«Оптика атмосферы и океана», **27**, № 1 (2014)

УДК 519.23, 551.551.8

Применение нового статистического метода для оценки адекватности моделей распространения примесей в пограничном слое атмосферы

В.А. Жуков[1,2], Б.М. Десятков[1], Н.А. Лаптева[1]*

[1]ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора
630559, Новосибирская область, пос. Кольцово
[2]ЗАО «Вектор-Бест»
630128, г. Новосибирск, ул. Пасечная, 3

Поступила в редакцию 24.04.2013 г.

Важным этапом создания математических моделей распространения примесей в пограничном слое атмосферы является проверка их на экспериментальном материале или на данных наблюдений – оценка адекватности модели. Однако невозможность повторения эксперимента при неизменных метеорологических условиях и как следствие малый объем выборки данных измерений не позволяют оценить адекватность корректным образом.

Предложен новый, основанный на интегральном преобразовании, метод оценки адекватности математических моделей. Он может быть использован даже в тех случаях, когда имеются только однократные измерения концентрации, в том числе с различными законами распределения в каждой точке пространства. Возможно объединение данных отдельных экспериментов, проведенных в различных метеорологических условиях. Эффективность метода иллюстрируется примером.

Ключевые слова: пограничный слой атмосферы, распространение примесей, математическая модель, адекватность, интегральное преобразование; boundary air layer, admixture distribution, mathematical model, adequacy, integral transform.

Введение

Одним из основных этапов разработки математических моделей распространения примесей в пограничном слое атмосферы является проверка их на экспериментальном материале или на данных наблюдений с целью оценки адекватности модели. Обычно это выполняется путем прямого сравнения теоретических результатов, полученных с помощью модели, с экспериментальными данными, например [1–3]. При этом чаще всего теоретические результаты представляют собой математическое ожидание анализируемой величины, а экспериментальные данные или данные наблюдений оказываются одной из возможных реализаций случайного процесса либо некоторым средним значением, полученным по относительно небольшой выборке. Таким образом, сравнивают математическое ожидание с одной из возможных реализаций случайного процесса.

В этом случае для убедительной оценки адекватности модели необходимо использовать методы математической статистики. Суть этих методов заключается в проверке выдвинутой гипотезы (в данном случае – об адекватности модели) на основе некоторых статистических критериев. Для этого необходимо иметь достаточно большие выборки данных наблюдений, полученных в большом количестве экспериментов при неизменных внешних условиях, в частности метеорологических. Сформированные таким образом выборки будут принадлежать одной генеральной совокупности, и применение статистических критериев будет корректно.

Однако чаще всего проведение таких экспериментов практически невозможно. Поэтому модели проверяются на относительно небольшом объеме данных наблюдений или на данных, взятых в относительно близких, но, вообще говоря, различных метеорологических условиях. При этом объединение экспериментальных данных, относящихся к различным экспериментам, в одну большую выборку является некорректным.

В данной работе предложен универсальный метод оценки адекватности математических моделей. Он основан на следующем известном свойстве случайных величин [4]: если $f(\xi)$ и x – соответственно плотность вероятности и реализация случайной величины ξ, то интеграл $\int_x^\infty f(\xi)d\xi$ также является случайной величиной, причем распределенной равномерно на отрезке [0, 1] независимо от

* Владимир Александрович Жуков (vzhukov@vector-best.ru); Борис Михайлович Десятков (dbm@vector.nsc.ru); Наталья Александровна Лаптева (lapteva@vector.nsc.ru).

вида исходного распределения. Метод проверен на разработанной нами ранее математической модели распространения примесей в атмосфере.

Обоснование метода

Пусть в некоторой точке пространства x_i (i = 1, ..., N) имеем экспериментальное значение C_{ie} концентрации C_i с функцией плотности вероятности $f_i(\theta_i, C_i)$, где θ_i — параметры закона распределения концентрации, подлежащие оценке с помощью некоторой модели. Адекватность модели означает, что экспериментальные значения концентрации C_{ie} могут считаться реализациями соответствующих случайных величин с функциями плотности вероятности $f_i(\theta_{iC}, C_i)$, где θ_{iC} — оценки параметров распределения, рассчитанные по модели. Справедливость последнего означает, что вероятности

$$P_i = \int_{C_{ie}}^{\infty} f_i(\theta_{iC}, C_i) dC_i$$

образуют выборку объема N из равномерно распределенной на отрезке [0, 1] генеральной совокупности. Гипотеза о том, что полученная выборка имеет равномерное распределение, может быть проверена с помощью критерия согласия, например критерия χ^2 [5]. Если выдвинутая гипотеза подтверждается, то можно сделать положительный вывод об адекватности модели. В противном случае вопрос об адекватности модели остается открытым.

Материалы и методы

Для иллюстрации предлагаемого метода воспользуемся данными, полученными при аэрозольной обработке сельскохозяйственных полей раствором дендробациллина с использованием установленного на автомобиле аэрозольного генератора АГ-УД-2 [6], который формирует линейный источник аэрозолей, а также моделью [7]. В этих экспериментах с помощью пробоотборника ПДК-50 и 4-ступенчатых каскадных импакторов с последующим микробиологическим анализом определялись значения интегральной концентрации (дозы) в приземном слое. Кроме этого измерялась плотность отложений препарата на листьях капусты и полиэтиленовых подложках.

В табл. 1 приведены рассчитанные по модели [7] и экспериментальные значения доз в воздухе на высоте 1 м, относящиеся к двум различным по метеорологическим условиям экспериментам. Контрольные точки располагались рядами поперек направления ветра. В первом столбце указан номер точки (ij), в котором первая цифра — номер ряда, вторая цифра — номер точки в ряде. Расстояние между рядами 100 м, расстояние между точками в ряде 50 м.

