Анализ и прогнозирование гидрологических данных

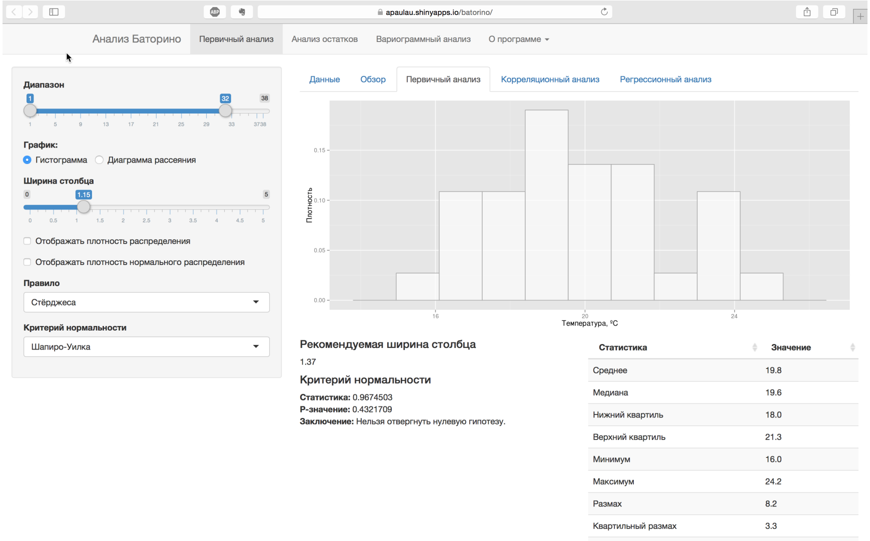
Павлов Александр Сергеевич

26 мая 2015 г.

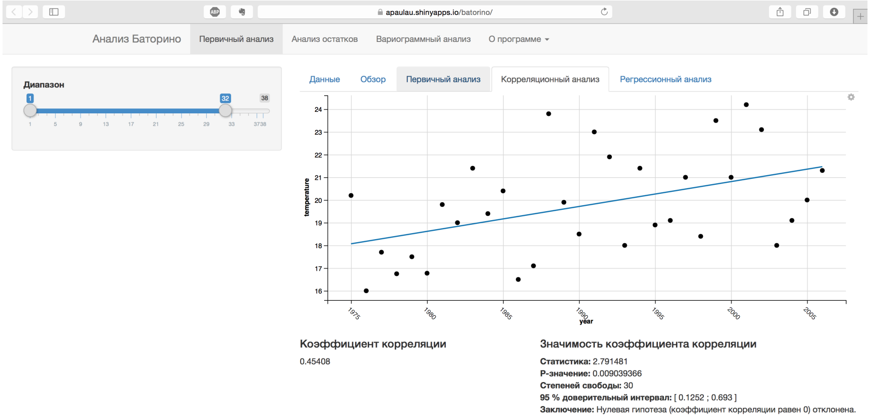
Постановка задачи

- Предварительный статистический анализ данных
- Исследование статистических свойств оценки вариограммы
- Вариограммный анализ
- Прогнозирование значений временного ряда методом Кригинг

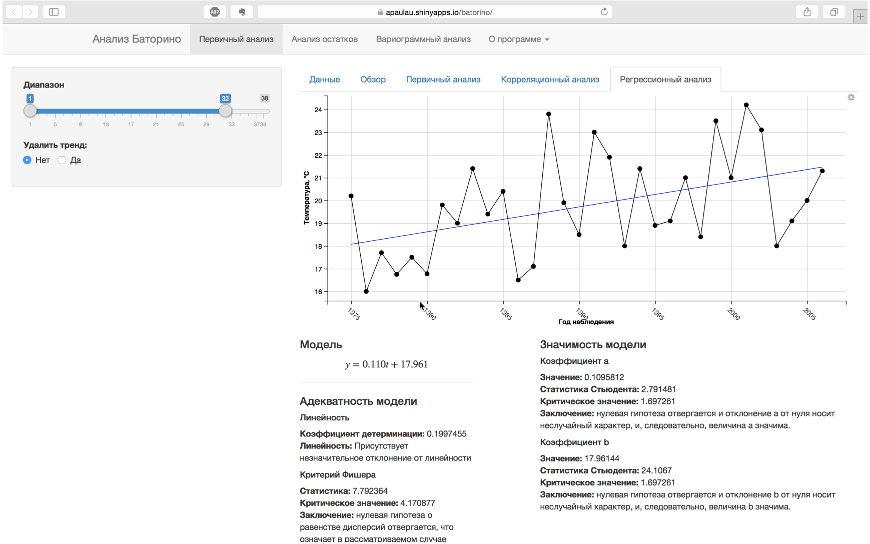
Первичный анализ и описательные статистики



