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

Содержание

Особенности

- 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

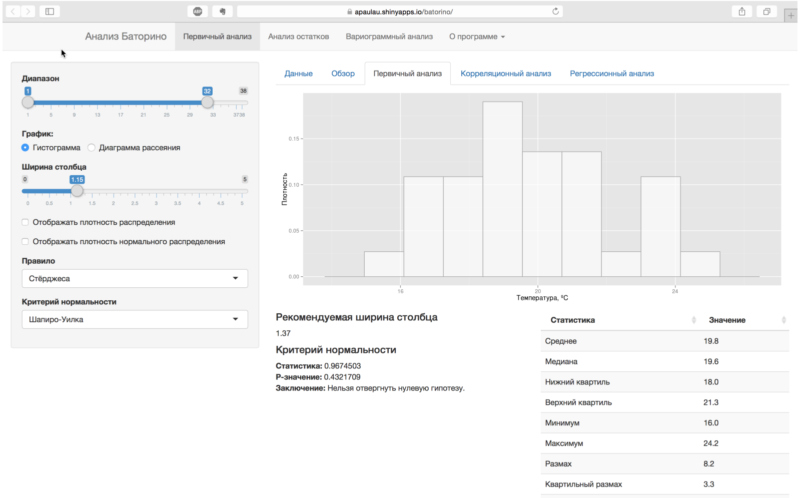


Рис. 1: Первичный анализ и описательные статистики

Модуль предварительного анализа

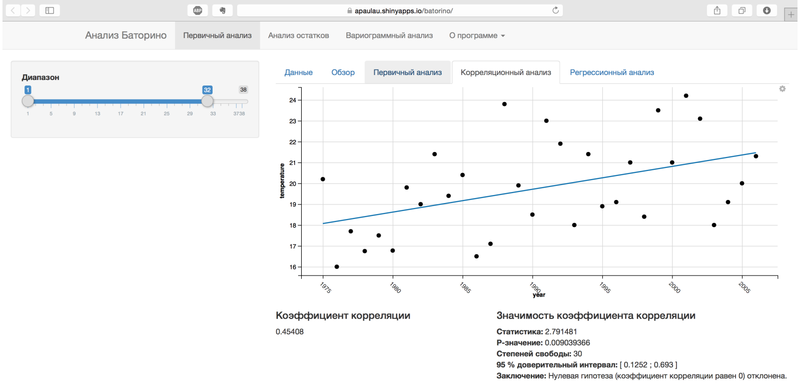


Рис. 2: Корреляционный анализ

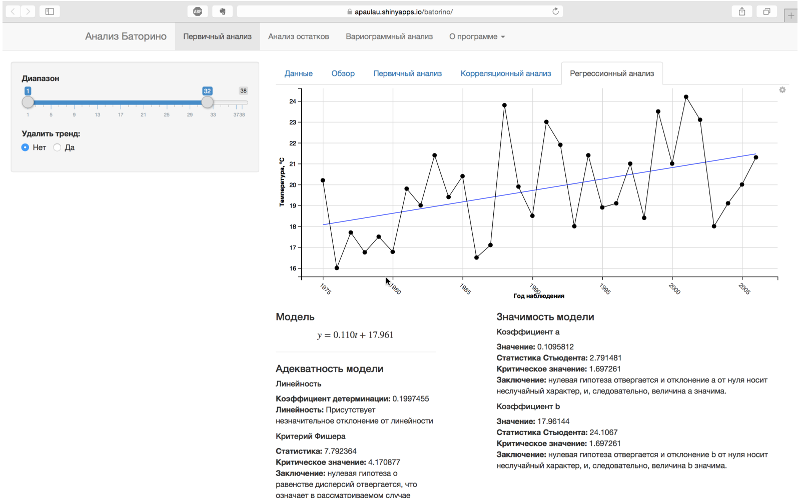


Рис. 3: Регрессионный анализ

Модуль анализа остатков

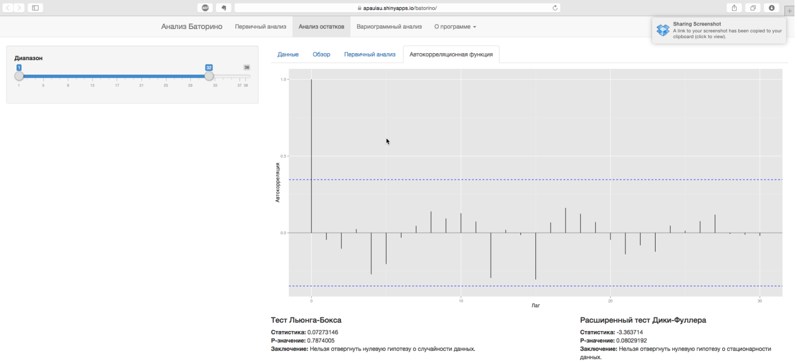


