

Е.Г. РЕВУНОВА

СРАВНЕНИЕ КРИТЕРИЕВ ВЫБОРА МОДЕЛИ В ЗАДАЧАХ АППРОКСИМАЦИИ С ЕСТЕСТВЕННЫМ БАЗИСОМ

Abstract: A comparative analysis of linear model selection criteria applied to the problem of assessing the radionuclide content in environmental objects is performed. Detector outputs obtained for gamma-quanta of particular energy are used as basis functions. Complexity and quality of models vs noise level for various model selection criteria is investigated.

Key words: linear approximation, model selection criteria, spectrometry.

Анотація: Проведено порівняльний аналіз критеріїв вибору лінійної моделі на прикладі задачі визначення складу радіонуклідів в об'єктах навколишнього середовища. За базисні функції було взято функції відгуку детектора на вплив гамма-квантів з визначеною енергією. Досліджено залежність складності й адекватності (якості) моделей від рівня шуму для різних критеріїв вибору моделі.

Ключові слова: лінійна апроксимація, критерії вибору моделі, задача спектрометрії.

Аннотация: Проведен сравнительный анализ критериев выбора линейной модели на примере задачи определения содержания радионуклидов в объектах окружающей среды. В качестве базисных функций использованы функции отклика детектора на воздействие гамма-квантов с определенной энергией. Исследована зависимость сложности и адекватности (качества) моделей от уровня шума для разных критериев выбора модели.

Ключевые слова: линейная аппроксимация, критерии выбора модели, задача спектрометрии.

1. Введение

Для многих технических приложений типичной является следующая ситуация. Имеется система «датчик – измерительный прибор», на вход которой осуществляются некоторые воздействия. Исследователь располагает априорной информацией о наборе возможных воздействий. Задача – определить, какие именно воздействия и с какими весовыми коэффициентами (параметрами) сформировали измеряемый выход прибора. Примерами таких задач являются определение активностей радионуклидов по измеренному спектру их гамма-излучения [1, 2], подавление активных радио помех [3, 4] и др. [5, 6].

В рамках аппроксимационного подхода к решению таких задач сигнал $f(z)$ на выходе прибора представляют моделью, линейной относительно параметров

$$f(z) = \sum_{j=1, \mu} \theta_j \varphi_j(z) + \varepsilon, \quad (1)$$

где z – независимая переменная (например, время, значение энергии и т.п.); $\varphi_j(z)$ – некоторая система базисных функций (каждая функция – отклик системы на определенное воздействие); $\theta = (\theta_1, \theta_2, \dots, \theta_\mu)^T$ – параметры (весовые коэффициенты воздействий); ε – собственный шум системы.

Традиционно задача аппроксимации состоит в оценке вектора параметров θ^* по известному набору L пар данных вход-выход $D_L = \{(x_i, y_i)\}_{i=1, L}$, где y_i – измеренное значение сигнала на выходе прибора; $x_{ij} = \varphi_j(z_i), i = 1 \dots L, j = 1 \dots \mu$. Матрицу значений базисных функций $X = \{x_{ij}\}$

называют матрицей плана. Для вычисления оценки параметров θ^* , как правило, применяют обычный или регуляризованный метод наименьших квадратов (МНК – см. также [4]). Для интересующего нас случая с неизвестным составом воздействий и, соответственно, базисных функций задача усложняется, так как перед этапом оценки параметров модели необходимо выбрать ее структуру, т.е. некоторое подмножество μ_1 базисных функций из их полного набора, включающего μ функций.

В настоящее время существует целый ряд критериев выбора структуры модели, предложенных исходя из разных принципов и теоретических посылок. Наиболее известные из них можно разделить на следующие группы. Критерии на основе минимизации ошибки обобщения (generalization error) – SIC [7], AIC [8], ошибки прогнозирования на обучающей выборке (predictive training error) – Cp, CI [9, 10], минимальной длины описания (minimum description length MDL) – gMDL [11]. При создании каждого критерия делается ряд предположений или упрощений, поэтому априори нельзя сказать, какой из них окажется лучшим для конкретной практической задачи.

