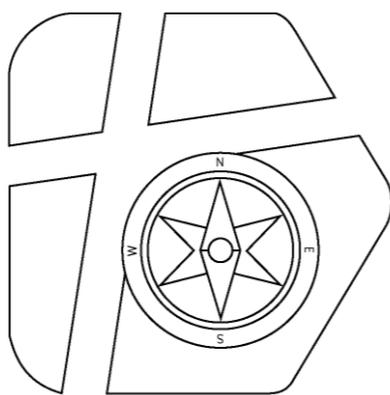


**Я**ндекс

карты



# Математическая модель для прогнозирования пробок

Михаил Хохлов  
Разработчик

# Содержание

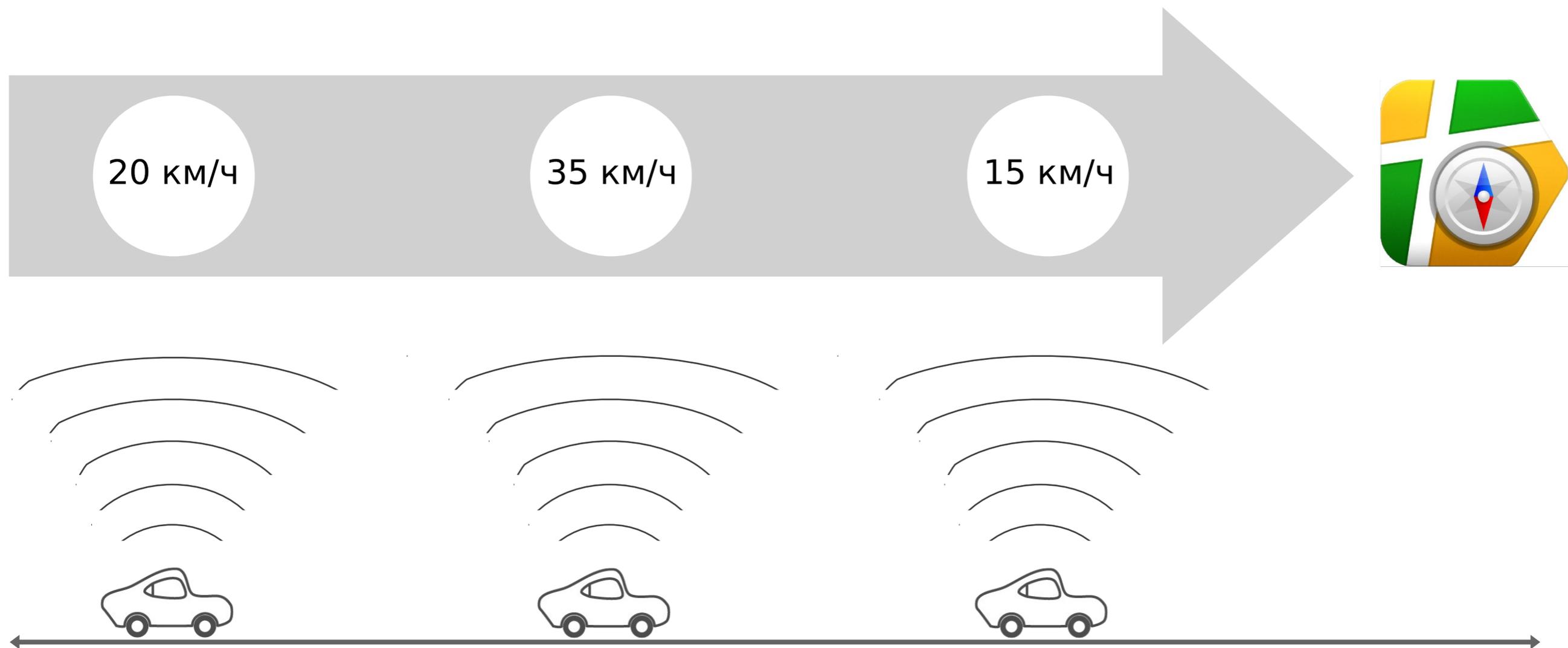
Входные данные

Постановка задачи

Методы решения

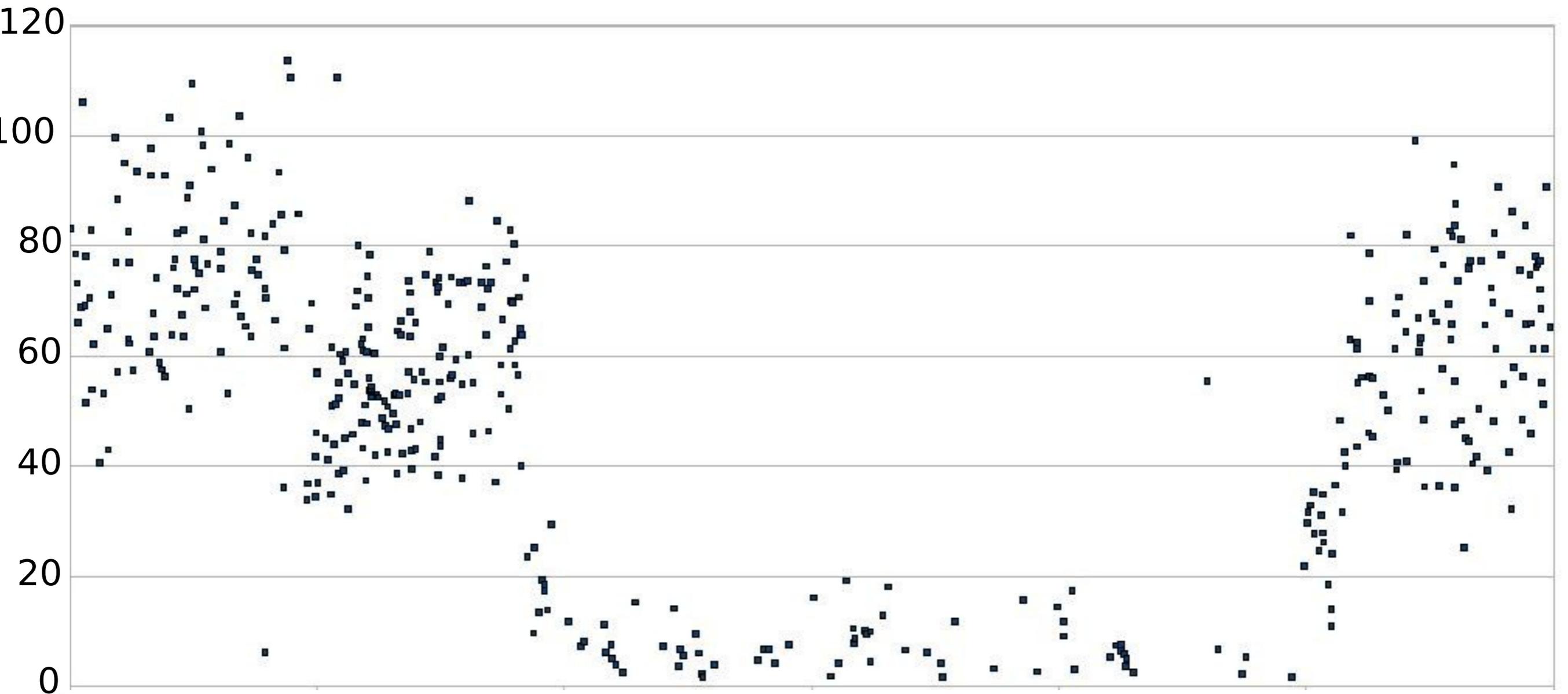
Оценка качества

