)
- Пусть y(x) - реализация гауссовского процесса, среднее значение гауссовского процесса равно нулю, ковариационная функция имеет вид: cov(y(x), y(x’)) = k(x, x’)
- Тогда: $$\mathbf{y} \propto \mathcal{N}(\mathbf{0}, K),$$ $$K = \{k(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^n$$

Ковариационные функции:

- RBF (Radial Basis Function)
- Exponential
- Linear



Регрессия на основе гауссовских процессов.

- Среднее и дисперсия равны:

(x* - новая точка)

k(x*) = [k(x*, x1), … , k(x*, xn)]

$$\begin{aligned} \mu(\mathbf{x}^*) &= \mathbf{k}(\mathbf{x}^*)^T \, \mathbf{K}^{-1} \, \mathbf{y} \\ \sigma^2(\mathbf{x}^*) &= k(\mathbf{x}^*) \; - \; \mathbf{k}(\mathbf{x}^*)^T \, K^{-1} \, \mathbf{k}(\mathbf{x}^*) \end{aligned}$$

Приближённый Байесовский вывод.

- Байесовский вывод — статистический вывод, в котором свидетельство и/или наблюдение используются, чтобы обновить или вновь вывести вероятность того, что гипотеза может быть верной.
- Теорема Байеса:
  (H - гипотеза, E - свидетельство)

$$P(H|E) = \frac{P(E|H)\ P(H)}{P(E)}$$

- прогнозируемое распределение:

$$p(f(\mathbf{x}^*)|\mu_{x^*}, \Sigma_{x^*}) = \int p(f(\mathbf{x}^*)|\mathbf{x}^*, \mathcal{D})p(\mathbf{x}^*)d\mathbf{x}^*$$

- Подход Монте-Карло:

$$p(f(\mathbf{x}^*)|\mu_{x^*}, \mathbf{\Sigma}_{x^*}) = \int p(f(\mathbf{x}^*)|\mathbf{x}^*)p(\mathbf{x}^*)d\mathbf{x}^* \simeq \frac{1}{T}\sum_{t=1}^{T} p(f(\mathbf{x}^*)|\mathbf{x}^{*t})$$

Модель ARIMA.
(autoregressive integrated moving average)

- Модель ARIMA(p, d, q) для нестационарного временного ряда Xt имеет вид:

$$\triangle^d X_t = c + \sum_{i=1}^{p} a_i \triangle^d X_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t$$

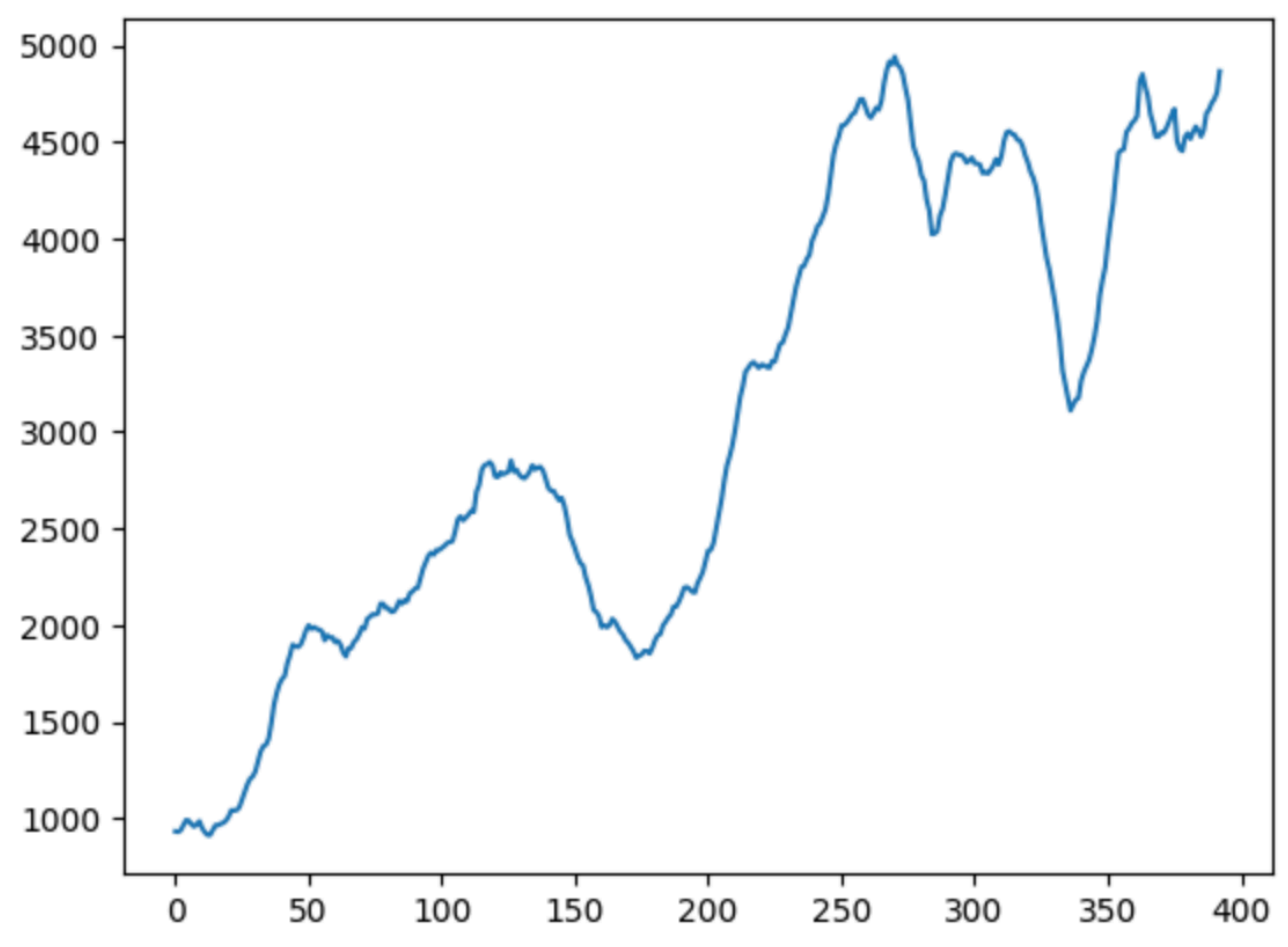
- стационарный временной ряд - ε_t
- параметры модели - c, a_i, b_j
- оператор разности временного ряда порядка d - $\triangle^d$

