

Алгоритм вычисления М-оценок параметров авторегрессионного поля

07, июль 2013

DOI: 10.7463/0713.0571094

Горяинов В. Б.

УДК 519.234.3

Россия, МГТУ им. Н.Э. Баумана
vb-goryainov@mail.ru

Введение

Рассмотрим модель пространственной авторегрессии — стационарное поле X_{ij} , описываемое уравнением

$$X_{ij} = a_{10}X_{i-1,j} + a_{01}X_{i,j-1} + a_{11}X_{i-1,j-1} + \varepsilon_{ij}, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad (1)$$

где $a = (a_{10}, a_{01}, a_{11})^T$ — авторегрессионные коэффициенты, а ε_{ij} — независимые одинаково распределенные случайные величины с нулевым математическим ожиданием $E\varepsilon_{ij} = 0$. Такие поля широко используются в естественных, технических и гуманитарных науках (см., например, [1, 2, 3, 4]).

Одной из важных задач является построение оценок коэффициентов a по наблюдениям авторегрессионного поля X_{ij} . Наиболее распространенным методом оценивания до сих пор является метод максимального правдоподобия [5]. Метод максимального правдоподобия предполагает известным тип функции распределения $F(x)$ обновляющего поля ε_{ij} . В частности, если ε_{ij} имеют нормальное распределение, то оценки максимального правдоподобия совпадают с оценками наименьших квадратов и с асимптотически эквивалентными им оценками типа Юла — Уолкера [6]. Если предположения о типе $F(x)$ выполняются абсолютно точно, то метод максимального правдоподобия является оптимальным. Однако при решении практических задач тип распределения $F(x)$ обновляющего поля ε_{ij} известен лишь приближенно или неизвестен совсем. В этом случае оценки максимального правдоподобия теряют свою эффективность [7].

Одной из альтернатив оценкам максимального правдоподобия являются М-оценки (оценки, обобщающие оценки максимального правдоподобия). М-оценки хорошо зарекомендовали себя в более простых моделях (одно- и двухвыборочная задача и линейная регрессия [8, 9],

процессы авторегрессии-скользящего среднего [10]). В этих моделях М-оценки оказались робастными, т.е. слабо чувствительными к нарушению предположения о виде распределения наблюдений. Свойства М-оценок коэффициентов авторегрессионного уравнения (1) изучались в [7], где была доказана асимптотическая нормальность М-оценок и вычислена асимптотическая относительная эффективность М-оценок по отношению к оценкам наименьших квадратов. Там же было показано, что уже при небольшом нарушении предположения о нормальности ε_{ij} использование М-оценок предпочтительнее оценок наименьших квадратов.

Нахождение М-оценки коэффициентов a в модели (1) сводится к решению системы уравнений, для чего формально можно использовать метод Ньютона (также известный как метод касательных) — итерационный численный метод нахождения нуля заданной функции. Однако специфика этой системы уравнений делает процедуру вычисления М-оценок методом касательных Ньютона ненадежной. Возникает потребность в алгоритме, который является заведомо сходящимся.

В данной работе построен алгоритм вычисления М-оценок коэффициентов a в модели (1) и доказана его сходимости. Алгоритм представляет собой итерационный вариант взвешенного метода наименьших квадратов с пересчитываемыми на каждом шаге весами и основан на идее из [11].

1. Определение М-оценок

Рассмотрим поле (1), где $a = (a_{10}, a_{01}, a_{11})$ — неизвестный вектор параметров, а ε_{ij} — независимые одинаково распределенные случайные величины с плотностью $f(x)$. Как показано в [6], достаточным условием стационарности X_{ij} является отсутствие корней уравнения

$$1 - a_{10}z_1 - a_{01}z_2 - a_{11}z_1z_2 = 0$$

внутри единичного полидиска $|z_1| \leq 1, |z_2| \leq 1$, что равносильно выполнению условий [12]

$$\begin{aligned} |a_{pq}| &< 1, \quad (p, q) \in \mathcal{I}; \\ (1 + a_{10}^2 - a_{01}^2 - a_{11}^2)^2 - 4(a_{10} + a_{01}a_{11})^2 &> 0; \\ 1 - a_{01}^2 &> |a_{10} + a_{01}a_{11}|. \end{aligned}$$

Здесь $\mathcal{I} = \{(0, 1), (1, 0), (1, 1)\}$.