# Способ получения данных



# Данные от пользователей

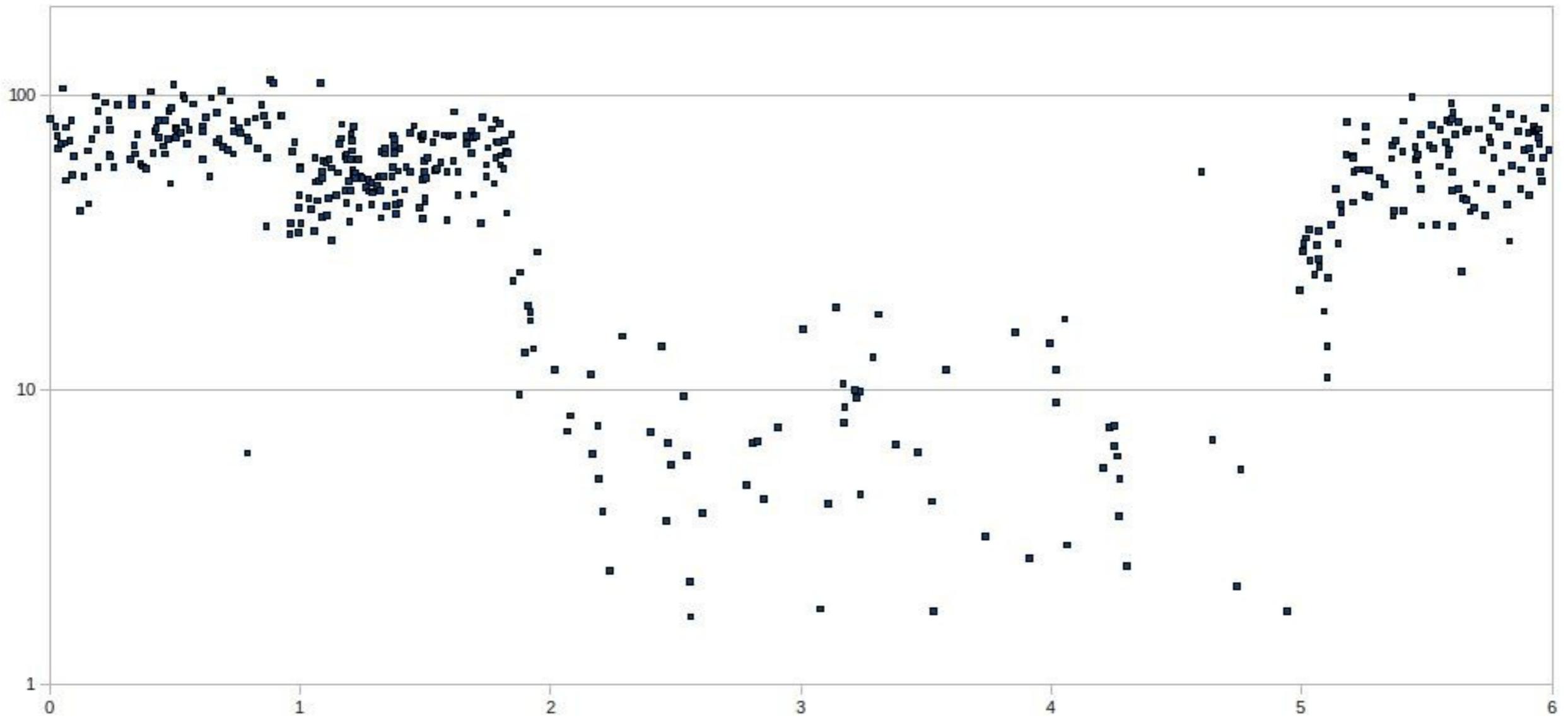
Скорость



Время

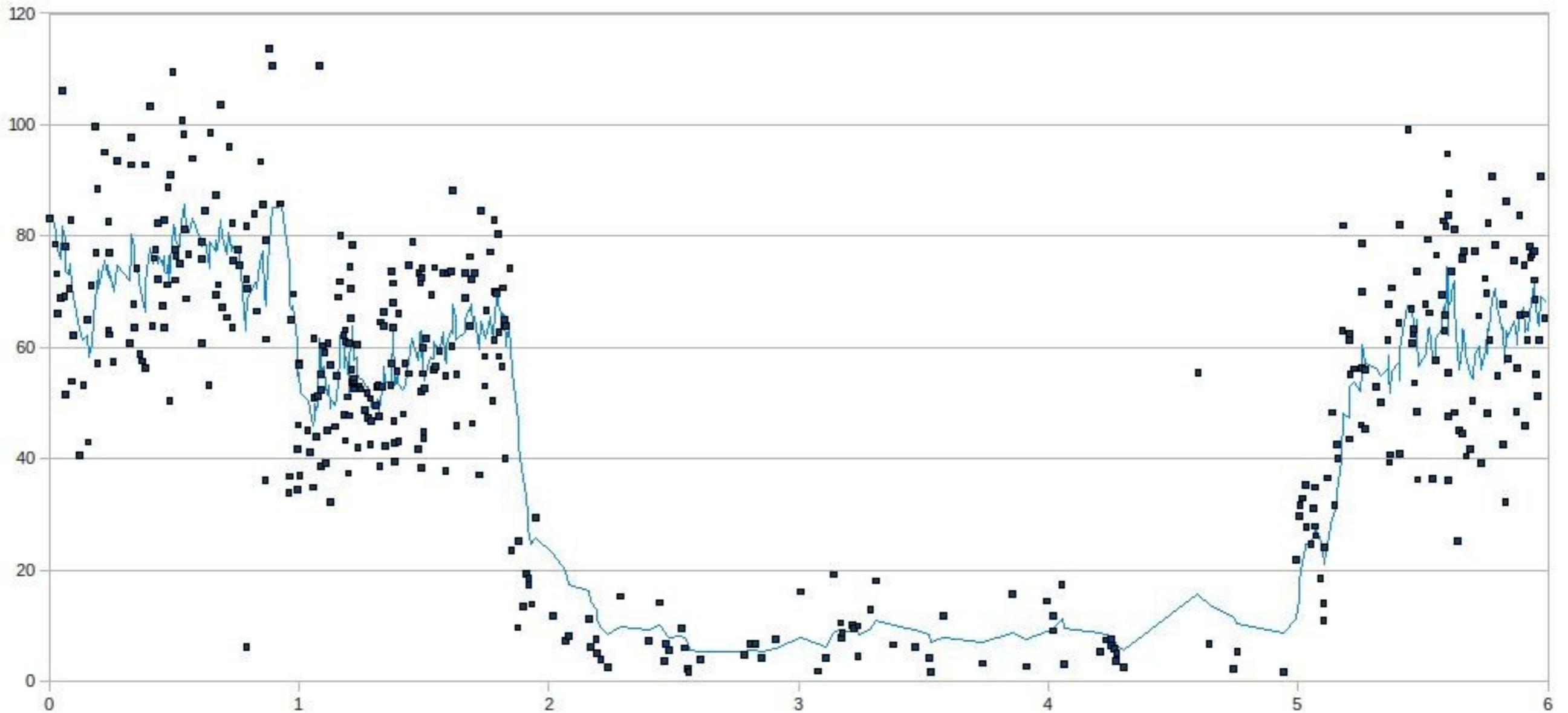
# Данные от пользователей

Скорость

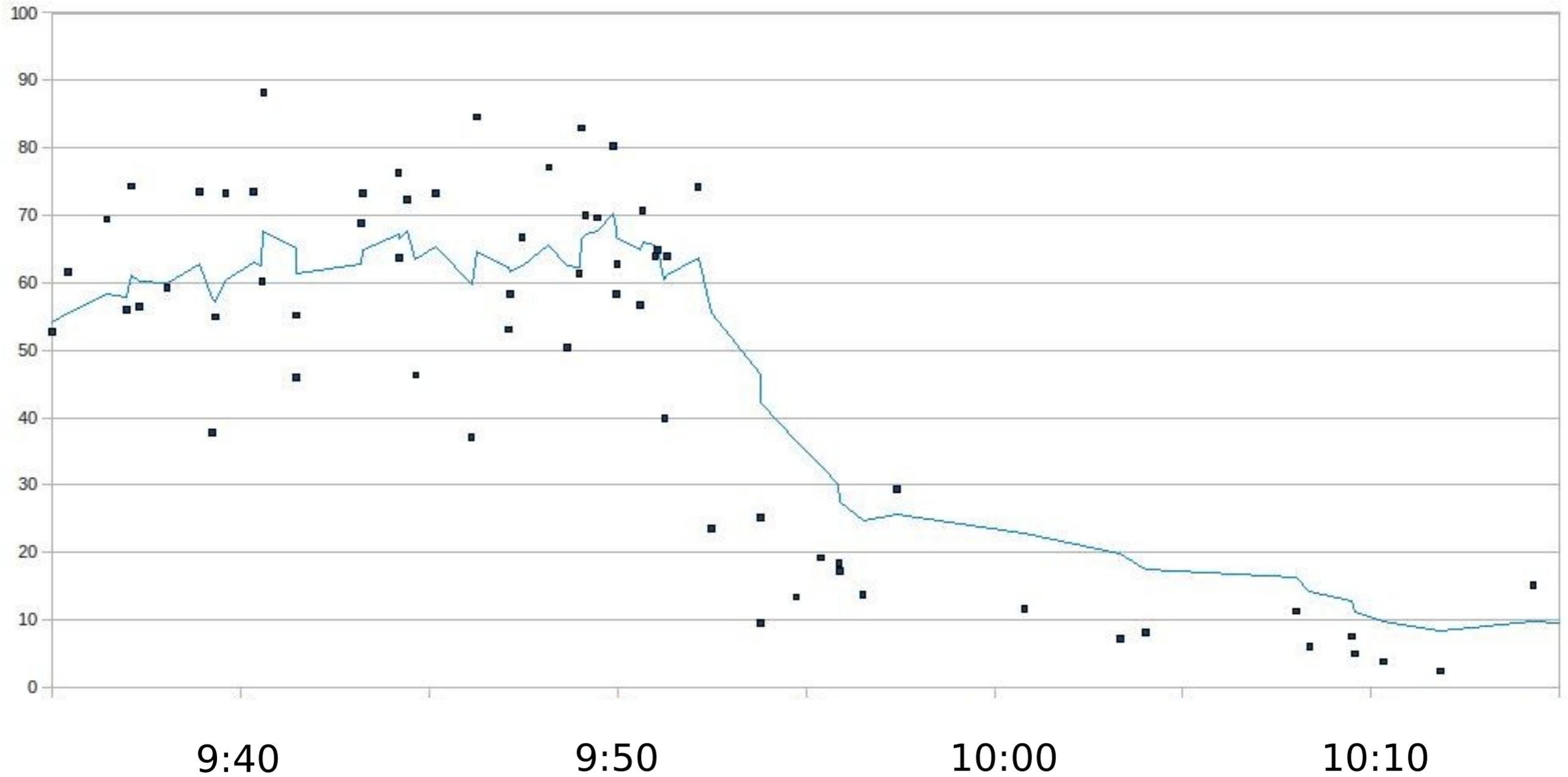


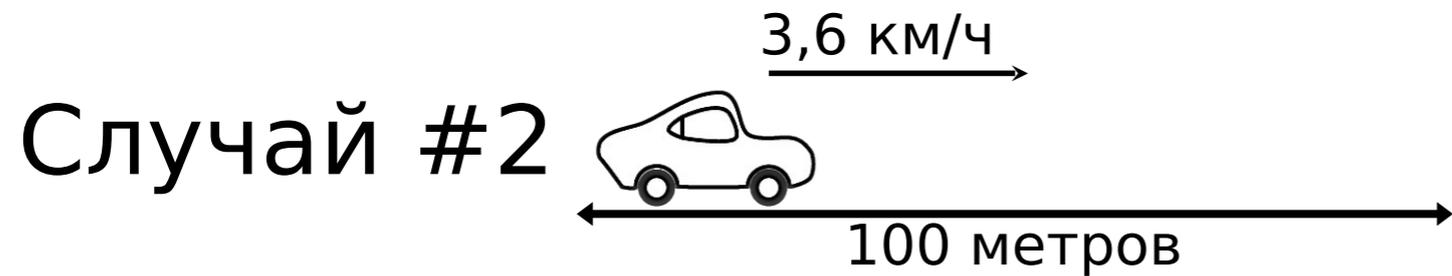
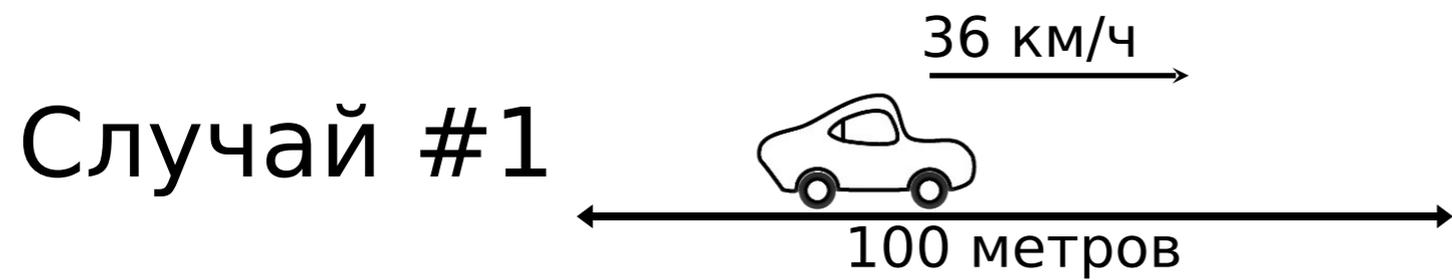
Время

