

# ОЦЕНКИ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК НЕОДНОРОДНЫХ МЕТЕОРОЛОГИЧЕСКИХ ПОЛЕЙ ПО ВЫБОРКАМ МАЛОГО ОБЪЕМА

С. С. Скворцов<sup>1</sup>, О. В. Сересева<sup>1</sup>, Н. А. Каргаполова<sup>1,2</sup>

<sup>1</sup> *Институт вычислительной математики и математической геофизики СО РАН, 630090, Новосибирск*

<sup>2</sup> *Новосибирский государственный университет, 630090, Новосибирск*

УДК 519.6, 551.501

В работе приведены результаты анализа по выборкам малого объема реальных данных статистической структуры случайного поля среднесуточной приземной температуры воздуха на территории Новосибирской области. Приведены результаты исследования одномерного распределения рассматриваемого поля, построены соответствующие аппроксимации аналитическими функциями. Для аппроксимации распределений температуры использованы смеси нормальных распределений. На основе специально разработанного критерия были проверены гипотезы об однородности и изотропности поля по корреляциям. Полученные результаты являются входными параметрами стохастической модели совместных полей среднесуточной температуры и относительной влажности.

В приоритет была поставлена высокая плотность станций на территории, а также отсутствие пропусков в наблюдениях. Таким образом, данные, удовлетворяющие условиям, представляют наблюдения на 30 метеорологических станциях Новосибирской области в период с 2001 по 2004 год. Наблюдаемый месяц — Январь.

**Ключевые слова:** стохастическая модель, однородные поля, изотропные поля, метод обратной функции, смесь нормальных распределений.

## Введение

Данная работа связана с разработкой численных стохастических параметрических моделей метеорологических пространственных двумерных полей с использованием данных наблюдений с метеорологических станций. При построении моделей такого типа обычно учитывается одномерное распределение вероятностей и корреляционная функция. Во многих случаях одномерное распределение может зависеть от пространственных координат, а значения корреляционной функции от пространственных координат соответствующих двух точек области. На первом этапе построения модели необходимо выбрать подходящее приближение, то есть определить, зависит ли одномерное распределение от пространственных координат и является ли поле однородным изотропным, однородным или неоднородным по корреляциям.

Целью работы является проверка предположения однородности и изотропности поля среднесуточной температуры по корреляциям, по имеющимся наблюдениям на 30 метеорологических станциях Новосибирской области в период с 2001 по 2004 год (наблюдаемый месяц — Январь). По этим данным в работе исследуется алгоритм проверки степени однородности поля, предложенный в работе [1], рассматривается также алгоритм проверки степени однородности изотропности поля.

Полученные результаты будут использованы для построения стохастической модели совместных полей среднесуточной температуры и относительной влажности.

---

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (коды проектов: гранты РФФИ № 15-01-01458-а, 16-31-00038-мол-а, 16-31-00123-мол-а, 17-41-543338-р-мол-а и Грант президента РФ МК-659.2017).

# 1 Аппроксимация плотности распределения

Для нормализации поля среднесуточной температуры воздуха необходимо знать одномерное распределение этой величины. Метеорологические станции расположены не равномерно на рассматриваемой территории, половина из них расположена вблизи Обского моря (подобласть А), вторая половина на оставшейся территории (подобласть В). Поэтому в работе вся область была поделена на эти две части и по соответствующим группам станций отдельно оценивались одномерные распределения.

Поле среднесуточной температуры полагается однородным по распределению, в рамках каждой подобласти, и, вообще говоря, неоднородным по всей области. Функции распределения температуры будут аппроксимированы смесями нормальных распределений для каждой подобласти при помощи ЕМ-алгоритма [3].