М-оценку a_{mn}^* параметра a по наблюдениям X_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$, поля (1) определим как точку минимума функции

$$\mathcal{L}(a) = \sum_{i=2}^m \sum_{j=2}^n \rho(X_{ij} - a_{10}X_{i-1,j} - a_{01}X_{i,j-1} - a_{11}X_{i-1,j-1}), \quad (2)$$

где ρ — подходящим образом выбранная функция. Обычно $\rho(x)$ — выпуклая четная гладкая функция.

В [7] доказана следующая теорема.

Теорема 1. Пусть поле X_{ij} , описываемое уравнением (1), является стационарным, $\rho(x)$ — выпуклая функция, $\rho''(x)$ непрерывна почти всюду и ограничена, а плотность $f(x)$ независимых одинаково распределенных случайных величин ε_{ij} в (1) удовлетворяет следующим условиям: $E\varepsilon_{11} = 0$, $E\rho'(\varepsilon_{11}) = 0$, $E(\rho'(\varepsilon_{11}))^2 > 0$, $\sigma^2 = D\varepsilon_{11} < \infty$, $0 < E\rho''(\varepsilon_{11}) < \infty$.

Тогда при $m, n \rightarrow \infty$ случайный вектор $\sqrt{mn}(a_{mn}^* - a)$ является асимптотически нормальным с нулевым математическим ожиданием и ковариационной матрицей

$$R^{-1} \frac{E[\rho'(\varepsilon_{11})^2]}{(E[\rho''(\varepsilon_{11})])^2},$$

где R — ковариационная матрица вектора $(X_{10}, X_{01}, X_{11})^T$.

Заметим, что если $f(x)$ — четная функция (а это типичная ситуация), то производная $\rho'(x)$ четной функции $\rho(x)$ является нечетной и условие $E\rho'(\varepsilon_{11}) = 0$ будет выполнено.

В [13] асимптотическая нормальность М-оценок и их обобщений доказана и для невыпуклой функции $\rho(x)$, если только существует $\rho'''(x)$.

Если функция $\rho(x)$ является выпуклой и дифференцируемой, то минимизация (2) равносильна решению системы уравнений

$$\Lambda_{pq}(a) = 0, \quad p, q \in \mathcal{I}, \quad (3)$$

где

$$\Lambda_{pq}(a) = \sum_{i=2}^m \sum_{j=2}^n \psi(X_{ij} - a_{10}X_{i-1,j} - a_{01}X_{i,j-1} - a_{11}X_{i-1,j-1})X_{i-p,j-q} = 0, \quad p, q \in \mathcal{I}, \quad (4)$$

а $\psi(x) = \rho'(x)$.

Из (1) следует, что функция правдоподобия параметра a имеет вид

$$L(a) = \prod_{i=2}^m \prod_{j=2}^n f(X_{ij} - a_{10}X_{i-1,j} - a_{01}X_{i,j-1} - a_{11}X_{i-1,j-1}).$$

Оценка максимального правдоподобия параметра a определяется как точка максимума $L(a)$, или, что равносильно, как точка минимума функции

$$l(a) = -\ln L(a) = \sum_{i=2}^m \sum_{j=2}^n (-\ln f(X_{ij} - a_{10}X_{i-1,j} - a_{01}X_{i,j-1} - a_{11}X_{i-1,j-1})).$$

Таким образом, оценка максимального правдоподобия является частным случаем М-оценки с ρ -функцией $\rho(x) = -\ln f(x)$, или ψ -функцией

$$\psi(x) = -\frac{f'(x)}{f(x)}.$$

Рекомендации по выбору $\rho(x)$ имеются в [8, 9]. Наиболее распространенной ρ -функцией является ρ -функция Хьюбера — параметрическое семейство, задаваемое уравнением

$$\rho_H(x) = \begin{cases} x^2, & |x| \leq k; \\ 2k|x| - k^2, & |x| > k. \end{cases}$$

Из определения ρ -функции Хьюбера следует, что если невязка

$$X_{ij} - a_{10}X_{i-1,j} - a_{01}X_{i,j-1} - a_{11}X_{i-1,j-1} \quad (5)$$

в минимизируемой сумме (2) принимает достаточно большие значения, то ее вклад в эту сумму линеен, т.е. ограничивается по сравнению с методом наименьших квадратов. Если же эта величина принимает небольшие значения, то ее вклад в сумму (2) такой же, как и в методе наименьших квадратов.