В данной работе экспериментально исследована адекватность разных типов критериев выбора модели для задач с "естественным" базисом – дискретно заданными функциями, каждая из которых является результатом измерений (или реконструкции) одного определенного воздействия в L точках. В качестве примера рассмотрена актуальная задача спектрометрии – определение активности отдельных радиоактивных элементов по измеренному суммарному спектру. В качестве базисных использованы функции отклика детектора на воздействие гамма-квантов с определенной энергией.

С помощью установки ПРС "Вектор" [12] получены результаты измерений спектров, для которых известен состав базисных функций, их веса и уровень шума. Это позволило оценить качество моделей, сконструированных в соответствии с разными критериями, путем сравнения получаемых результатов с "эталонными" как по точности восстановления весовых коэффициентов (активностей) отдельных элементов, так и по точности восстановления самой функции, т.е. суммарного спектра.

2. Критерии выбора модели

Рассмотрим ряд критериев, которые будут далее использованы для построения моделей. Оптимальной по некоторому критерию считается модель, для которой данный критерий принимает оптимальное (обычно, минимальное) значение.

2.1. Критерии, основанные на ошибке обобщения

Ошибкой обобщения для квадратичной функции потерь называют

$$E_\varepsilon \|f_\theta^* - f_0\|^2, \quad (2)$$

где E_ε – матожидание по всем возможным реализациям шума; f_0 – идеальная оценка – функция, минимизирующая средний риск на классе функций F ; f_θ^* – оценка, полученная по модели θ .

Критерием, основанным на наихудшей ошибке обобщения, является минимизация структурного риска [14]. К критериям, основанным на средней ошибке обобщения, относятся информационный критерий подпространств SIC (subspace information criterion) [7] и информационный критерий Акаике (AIC) [8, 13]. SIC позволяет получить несмещенную оценку ошибки обобщения при условии наличия аддитивного шума ε ($y = y_0 + \varepsilon$) с нулевым средним и известной дисперсией σ^2 :

$$E_\varepsilon \text{SIC} = E_\varepsilon \|f_\theta^* - f_0\|^2. \quad (3)$$

Для обучения, минимизирующего среднеквадратичную ошибку, выражение для критерия SIC имеет следующий вид:

$$\text{SIC} = \langle \mathbf{T}y, y \rangle - \sigma^2 \text{tr}(\mathbf{T}) + \sigma^2 \text{tr}(\mathbf{T}_S^+), \quad (4)$$

$$\mathbf{T} = \mathbf{T}_S^+ - \mathbf{T}_F^+ \mathbf{T}_S \mathbf{T}_S^+ - \mathbf{T}_S^+ \mathbf{T}_S \mathbf{T}_F^+ + \mathbf{T}_F^+. \quad (5)$$

$(\cdot)^+$ обозначает псевдоинверсию; \mathbf{T}_S и \mathbf{T}_F – матрицы $L \times L$, элементы которых есть воспроизводящие ядра [15]. Модификация SIC, учитывающая неотрицательность $\langle \mathbf{T}y, y \rangle - \sigma^2 \text{tr}(\mathbf{T})$, имеет вид [7]:

$$\text{cSIC} = \left[\langle \mathbf{T}y, y \rangle - \sigma^2 \text{tr}(\mathbf{T}) \right]_+ + \sigma^2 \text{tr}(\mathbf{T}_S^+). \quad (6)$$

$[\cdot]_+$ определяется как $[t]_+ = \max(0, t)$. Моделью оптимальной сложности является модель с минимальным значением SIC. Для вычисления \mathbf{T}_S и \mathbf{T}_F может быть использовано выражение для воспроизводящих ядер конечномерных классов [15]. Тогда, для матриц плана $\mathbf{X}_S (L \times \mu_1)$ и $\mathbf{X}_F (L \times \mu)$ ($\mu_1 < \mu$)

$$\mathbf{T}_S = ((\mathbf{X}_S \mathbf{X}_S^T)^{-1})^T \mathbf{X}_S \mathbf{X}_S^T, \quad \mathbf{T}_F = ((\mathbf{X}_F \mathbf{X}_F^T)^{-1})^T \mathbf{X}_F \mathbf{X}_F^T. \quad (7)$$