# Итерационное сглаживание



# Итерационное сглаживание





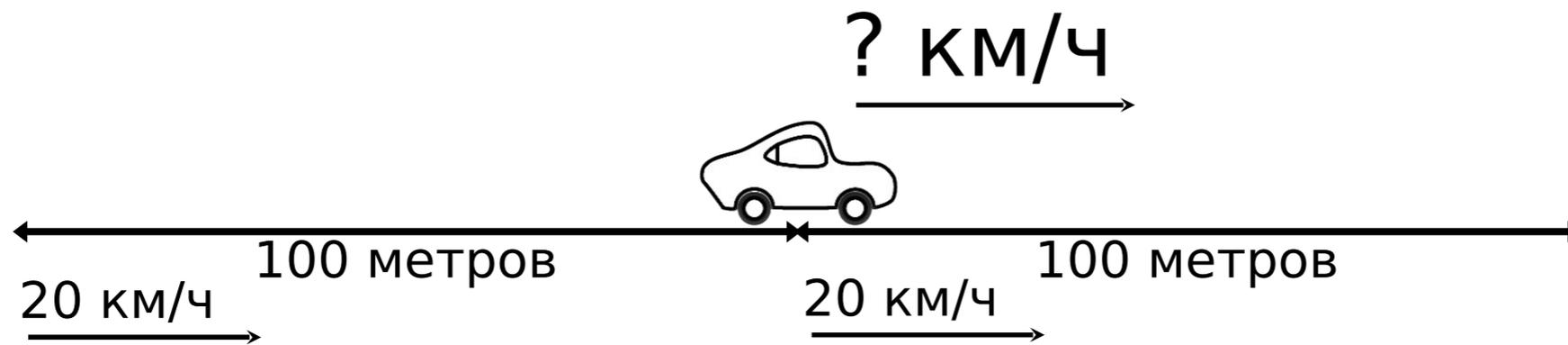
Время проезда

Скорость

Случай #1 10 сек. 36 км/ч

Случай #2 100 сек. 3,6 км/ч

Среднее 55 сек. ( ~ 7 км/ч) ~ 20 км/ч



Время проезда

Скорость

Случай #1 10 сек. + 10 сек. 36 км/ч

Случай #2 100 сек. + 10 сек. 6,5 км/ч

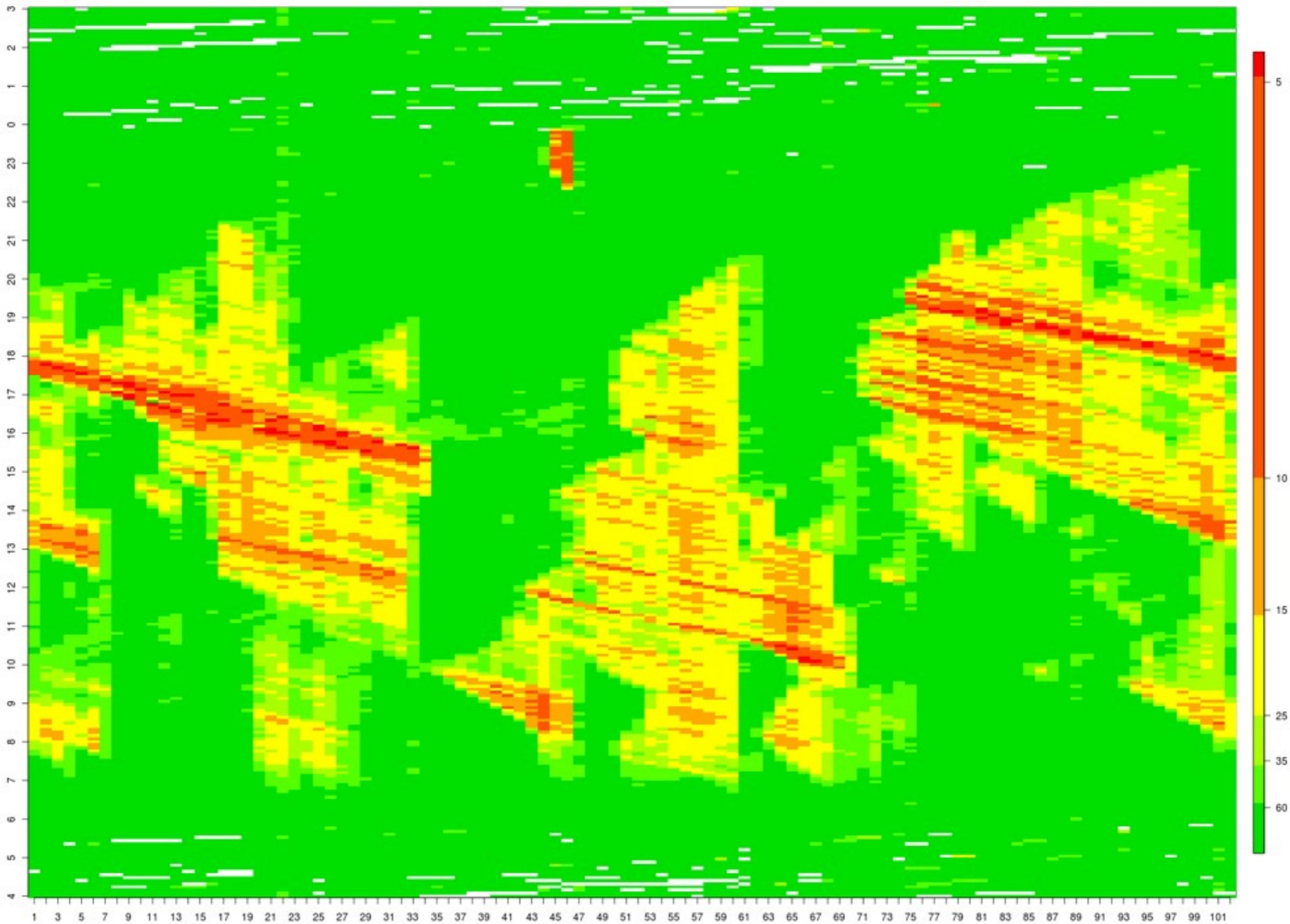
Случай #3 10 сек. + 100 сек. 6,5 км/ч

Случай #4 100 сек. + 100 сек. 3,6 км/ч

Среднее 110 сек. ( ~ 7 км/ч) ~ 13 км/ч

# Постановка математической задачи

Предсказать время проезда по каждому участку дороги через 15–30–45–60 мин, наблюдая времена проезда сейчас и имея историю таких наблюдений в прошлом.



