

Анализ и прогнозирование гидрологических данных

Павлов Александр Сергеевич

Научный руководитель: Цеховая Татьяна Вячеславовна

Рецензент: Рафеенко Екатерина Дмитриевна

Кафедра Теории Вероятностей и Математической Статистики

Факультет Прикладной Математики и Информатики

Белорусский Государственный Университет

Минск, 2015

① Обзор реализованного программного обеспечения

- Модуль предварительного анализа

- Модуль анализа остатков

- Модуль вариограммного анализа

② Детерминированный подход

- Проверка на нормальность

- Корреляционный анализ

- Регрессионный анализ

 - Регрессионная модель

 - Качество регрессионной модели

- Анализ остатков

③ Геостатистический подход

- Введение

- Вариограммный анализ

- Кригинг

- Автоматический подбор



- Доступно с любого устройства, имеющего доступ в интернет, по адресу apaulau.shinyapps.io/batorino
- Реализовано на языке программирования **R**
- Логически разделёно на три модуля
- Имеет простой, быстро расширяемый гибкий интерфейс
- Широкие графические возможности
- Проверка тестов и критериев
- Мгновенный отклик на изменение параметров
- Быстрая проверка различных моделей



Модуль предварительного анализа

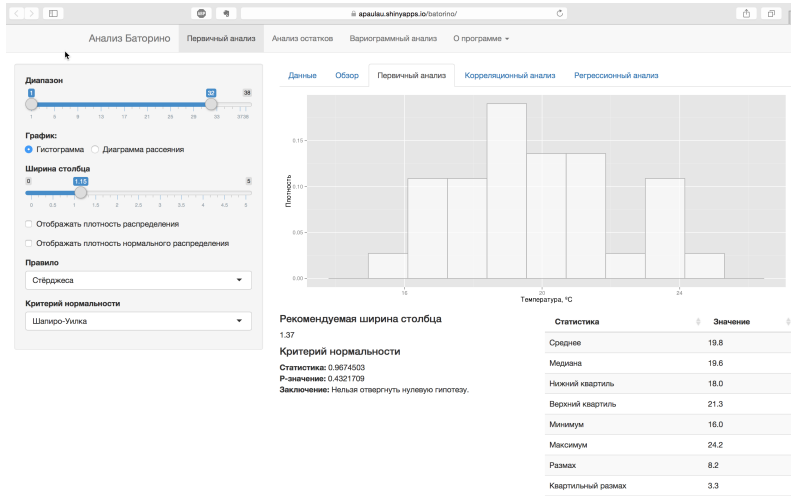


Рис. 1: Первичный анализ и описательные статистики



Модуль предварительного анализа

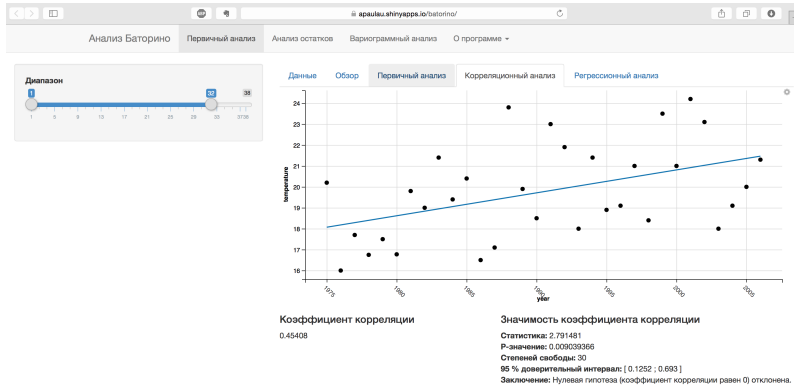


Рис. 2: Корреляционный анализ



Модуль предварительного анализа

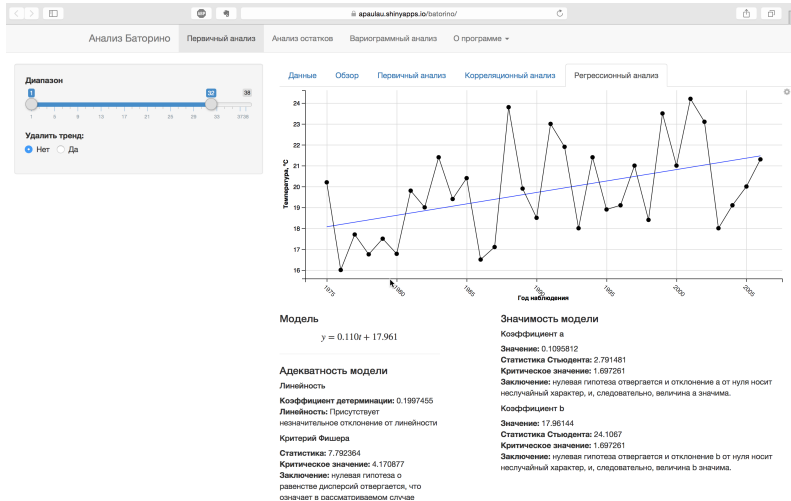


Рис. 3: Регрессионный анализ



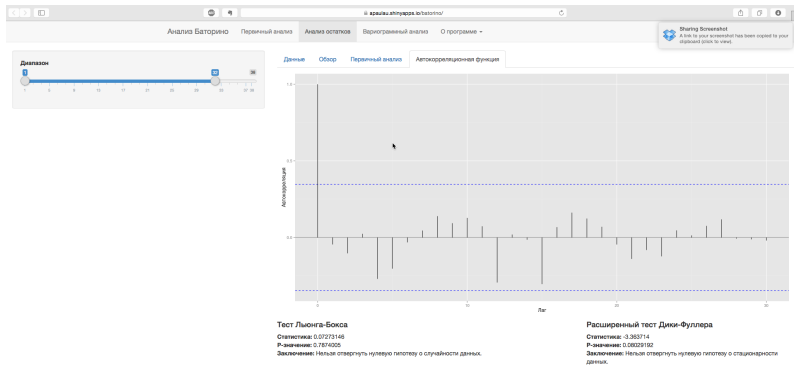


Рис. 4: Автокорреляционная функция



Модуль вариограммного анализа

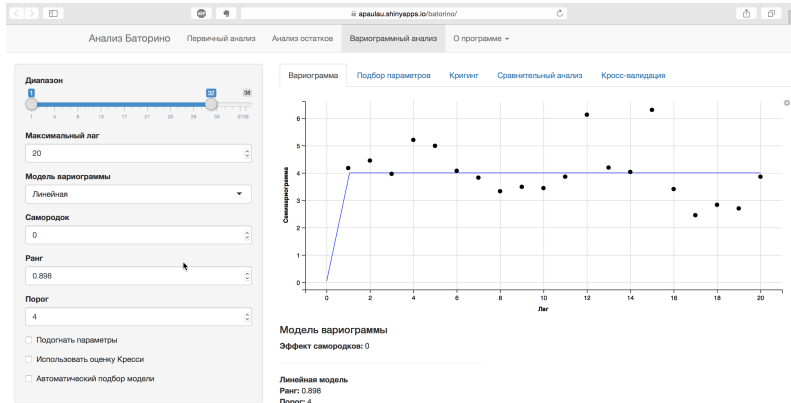


Рис. 5: Возможности по подбору модели вариограммы



Модуль вариограммного анализа

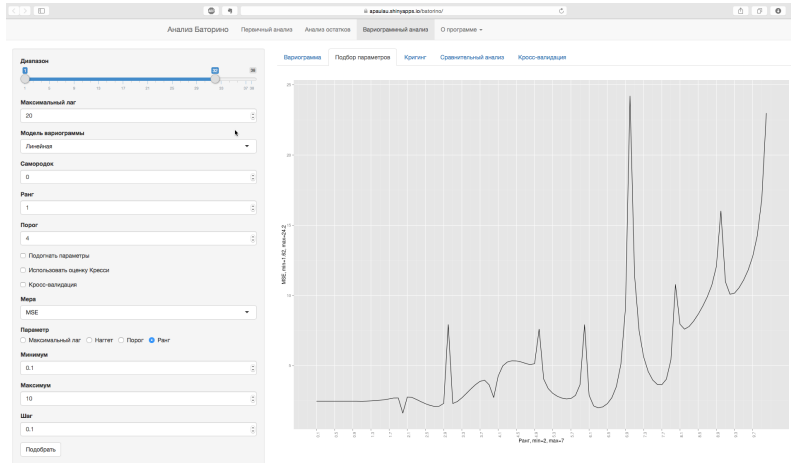