Критерии Акаике (AIC, cAIC) [8, 16] основаны на асимптотической несмещенной оценке среднего правдоподобия (при $L \rightarrow \infty$). Выражение для cAIC имеет вид

$$\text{cAIC} = (L/2) \log \text{RSS} + (L/2)(1 + \mu_1/L)/(1 - (\mu_1 + 2))/L; \quad (8)$$

$$\text{RSS} = \|y - y^*\|^2 = \|y - \mathbf{X}_S (\mathbf{X}_S^T \mathbf{X}_S)^{-1} \mathbf{X}_S^T y\|^2, \quad (9)$$

где y^* – оценка y , полученная моделью; $\mathbf{B} = \mathbf{X}_S (\mathbf{X}_S^T \mathbf{X}_S)^{-1} \mathbf{X}_S^T$ – оператор, отображающий y в y^* ; $\mathbf{X}_S - (L \times \mu_1)$ матрица плана, ассоциированная с выбранными базисными функциями.

2.2. Критерии, основанные на ошибке прогнозирования для обучающей выборки

Ошибкой прогнозирования для обучающей выборки (predictive training error) J_{PTE} [17] называют

$$J_{PTE} = E_\varepsilon \|y^* - y_0\|^2 = E_\varepsilon \|\mathbf{B}y - y_0\|^2 = \|\mathbf{B}y_0 - y_0\|^2 + \text{tr}(\mathbf{BQB}^T). \quad (10)$$

К критериям, основанным на ошибке прогнозирования для обучающей выборки, относятся критерии Маллоуза C_P , C_L [9, 10], и (обобщенная) кросс-валидация [18]. В данной работе используется C_P , который дает несмещенную оценку J_{PTE} :

$$C_P = RSS + 2\sigma^2\mu_1. \quad (11)$$

2.3. Критерии, основанные на минимизации длины описания

Предложен ряд критериев, основанных на минимизации длины описания [19]

$$L(y | \mathbf{X}_S, s) + L(s), \quad (12)$$

где $L(y | \mathbf{X}_S, s)$ – длина описания данных y моделью s ; $L(s)$ – длина описания модели s .

Таковыми критериями выбора модели являются $L(y)$, gMDL, критерий нормализованного максимума правдоподобия (nMDL) [11], критерий стохастической сложности [19]. Критерий Бина Ю $L(y)$ основан на так называемой смешанной форме принципа минимизации длины описания (mixture form MDL). Длина кода $L(y)$, т.е. число бит, необходимое для передачи y , есть

$$L(y) = -\log_2 m(y | X), \quad (13)$$

где $m(y | X)$ – смешанное распределение, которое объединяет все члены класса моделей.

$L(y)$ вычисляется следующим образом:

$$L(y) = 0,5(y^T y - FSS) / \sigma^2 + 0,5\mu_1 [1 + \log(FSS / (\sigma^2 \mu_1))] + 0,5 \log L \text{ при } FSS > \mu_1 \sigma^2; \\ L(y) = 0,5(y^T y) \sigma^2 - \text{иначе, где } FSS = y^T \mathbf{X}_S (\mathbf{X}_S^T \mathbf{X}_S)^{-1} \mathbf{X}_S^T y. \quad (14)$$

В [11] также предложен критерий gMDL для случая неизвестной дисперсии шума:

$$\text{gMDL} = (L/2) \log S + 0,5\mu_1 \log F + \log L \text{ при } R^2 \geq \mu_1 / L; \\ \text{gMDL} = (L/2) \log(y^T y / L) + 0,5 \log L - \text{иначе,} \quad (15)$$

где $S = RSS / (L - \mu_1)$, $F = (y^T y - RSS) / \mu_1 S$, R – множественный коэффициент корреляции.

3. Схема экспериментального исследования

Общая схема эксперимента включает этапы выбора структуры и построение модели, оптимальной с точки зрения исследуемых критериев, вычисление показателей качества полученных моделей, оценку адекватности критериев для тестовой задачи по показателям качества полученных ими моделей.