# Методы решения

**Физическое моделирование**

Потоковые модели

Имитационные модели

**Машинное обучение**

Авторегрессия

К ближайших соседей

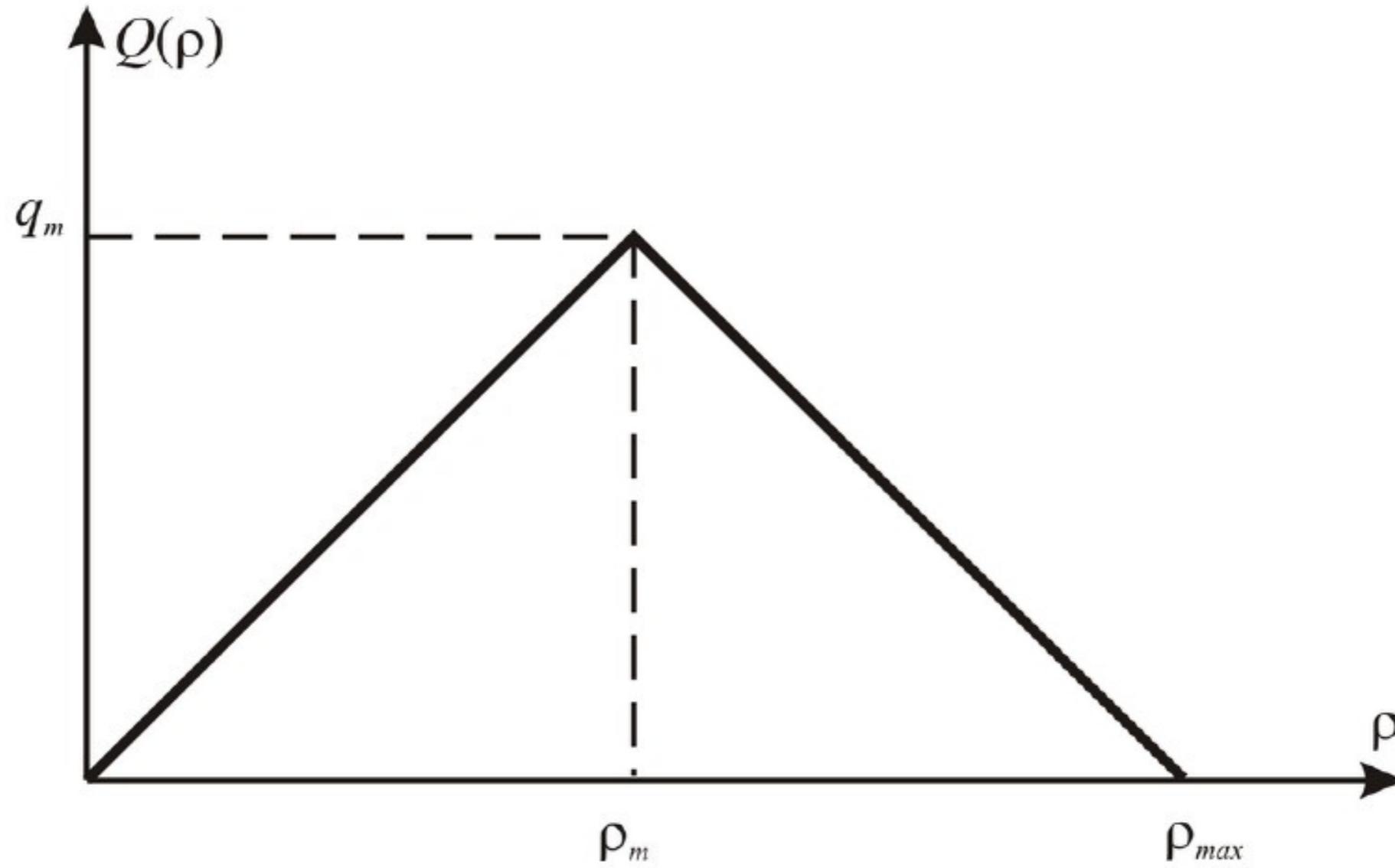
# Имитационные модели



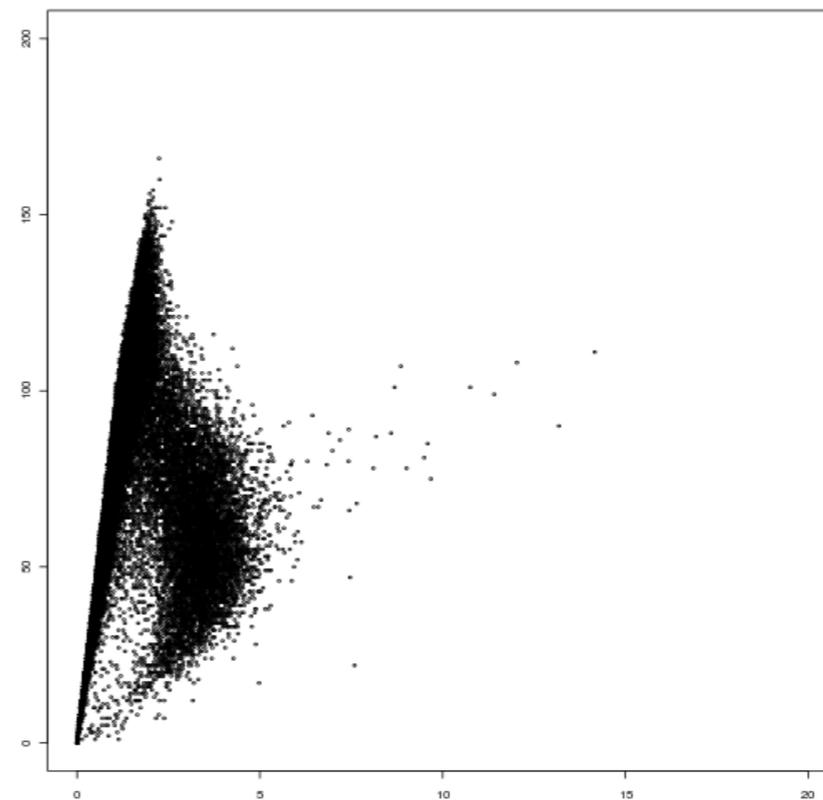
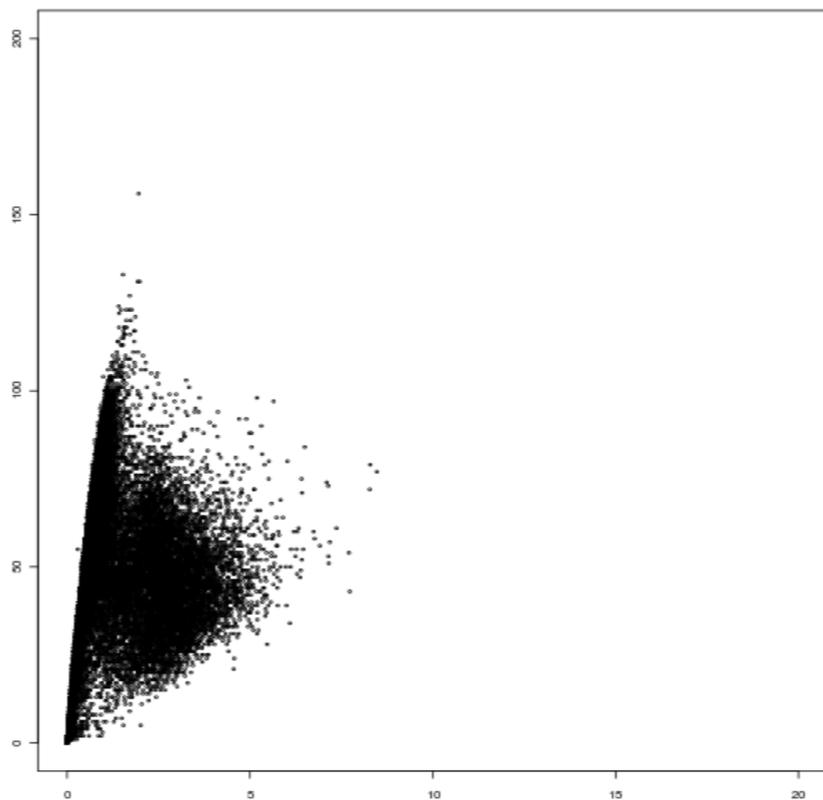
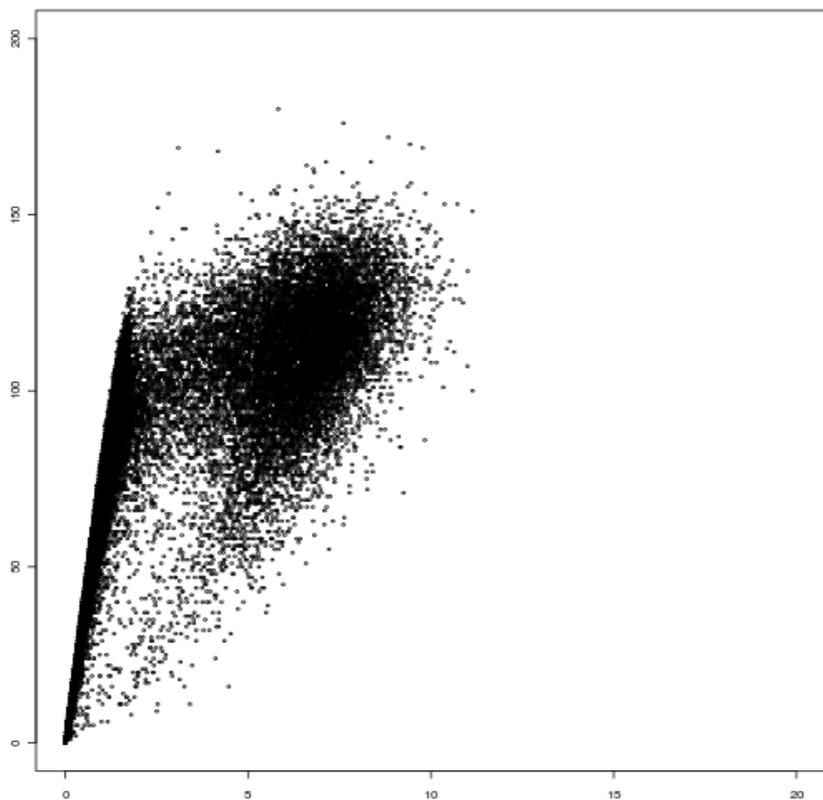
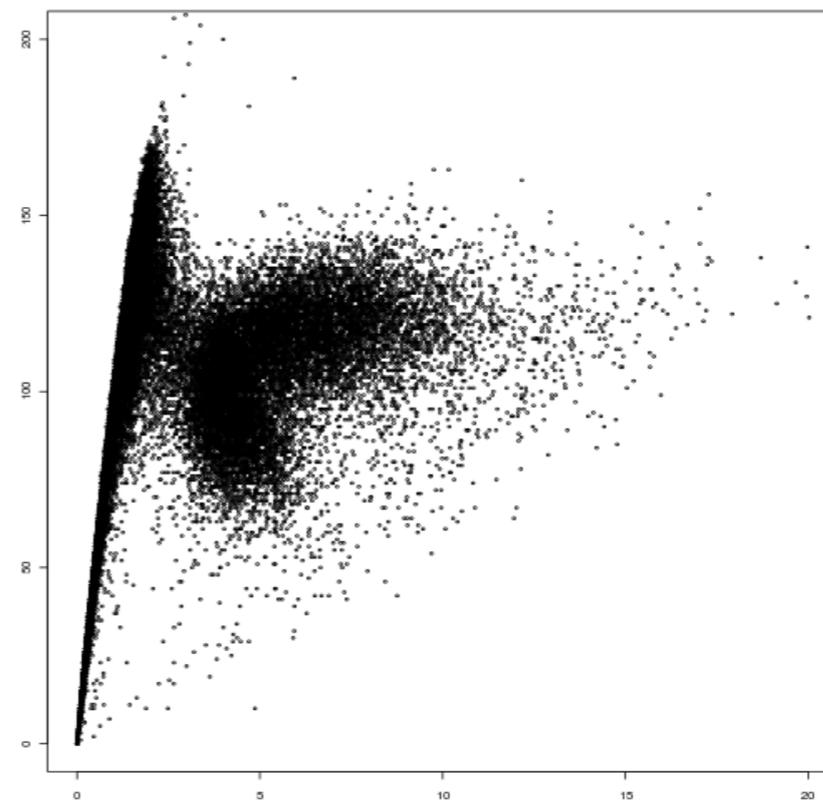
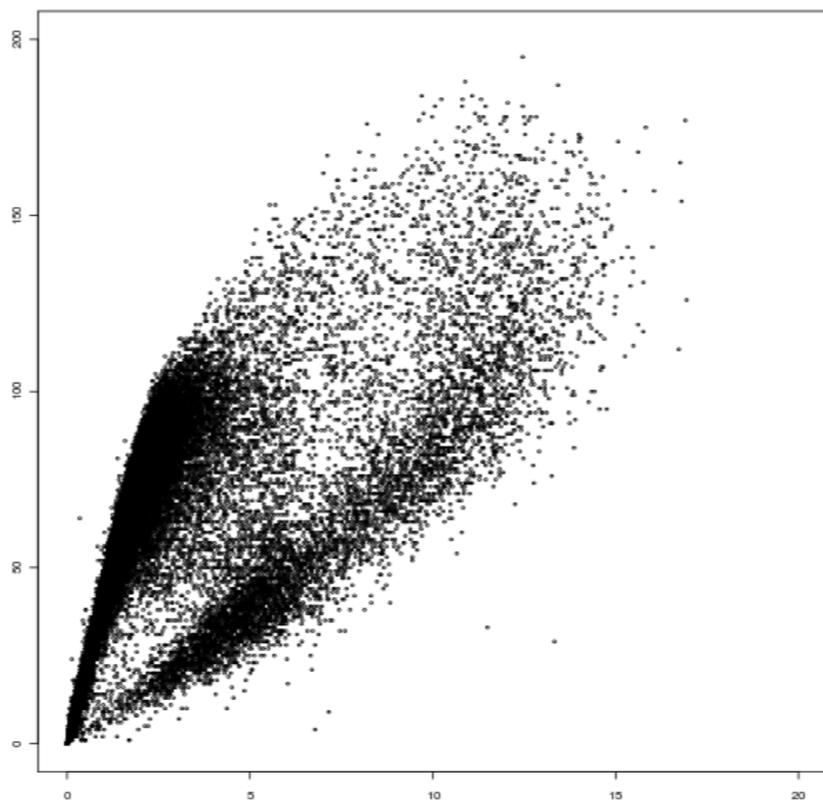
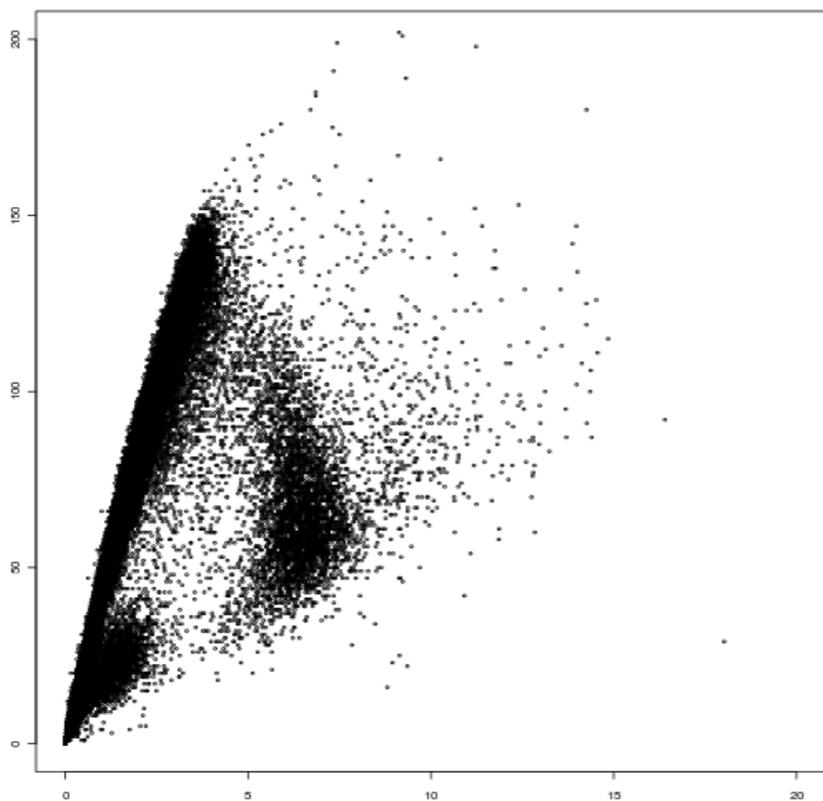
# Потоковые модели