Плотность распределения, выбранная для среднесуточной температуры подобласти А, является смесью из трех нормальных компонент:

$$f_A(x) = \sum_{i=1}^3 \frac{a_i}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}},$$

где

$$\begin{aligned} a_1 &= 0.30665732, \mu_1 = -5.81991733, \sigma_1^2 = 9.20264129, \\ a_2 &= 0.03351083, \mu_2 = -37.93288998, \sigma_2^2 = 16.07278080, \\ a_3 &= 0.65983184, \mu_3 = -19.75634399, \sigma_3^2 = 25.47740525. \end{aligned}$$

В таблице 1 приведены выборочные значения средних, дисперсий, коэффициентов асимметрии и эксцесса для поля среднесуточной температуры воздуха в среднем по всем станциям подобласти А и соответствующие им аналогичные характеристики, вычисленные по плотности  $f_A(x)$

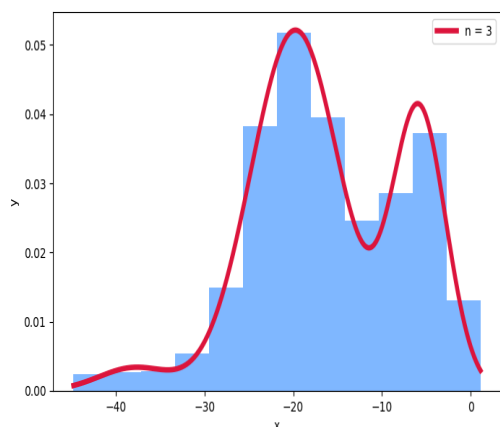


Рис. 1: Эмпирическая и модельная плотности распределения поля среднесуточной температуры воздуха (подобласть А).

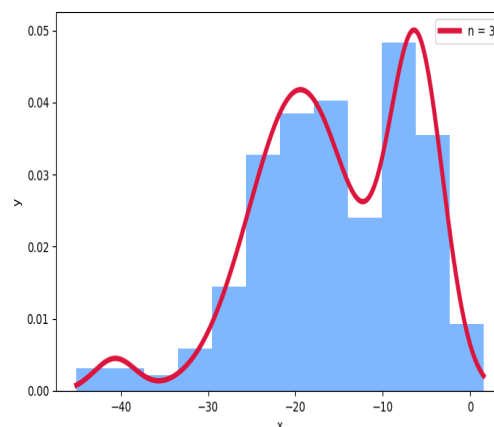


Рис. 2: Эмпирическая и модельная плотности распределения поля среднесуточной температуры воздуха (подобласть В).

Таблица 1: Характеристики реального и модельного распределения поля среднесуточной температуры воздуха (подобласть А).

Характеристика	Реальные данные	Модель
Среднее	-16.0917	-16.1047
Дисперсия	77.3740	77.6368
Асимметрия	-0.2684	-0.2622
Эксцесс	-0.1381	-0.1439

Плотность распределения  $f_B(x)$ , выбранная для среднесуточной температуры из подобласти В, так же является смесью из трех нормальных компонент вида  $f_A(x)$  с параметрами:

$$\begin{aligned} a_1 &= 0.35918038, \mu_1 = -6.14431458, \sigma_1^2 = 9.43806469, \\ a_2 &= 0.02652844, \mu_2 = -40.66030923, \sigma_2^2 = 5.65727484, \\ a_3 &= 0.61429118, \mu_3 = -19.42359541, \sigma_3^2 = 34.42618248. \end{aligned}$$

В таблице 2 приведены характеристики, аналогичные характеристикам из таблицы 1, для подобласти В.

Таблица 2: Характеристики реального и модельного распределения поля среднесуточной температуры воздуха (подобласть В).

Характеристика	Реальные данные	Модель
Среднее	-15.2173	-15.2310
Дисперсия	82.2968	82.3375
Ассиметрия	-0.5435	-0.5394
Эксцесс	0.0745	0.0312

## 2 Нормализация данных, оценка корреляционной матрицы и ее статистической погрешности

Под нормализацией негауссовской случайной величины в данной работе понимается преобразование вида  $T_{Nor} = \Phi^{-1}(F(T))$ , где  $F(x) = \int_{-\infty}^x f(x)dx$  — функция распределения среднесуточной температуры,  $f(x)$  — функция плотности распределения среднесуточной температуры, полученная аппроксимацией наблюдений, а  $\Phi(x)$  — функция стандартного нормального распределения.