3.1. Выбор структуры модели

Исчерпывающим методом перебора моделей-претендентов для определения структуры модели оптимальной сложности является полный перебор всех возможных подмножеств базисных функций [20]. Для каждой из моделей-претендентов оцениваются параметры (в нашем случае, с помощью МНК)

и определяется значение критерия выбора модели. В качестве модели-победителя отбирается модель с оптимальным (обычно, наименьшим) значением критерия. Используя полученные на этапе перебора моделей оценки параметров модели-победителя, вычисляются значения аппроксимируемой функции в точках обучающей выборки и/или в точках тестирующей выборки.

Полный перебор моделей реален только для размерности полного базиса не больше 20–30 [20]. В рассматриваемых далее задачах размерность базиса может составлять 256 и более, что исключает полный перебор. Поэтому в данной работе использовалась следующая процедура "жадного" перебора моделей-претендентов и выбора модели-победителя.

Начиная с моделей сложности $c = 1$ (с одной базисной функцией), для всех μ моделей вычисляется значение критерия выбора модели CR. Выбирается базисная функция, соответствующая оптимальному (обычно, минимальному) значению критерия, которая и включается в модель. Среди оставшихся $(\mu - c)$ базисных функций выбирается одна, которая в совокупности с отобранными ранее минимизирует значение критерия. Она включается в модель сложностью $c + 1$. Аналогично, с добавлением на каждом шаге только одной базисной функции, проверяются модели сложностью до μ . Для каждой сложности модели фиксируются минимальное значение критерия $CR(c)$ и соответствующий ему набор базисных функций. На выходе получаем набор базисных функций, при котором достигнуто минимальное по всем c значение критерия $CR(c^*)$, их число $\mu_1 = c^*$ и вектор оценок параметров θ^* . При таком алгоритме проверяется (перебирается) всего лишь $O(\mu(\mu + 1)/2)$ моделей, однако, очевидно, не гарантируется выбор глобально оптимальной модели.

3.2. Показатели качества построенной модели

Для экспериментального исследования (разд. 4) качества модели используются данные, для которых известны незашумленные ("эталонные") значения функции на выходе, сформировавший ее состав базисных функций, их эталонные весовые коэффициенты и уровень шума. Эту информацию можно использовать для оценки качества модели независимо от критерия, по которому выбран состав базисных функций. В экспериментах использовались два показателя качества моделей.

1. $J_m(s)$ – дисперсия ошибки восстановления эталонного сигнала по m реализациям зашумленного:

$$J_m(s) = E_m \|y_0 - f_{\theta^*}\|^2. \quad (16)$$

Здесь E_m – матожидание по m реализациям; y_0 – известное значение эталонного вектора выходных данных y в L точках; f_{θ^*} – соответствующие значения функции, полученные моделью s с параметрами θ^* . Отличие от "идеального" критерия $J(s)$ [10] заключается в том, что в последнем матожидание E_ϵ берется по всем возможным реализациям шума.

2. $J_\theta(s)$ – дисперсия ошибки восстановления эталонного вектора параметров:

$$J_{\theta}(s) = E_m \left\| \theta_0 - \theta^* \right\|^2, \quad (17)$$

где θ_0 – эталонный вектор параметров; θ^* – оценка вектора параметров, давшая f_{θ^*} .

3.3. Прикладная задача и план эксперимента

В качестве задачи аппроксимации с естественным базисом и выходом, формируемым линейной моделью вида (1), рассмотрим задачу определения содержания радионуклидов в объектах окружающей среды. Эта задача заключается в том, что по измеренному суммарному спектру требуется определить состав и параметры (весовые коэффициенты, активности) сформировавших его элементов (радионуклидов).

Традиционный метод окон и его модификации [12] основан на определении состава спектра в так называемых *зонах интереса* (участках спектра), которым соответствуют лишь некоторые радиоактивные элементы. Работа в зонах интереса не позволяет исследовать пробы произвольного состава и понижает точность определения активностей. Кроме того, для относительно недорогих и широко распространенных сцинтилляционных детекторов характерно сильное перекрытие зон интереса, соответствующих пикам полного поглощения радионуклидов. Задача исследования проб произвольного состава для таких детекторов ранее не рассматривалась. Предлагаемая схема решения этой задачи, в отличие от традиционных методов, позволяет определять (радио)активность элементов не только в зонах интереса, а во всем диапазоне энергий спектрометра.