Прогнозирование временного ряда.

- дана некоторая зависимость, состоящая из 393 точек:

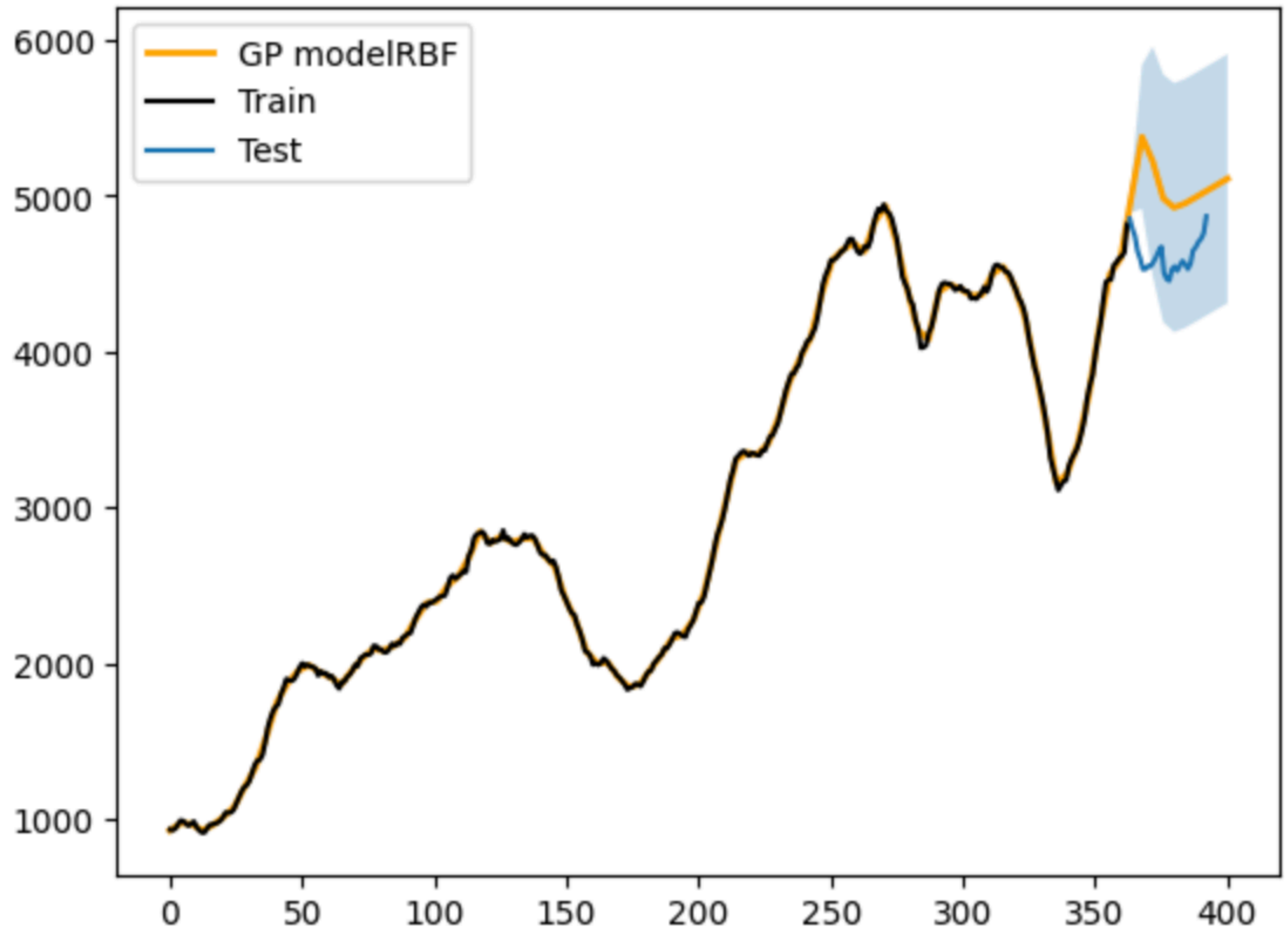
1. возьмём 363 точки для обучения моделей