Еще в большей степени минимизировать влияние больших значений невязки (5) на сумму (2) можно, взяв в качестве $\rho_T(x)$ функцию, называемую бивес Тьюки:

$$\rho_T(x) = \begin{cases} 1 - \left(1 - \left(\frac{x}{k}\right)^2\right)^3, & |x| \leq k; \\ 1, & |x| > k. \end{cases}$$

Из определения ρ -функции Тьюки видно, что если значение невязки (5) становится больше, чем значение параметра k , то ее вклад в сумму (2) практически игнорируется, заменяясь единицей.

Таким образом, по сравнению с оценками наименьших квадратов, М-оценки с ρ -функциями Хьюбера и Тьюки должны быть менее чувствительными к большим значениям слагаемых в (2), которые индуцируются большими ошибками в наблюдениях поля X_{ij} . Смысл понятий «небольшие значения» и «достаточно большие значения» регулируется параметром k . В частности, если $\rho(x) = x^2$, то получается оценка наименьших квадратов, а если $\rho(x) = |x|$ — оценка наименьших модулей.

2. Построение алгоритма

При решении системы уравнений (3) формально можно использовать алгоритм Ньютона [14], суть которого состоит в линеаризации уравнений этой системы, т.е. в замене функций $\Lambda_{pq}(a)$ их разложением по формуле Тейлора в точке текущего приближения решения. Таким образом, если k -ю аппроксимацию решения обозначить a_k^* , то $(k+1)$ -я итерация является решением системы уравнений

$$\Lambda(a_k^*) + \frac{\partial \Lambda(a_k^*)}{\partial a}(a_{k+1}^* - a_k^*) = 0.$$

К сожалению, этот алгоритм может расходиться, например, при неудачном выборе начального приближения. Кроме того, если определитель матрицы $\frac{\partial \Lambda(a_k^*)}{\partial a}$ может быть сколь угодно малым, то скорость сходимости последовательности a_k^* не будет квадратичной, а сам метод может преждевременно прекратить поиск и дать неверное для заданной точности приближение.

Для того чтобы М-оценки были робастными (устойчивыми к нарушению предположения о распределении обновляющего поля ε_{ij}), функция $\psi(x)$ в (4) должна быть ограничена. В этом случае ее производная $\psi'(x)$ будет стремиться к нулю на бесконечности, и, следовательно, определитель матрицы $\frac{\partial \Lambda(a_k^*)}{\partial a}$ не будет отделен от нуля, что делает процедуру

вычисления М-оценок методом Ньютона ненадежной. По этой причине построим алгоритм, который являются заведомо сходящимся.

Предположим, что функция ψ дифференцируема. Определим функцию

$$w(x) = \begin{cases} \frac{\psi(x)}{x}, & x \neq 0; \\ \psi'(0), & x = 0. \end{cases}$$

В этом случае решение системы (3) равносильно решению системы

$$\sum_{i=2}^m \sum_{j=2}^n w(\varepsilon_{ij}(a)) \varepsilon_{ij}(a) X_{\alpha\beta} = 0, \quad \alpha, \beta \in \mathcal{I}, \quad (6)$$

где $\varepsilon_{ij}(a) = X_{ij} - a_{10}X_{i-1,j} - a_{01}X_{i,j-1} - a_{11}X_{i-1,j-1}$. Запишем систему (6) в векторном виде:

$$\sum_{i=2}^m \sum_{j=2}^n w(\varepsilon_{ij}(a)) \varepsilon_{ij}(a) \tilde{X}_{ij} = 0, \quad (7)$$