Будем строить модели, оптимальные с точки зрения одного из критериев разд. 2, с помощью процедуры "жадного" перебора моделей разд. 3.1. В качестве μ базисных функций (формирующих исходную матрицу плана) будем использовать функции отклика детектора на воздействие гамма-квантов с определенной энергией. Для оптимальной модели получаем подмножество полного набора базисных функций, соответствующих радионуклидам в пробе, и вектор оценок их параметров θ^* (то есть активностей). Построим оптимальные модели для целого ряда реализаций шума определенного уровня σ^2 и вычислим для них показатели качества (разд. 3.2). По зависимостям показателей качества моделей от уровня шума σ^2 определим наиболее адекватный критерий как такой, который формирует модели с наилучшими показателями качества в наибольшем диапазоне шумов.

4. Экспериментальные исследования

Для сравнения поведения моделей использовались как реально измеренные спектры, так и спектры, искусственно сформированные линейной моделью.

4.1. Реальные данные

В качестве исходных данных использовались результаты измерения спектров радионуклидов спектрометрическим трактом, включающим детектор типа NaJ(Tl) 63×63 и ПРС "Вектор". Измеряемый

спектр представлял собой вектор из $L = 256$ чисел, каждое из которых – количество гамма-квантов, обладающих определенной энергией в диапазоне 120...2000 КэВ. Набор из 256 базисных функций (по числу измеряемых энергий) формировался следующим образом. Спектрометрическим трактом измерялись функции отклика нескольких реально существующих монохромных источников гамма-квантов, и такие функции отклика включались в набор базисных функций как «первичные». Остальные базисные функции формировались из первичных алгоритмически [12].

Для тестирования из базы данных ПРС «Вектор» были отобраны спектры, в состав которых входили радионуклиды Co^{60} , Cs^{137} , Cs^{134} , Mn^{58} с известными активностями. Для них была произведена оценка уровня собственного шума спектрометрического тракта, и спектры были упорядочены по уровню шума $\sigma^2 = 0,0...9,0$.

Спектры включали пять спектральных линий с энергиями 662, 795, 834, 1173, 1332 кэВ (γCo^{60} две линии), имеющих близкие пики полного поглощения. Кроме пяти вышеперечисленных, в качестве базисных функций использовались еще пять с энергиями 863, 883, 1120, 1461, 1764 кэВ. Такие спектры были выбраны потому, что для них традиционный метод окон и его модификации [12], а также метод редукции модели [12] не обеспечивают достаточную точность оценки параметров.

Эксперименты на реальных данных показали, что при изменении уровня шума меняется сложность модели, что влечет за собой соответствующие изменения значений показателей качества модели J_m и J_θ . На рис. 1 показана зависимость сложности оптимальной модели от уровня шума для критериев выбора модели Cp, gMDL, LyMDL, SIC (разд. 2) и «жадной» процедуры выбора модели (разд. 3.1). TRUE – истинная модель, OLS (ordinary least squares) – МНК с десятью базисными функциями. Для всех исследуемых критериев сложность модели не превышает истинной, т.е. критерии не включают в модель избыточных (реально не входящих в модель) регрессоров. Исключением является критерий LyMDL, дающий при уровне шума 0 сложность модели на один большую, чем истинная.

На рис. 2 представлена зависимость значения показателя качества модели J_m от уровня шума. Истинная модель TRUE дает минимальные значения J_m относительно исследуемого набора критериев практически во всем диапазоне шумов, OLS – максимальные. Остальные критерии при нарастании уровня шума дают примерно одинаковый результат, приближающийся к OLS. Зависимость значения показателя качества модели J_θ от уровня шума (рис. 3) ведет себя аналогично.

