

Прогнозирование временного ряда при помощи авторегрессионных моделей

Цели темы

- Понять, что такое авторегрессионные модели
- Рассмотреть модели:
 - экспоненциального сглаживания
 - ARIMA
 - TBATS
 - Prophet

Прогнозирование временного ряда при помощи авторегрессионных моделей

Авторегрессионная модель

Прогноз на основе только предыдущих значений.

Экспоненциальное сглаживание

Метод, который прогнозирует на основе предыдущих значений.

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha) \hat{y}_{t|t-1}$$

Где $\hat{y}_{t+1|t}$ – прогноз y_{t+h} в момент времени t

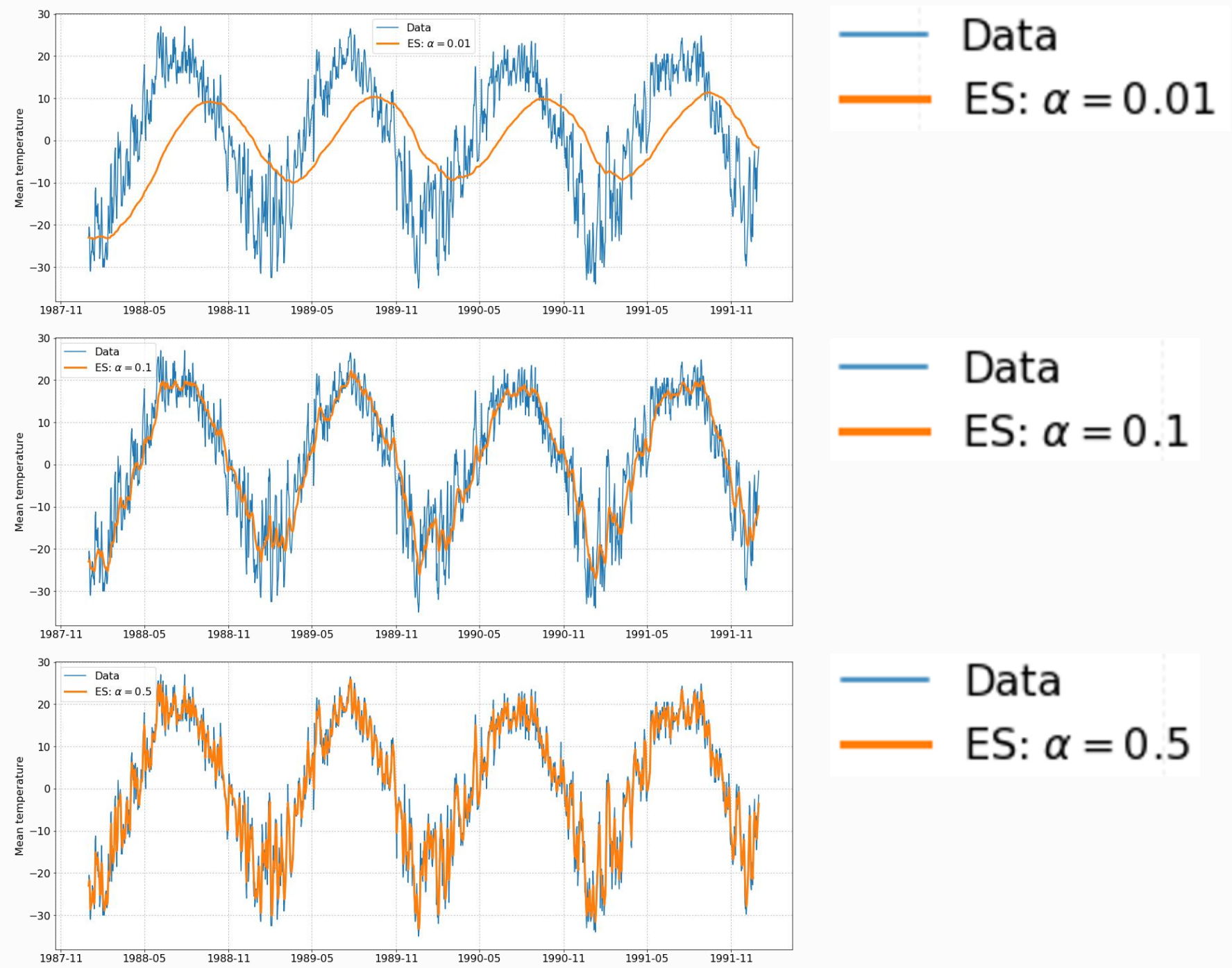
α – параметр сглаживания

Рекурсивная формула — то есть задаёт значение исходя из значения предыдущей.