где $\tilde{X}_{ij} = (X_{i-1,j}, X_{i,j-1}, X_{i-1,j-1})$. Если бы величины $w_{ij} = w(\varepsilon_{ij}(a))$ были известны (не зависели от a), то на решение системы (7) относительно a можно было бы смотреть как на оценки взвешенного метода наименьших квадратов, поскольку они получались бы как точка минимума функции

$$L_w(a) = \sum_{i=2}^m \sum_{j=2}^n w_{ij} \varepsilon_{ij}^2(a). \quad (8)$$

Отметим, что можно было бы свести минимизацию (8) и к нахождению обычных оценок наименьших квадратов, если в качестве «наблюдений» взять $\sqrt{w_{ij}}X_{ij}$, а в качестве векторов-столбцов регрессионной матрицы величины $\sqrt{w_{ij}}\tilde{X}_{ij}$. Тогда решение (7) — точка минимума функции

$$\tilde{L}_w(a) = \sum_{i=2}^m \sum_{j=2}^n (\sqrt{w_{ij}} X_{ij} - \sqrt{w_{ij}} \tilde{X}_{ij}^T a)^2.$$

Эти соображения наводят на мысль нахождения минимума (2) итерационным способом. Соответствующий алгоритм будет иметь следующий вид.

1. Определяется начальное приближение a_0^* М-оценки a^* параметра a .
2. По известному k -му приближению a_k^* вычисляются

$$\varepsilon_{ijk} = X_{ij} - a_k^{*T} \tilde{X}_{ij}, \quad w_{ijk} = w(\varepsilon_{ijk}), \quad i = 2, \dots, m, \quad j = 2, \dots, n.$$

3. Вычисляется следующее $(k+1)$ -е приближение a_{k+1}^* как решение системы линейных относительно a уравнений

$$\sum_{i=2}^m \sum_{j=2}^n w_{ijk} (X_{ij} - a^T \tilde{X}_{ij}) \tilde{X}_{ij} = 0,$$

или

$$\sum_{i=2}^m \sum_{j=2}^n \tilde{X}_{ij} \tilde{X}_{ij}^T a = \sum_{i=2}^m \sum_{j=2}^n w_{ijk} (X_{ij} \tilde{X}_{ij}). \quad (9)$$

4. Процесс заканчивается, когда

$$|a_{k+1}^* - a_k^*| < \delta,$$

где постоянная δ задает точность вычисления М-оценки a^* .

Обозначив

$$h(a) = \operatorname{argmin}_{\beta} \sum_{i=2}^m \sum_{j=2}^n w(\varepsilon_{ij}(a)) (X_{ij} - \beta^T \tilde{X}_{ij}),$$

алгоритм можно записать в виде

$$a_{k+1}^* = h(a_k^*), \quad (10)$$

причем решение a_{∞}^* системы (3) удовлетворяет условию

$$a_{\infty}^* = h(a_{\infty}^*).$$

3. Сходимость алгоритма

Докажем сходимость этого алгоритма.

Теорема 2. Пусть функция $\rho(x)$ не убывает и имеет производную $\rho'(x)$, а $w(x)$ не возрастает при $|x| \rightarrow \infty$. Пусть существует единственное решение a^* системы (3). Тогда алгоритм сходится к a^* .

Доказательство. Определим функцию $g(x)$ по формуле

$$g(x) = \rho(\sqrt{x}).$$

Тогда $\rho(x) = g(x^2)$, $\rho'(x) = 2xg'(x^2)$ и $w(x) = 2g'(x^2)$ при $x \neq 0$.

Применив теорему о среднем, получим

$$L(a_{k+1}^*) - L(a_k^*) = \sum_{i=2}^m \sum_{j=2}^n (g(\varepsilon_{ij}^2(a_{k+1}^*)) - g(\varepsilon_{ij}^2(a_k^*))) = \sum_{i=2}^m \sum_{j=2}^n g'(\tau) (\varepsilon_{ij}^2(a_{k+1}^*) - \varepsilon_{ij}^2(a_k^*)),$$

где τ лежит между $\varepsilon_{ij}^2(a_k^*)$ и $\varepsilon_{ij}^2(a_{k+1}^*)$.