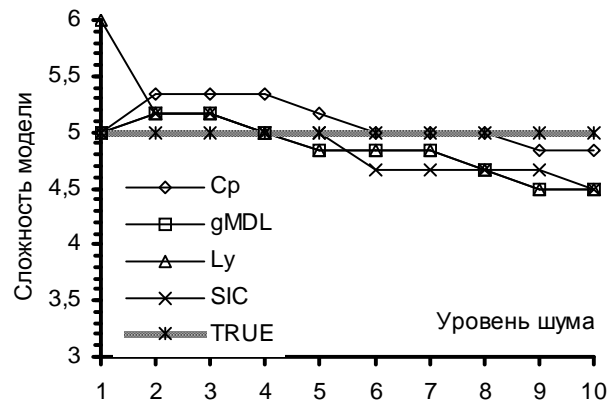


Рис. 1. Зависимость сложности модели от σ^2 для реальных данных

4.2. Искусственные данные

Для формирования искусственных тестовых данных с использованием в качестве базисных функций реальных функций отклика детектора использовались те же $\mu = 10$ базисных функций, что и для реальных данных, число реализаций случайного процесса для каждого уровня шума $m = 6$, $\theta_0 = (0, 3, 0, 3,5, 0, 4, 7, 9, 0, 0)$, ε – нормально распределенная центрированная случайная величина с дисперсией σ^2 в широком диапазоне 0,0-104,5 с шагом 5,5. Все $L = 256$ значений y_i для конкретной реализации шума вычислялись по (1).

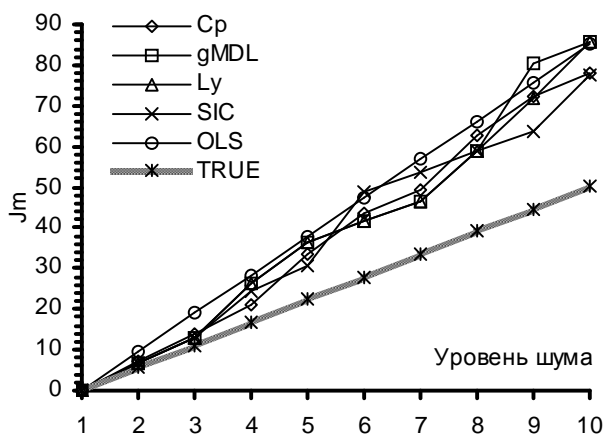


Рис. 2. Зависимость J_m от σ^2 для реальных данных

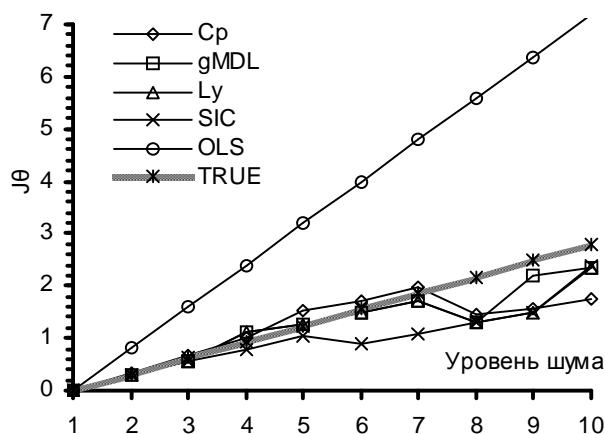


Рис. 3. Зависимость J_θ от σ^2 для реальных данных

Рассмотрим результаты экспериментов на искусственных данных. На рис.4 показана зависимость сложности оптимальной модели от уровня шума для тех же критериев и процедуры выбора модели, что и ранее. С ростом уровня шума сложность модели для всех критериев понижается, при этом кривая для Cp находится выше других, что свидетельствует о том, что Cp обеспечивает сложность модели ближе к истинной.