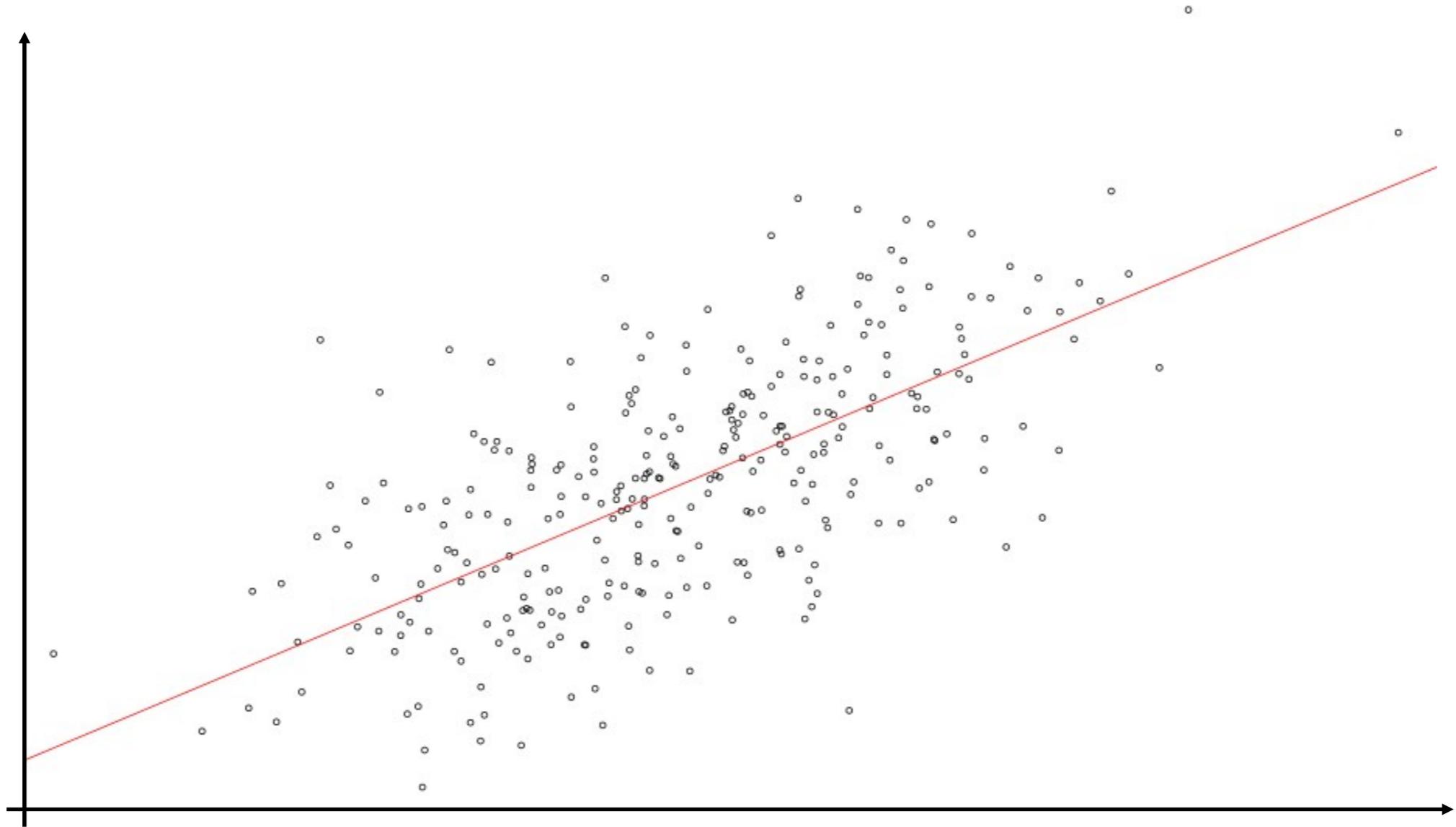
# ПОТОК ОТ ПЛОТНОСТИ



# ПОТОК ОТ ПЛОТНОСТИ



# Метод авторегрессии



Прогноз: 
$$x_i^t = \sum_{j \in N(i)} \sum_{l=1}^L a_{ij}^l x_j^{t-l}$$

# Метод авторегрессии

Минимизируемые критерии:

**Сумма квадратов** подразумевает нормальное распределение ошибок, очень чувствительна к выбросам.

**Сумма модулей** менее чувствительна к выбросам, но гораздо вычислительно сложнее.

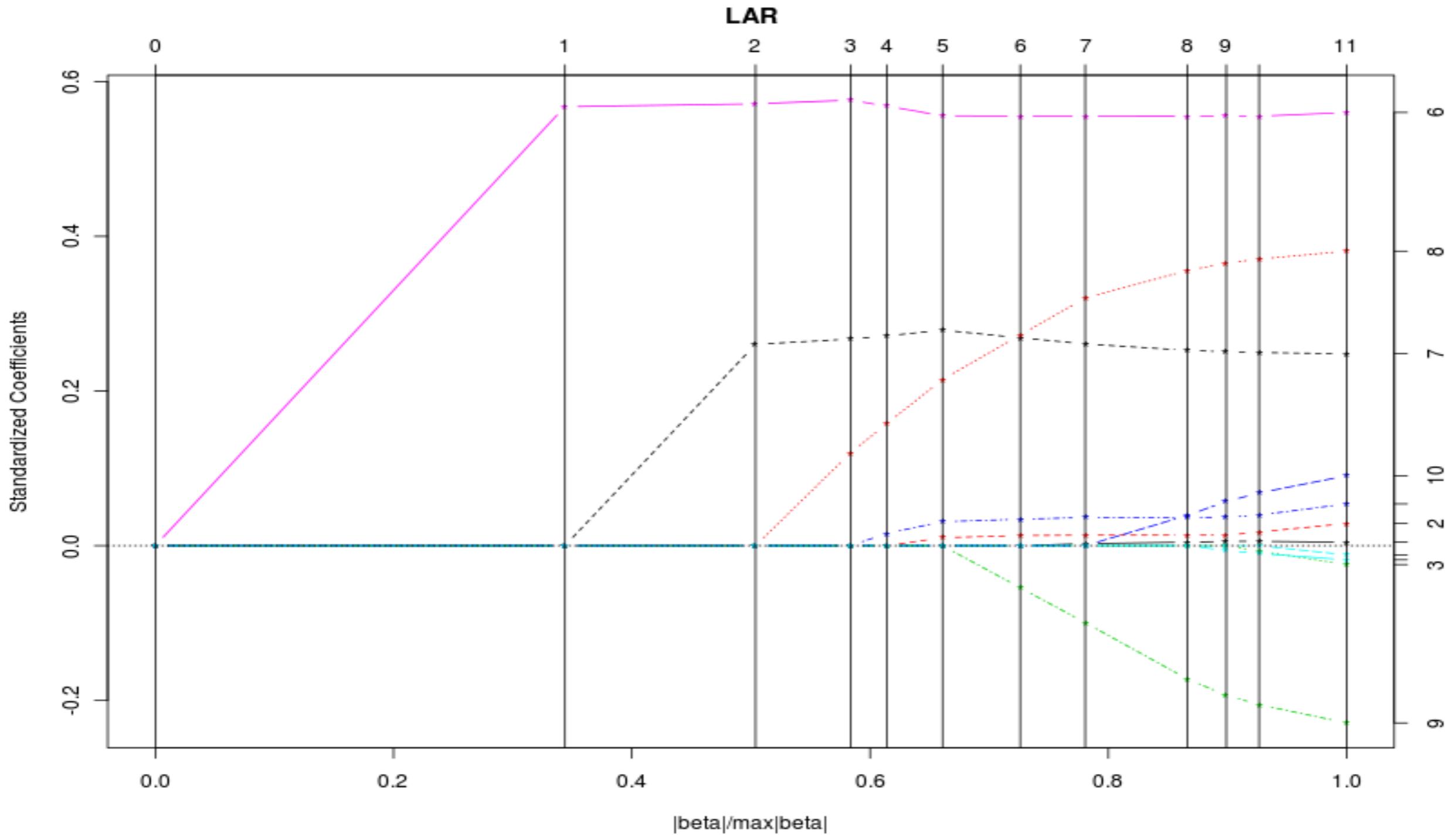
Выбор предикторов:

**Геометрические соседи** могут включать в себя много лишних ребер (например, в центре Москвы).

**Топологические соседи** могут упустить что-то важное (например, проспект и дублер).

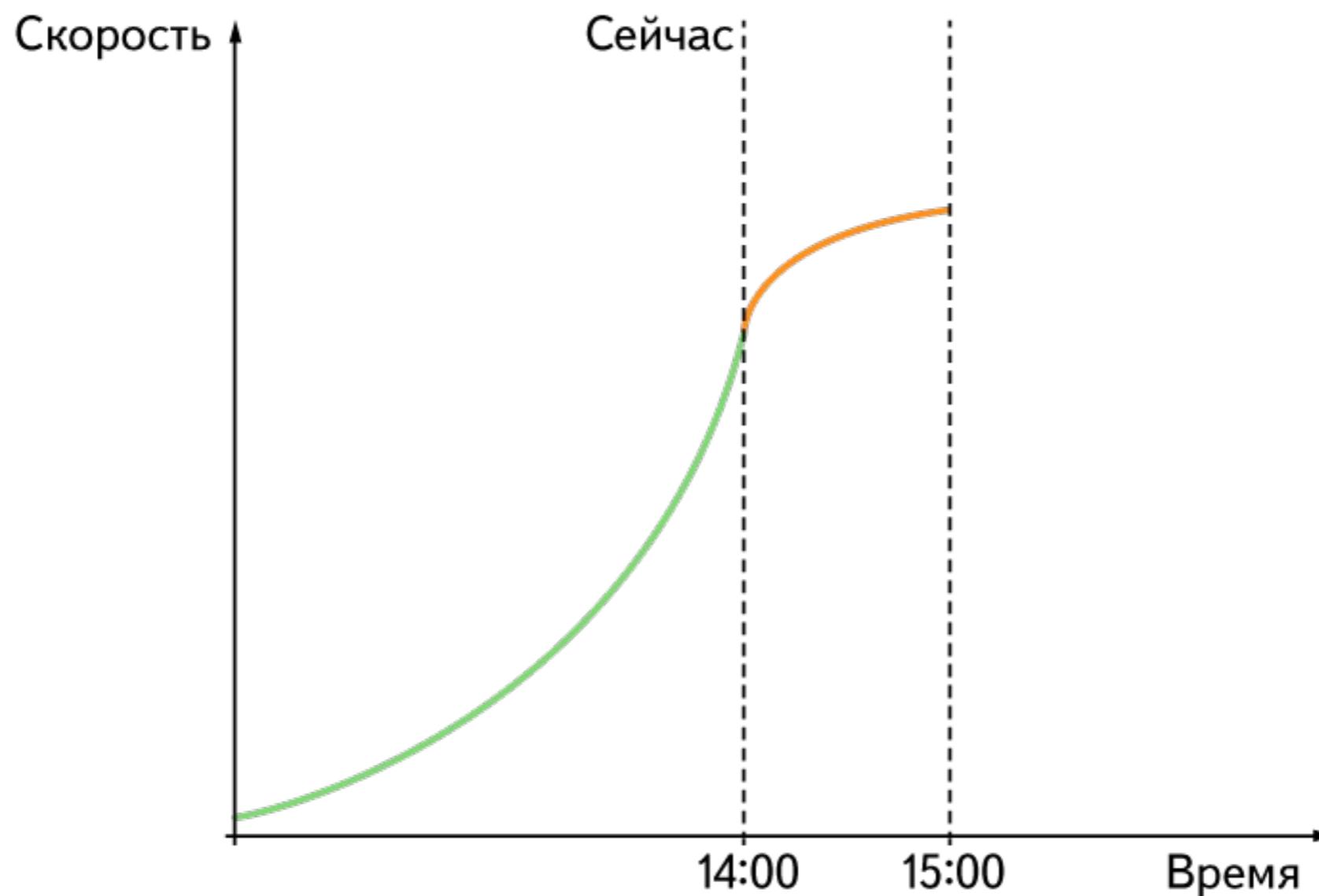
**Автоматический выбор** предикторов выглядит перспективно.