2. возьмём 325 точек для обучения моделей

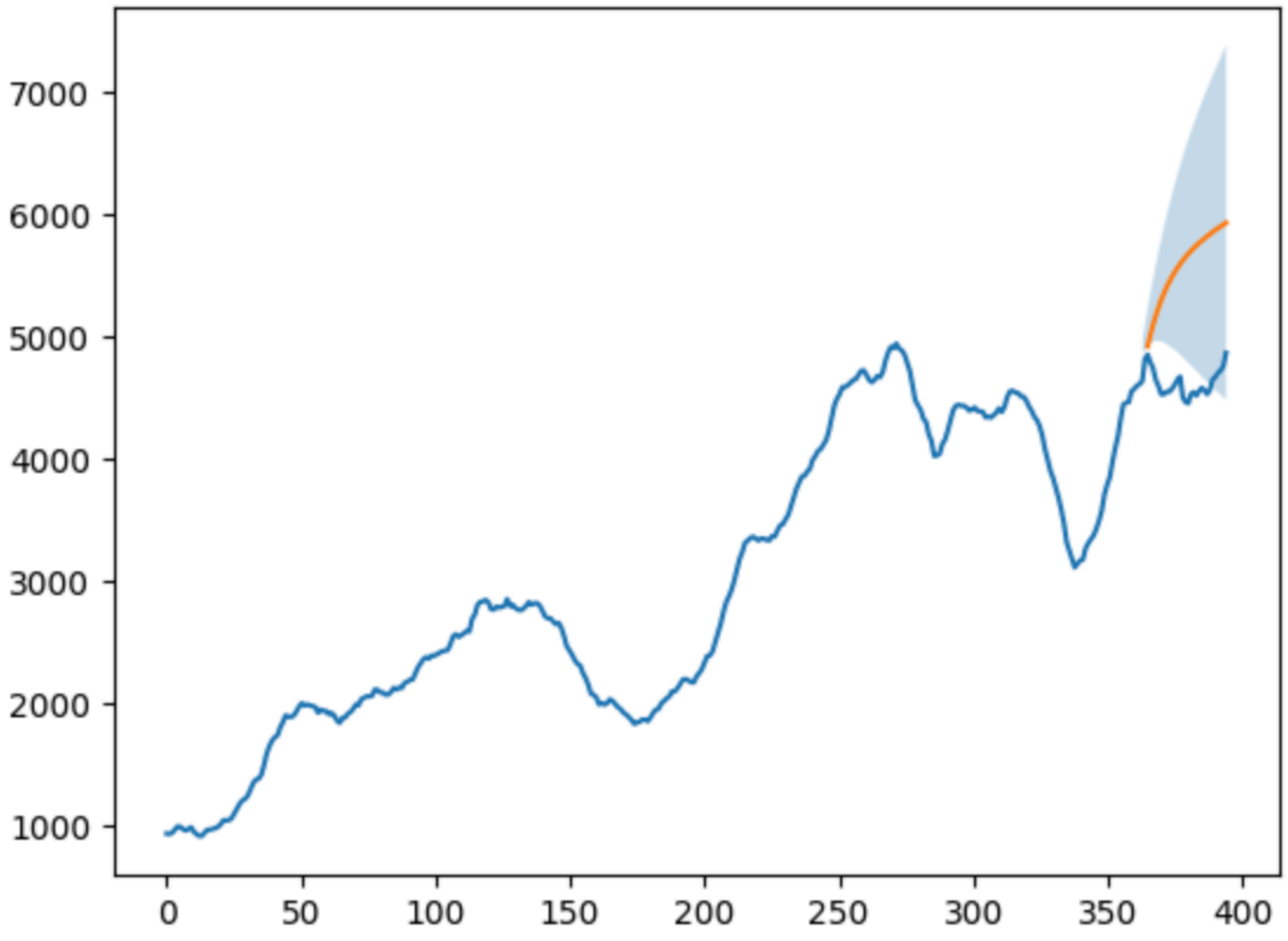