На рис. 5 представлена зависимость J_m от уровня шума. Истинная модель TRUE дает минимальные значения J_m относительно исследуемого набора критериев практически во всем диапазоне шумов. Для других критериев при росте уровня шума значения J_m достигают и превышают уровень для OLS, однако Cp в большом диапазоне параметров шума ведет себя

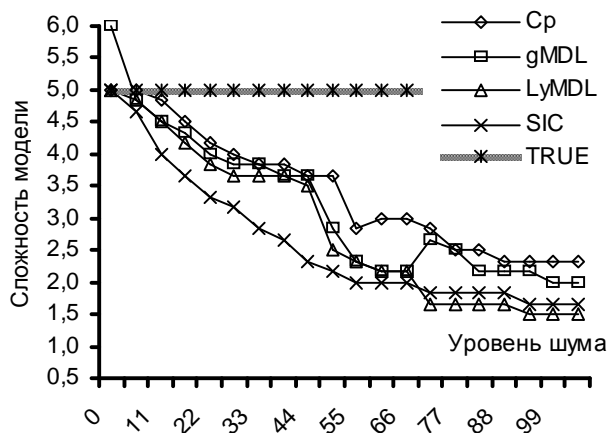


Рис. 4. Зависимость сложности модели от σ^2 для искусственных данных

лучше OLS и остальных исследованных критериев.

Зависимость J_θ от уровня шума приведена на рис. 6. Ее вид несколько иной, чем для J_m . Все критерии дают модель хуже TRUE, но лучше OLS. В большом диапазоне параметров шума Cp ведет себя близко к TRUE и, как и в предыдущих случаях, лучше остальных критериев.

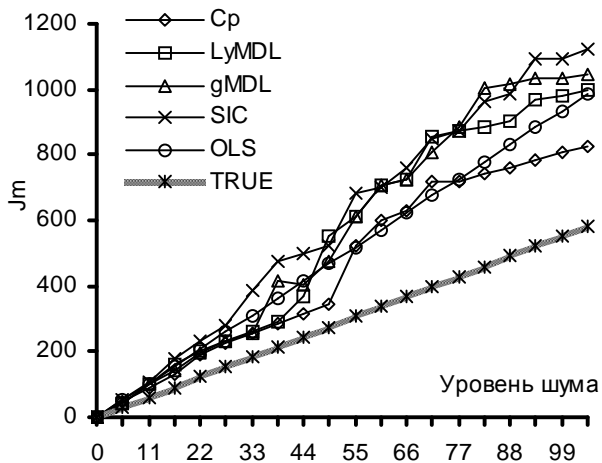


Рис. 5. Зависимость J_m от σ^2 для искусственных данных

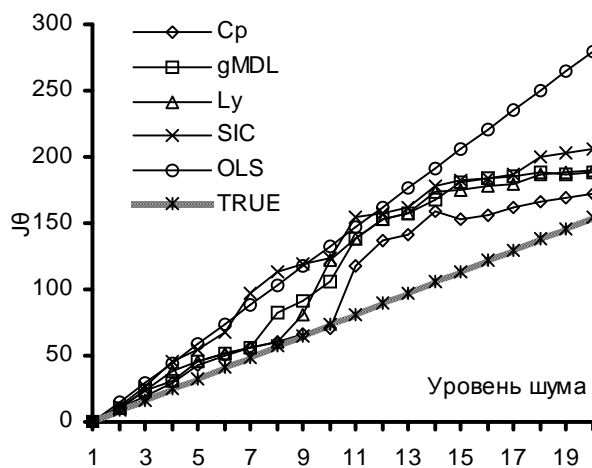


Рис. 6. Зависимость J_θ от σ^2 для искусственных данных

5. Заключение

Проведен сравнительный анализ критериев выбора модели для линейных по параметрам аппроксимационных моделей. Исследовалось поведение моделей, построенных с помощью недавно предложенных критериев: основанного на несмещенной оценке ошибки обобщения SIC, на принципе минимальной длины описания gMDL, LyMDL, а также классического критерия Маллоуза Cp. В качестве показателя качества модели использовался квадрат Евклидова расстояния между известными незашумленными значениями функции в точках обучающей выборки и значениями на выходе модели, а также непосредственно между известными и полученными моделью векторами параметров.

В качестве тестовой была выбрана актуальная задача определения содержания радионуклидов в объектах окружающей среды с помощью спектрометров со сцинтилляционными детекторами. Применение аппроксимационного подхода с использованием критериев выбора позволило впервые решить задачу определения активности элементов в пробах с произвольным радионуклидным составом для такого типа спектрометров.