# LARS



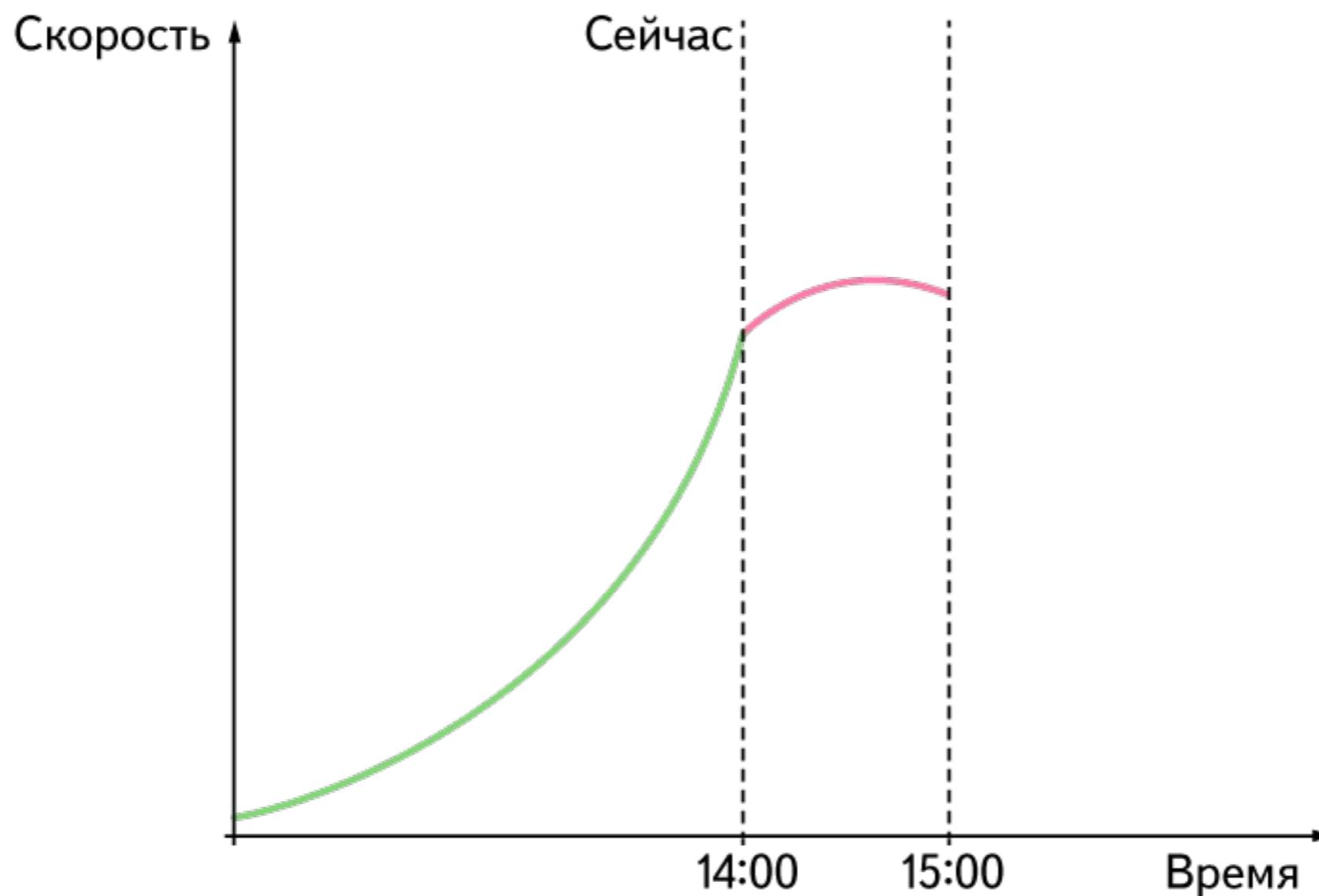
Процедура регуляризации LARS.