Рис. 4: Автокорреляционная функция

Модуль вариограммного анализа

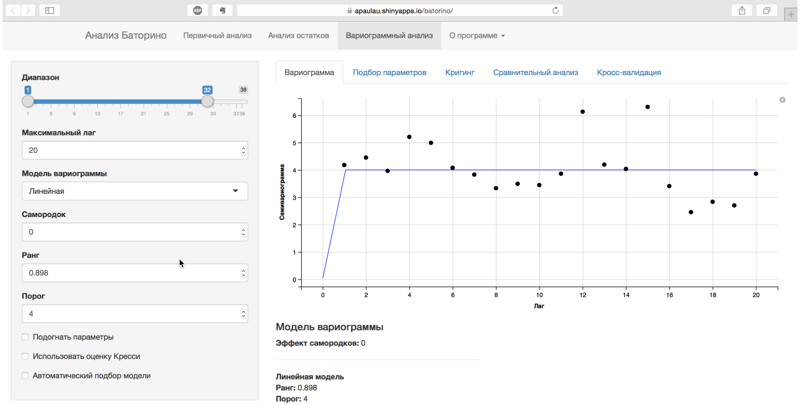


Рис. 5: Возможности по подбору модели вариограммы

Модуль вариограммного анализа

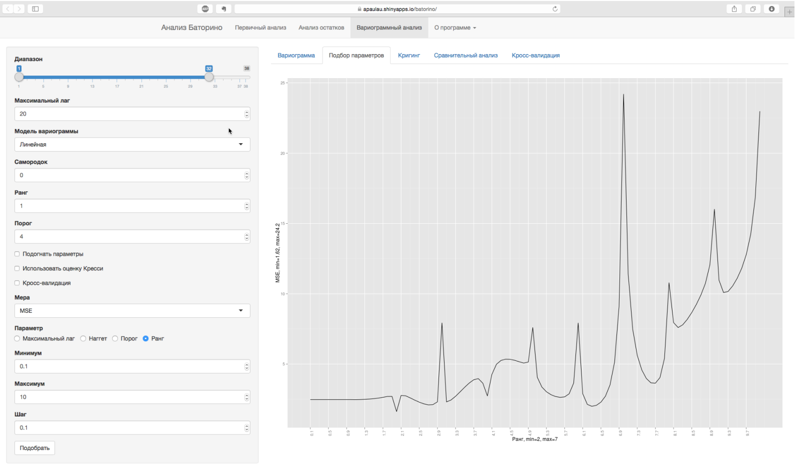


Рис. 6: Подбор параметров модели вариограммы

Модуль вариограммного анализа

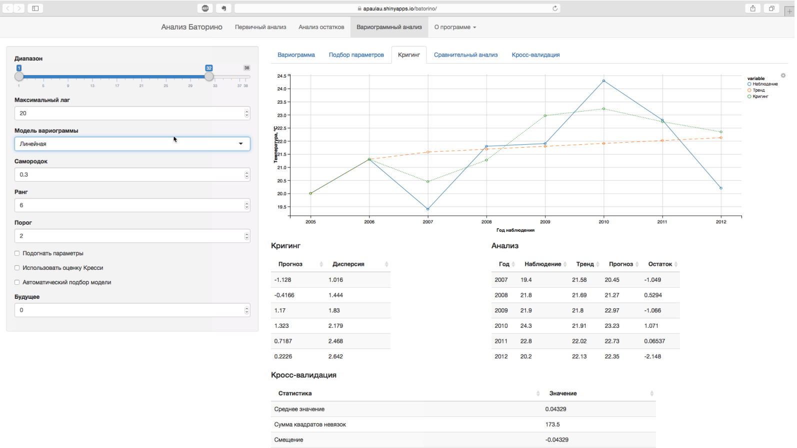


Рис. 7: Сравнение прогнозных значений

Исседуемые 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.