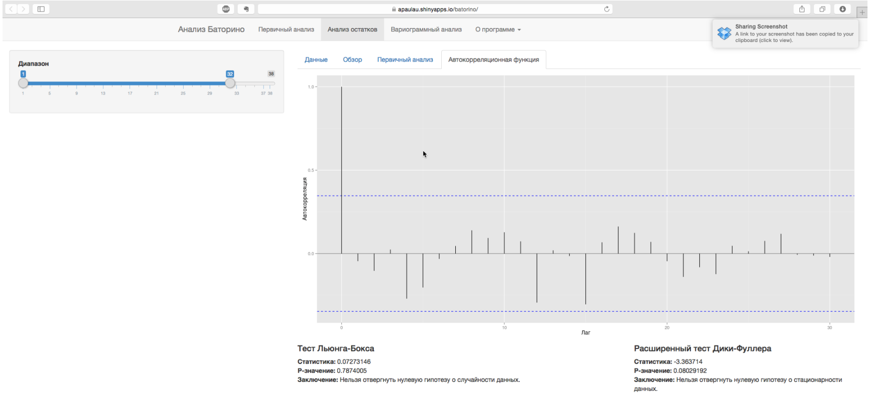
Корреляционный анализ



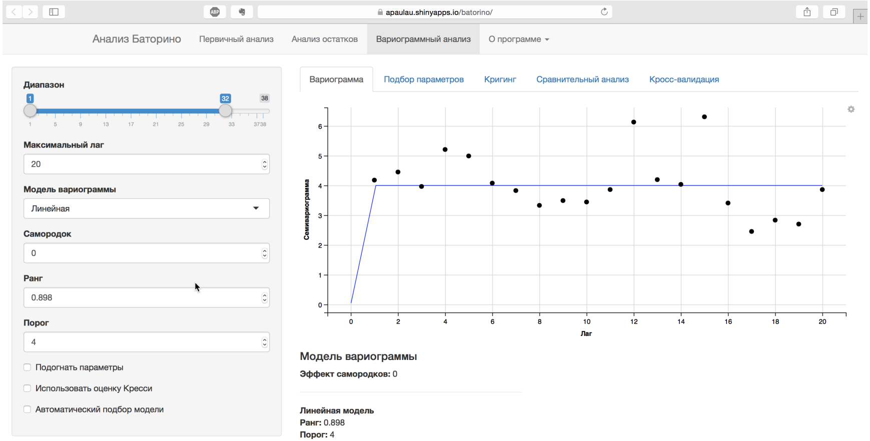
Регрессионный анализ



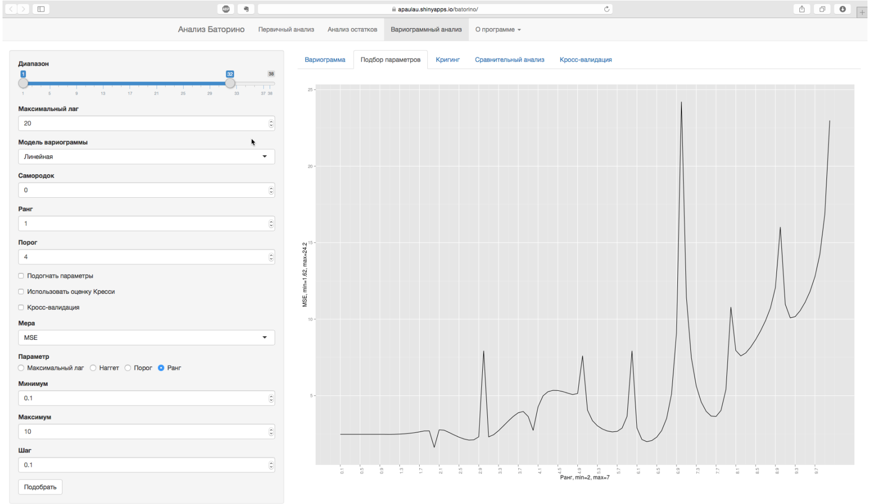
Автокорреляционная функция



Возможности по подбору модели вариограммы



Подбор параметров модели вариограммы



Сравнение прогнозных значений

apaulau.shinyapps.io/batorino/

Анализ Баторино | Первичный анализ | Анализ остатков | Вариограммный анализ | О программе