Свойства модели

- $\alpha \approx 1 \rightarrow$ большой вес последним точкам: $\hat{y}_{T+1|T} \approx y_T$
 - $\alpha \approx 0 \rightarrow$ большее сглаживание: $\hat{y}_{T+1|T} \approx \bar{y}$
 - Оптимальное $\alpha^* : \sum_{t=t_0}^T (\hat{y}_t(\alpha) - y_t)^2 \rightarrow \min_{\alpha}$
-
- если $\alpha^* \in (0, 0.3)$ то ряд стационарен, можно применять экспоненциальное сглаживание
 - если $\alpha^* \in (0.3, 1)$ то ряд нестационарен, нужно применять модель тренда

Пример прогноза для разных α



Модель ARIMA

Модель строится итеративно и состоит из объединённого набора более простых моделей.

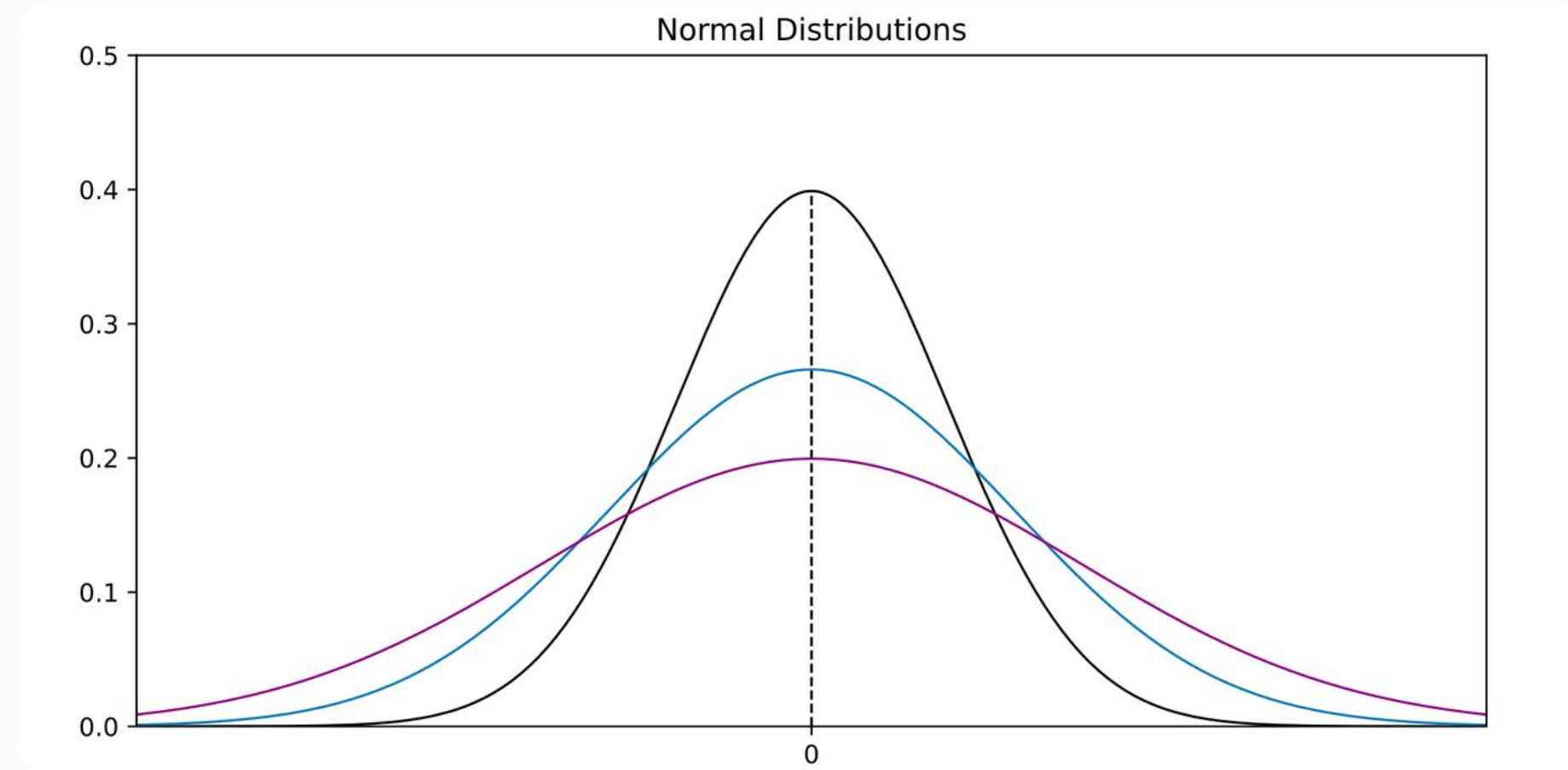
1. MA → AR → ARMA → ARIMA

2. SARIMA → ARIMAX → SARIMAX

Модель скользящего среднего MA(q)