Рис. 6: Подбор параметров модели вариограммы



Модуль вариограммного анализа

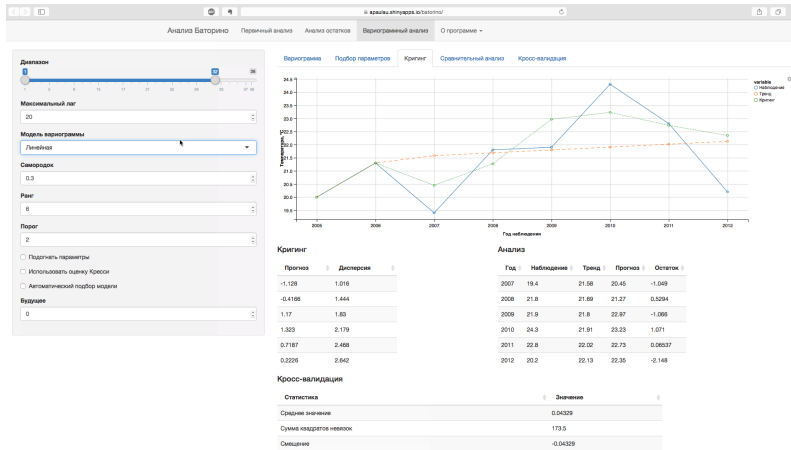


Рис. 7: Сравнение прогнозных значений



Исследуемые данные получены от учебно-научного центра «Нарочанская биологическая станция им. Г.Г.Винберга». Исходные данные представляют собой выборку $X(t)$, состоящую из значений средней температуры воды в июле месяце каждый год в период с 1975 по 2012 годы.

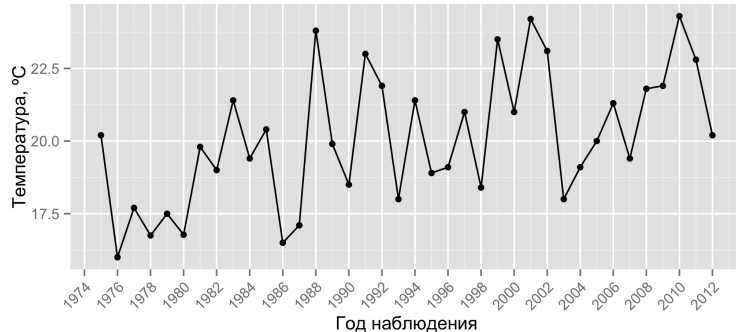


Рис. 8: Исходные данные



Визуально и проверкой критериев Шапиро-Уилка, χ^2 -Пирсона и Колмогорова-Смирнова была показана близость выборочного распределения к нормальному с параметрами $\mathcal{N}(19.77, 5.12)$. При этом выборочное распределение характеризуется небольшой скошенностью вправо (коэффициент асимметрии 0.30) и пологостью пика кривой распределения (-0.746) относительного нормального.