Диапазон

1 — 32 — 38

Максимальный лаг

20

Модель вариограммы

Линейная

Самородок

0.3

Ранг

6

Порог

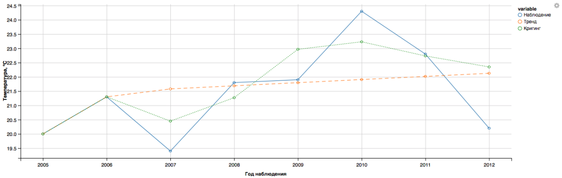
2

- [ ] Подогнать параметры
- [ ] Использовать оценку Кресси
- [ ] Автоматический подбор модели

Будущее

0

Вариограмма | Подбор параметров | Кригинг | Сравнительный анализ | Кросс-валидация



Кригинг

Прогноз	Дисперсия
-1.128	1.016
-0.4166	1.444
1.17	1.83
1.323	2.179
0.7187	2.468
0.2226	2.642

Анализ

Год	Наблюдение	Тренд	Прогноз	Остаток
2007	19.4	21.58	20.45	-1.049
2008	21.8	21.69	21.27	0.5294
2009	21.9	21.8	22.97	-1.066
2010	24.3	21.91	23.23	1.071
2011	22.8	22.02	22.73	0.06537
2012	20.2	22.13	22.35	-2.148

Кросс-валидация

Статистика	Значение
Среднее значение	0.04329
Сумма квадратов невязок	173.5
Смещение	-0.04329

Исходные данные

Исходные данные получены от учебно-научного центра «Нарочанская биологическая станция им. Г.Г.Винберга». На рисунке представлена выборка, состоящая из наблюдений за температурой воды в июле месяце в период с 1975 по 2012 годы.

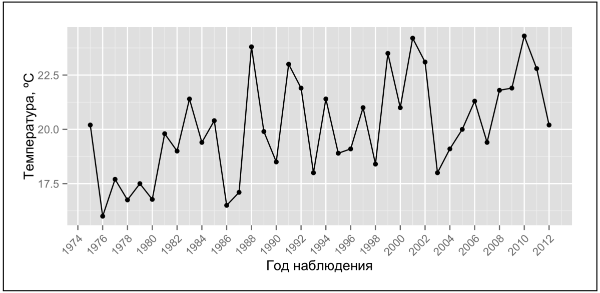


График квантилей

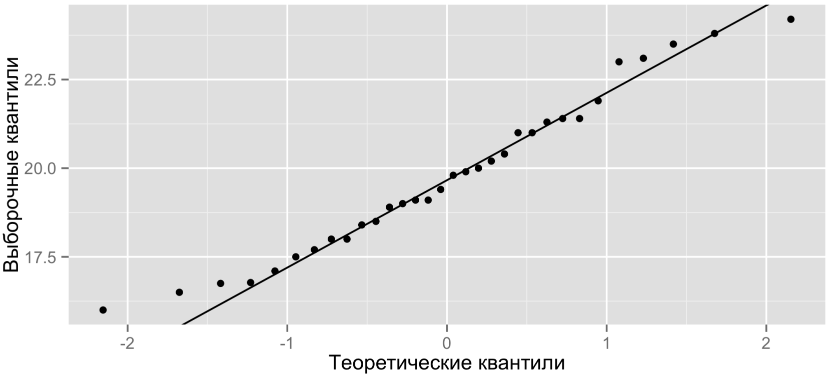
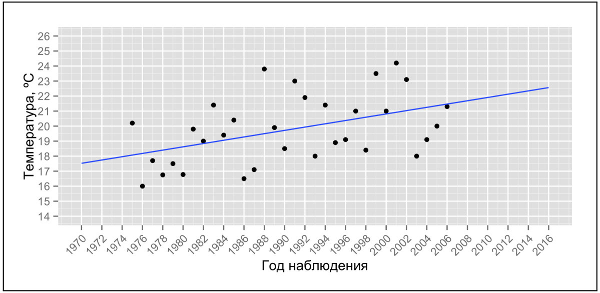