Прогнозирование временного ряда.

GPR (kern == RBF)

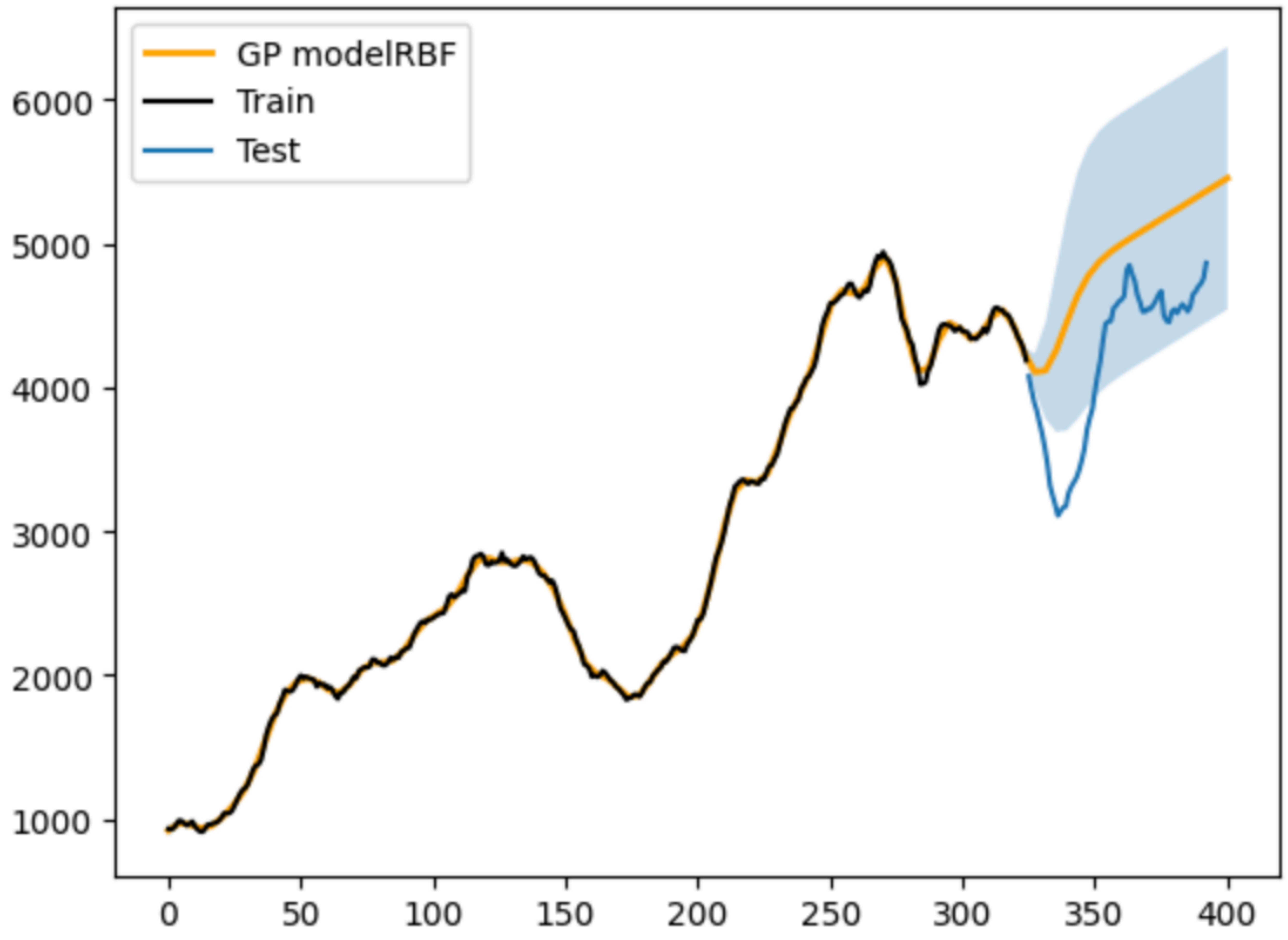


ARIMA(1, 1, 1)