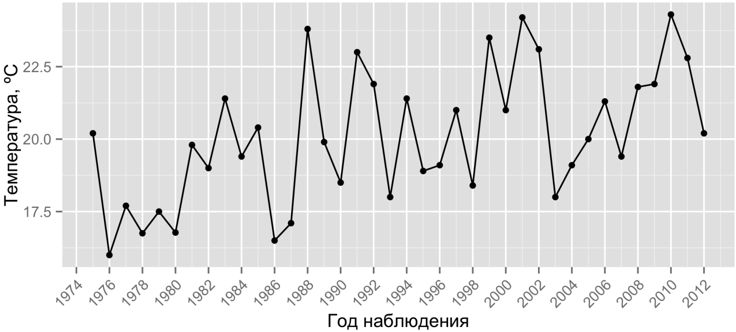


Рис. 8: Исходные данные

Визуально и проверкой критериев Шапиро-Уилка, χ^2-Пирсона и Колмогороваа-Смирнова была показана близость выборочного распределения к нормальному с параметрами $\mathcal{N}(19.77, 5.12)$.

При этом выборочное распредлеение характеризуется небольшой скошенностью вправо (коэффициент асимметрии 0.30) и пологостью пика кривой распределения (-0.746) относительного нормального.

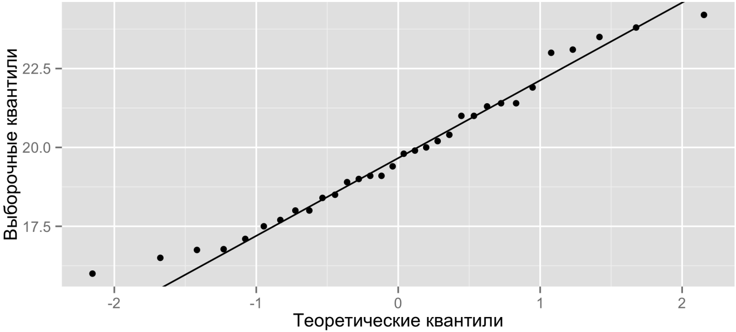


Рис. 9: График квантилей

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.

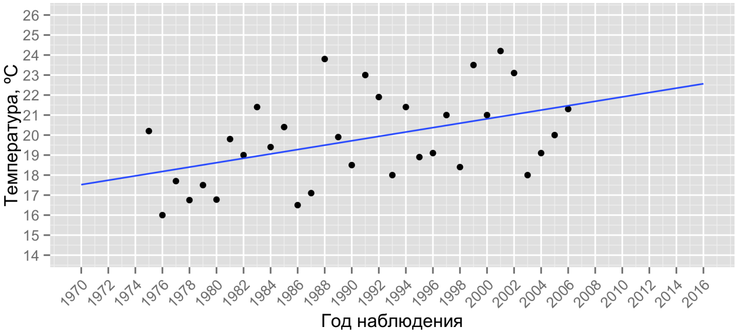


Рис. 10: Диаграмма рассеяния

Регрессионная модель

Выявлено, что исследуемый временной ряд является аддитивным:

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

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

Найдена модель тренда: $y(t) = at + b = 0.1014t + 18.0521$

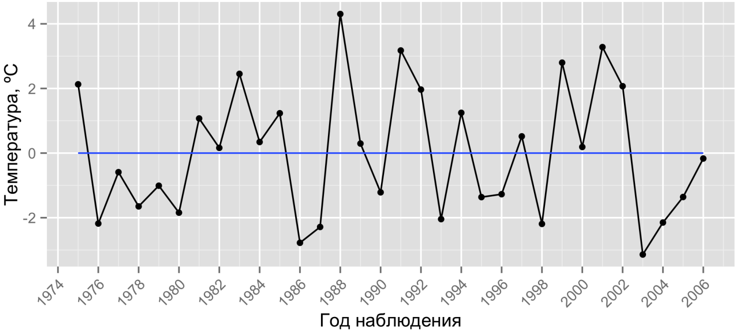


Рис. 11: Ряд остатков $\varepsilon(t)$

- 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 $\eta^2_{x(t)} = 0.275$

	$X(t)$	$y(t)$	$X(t) - y(t)$
2007	19.400	18.071	1.329
2008	21.800	18.181	3.619
2009	21.900	18.290	3.610
2010	24.300	18.400	5.900
2011	22.800	18.509	4.291
2012	20.200	18.619	1.581

Таблица 1: Сравнение прогнозных значений (модель $y(t)$)

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

По графику было сделано предположение об отсутствии значимых автокорреляций. Проведённый тест Льюнга-Бокса подтвердил данное замечание.

Также было отмечено, что значения имеют небольшую амплитуду и имеют тенденцию к затуханию. Это говорит о стационарности в широком смысле, что показал расширенный тест Дики-Фуллера.