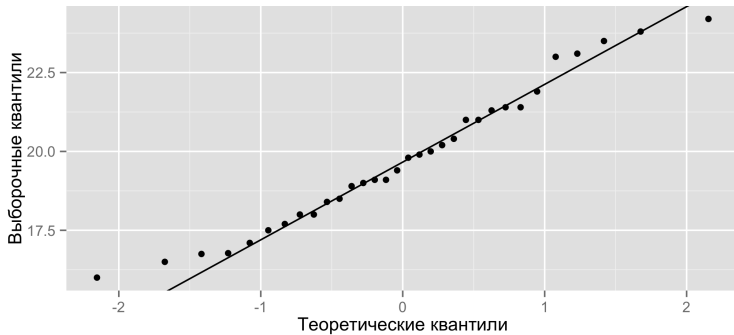


Рис. 9: График квантилей



С помощью критерия Граббса
показано отсутствие выбросов в
исходных данных.
Вычислен выборочный
коэффициент корреляции:
 $r_{xt} = 0.454$.
При уровне значимости $\alpha = 0.05$
доказана его значимость.

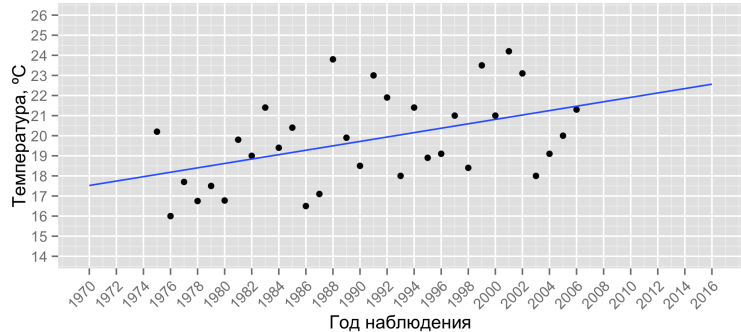


Рис. 10: Диаграмма рассеяния



Выявлено, что исследуемый временной ряд является аддитивным:

$$X(t) = y(t) + \varepsilon(t), \quad (1)$$

где $y(t)$ — тренд, $\varepsilon(t)$ — нерегулярная составляющая.

Найдена модель тренда:

$$y(t) = at + b = 0.1014t + 18.0521$$

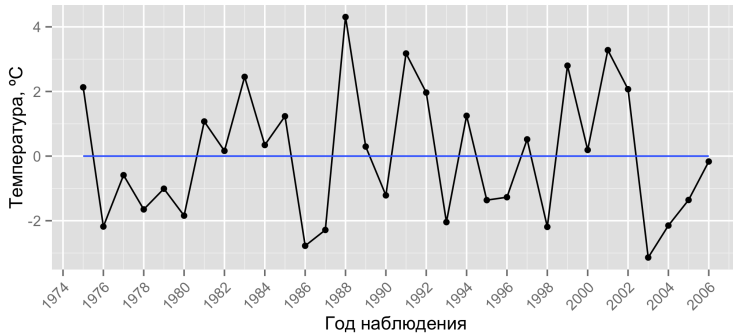


Рис. 11: Ряд остатков $\varepsilon(t)$



- С помощью критерия Стьюдента, при уровне значимости $\alpha = 0.05$, доказана значимость коэффициентов регрессионной модели
- F-критерий Фишера при уровне значимости $\alpha = 0.05$ показал адекватность модели
- Точность модели невысока, поскольку коэффициент детерминации $\eta_{x(t)}^2 = 0.275$



Визуально и проверкой тестов показана близость выборочного распределения к нормальному $\mathcal{N}(0.00, 4.07)$.

На графике видно, что значения автокорреляций не выходят за интервал, обозначенный пунктиром. Что значит отсутствие значимых автокорреляций. Проведённый тест подтвердил данное замечание.

Также было отмечено, что значения имеют небольшую амплитуду и затухают с ростом лага. Это говорит о стационарности в широком смысле, что подтвердил расширенный тест Дики-Фуллера.