В качестве базисных функций использовались функции отклика детектора на воздействие гамма-квантов с определенной энергией. Исследованные зависимости сложности и адекватности (качества) моделей от уровня шума для разных критериев выбора модели показали, что для данных тестовой задачи при умеренном уровне шума (порядка 0,0...9,0, что соответствует уровню шума при реальных измерениях спектра) все критерии функционируют примерно одинаково, с незначительным

преимуществом C_p . При дальнейшем повышении уровня шума преимущество C_p становится более заметным. Предположительно, это связано с тем, что C_p основан на минимизации ошибки прогнозирования в точках обучающей выборки, а остальные три исследованных критерия ориентированы на минимизацию ошибки и в других точках входного пространства. Исследование этого вопроса является темой дальнейшей работы.

СПИСОК ЛИТЕРАТУРЫ

1. Козлов В.Ф. Справочник по радиационной безопасности. – Москва: Энергоатомиздат, 1991. – 208 с.
2. Забулонов Ю.Л., Лісиченко Г.В., Ревунова О.Г. Визначення ступеню радіаційного забруднення об'єктів в нефіксованих геометріях // Збірник наукових праць ІПМС НАН України "Моделювання та інформаційні технології". – 2004. – № 27. – С. 95–101.
3. Ширман Я.Д. Радиозлектронные системы: основы построения и теория. – Москва, 1998. – 828 с.
4. Revunova E.G., Rachkovskij D.A. Jamming cancellation based on a stable LSP solution // International Journal "Information Theories & Applications". – 2005. – N 1. – P. 578–583.
5. Пытьев Ю.П. Математические методы интерпретации эксперимента. – М.: Высшая школа, 1989. – 351 с.
6. Старков В.Н. Конструктивные методы вычислительной физики в задачах интерпретации. – Киев: Наукова думка, 2002. – 263 с.
7. Sugiyama M., Ogawa H. Subspace Information Criterion for Model Selection // Neural Computation. – 2001. – Vol. 13. – N 8. – P. 1863–1889.
8. Akaike H. A new look at the statistical model identification // IEEE Transactions on Automatic Control. – 1974. – Vol. 19. – N 6. – P. 716–723.
9. Mallows C.L. Some comments on C_p // Technometrics. – 1973. – Vol. 15. – N 4. – P. 661–675.
10. Степашко В.С. Теоретические аспекты МГУА как метода индуктивного моделирования // УСИМ. – 2003. – №2. – С. 31–38.
11. Hansen M., Yu B. Model selection and minimum description length principle // In J. Amer. (ed.) Statist. Assoc. – 2001. – Vol. 96. – P. 746–774.
12. Ревунова Е.Г. Институт экологических и медицинских систем: Техн. отчет. – 1998. – С. 1–20.
13. Akaike H. Likelihood and the Bayes procedure // In N. J. Bernardo, M. H. DeGroot, D. V. Lindley, A. F. M. Smith (eds.) Bayesian Statistics. – Valencia: University Press, 1980. – P. 141–166.
14. Vapnik V.N. Statistical Learning Theory. – John Wiley & Sons, Inc.: New York, 1998. – 736 p.
15. Aronszajn N. Theory of reproducing kernels // Trans. of the American Mathematical Soc. – 1950. – Vol. 68. – P. 337–404.
16. Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections // Communications in Statistics. Theory and Methods. – 1978. – Vol. 7. – N 1. – P. 13–26.
17. Murata N., Yoshizawa S., Amari S. Network information criterion – Determining the number of hidden units for an artificial neural network model // IEEE Transactions on Neural Networks. – 1994. – Vol. 5, N 6. – P. 865–872.
18. Craven P., Wahba G. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation // Numerische Mathematik. – 1979. – N 31. – P. 377–403.
19. Rissanen J. Stochastic complexity // J. of the Royal Statistical Society, Series B. – 1987. – Vol. 49, N 3. – P. 223–239.
20. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наукова думка, 1985. – 216 с.