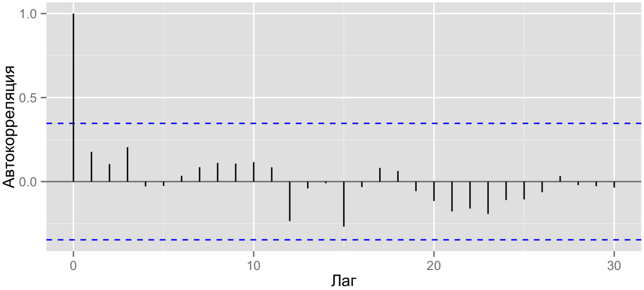


Рис. 12: Автокорреляционная функция

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

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

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

В качестве оценки вариограммы рассматривается статистика, предложенная Матероном:

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

Теорема 1

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

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

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

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

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

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

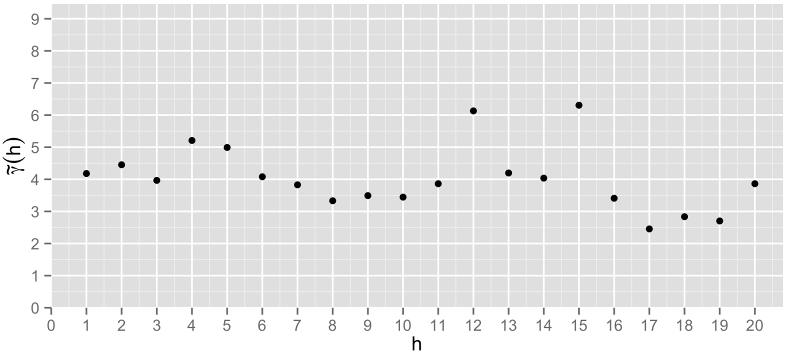
$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty, \text{ то}$$

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

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

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



$$\widehat{\gamma}(h) = c_0 + Lin(h) = \begin{cases} c_0 + b \cdot h, & h > 0, \\ c_0, & h \leq 0, \end{cases} \quad (4)$$

где b – параметр, отвечающий за угол наклона, c_0 — эффект самородков.

Подобранная модель:

$$\widehat{\gamma}_1(h) = Lin(h), \quad b = 4, \quad (5)$$

Показатели качества

$$r_{\varepsilon\varepsilon^*} = -0.09129, \quad MSE = 6.324$$

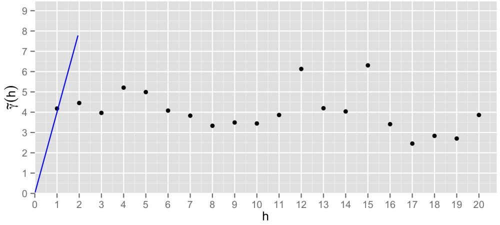


Рис. 13: Модель семивариограммы $\widehat{\gamma}_1(h)$

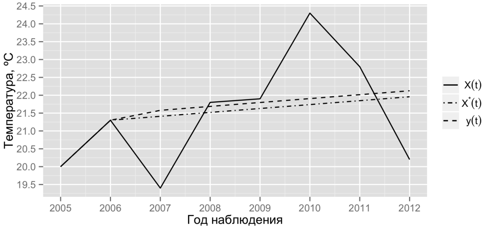


Рис. 14: Прогноз по модели $\widehat{\gamma}_1(h)$

$$\widehat{\gamma}(h) = c \cdot Nug(h) = \begin{cases} 0, & h = 0, \\ c, & h \neq 0, \end{cases} \qquad (6)$$

где b – параметр, отвечающий за угол наклона, c_0 — эффект самородков.

Подобранная модель:

$$\widehat{\gamma}_2(h) = 4.04 \cdot Nug(h). \qquad (7)$$

Показатели качества

$$r_{\varepsilon\varepsilon^*} = -1, \quad MSE = 4.199$$

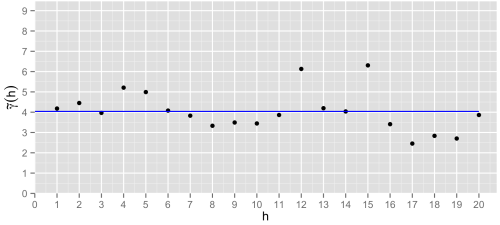


Рис. 15: Модель семивариограммы $\widehat{\gamma}_1(h)$

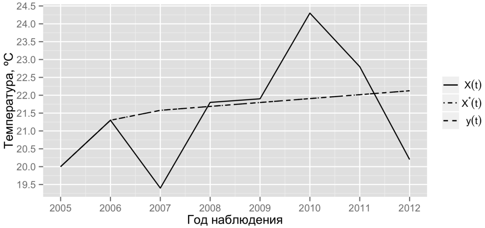


Рис. 16: Прогноз по модели $\widehat{\gamma}_1(h)$

$$\widehat{\gamma}(h) = c_0 + c \cdot Lin(h, a) = \tag{8}$$
$$= \begin{cases} c_0 + c \cdot \frac{h}{a}, & 0 \le h \le a, \\ c_0 + c, & h > a, \end{cases}$$

где c_0 – эффект самородков, c – порог, a – ранг.

Подобранная модель:

$$\widehat{\gamma}_4(h) = 4 \cdot Lin(h, 2). \tag{9}$$

Показатели качества

$$r_{\varepsilon\varepsilon^*} = 0.152, \quad MSE = 18.69$$

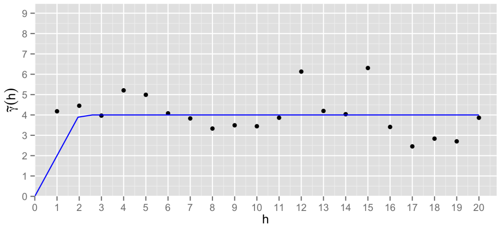


Рис. 17: Модель семивариограммы $\widehat{\gamma}_4(h)$

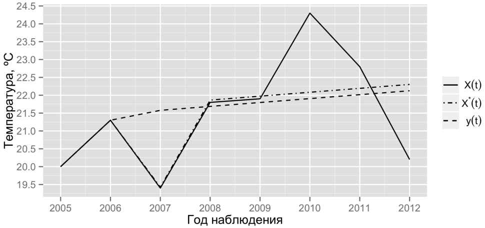


Рис. 18: Прогноз по модели $\widehat{\gamma}_4(h)$

$$\widehat{\gamma}(h) = c_0 + c \cdot Sph(h, a) =$$
$$= \begin{cases} c_0 + c \cdot \left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right), & h \le a, \\ c_0 + c, & h \ge a, \end{cases} \tag{10}$$

где c_0 – эффект самородков, c – порог, a – ранг.

Подобранная модель:

$$\widehat{\gamma}_5(h) = 0.9 + 4Sph(h, 6.9), \tag{11}$$

Показатели качества

$$r_{\varepsilon\varepsilon^*} = -0.009, \quad MSE = 5.396$$

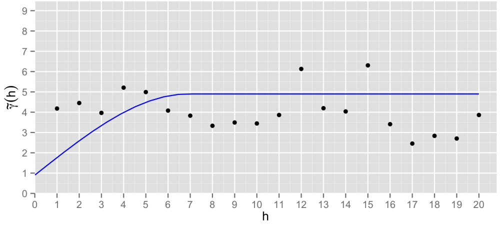


Рис. 19: Модель семивариограммы $\widehat{\gamma}_5(h)$

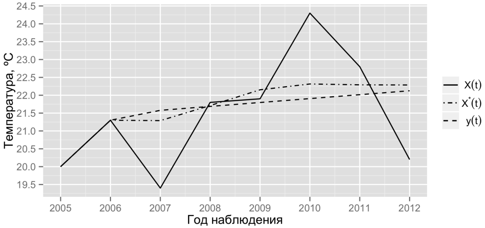


Рис. 20: Прогноз по модели $\widehat{\gamma}_5(h)$

Периодическая модель

$$\widehat{\gamma}(h) = c_0 + c \cdot Per(h, a) = 1 - cos(\frac{2\pi h}{a}), \qquad (12)$$

где c_0 – эффект самородков, c – порог, a – ранг.