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

где y_t — стационарный ряд со средним μ
 ε_t — гауссовский белый шум, т.е. $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ и независимы



Модель авторегрессии AR(p)

$$y_t = \alpha + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

где y_t – стационарный ряд,
 ε_t – гауссовский белый шум, т.е. $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ и независимы

Это модель линейной регрессии с p лагами, для которой:

- отклик: y_t — значения ряда в момент времени t
- признаки: y_{t-1}, \dots, y_{t-p} — значения ряда в предыдущий момент времени

Пример как выглядит модель для $p=1$:

$$y_t = \alpha + \varphi y_{t-1} + \varepsilon_t, \text{ где } \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Модель ARMA

ARMA рассматривает только стационарные временные ряды.

Соединяете AR(p) и MA(q):

$$y_t = \alpha + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Модель ARIMA

Рассматривает не только стационарные временные ряды, но и интегрированные ряды путём добавления ещё одной компоненты.

Интегрированные ряды — те ряды, которые можно привести к стационарным путём дифференцирования 1 и более раз.

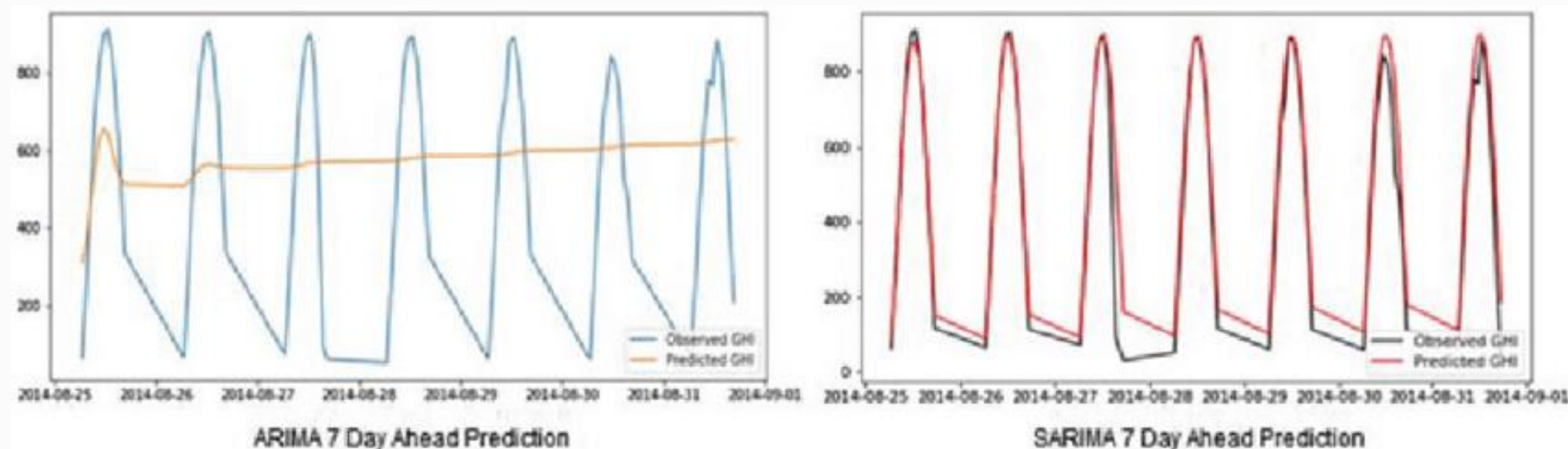
I в названии означает **integrated** — то есть интегрированность ряда.

Гиперпараметры модели — p, d, q .

p — от AR, d — MA, q — параметр, отвечающий за интегрирование.