Ввиду того, что для моделирования используются нормализованные данные, исследование корреляционной структуры поля так же будет проведено на этих нормализованных данных.

$$T_{Nor} = \Phi^{-1}(F(T)),$$

где в зависимости от подобласти,  $F(x)$  принимает вид либо  $F_A(x)$ , либо  $F_B(x)$ .

По нормализованным данным, соответствующим реальным данным о среднесуточной температуре воздуха на всех станциях Новосибирской области оценивается корреляционная матрица

$$R_{Nor} = (r_{ij}) = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1S} \\ r_{12} & r_{22} & \dots & r_{2S} \\ \dots & \dots & \dots & \dots \\ r_{1S} & r_{2S} & \dots & r_{SS} \end{pmatrix}.$$

Для оценки используются стандартные формулы для выборочной корреляционной матрицы по данным объема  $N = MY$  где  $M = 31$  — число дней в месяце,  $Y = 4$  — число лет наблюдений,  $S$  — число станций во всей области.

Оценка статистической погрешности выборочных коэффициентов корреляции (элементов матрицы  $R_{Nor}$ ) осуществляется следующим образом:

1. Моделируется  $N$  нормальных векторов размерности  $S$  с корреляционной матрицей  $R_{Nor}$ , где  $N = MY$  определенный выше объем выборки.
2. По полученным модельным данным (совпадающим по структуре и объему с реальными данными) оценивается модельная корреляционная матрица  $R^m$
3. Процедура 1. – 2. повторяется  $K$  раз, образуя ряд модельных нормализованных корреляционных матриц  $R_1^m, R_2^m, \dots, R_K^m$ ,  $K = 10^5$ , ( $R_k^m = (r_{k,ij}^m)$ ).

По полученному ряду оценивается стандартное отклонение каждого выборочного коэффициента корреляции из матрицы  $R_{Nor}$  по формуле

$$\Sigma_{Nor} = (\sigma_{ij}), \quad \sigma_{ij} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (r_{ij,k}^m - \bar{r}_{ij}^m)^2}, \quad \bar{r}_{ij}^m = \frac{1}{K} \sum_{k=1}^K r_{ij,k}^m, \quad i, j = 1, \dots, S.$$

По реальным данным объем выборки  $N$  равен 119, а число выборок  $K$  в расчетах, как отмечалось, бралось равным 100000.

### 3 Исследование корреляционной структуры поля нормализованной среднесуточной температуры воздуха.

Каждому коэффициенту корреляции из матрицы  $R_{Nor}$  соответствуют две станции  $i$  и  $j$  с координатами  $x_i, y_i$  и  $x_j, y_j$ . Поэтому для каждой пары станций имеется следующий вектор значений:  $(r_{ij}, x_i - x_j, y_i - y_j, \sigma_{ij})$ , где  $r_{ij}$  — оцененный коэффициент корреляции матрицы  $R_{Nor}$  между  $i$ -ой и  $j$ -ой станциями,  $x_i, y_i$  — координаты соответствующей станции,  $\sigma_{ij}$  — оцененная выше соответствующая статистическая погрешность коэффициента корреляции.

Таким образом элементы  $r_{ij}$  корреляционной матрицы  $R_{Nor}$  являются значениями корреляционной функции  $r_{ij} = r(x_i, y_i, x_j, y_j)$  дискретного аргумента. Поскольку станции в области расположены не равномерно, то по структуре матрицы  $R_{Nor}$  без специального анализа трудно определить, является ли поле однородным изотропным, однородным или неоднородным. В работе [1] рассмотрен специальный способ оценки степени однородности поля по реальным данным, который состоит в том, что среди всевозможных двойных пар станций (точек) в рассматриваемой сети, отбираются те двойные пары, в которых по заранее выбранному критерию, отрезки, соединяющие точки в каждой паре, можно считать параллельными и равными.