Заметим, что если $w(x)$ не возрастает при $x > 0$, то $g'(x)$ также не возрастает. Поэтому $g'(\tau) \leq g'(x)$ при $\tau > x$. Следовательно, если $x < y$, то

$$g(y) - g(x) = g'(\tau)(y - x) \leq g'(x)(y - x), \quad x < \tau < y.$$

Если же $x > y$, то

$$g(x) - g(y) = g'(\tau)(x - y) \geq g'(x)(x - y), \quad x < \tau < y,$$

и поэтому также

$$g(y) - g(x) = g'(\tau)(y - x) \leq g'(x)(y - x), \quad x < \tau < y.$$

Таким образом,

$$L(a_{k+1}^*) - L(a_k^*) \leq \sum_{i=2}^m \sum_{j=2}^n g'(\varepsilon_{ij}^2(a_k^*)) (\varepsilon_{ij}^2(a_{k+1}^*) - \varepsilon_{ij}^2(a_k^*)).$$

Так как

$$\varepsilon_{ij}^2(a_{k+1}^*) - \varepsilon_{ij}^2(a_k^*) = (a_k^* - a_{k+1}^*)^T \tilde{X}_{ij}, \quad \varepsilon_{ij}^2(a_{k+1}^*) + \varepsilon_{ij}^2(a_k^*) = 2X_{ij} - (a_k^* + a_{k+1}^*)^T \tilde{X}_{ij},$$

то

$$\begin{aligned} L(a_{k+1}^*) - L(a_k^*) &\leq \sum_{i=2}^m \sum_{j=2}^n g'(\varepsilon_{ij}^2(a_k^*)) (\varepsilon_{ij}(a_{k+1}^*) - \varepsilon_{ij}(a_k^*)) (\varepsilon_{ij}(a_{k+1}^*) + \varepsilon_{ij}(a_k^*)) = \\ &= \sum_{i=2}^m \sum_{j=2}^n g'(\varepsilon_{ij}^2(a_k^*)) (a_k^* - a_{k+1}^*)^T \tilde{X}_{ij} (2X_{ij} - (a_k^* + a_{k+1}^*)^T \tilde{X}_{ij}). \end{aligned}$$

Так как

$$g'(\varepsilon_{ij}^2(a_k^*)) = \frac{1}{2} w(\varepsilon_{ij}^2(a_k^*)),$$

и a_{k+1}^* является решением (9), то

$$\sum_{i=2}^m \sum_{j=2}^n w(\varepsilon_{ij}^2(a_k^*)) a_{k+1}^T \tilde{X}_{ij} = \sum_{i=2}^m \sum_{j=2}^n w(\varepsilon_{ij}^2(a_k^*)) X_{ij} \tilde{X}_{ij}.$$

Поэтому

$$\begin{aligned} L(a_{k+1}^*) - L(a_k^*) &\leq \frac{1}{2} \sum_{i=2}^m \sum_{j=2}^n (a_k^* - a_{k+1}^*)^T \tilde{X}_{ij} w(\varepsilon_{ij}^2(a_k^*)) (2a_{k+1}^* \tilde{X}_{ij} - (a_k^* + a_{k+1}^*)^T \tilde{X}_{ij}) = \\ &= \frac{1}{2} \sum_{i=2}^m \sum_{j=2}^n w(\varepsilon_{ij}^2(a_k^*)) (a_k^* - a_{k+1}^*)^T \tilde{X}_{ij} \tilde{X}_{ij}^T (a_{k+1}^* - a_k^*). \end{aligned}$$

Матрица $\tilde{X}_{ij} \tilde{X}_{ij}^T$ положительно определена. Следовательно,

$$(a_k^* - a_{k+1}^*)^T \tilde{X}_{ij} \tilde{X}_{ij}^T (a_{k+1}^* - a_k^*) < 0.$$

Так как $\rho(x)$ возрастает при $x > 0$, то $\psi(x) > 0$ при $x > 0$ и, следовательно, $w(x) > 0$ при $x > 0$. Поэтому $w(\varepsilon_{ij}^2(a_k^*)) > 0$ и, следовательно, $L(a_{k+1}^*) - L(a_k^*) \leq 0$.