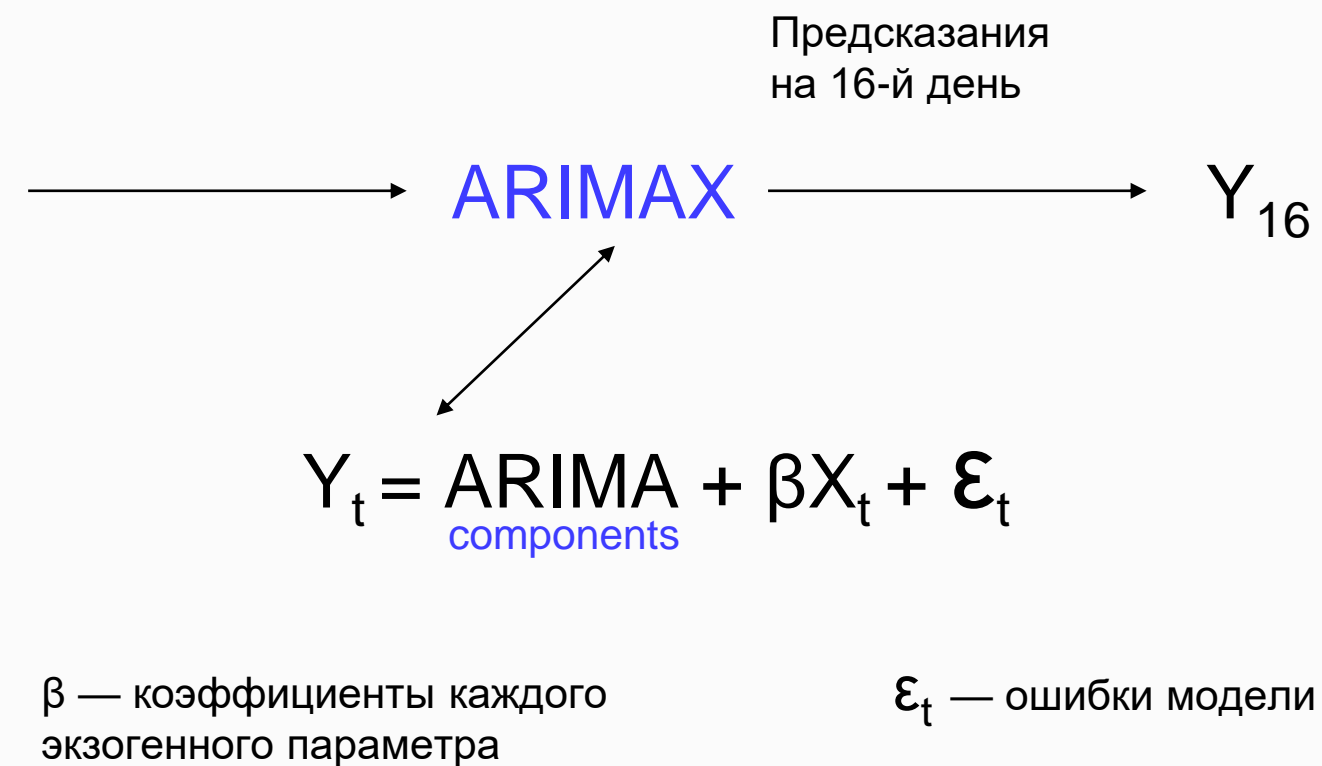
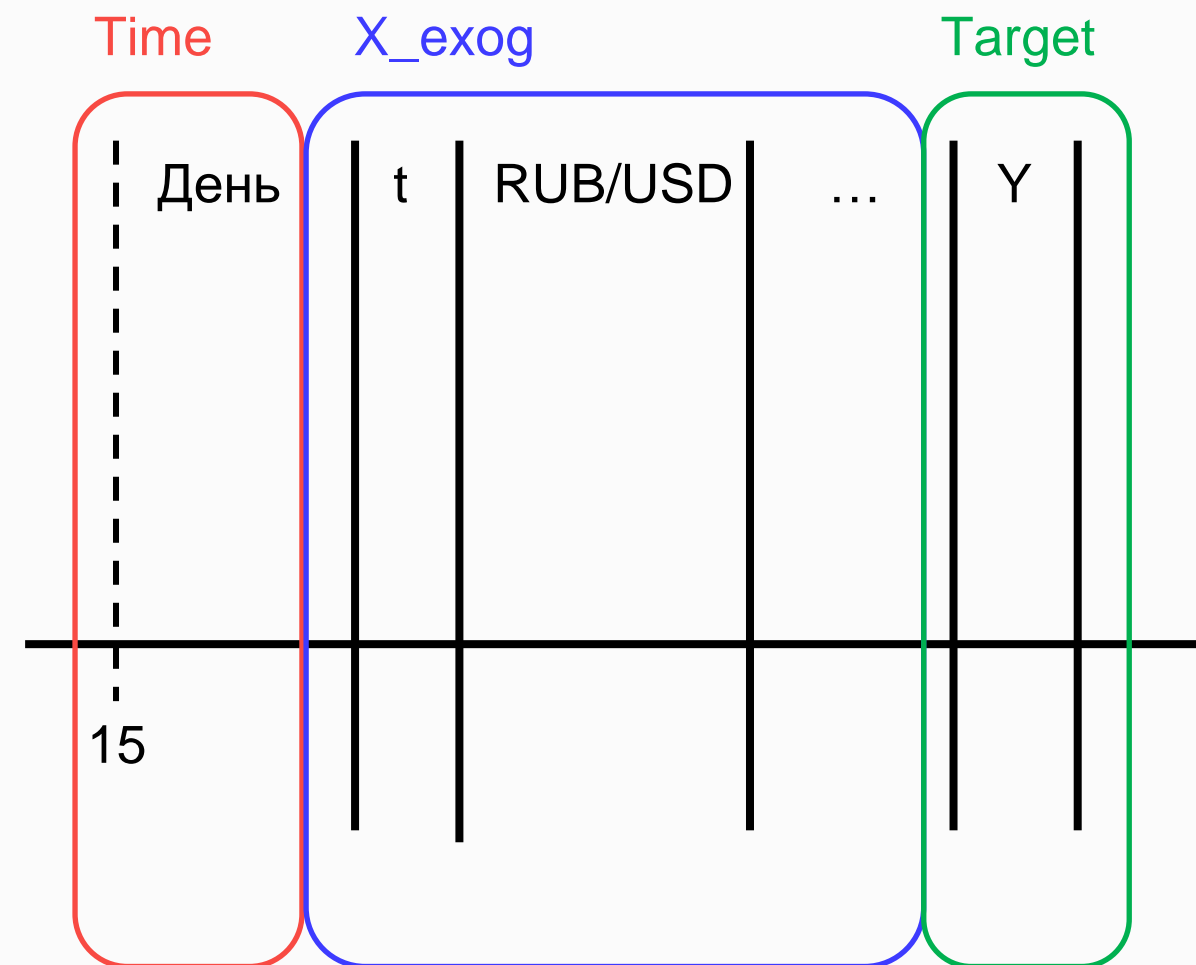
Модель SARIMA