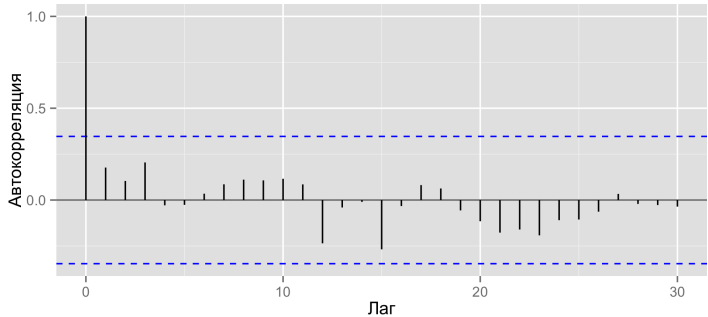


Рис. 12: Автокорреляционная функция



Рассматривается стационарный в широком смысле гауссовский случайный процесс с дискретным временем $X(t)$, $t \in \mathbb{Z}$, нулевым математическим ожиданием, постоянной дисперсией и неизвестной вариограммой $2\gamma(h)$, $h \in \mathbb{Z}$.

Определение 1

Вариограммой случайного процесса $X(t)$, $t \in \mathbb{Z}$, называется функция вида

$$2\gamma(h) = V\{X(t+h) - X(t)\}, \quad t, h \in \mathbb{Z}. \quad (2)$$

При этом функция $\gamma(h)$, $h \in \mathbb{Z}$, называется семивариограммой.

В качестве оценки вариограммы рассматривается статистика, предложенная Матероном:

$$2\tilde{\gamma}(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (X(t+h) - X(t))^2, \quad h = \overline{0, n-1}, \quad (3)$$



Теорема 1

Для оценки $2\tilde{\gamma}(h)$ имеют место следующие соотношения:

$$E\{2\tilde{\gamma}(h)\} = 2\gamma(h),$$

$$\text{cov}(2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)) =$$

$$= \frac{2}{(n-h_1)(n-h_2)} \sum_{t=1}^{n-h_1} \sum_{s=1}^{n-h_2} (\gamma(t-h_2-s) + \gamma(t+h_1-s) - \gamma(t-s) - \gamma(t+h_1-s-h_2))^2,$$

$$V\{2\tilde{\gamma}(h)\} = \frac{2}{(n-h)^2} \sum_{t,s=1}^{n-h} (\gamma(t-h-s) + \gamma(t+h-s) - 2\gamma(t-s))^2,$$

где $\gamma(h)$, $h \in \mathbb{Z}$, — семивариограмма процесса $X(t)$, $t \in \mathbb{Z}$, $h, h_1, h_2 = \overline{0, n-1}$.



Теорема 2

Если имеет место соотношение

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty, \text{ то}$$

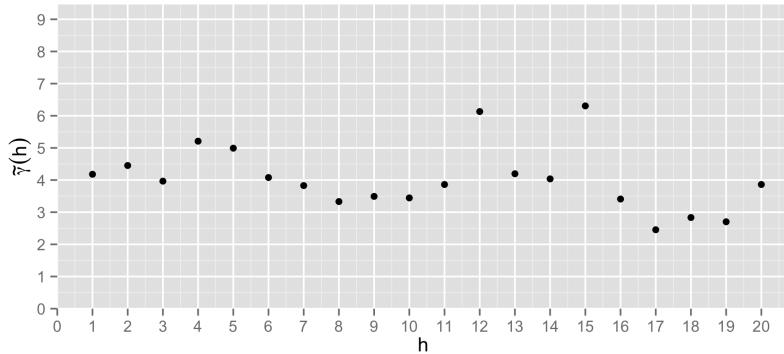
$$\lim_{n \rightarrow \infty} (n - \min\{h_1, h_2\}) \operatorname{cov}\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} = 2 \sum_{m=-\infty}^{+\infty} \gamma(m - h_2) + \gamma(m + h_1) - \gamma(m) - \gamma(m + h_1 - h_2))^2,$$

$$\lim_{n \rightarrow \infty} (n - h) V\{2\tilde{\gamma}(h)\} = 2 \sum_{m=-\infty}^{+\infty} \gamma(m - h) + \gamma(m + h) - 2\gamma(m))^2.$$

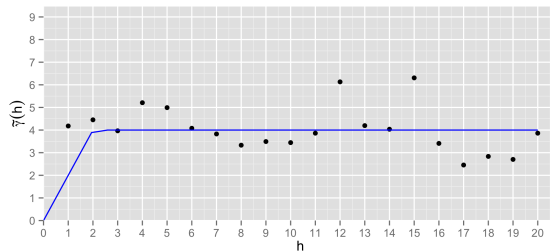
где $\gamma(h)$, $h \in \mathbb{Z}$, — семивариограмма процесса $X(t)$, $t \in \mathbb{Z}$, $h, h_1, h_2 = \overline{0, n-1}$.