Докажем сходимость последовательности a_k^* к М-оценке a^* параметра a . Так как последовательность $L(a_k^*)$ не возрастает и ограничена снизу, то она имеет конечный предел. Следовательно, последовательность a_k^* ограничена. Действительно, в противном случае существовала бы подпоследовательность $a_{k_i}^*$, стремящаяся к бесконечности, и тогда подпоследовательность $L(a_{k_i}^*)$ в силу возрастания функции $\rho(x)$ при $x > 0$ также стремилась бы к бесконечности.

Так как a_k^* ограничена, то у нее существует сходящаяся подпоследовательность, предел которой обозначим a_∞^* . Перейдя в обеих частях равенства (10) к пределу, получим, что $a_\infty^* = h(a_\infty^*)$, т.е. a_∞^* — стационарная точка преобразования h .

Если a_∞^* — единственная стационарная точка h , то $a_k^* \rightarrow a_\infty^*$. Действительно, в противном случае существовала бы подпоследовательность $b_{k_i}^*$ последовательности a_k^* , для которой для некоторого $\delta > 0$ $|b_{k_i}^* - a_\infty^*| \geq \delta$. Последовательность $b_{k_i}^*$ в свою очередь имела бы сходящуюся подпоследовательность, стремящуюся к $b_\infty^* \neq a_\infty^*$, и для которой бы выполнялось бы $b_\infty^* = h(b_\infty^*)$. Теорема доказана.

4. Выводы

В работе построен алгоритм вычисления М-оценок коэффициентов уравнения авторегрессионного поля и доказана сходимость этого алгоритма. Алгоритм представляет собой итерационный вариант взвешенного метода наименьших квадратов с пересчитываемыми на каждом шаге весами. В отличие от метода Ньютона (метода касательных) алгоритм является сходящимся при любом выборе начального приближения.

Список литературы

1. Ripley B.D. Spatial statistics. Hoboken: Wiley, 1981. DOI: 10.1002/0471725218.
2. Gaetan C., Guyon X. Spatial statistics and modeling. New York: Springer, 2010. (Springer Series in Statistics.) DOI: 10.1007/978-0-387-92257-7.
3. Handbook of spatial statistics / A.E. Gelfand, P.J. Diggle, M. Fuentes, P. Guttorp (eds.). Boca Raton: Taylor & Francis, 2010.
4. Haining R. Spatial Data Analysis: Theory and Practice. Cambridge: CUP, 2004.
5. Yao Q., Brockwell P.J. Gaussian Maximum Likelihood Estimation for ARMA Models II Spatial Processes // Bernoulli. 2006. Vol. 12, no. 3. P. 403–429.
6. Tjøstheim D. Statistical Spatial Series Modelling // Advances in Applied Probability. 1978. Vol. 10, no. 1. P. 130–154.
7. Горяинов В.Б. М-оценки пространственной авторегрессии // Автоматика и телемеханика. 2012. № 8. С. 119–129.
8. Хьюбер П.Дж. Робастность в статистике. М.: Мир, 1984.
9. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. М.: Мир, 1989.
10. Lee C.-H., Martin R.D. Ordinary and proper location M-estimates for autoregressive-moving average models // Biometrika. 1986. Vol. 73, no. 3. P. 679–686.
11. Stirling W.D. Iteratively reweighted least squares for models with a linear part // J. Roy. Statist. Soc. Ser. C. 1984. Vol. 33, no. 1. P. 7–17.
12. Basu S., Reinsel G.C. Properties of the Spatial Unilateral First-Order ARMA Model // Advances in Applied Probability. 1993. Vol. 25, no. 3. P. 631–648.
13. Горяинов В.Б. Обобщенные М-оценки коэффициентов авторегрессионного поля .. Автоматика и телемеханика. 2012. № 10. С. 42–51.
14. Амосов А.А., Дубинский Ю.А., Копченкова Н.В. Вычислительные методы для инженеров. М.: Высшая школа, 1994.