ARIMA с добавлением сезонной компоненты.



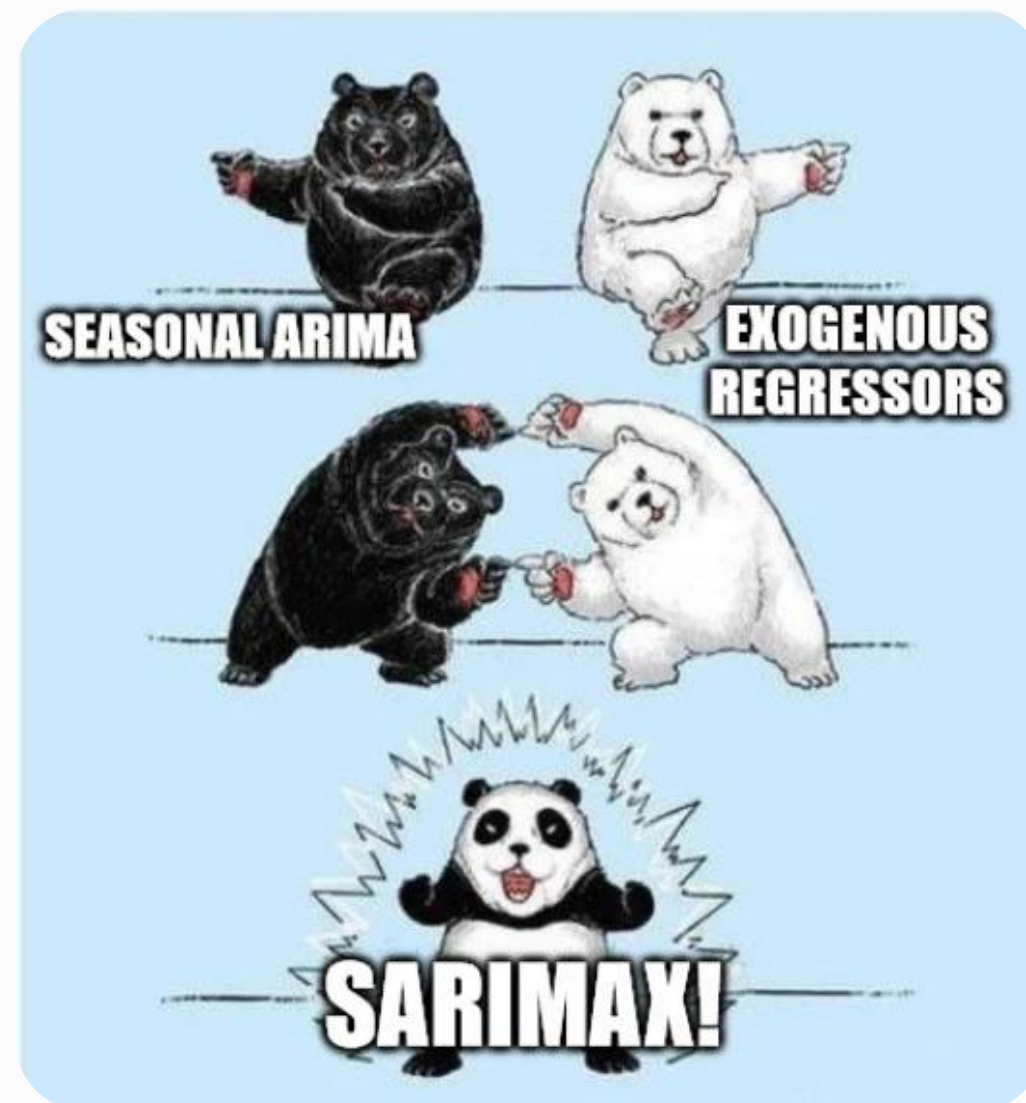
Модель ARIMAX

Компонента X отвечает за добавление других независимых переменных регрессоров к модели ARIMA.



Модель SARIMAX

В модель ARIMA добавлена сезонность и внешние регрессоры.