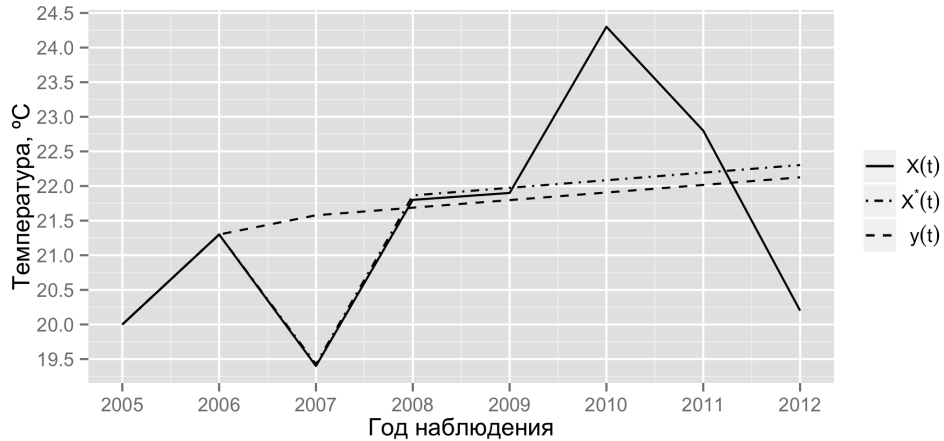
График экспериментальной вариограммы



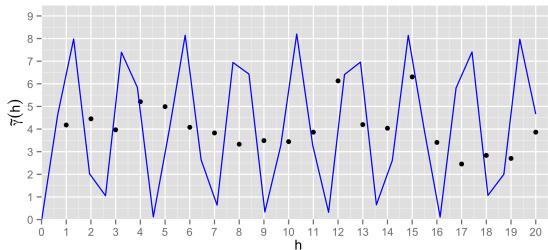
Подобранная модель:
 $0.3 + 4 \cdot \text{Lin}(h, 6.2)$



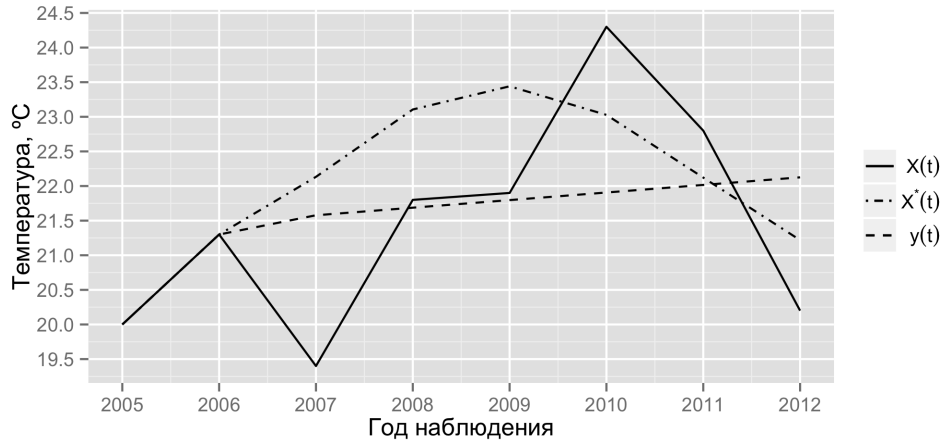
Прогнозирование методом ординарного кригинга