# Метод k ближайших соседей



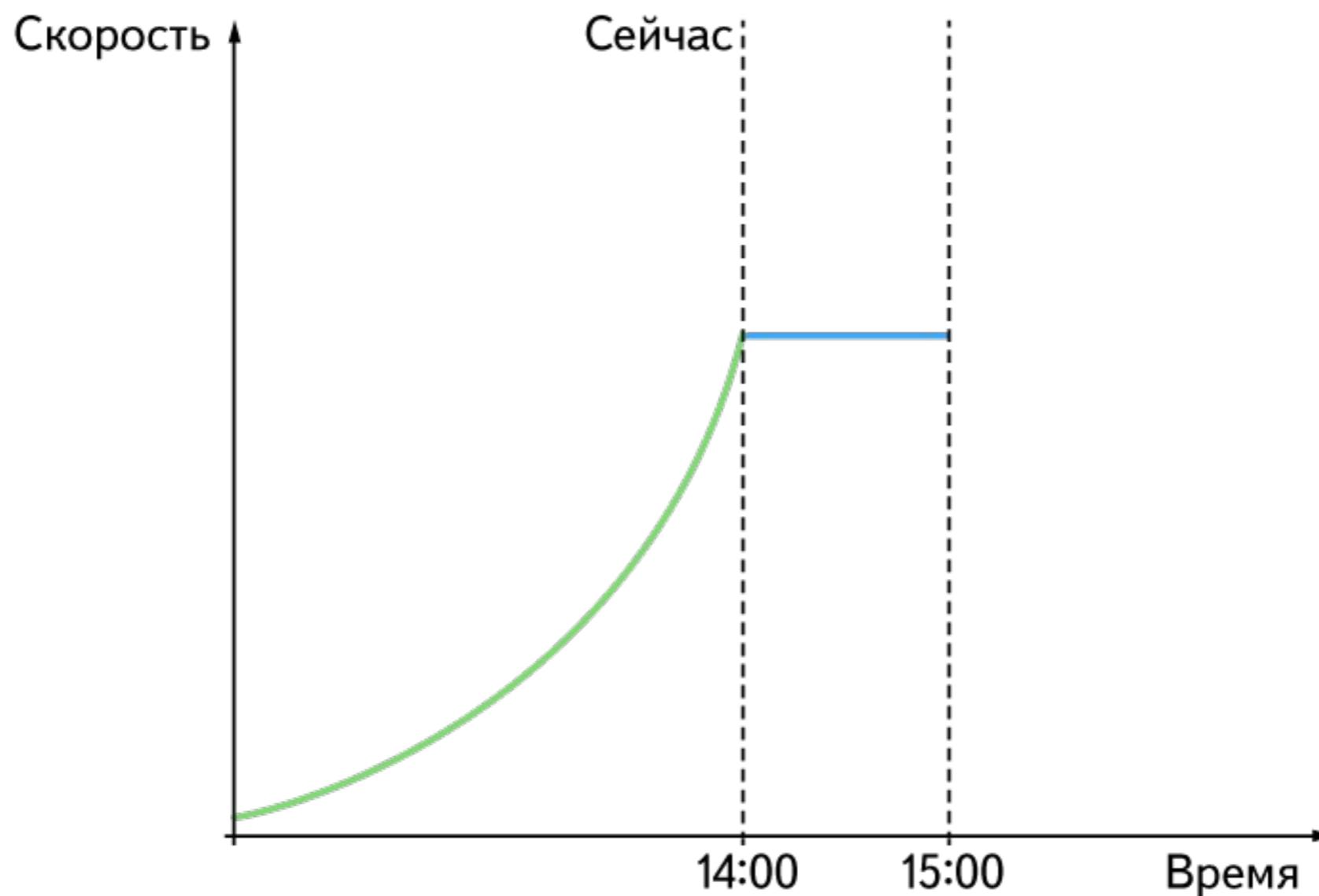
**Картина пробок #1**

# Метод k ближайших соседей



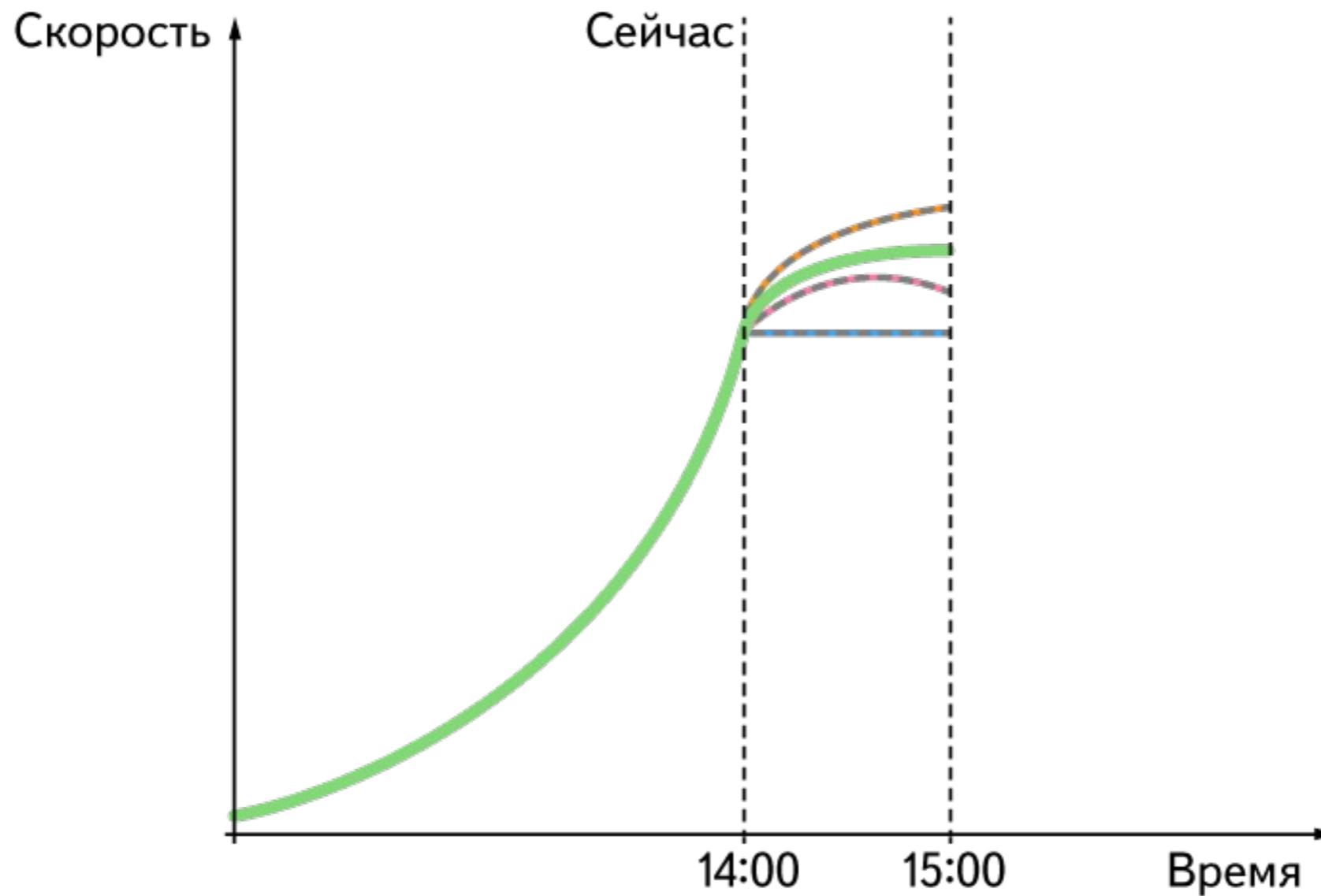
**Картина пробок #2**

# Метод k ближайших соседей



Картина пробок #3

# Метод k ближайших соседей



Прогноз: 
$$x_i^t = \frac{1}{k} \sum_{\tau \in N_k(x_i^t)} x_i^\tau$$

# Оценка качества

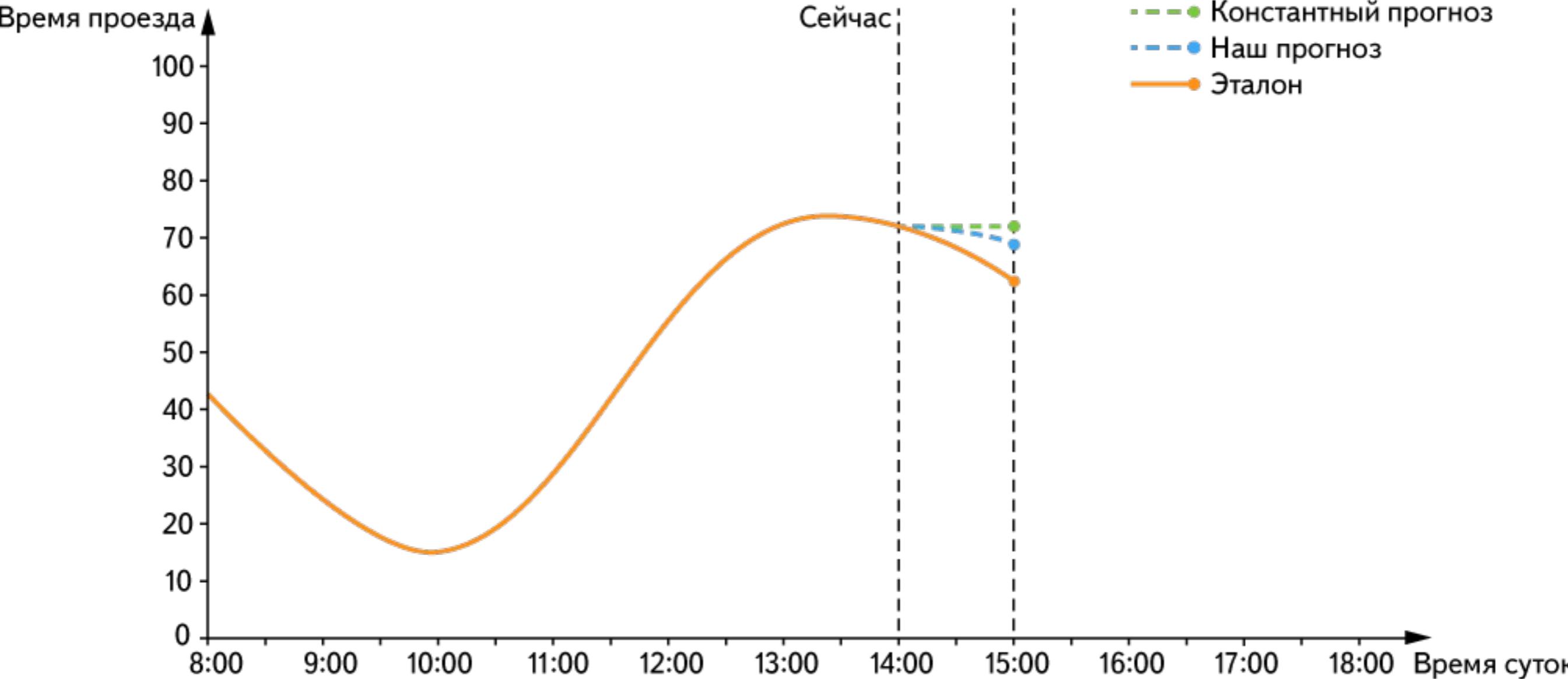
## Задача сравнить:

Разные математические модели.

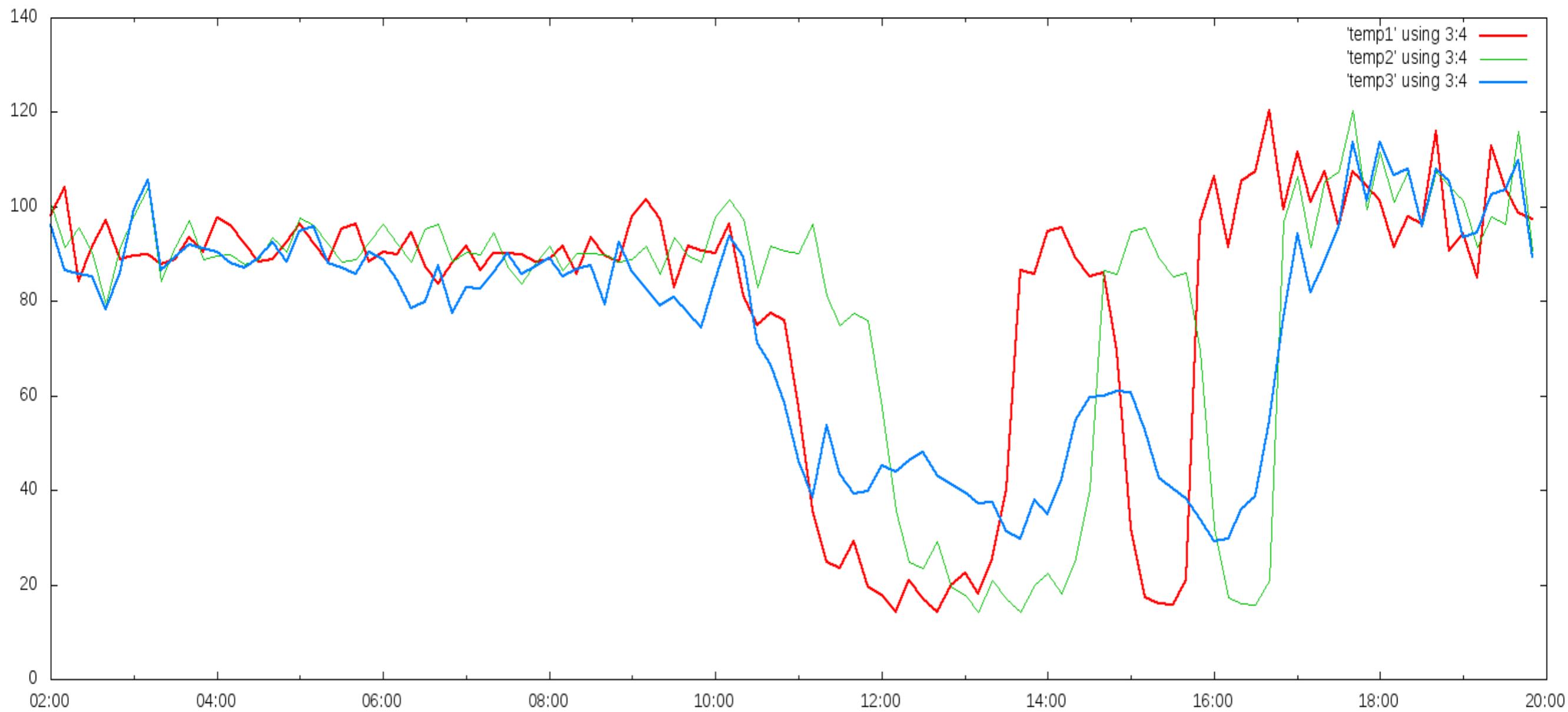
Дороги с разным количеством данных.

Дороги с разной степенью загруженности.