Подобранная модель:

$$\widehat{\gamma}_6(h) = 4 \cdot Per(h, 0.898), \qquad (13)$$

Показатели качества

$$r_{\varepsilon\varepsilon^*} = 0.404, \quad MSE = 4.369$$

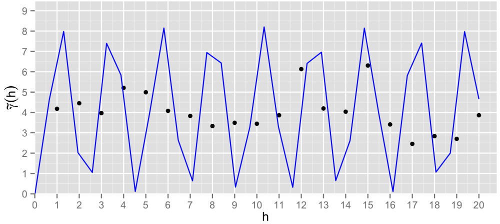


Рис. 21: Модель семивариограммы $\widehat{\gamma}_6(h)$

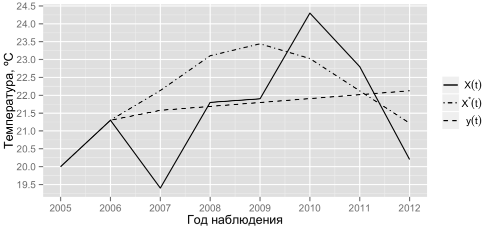
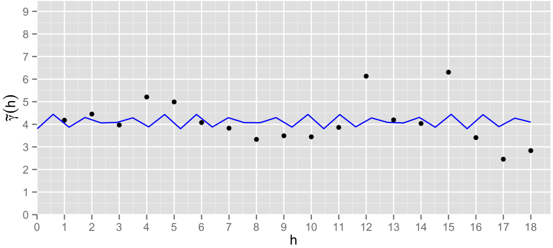
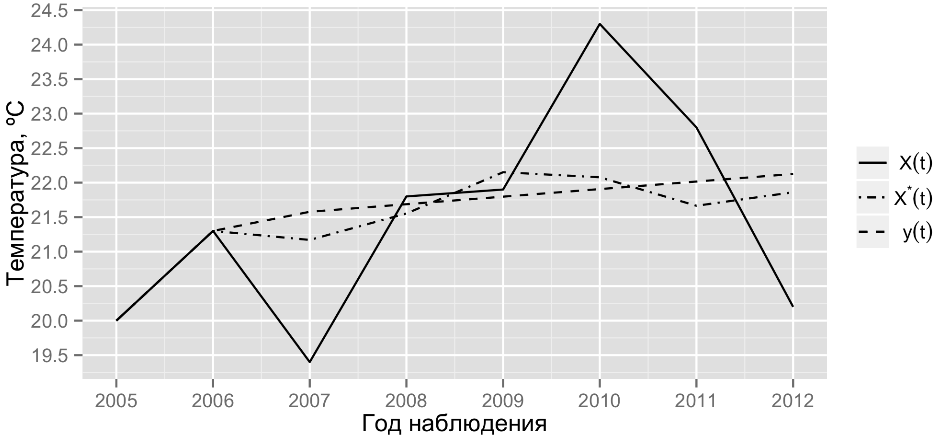


Рис. 22: Прогноз по модели $\widehat{\gamma}_6(h)$

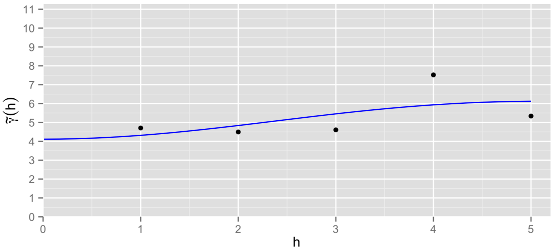
Подобранная модель:
$3.8 + 0.32 \cdot Per(h, 1.3)$



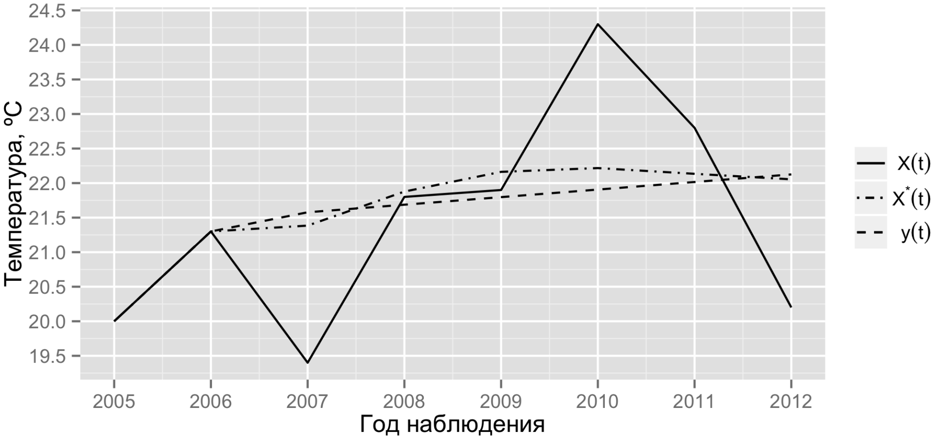
Прогнозирование методом ординарного кригинга



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



Прогнозирование методом ординарного кригинга



Заключение

Спасибо за внимание!