#### 3.1 Предположение изотропности поля

Для проверки предположения об изотропности поля необходимо среди всевозможных двойных пар точек найти такие пары, в которых с точки зрения некоторого критерия, отрезки, соединяющие точки в каждой паре, можно было бы считать равными по длине. В качестве такого критерия в данной работе выбирается следующее условие: пусть нам известна корреляционная функция  $r(x_i, x_j, y_i, y_j) = r(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2})$  рассматриваемого поля. Если для поля на станциях с такой корреляционной функцией выполняется условие:

$$\left| r\left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\right) - r\left(\sqrt{(x_k - x_l)^2 + (y_k - y_l)^2}\right) \right| \leq \varepsilon(\sigma_{ij} + \sigma_{kl}) \quad (1)$$

где  $i, j, k, l$  — индексы соответствующих станций,  $\sigma_{ij}$  — статистическая погрешность реального коэффициента корреляции  $r_{ij}$  нормализованного поля температуры, то отрезки, соединяющие соответствующие пары станций считаются равными. В качестве корреляционной функции  $r(x_i, x_j, y_i, y_j) = r(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2})$  выбирается функция вида  $r(x_i, x_j, y_i, y_j) = e^{-a\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}^q}$ , в которой параметры  $a$  и  $q$  выбираются таким образом, чтобы средний квадрат разности между ней и  $r_{ij} = r(x_i, y_i, x_j, y_j)$  был минимален. Значения этих параметров приведены в таблице 3.

Таблица 3: Параметры аппроксимирующей корреляционной функции изотропного поля

Параметр	Значение
$a$	0.0009380132975188
$q$	1.4127988376958411

Данное условие, по сути является условием равноудаленности точек в разных парах, то есть выбираются только те четверки станций, где расстояние между станциями в одной паре близко к расстоянию между станциями в другой паре (близость определяется величиной  $\varepsilon$ ). После того, как выбраны подходящие пары, необходимо проверить, близки ли на них коэффициенты корреляции.

Для этого используется условие:

$$|r_{ij} - r_{kl}| \leq \sigma_{ij} + \sigma_{kl}. \quad (2)$$

Выполнение условия (2) означает, что разница коэффициентов корреляций между значениями поля на этих двух парах станций, расстояния между которыми крайне близки, не больше статистической погрешности, что говорит в пользу предположения об изотропности поля.

Таблица 4: Показатели степени изотропности поля для различных значений  $\varepsilon$ .

$\varepsilon$	$N$ — количество четверок станций, удовлетворяющих условию	$M$ — количество четверок станций, удовлетворяющих условиям (1) и (2)	$\frac{N}{M} \times 100$
0.1	1286	367	28.54%
0.09	1165	330	28.33%
0.08	1026	292	28.46%
0.07	908	251	27.64%
0.06	789	221	28.01%
0.05	658	178	27.05%
0.04	513	143	27.88%
0.03	367	108	29.43%
0.02	238	66	27.73%
0.01	108	27	25.00%

Из таблицы 4 видно, что в пользу предположения изотропности нормализованного поля температуры говорит в среднем только 28% случаев, для которых были выполнены условия (1) и (2), причем этот показатель слабо зависит от выбора  $\varepsilon$ .

Таким образом приближение изотропности представляется весьма грубым.

### 3.2 Предположение однородности поля

Для проверки предположения об однородности поля, необходимо среди всевозможных двойных пар точек найти такие, в которых с точки зрения некоторого критерия, отрезки, соединяющие точки в каждой паре, можно было бы считать параллельными и равными по длине.

В качестве такого критерия в данной работе выбирается следующее условие: пусть нам известна корреляционная функция рассматриваемого поля  $r(x_i, x_j, y_i, y_j) = r(x_i - x_j, y_i - y_j)$ . Если для поля на станциях с такой корреляционной функцией выполняется условие:

$$|r(x_i - x_j, y_i - y_j) - r(x_k - x_l, y_k - y_l)| \leq \varepsilon(\sigma_{ij} + \sigma_{kl}), \quad (3)$$

где  $i, j, k, l$  — индексы соответствующих станций,  $\sigma_{ij}$  — статистическая погрешность реального коэффициента корреляции  $r_{ij}$  нормализованного поля температуры, то отрезки, соединяющие соответствующие пары станций считаются параллельными и равными по длине. В качестве корреляционной функции  $r(x_i, x_j, y_i, y_j) = r(x_i - x_j, y_i - y_j)$  выбирается функция вида

$$r = e^{-(a(x_i - x_j)^2 + b(x_i - x_j)(y_i - y_j) + c(y_i - y_j)^2)^q}, \quad (4)$$

в которой параметры  $a, b, c, q$  выбираются таким образом, чтобы средний квадрат разности между ней и  $r_{ij} = r(x_i, y_i, x_j, y_j)$  был минимален. Значения этих параметров приведены в таблице 5.