Algorithm for calculating M-estimates of autoregressive field's parameters

07, July 2013

DOI: 10.7463/0713.0571094

Goryainov V. B.

Bauman Moscow State Technical University
105005, Moscow, Russian Federation
vb-goryainov@mail.ru

A process of a two-dimensional autoregression of order (1, 1) is considered in this article. Distribution of the innovation field of the autoregressive model is assumed to be unknown. An algorithm for calculating M-estimates of coefficients of the autoregressive field was constructed. The convergence of this algorithm was proved. The algorithm is an iterative version of the weighted least squares method. The weights are recalculated at each step. Each iteration represents the process of solving the system of linear equations. In contrast to the method of Newton (method of tangents) the algorithm converges from any point of the initial approximation.

References

1. Ripley B.D. *Spatial statistics*. Hoboken, Wiley, 1981. DOI: 10.1002/0471725218.
2. Gaetan C., Guyon X. *Spatial statistics and modeling*. Springer New York, 2010. (*Springer Series in Statistics*). DOI: 10.1007/978-0-387-92257-7.
3. Gelfand A.E., Diggle P.J., Fuentes M., Guttorp P. (eds.). *Handbook of spatial statistics*. Boca Raton: Taylor & Francis, 2010.
4. Haining R. *Spatial Data Analysis: Theory and Practice*. Cambridge, CUP, 2004.
5. Yao Q., Brockwell P.J. Gaussian Maximum Likelihood Estimation for ARMA Models II Spatial Processes. *Bernoulli*, 2006, vol. 12, no. 3, pp. 403–429.
6. Tjøstheim D. Statistical Spatial Series Modelling. *Advances in Applied Probability*, 1978, vol. 10, no. 1, pp. 130–154.
7. Goriainov V.B. M-otsenki prostranstvennoi avtoregressii [M-estimates of the spatial autoregression coefficients]. *Avtomatika i telemekhanika*, 2012, no. 8, pp. 119–129. (Trans.

version: *Automation and Remote Control*, 2012, vol. 73, no. 8, pp. 1371–1379. DOI: 10.1134/S0005117912080103.)

8. Huber P.J. *Robust Statistics*. John Wiley & Sons, 1981. 320 p. (Russ. ed.: Huber P.J. *Robastnost v statistice*. Moscow, Mir, 1984. 304 p.).
9. Hampel F.R., Ronchetti E.M., Rausseau P.J., Stahel W.A. *Robust statistics: The approach based on influence functions*. Wiley, New York, 1986. 536 p. (*Wiley Series in Probability and Statistics*). (Russ. ed.: Khampel' F., Ronchetti E., Rausseau P., Shtael' V. *Robastnost v statistike. Podhod na osnove funktsii vplivaniia*. Moscow, Mir, 1989. 512 p.).
10. Lee C.-H., Martin R.D. Ordinary and proper location M-estimates for autoregressive-moving average models. *Biometrika*, 1986, vol. 73, no. 3, pp. 679–686.
11. Stirling W.D. Iteratively reweighted least squares for models with a linear part. *J. Roy. Statist. Soc. Ser. C*, 1984, vol. 33, no. 1, pp. 7–17.
12. Basu S., Reinsel G.C. Properties of the Spatial Unilateral First-Order ARMA Model. *Advances in Applied Probability*, 1993, vol. 25, no. 3, pp. 631–648.
13. Goriainov V.B. Obobshchennye M-otsenki koeffitsientov avtoregressionnogo polia [Generalized M-estimates of the autoregression field coefficients]. *Avtomatika i telemekhanika*, 2012, no. 10, pp. 42–51. (Trans. version: *Automation and Remote Control*, 2012, vol. 73, no. 10, pp. 1624–1631. DOI: 10.1134/S0005117912100049.)
14. Amosov A.A., Dubinskii Iu.A., Kopchenova N.V. Vychislitel'nye metody dlia inzhenerov [Computational Methods for Engineers]. Moscow, Vysshaia shkola, 1994. 544 p.