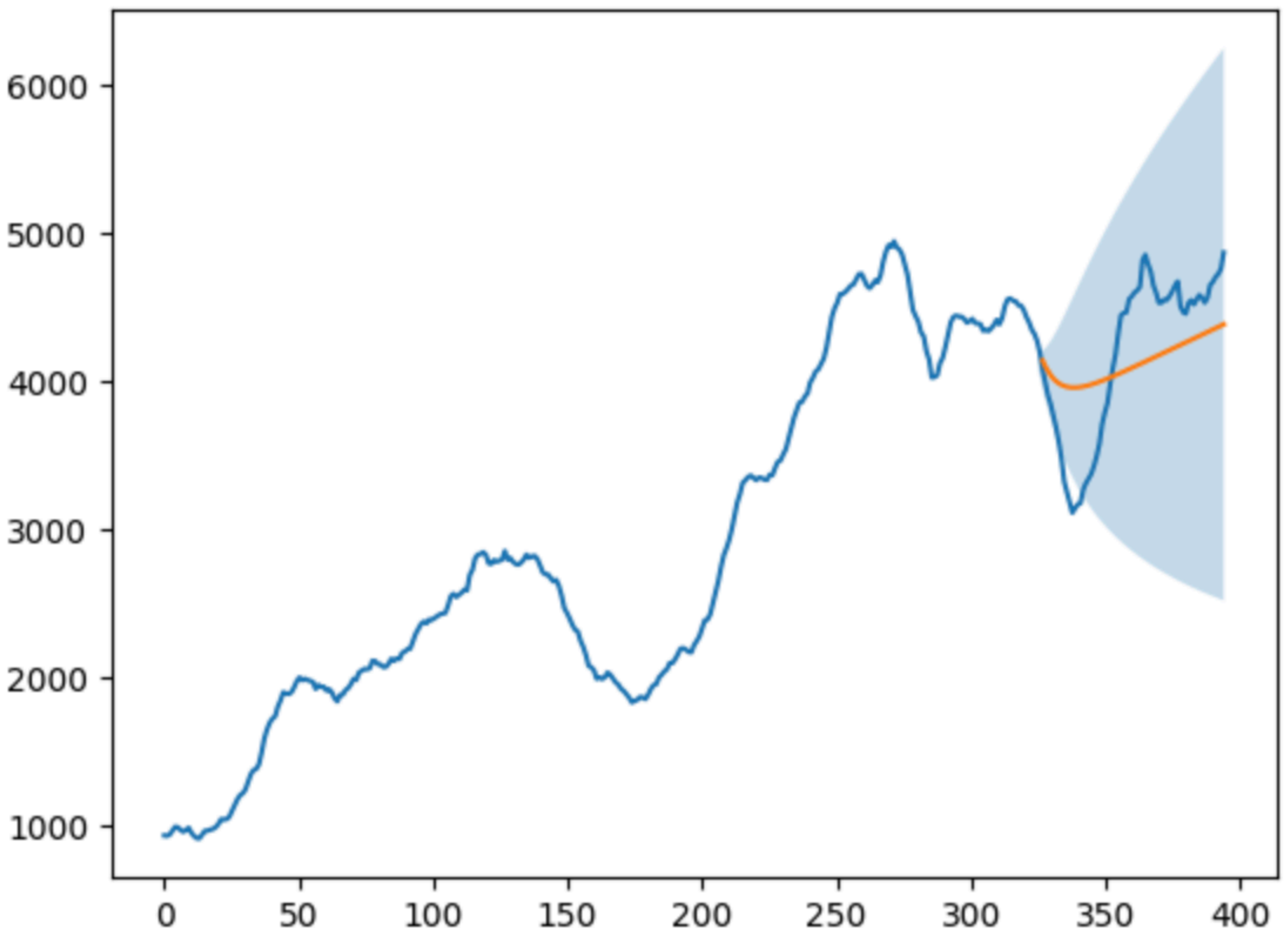


Прогнозирование временного ряда.

GPR (kern == RBF)



ARIMA(4, 1, 0)



Спасибо за внимание!