Модель TBATS

Ещё одно усложнение модели ARMA.

- Ряды Фурье (то есть тригонометрические функции) для прогнозирования сезонной компоненты
- TBATS — преобразование Бокса-Кокса

Сравнение SARIMAX и TBATS

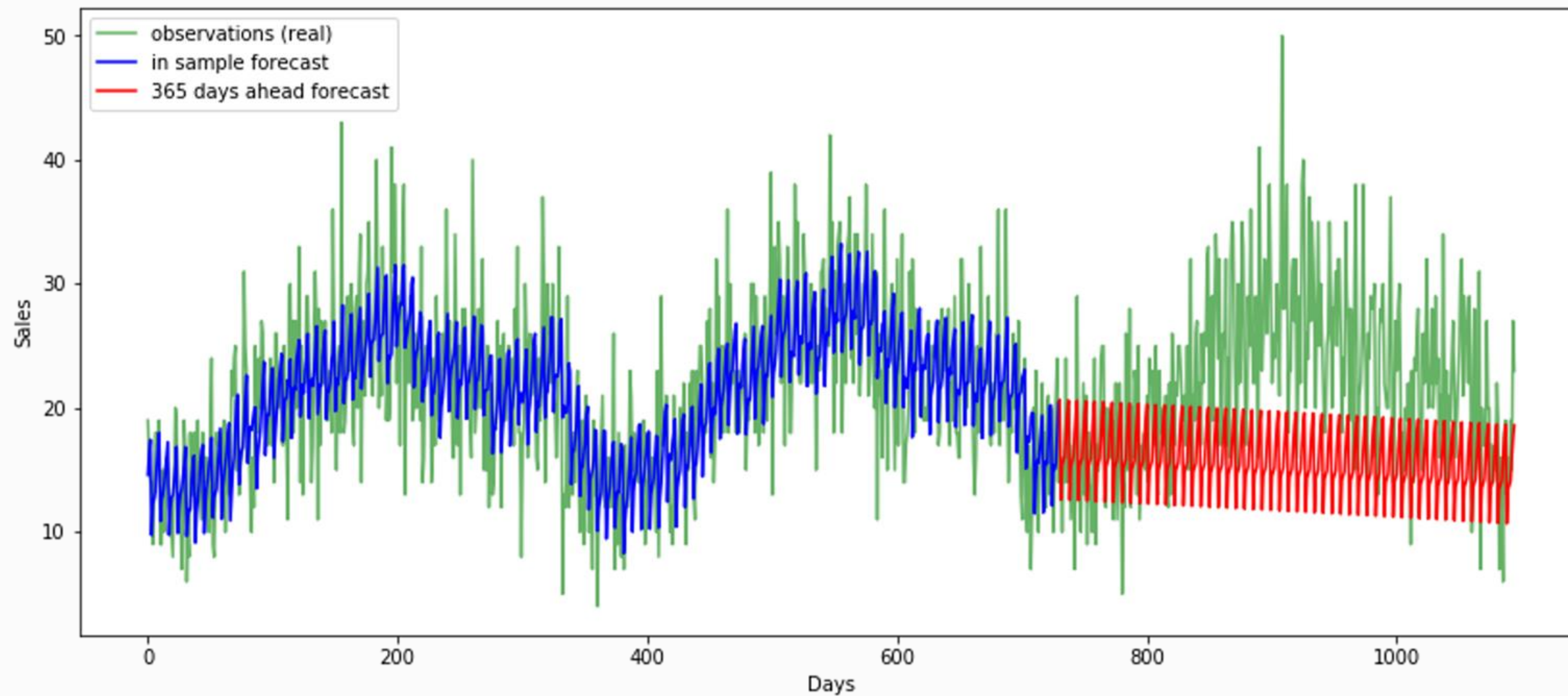
SARIMAX:

- для учёта сезонности использует сезонную авторегрессию (SAR) и сезонное скользящее среднее (SMA)
- для приведения ряда к стационарному использует дифференцирование
- позволяет тоньше настроить параметры

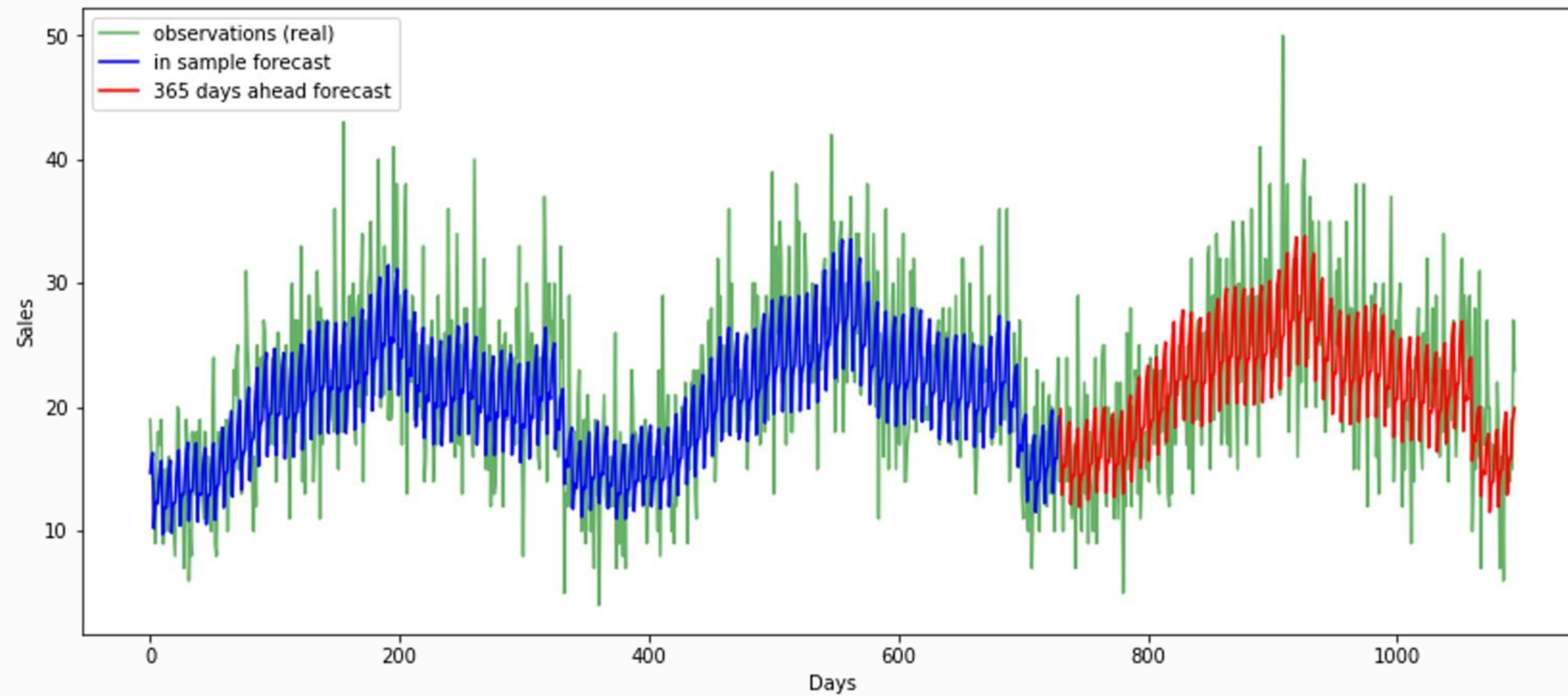
TBATS:

- для учёта сезонности использует ряды Фурье (то есть тригонометрические функции)
- для приведения ряда к стационарному использует преобразование Бокса-Кокса
- меньше параметров, и она проще в использовании

SARIMAX



TBATS



Модель Prophet

Модель от компании «Фейсбук»* (Forecasting at scale).