Данное условие является условием параллельности и равенства по длине отрезков, соединяющих пары станций в рассматриваемой четверке станций (здесь близость и параллельность определяются величиной  $\varepsilon$ ).

После того, как выбраны подходящие пары, необходимо проверить, близки ли на них коэффициенты корреляции.

Для этого также используется условие (2). Выполнение этого условия означает, что разница коэффициентов корреляций между значениями поля на этих двух парах станций, расстояния между которыми крайне близки, а соответствующие отрезки параллельны, не больше статистической погрешности. Это говорит в пользу предположения об однородности поля.

Таблица 5: Параметры аппроксимирующей корреляционной функции однородного поля

Параметр	Значение
$a$	4.65786848e-05
$b$	-5.46216340e-05
$c$	8.28846032e-05
$q$	7.26780874e-01

Таблица 6: Показатели степени однородности поля для различных значений  $\varepsilon$ .

$\varepsilon$	$N$ — количество четв- рок станций, удовлетво- ряющих условию (3)	$M$ — количество четв- рок станций, удовлетво- ряющих условиям (3) и (2)	$\frac{N}{M} \times 100$
0.1	1126	692	61.46%
0.09	1014	627	61.83%
0.08	906	561	61.92%
0.07	788	485	61.55%
0.06	650	400	61.54%
0.05	563	347	61.63%
0.04	454	283	62.33%
0.03	353	224	63.46%
0.02	228	144	63.16%
0.01	116	69	59.48%

Из таблицы 6 видно, что в пользу предположения однородности нормализованного поля температуры говорит в среднем только 62% случаев, для которых были выполнены условия (3) и (2), причем этот показатель слабо зависит от выбора  $\varepsilon$ .

Такое приближение принимается как удовлетворительное.

## Заключение

После выбора соответствующего приближения, в данном случае приближения однородного поля, численная параметрическая стохастическая модель полей температуры будет строиться с выбранными в данной работе одномерными распределениями и корреляционной функцией. В качестве продолжения данной работы предполагаются аналогичные исследования для полей среднесуточной относительной влажности по выборкам такого же объема, которые имеют более сложную корреляционную структуру. Отметим, что если при построении стохастических моделей реальных процессов, ни одно из рассмотренных приближений (однородности изотропности и однородности) не подходит для построения модели, то необходимо рассматривать подходы к моделированию неоднородных полей. Некоторые из приближенных способов построения неоднородных полей рассмотрены в работе [2].

## Список литературы

- [1] Огородников В.А., Сересева О.В. Мультипликативная численная стохастическая модель полей суточных сумм жидких осадков и ее использование для оценки статистических характеристик экстремальных режимов их выпадения. // Оптика атмосферы и океана. 2015. Т. 28. № 03. С. 238–245
- [2] Ogorodnikov V. and Sereseva O. Approximate numerical modeling of inhomogeneous stochastic fields of daily sums of liquid precipitation. // Russian Journal of Numerical Analysis and Mathematical Modelling. — 2014. — Vol. 29, No. 6. — P.375–382.

- [3] The EM algorithm // The Elements of Statistical Learning. — New York: Springer, 2001 — P. 236–243

*Степан Сергеевич Скворцов — аспирант Института  
вычислительной математики и математической геофизики СО РАН;  
e-mail: skvortsovstepan54@gmail.com;*

*Ольга Владимировна Сересева — к.ф.-м.н., мл. науч. сотрудник Института вычислительной  
математики и математической геофизики СО РАН;  
e-mail: sereseva@mail.ru;*

*Нина Александровна Каргаполова — к.ф.-м.н., науч. сотрудник Института вычислительной  
математики и математической геофизики СО РАН; Новосибирский государственный университет;  
e-mail: nkargapolova@gmail.com.*

*Дата поступления — 31 мая 2017 г.*