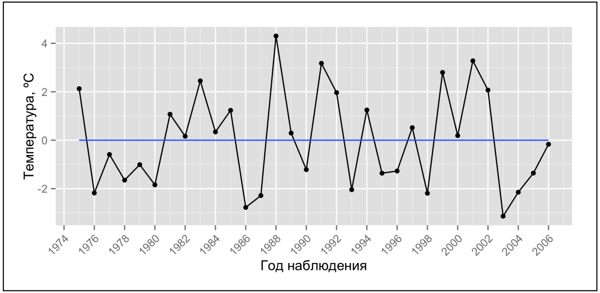
Диаграмма рассеяния

Выборочный коэффициент корреляции: $r_{xt} = 0.454$.
При уровне значимости $\alpha = 0.05$ коэффициент является значимым



Временной ряд

Вид исходного временного ряда:
$x(t) = y(t) + \varepsilon(t)$
Уравнение тренда:
$y(t) = at + b = 0.1014t + 18.0521$



Оценка модели

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

Оценка вариограммы

В качестве оценки вариограммы рассматривается статистика, предложенная Матероном:

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

Первые два момента оценки вариограммы

Теорема

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* $X(t), t \in \mathbb{Z}$, $h, h_1, h_2 = \overline{0, n-1}$.

Асимптотическое поведение оценки вариограммы

Теорема

Если имеет место соотношение

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

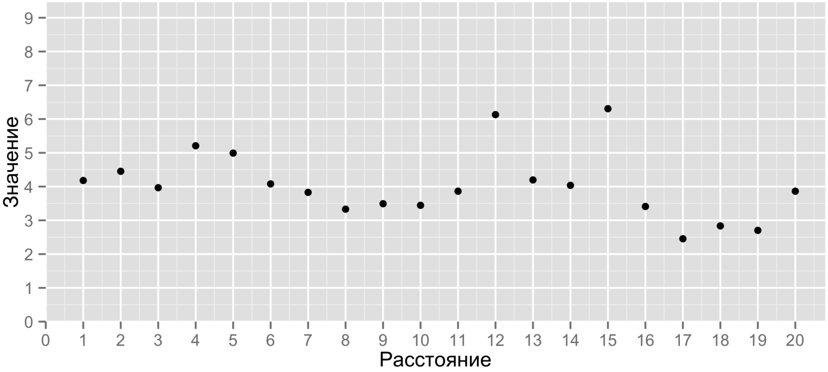
$$\lim_{n\to\infty} (n - \min\{h_1, h_2\}) cov\{2\tilde{\gamma}(h_1), 2\tilde{\gamma}(h_2)\} =$$

$$= 2 \sum_{m=-\infty}^{+\infty} \gamma(m - h_2) + \gamma(m + h_1) - \gamma(m) - \gamma(m + h_1 - h_2))^2,$$

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

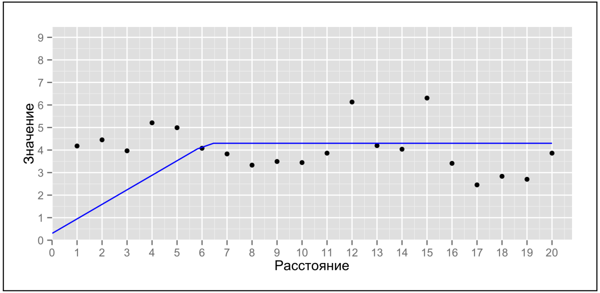
где $\gamma(h), h \in \mathbb{Z},$ *— семивариограмма процесса* $X(t), t \in \mathbb{Z},\ h, h_1, h_2 = \overline{0, n-1}.$