Принцип работы:

$$y_t = g_t + s_t + h_t + \varepsilon_t,$$

- s — сезонные компоненты. Моделируются рядами Фурье
- h — аномальные или нерегулярные дни: праздники, дни распродаж
- g — тренд
- ε — ошибки модели

Тренд в модели Prophet

Тренд в модели может быть двух видов:

- линейный

$$g_t = mt + b,$$

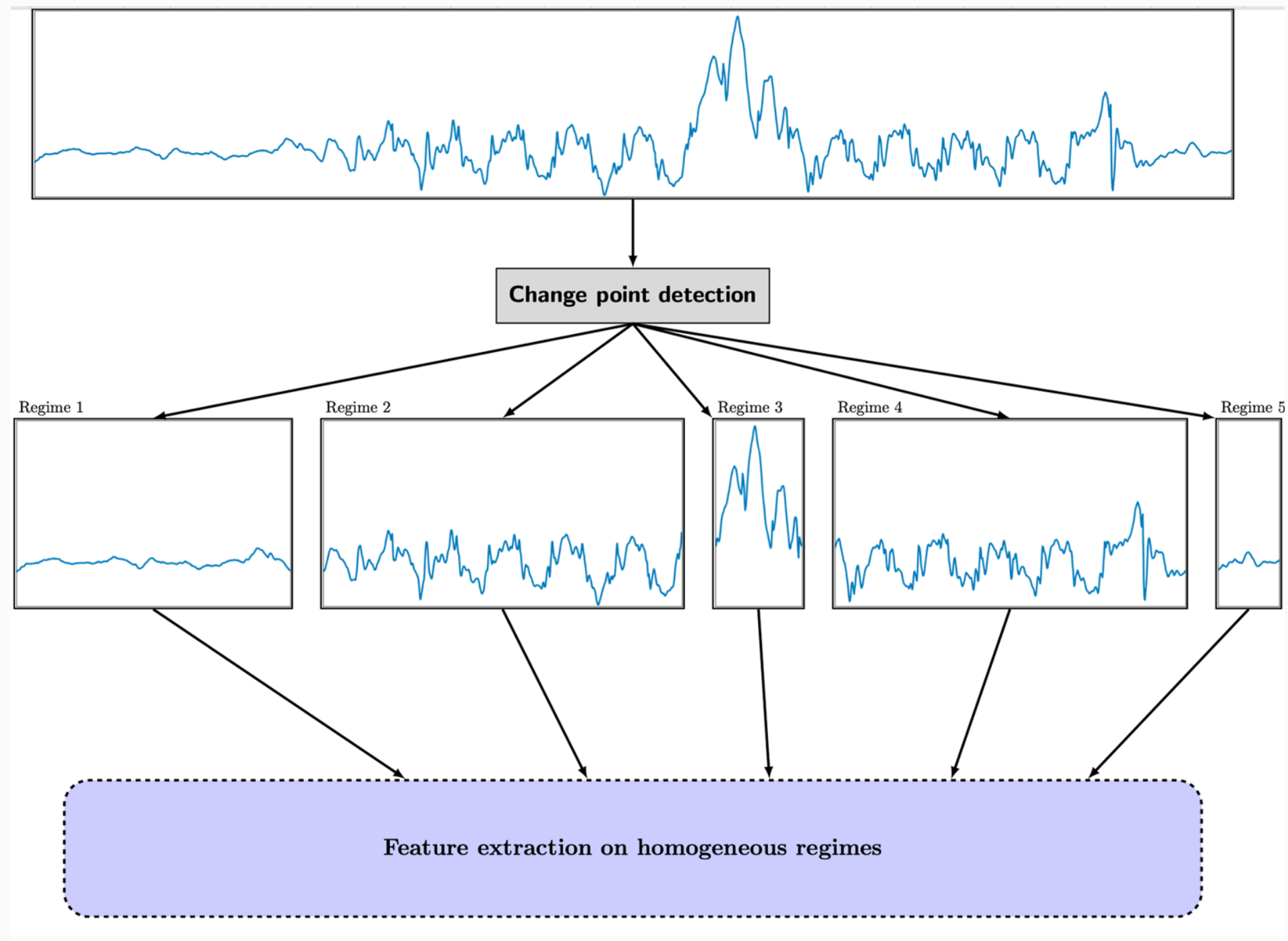
- логистический

$$g_t = \frac{C_t}{1 + e^{-kt}}.$$

Преимущества модели Prophet

1. Часто строит хорошие прогнозы по умолчанию
2. Учитывает несколько сезонностей
3. В сравнении с SARIMAX имеет более понятные гиперпараметры
4. Умеет по историческим данным выбирать оптимальные точки изменения тренда автоматическими методами

Работа с трендом в Prophet



Выводы темы

- Рассмотрели главный способ прогнозирования рядов — при помощи авторегрессионных моделей
- Рассмотрели устройство моделей:
 - экспоненциального сглаживания
 - ARIMA
 - TBATS
 - Prophet

Выводы модуля

- ✓ Задали важные определения и формулировки задачи прогнозирования временных рядов
- ✓ Узнали о способах декомпозиции ряда
- ✓ Познакомились с определением стационарности и способами приведения ряда к стационарному
- ✓ Разобрали классические способы прогнозирования рядов при помощи авторегрессионных моделей