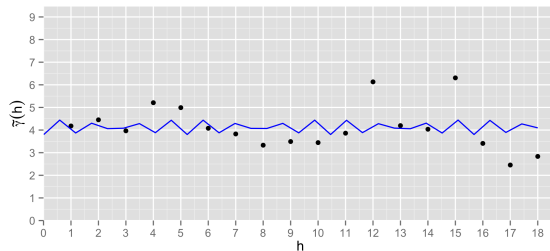
Подобранная модель:
 $4 \cdot \text{Per}(h, 0.898)$



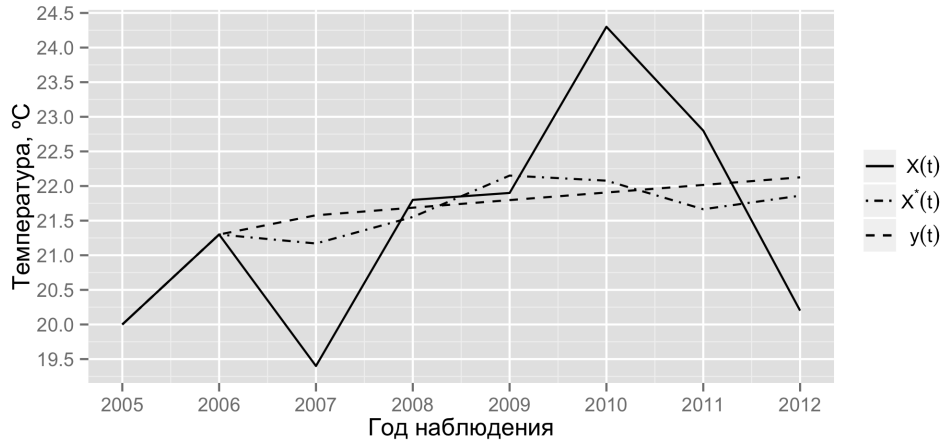
Прогнозирование методом ординарного кригинга



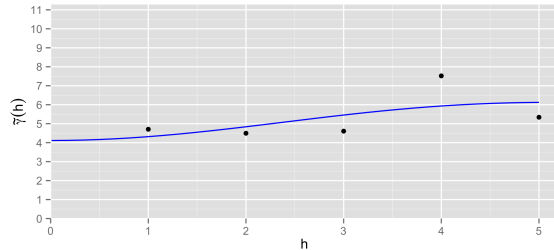
Подобранная модель:
 $3.8 + 0.32 \cdot \text{Per}(h, 1.3)$



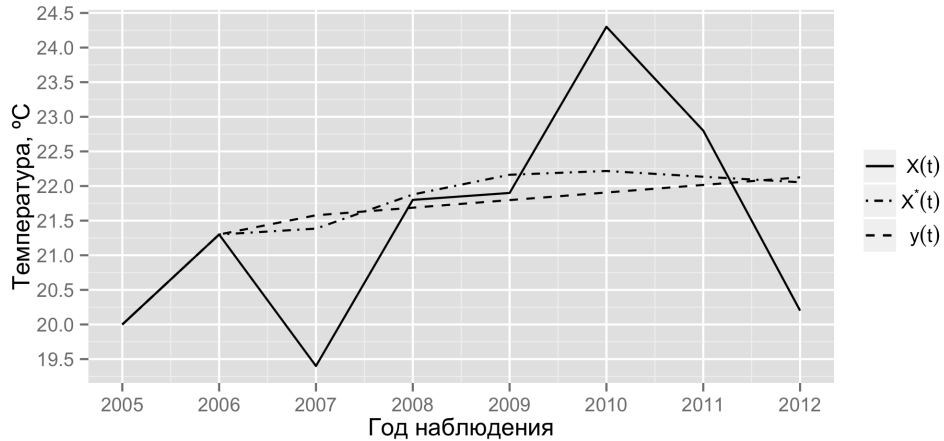
Прогнозирование методом ординарного кригинга



Подобранная модель:
 $4.11 + 1.65 \cdot Wav(h, 3.59)$



Прогнозирование методом ординарного кригинга





Спасибо за внимание!