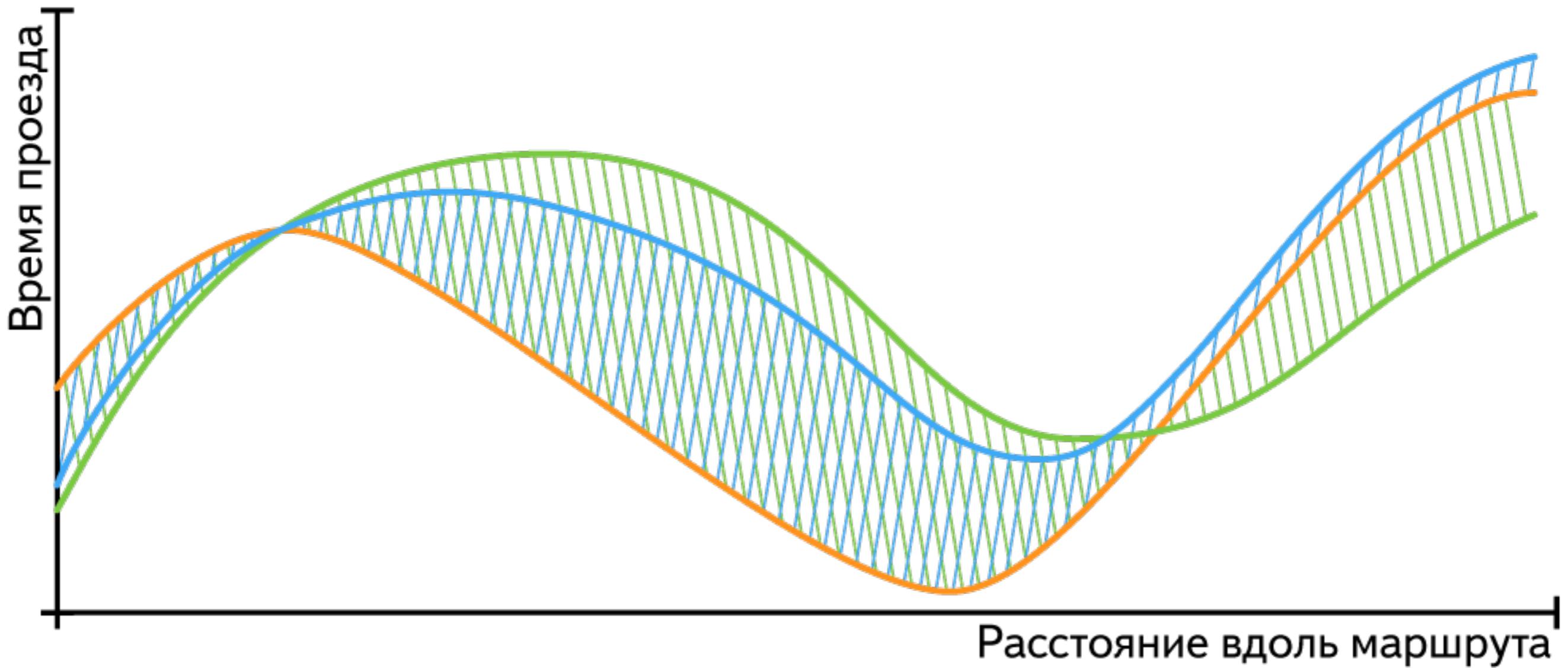
# Оценка качества



# Оценка качества

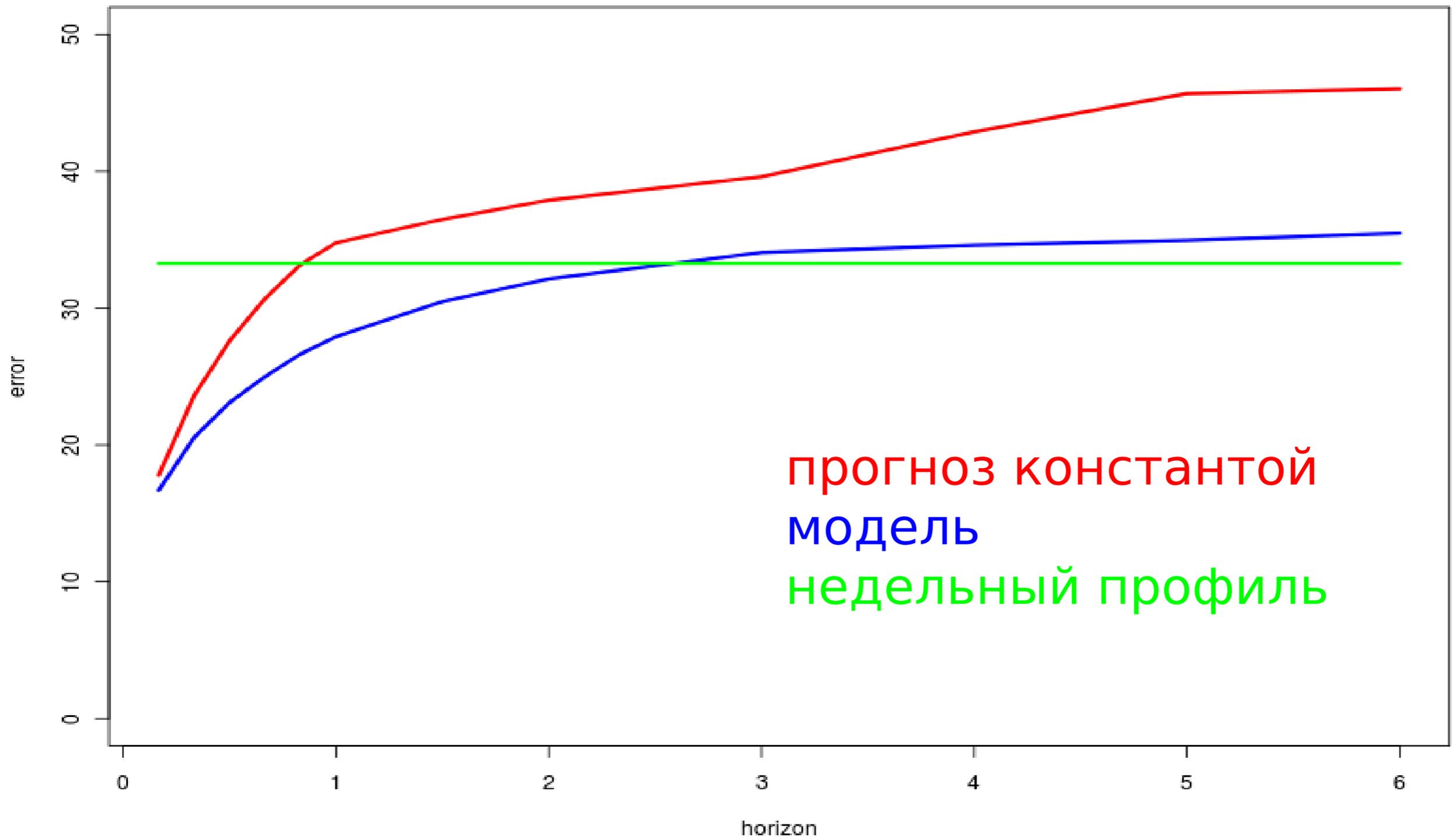


# Оценка качества

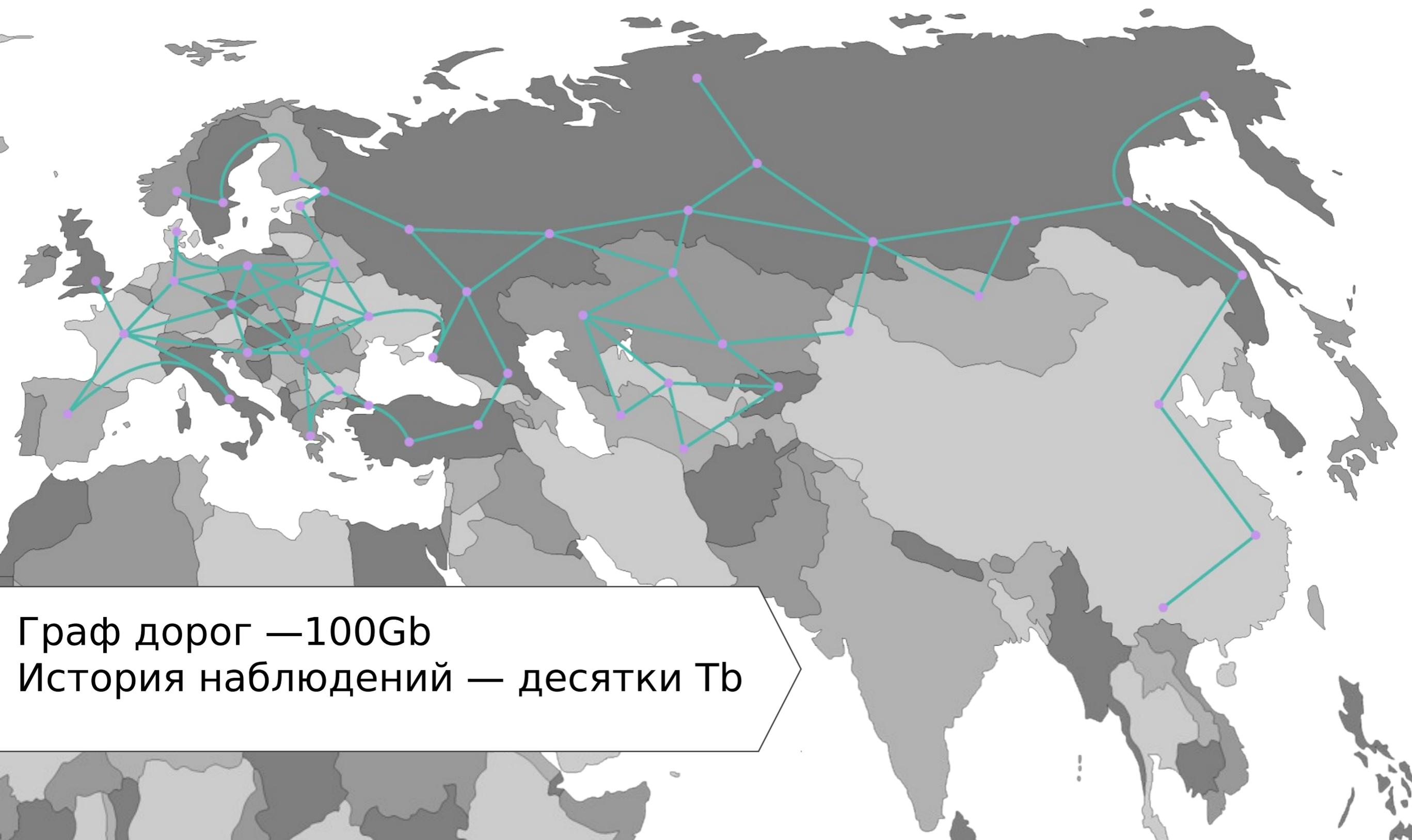


- Эталон    — Наш прогноз    — Константный прогноз
- ▨ Разница между эталоном и константным прогнозом
- ▨ Разница между эталоном и нашим прогнозом

# Оценка качества

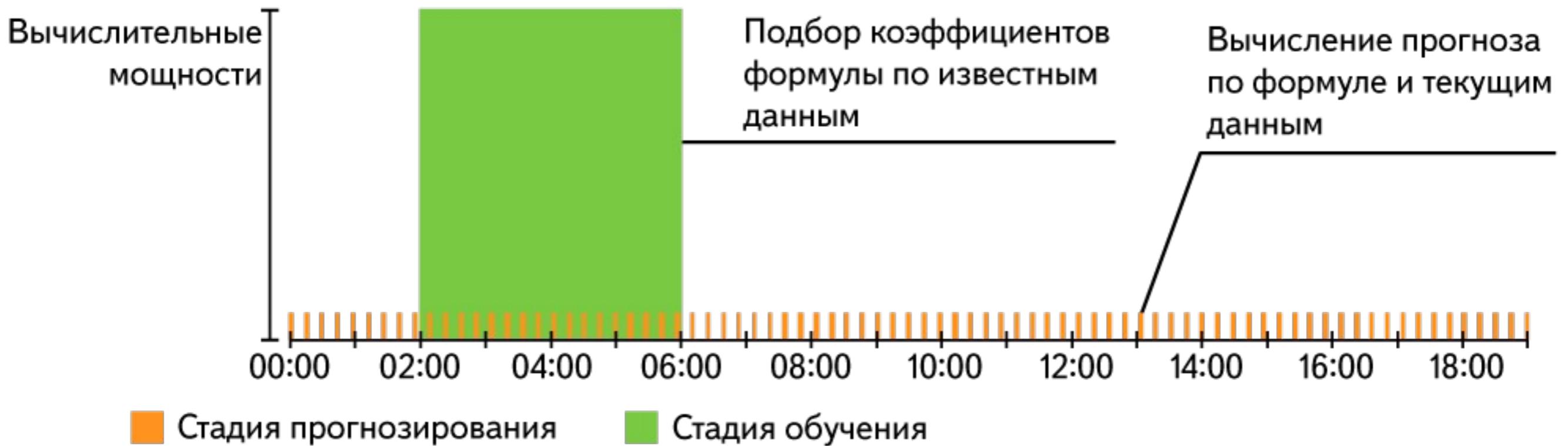


# Граф дорог

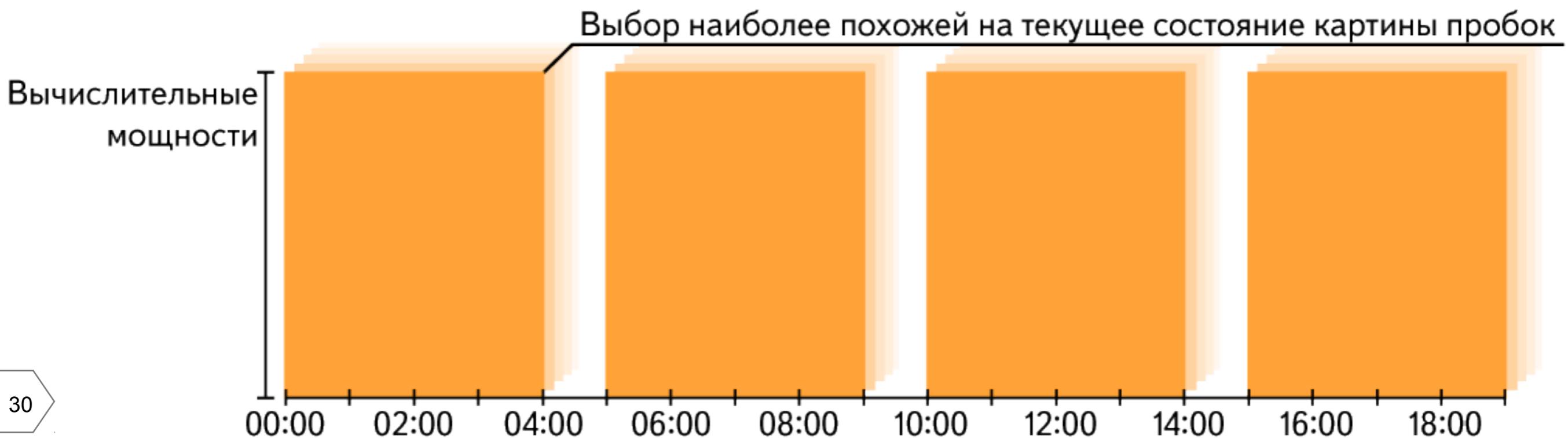


Граф дорог — 100Gb  
История наблюдений — десятки Tb

# Метод авторегрессии



# Метод k ближайших соседей



**Я**ндекс

Михаил Хохлов

Кандидат физико-  
математических наук

Разработчик

**Спасибо**