График экспериментальной вариограммы

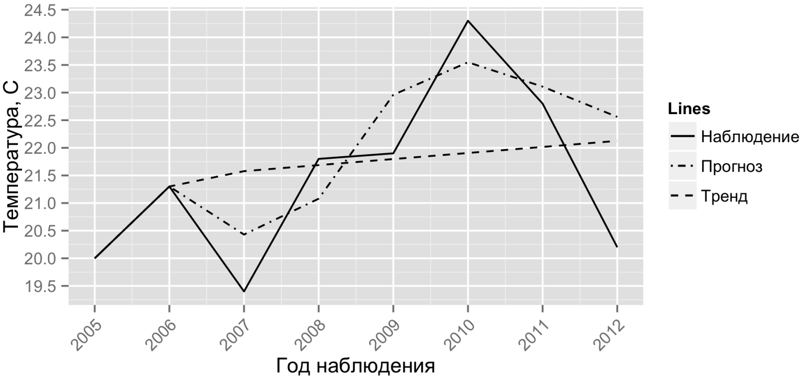


Линейная модель с порогом

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



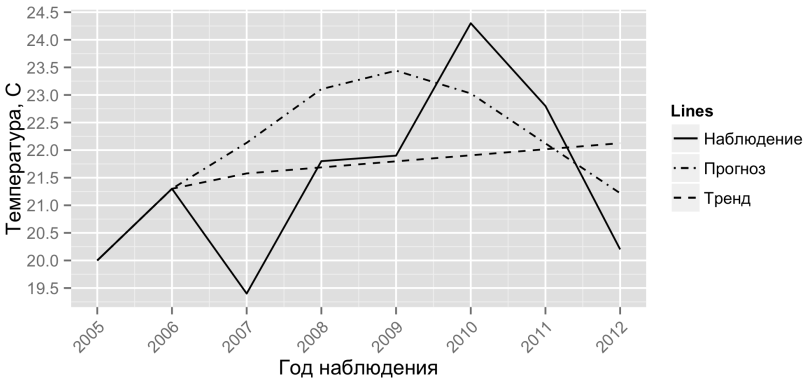
Прогнозирование методом ординарного кригинга



Периодическая модель

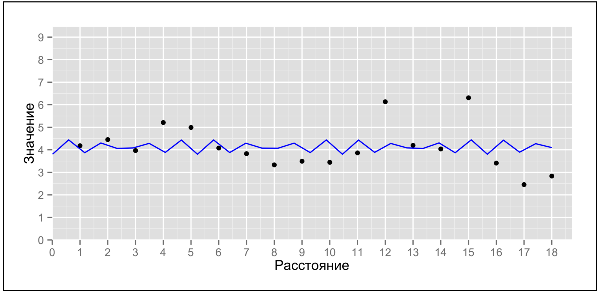
Подобранная модель:
$4 \cdot Per(h, 0.898)$

Прогнозирование методом ординарного кригинга

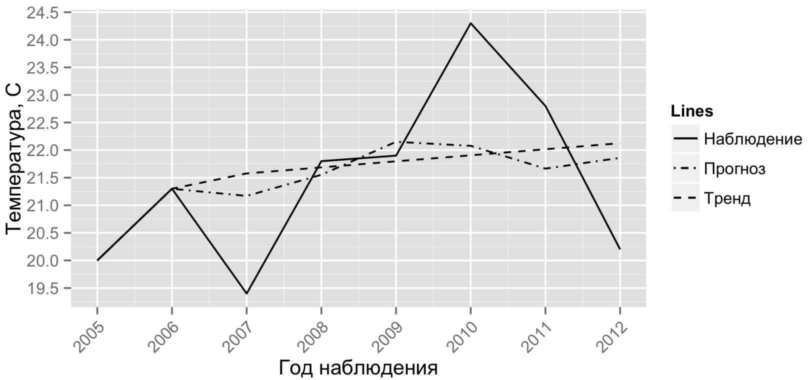


Периодическая модель

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

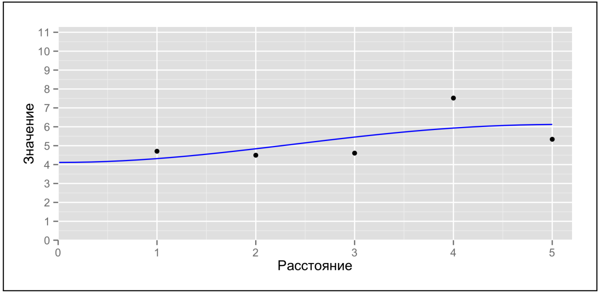


Прогнозирование методом ординарного кригинга



Волновая модель

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



Прогнозирование методом ординарного кригинга