Таблица 1

Измеренные в эксперименте ($\lg d_{eij}$) и рассчитанные по модели ($\lg d_{cij}$) значения логарифмов доз

Эксперимент № 1				Эксперимент № 2			
(ij)	$\lg d_{eij}$	$\lg d_{cij}$	σ_{cij}	(ij)	$\lg d_{eij}$	$\lg d_{cij}$	σ_{cij}
12	8,36	8,83	0,15	12	6,15	7,90	0,18
13	8,38	8,88	0,15	13	6,30	7,91	0,18
14	8,63	8,82	0,15	14	6,51	7,90	0,18
21	8,70	7,97	1,10	21	7,11	7,43	1,62
22	7,99	8,43	1,10	22	7,65	7,69	1,62
23	5,98	8,53	1,10	23	7,60	7,87	1,62
24	8,49	8,45	1,10	24	8,00	7,79	1,62
25	8,34	8,15	1,10	25	4,04	7,28	1,62
31	7,62	7,81	0,63	31	7,85	7,20	0,48
32	8,80	8,11	0,63	32	6,86	7,54	0,48
33	8,41	8,26	0,63	33	7,99	7,68	0,48
34	7,00	8,28	0,63	34	7,75	7,66	0,48
35	8,11	8,11	0,63	35	8,04	7,49	0,48
36	8,00	7,80	0,63	41	7,56	7,34	0,17
41	8,23	7,85	0,20	42	7,20	7,26	0,17
42	8,11	7,95	0,20	43	7,49	7,11	0,17
43	7,84	7,95	0,20	44	7,32	6,89	0,17
51	7,11	7,87	0,80	45	7,20	6,60	0,17
52	7,41	7,88	0,80	51	6,85	7,23	0,20
53	7,49	7,79	0,80	52	7,20	7,28	0,20
54	7,73	7,63	0,80	53	7,23	7,26	0,20
55	6,74	7,41	0,80	54	7,46	7,15	0,20
56	5,81	7,11	0,80	55	7,23	7,04	0,20
57	5,78	6,80	0,80				

Примечание. lg — логарифм по основанию 10; σ_{cij} — стандартные отклонения логарифмов прогнозируемых значений доз, (ij) — номер точки.

Согласно [8] предполагаем, что функция плотности распределения интегральной концентрации (дозы) C_{ij} в точке (ij) – $f_{ij}(\lg d_{cij}, \sigma^2_{cij}, c_{ij})$ – подчиняется логарифмически-нормальному закону с математическим ожиданием, равным десятичному логарифму рассчитанного значения дозы $\lg d_{cij}$, и дисперсией σ^2_{cij}, равной выборочной дисперсии s^2_{ei}, рассчитанной по измеренным значениям дозы в i-м ряде:

$$\sigma^2_{cij} = s^2_{ei} = \frac{\sum_{j=1}^{i_k} (\lg d_{eij} - D_i)^2}{i_k - 1},$$

где $D_i = \frac{\sum_{j=1}^{i_k} \lg d_{eij}}{i_k}$, i_k – количество точек в i-м ряде, d_{eij} – измеренное значение дозы в j-й точке i-го ряда.

Далее в соответствии с изложенным алгоритмом для каждой точки (ij) вычисляем вероятности $P_{ij} = \int_{C_{ije}}^{\infty} f_{ij}(\lg d_{cij}, \sigma^2_{cij}, C_{ij}) dC_{ij}$. Нуль-гипотезу о принадлежности выборки значений P_{ij} генеральной совокупности, распределенной по равномерному закону, проверяли по критерию χ^2 на уровне значимости $\alpha = 5\%$, при этом отрезок [0, 1] был разбит на 5 одинаковых интервалов.

Результаты и обсуждение

В результате расчетов получили, что вычисленное значение χ^2-статистики равно 7,25 в первом эксперименте и 7,65 во втором. Критическое значение статистики равно 9,49 (при уровне значимости 5% и для 4 степеней свободы). Если данные двух экспериментов объединить в одну выборку, то в этом случае значение χ^2-статистики будет равно 8,21.

Анализ экспериментальных значений доз, приведенных в табл. 1, показывает, что значения доз в точке 23 в первом эксперименте и в точке 25 во втором эксперименте выпадают из общего ряда значений. Если эти точки не учитывать при проверке адекватности модели, то результаты будут несколько лучше, а именно: значение χ^2-статистики равно 6,35 в первом эксперименте и 5,78 во втором. Если данные двух экспериментов объединить в одну выборку, то в этом случае значение χ^2-статистики будет равно 7,78.

Таким образом, во всех случаях вычисленные значения χ^2-статистики меньше критического значения, равного 9,49 (при уровне значимости 5% и для 4 степеней свободы). Это означает, что у нас нет оснований отвергнуть нуль-гипотезу и можно сделать положительный вывод об адекватности модели.

Дополнительно по данным первого эксперимента были проведены две серии тестовых расчетов. В первой из них к рассчитанному значению математического ожидания логарифма дозы прибавлялась величина ε_1 из интервала от –1 до 1 с шагом 0,2. Это можно интерпретировать как смещение рассчитанных значений доз, обусловленное наличием в модели систематической ошибки. Во второй серии к стандартному отклонению σ_{cij} прибавлялась величина ε_2 из интервала от 0 до 1 с шагом 0,1, что соответствует увеличению дисперсии и ширины доверительного интервала вычисленных значений доз. В этом случае предполагалось, что ошибки измерений и данные по дозам, необходимые для тестирования адекватности, определялись в независимых двух сериях экспериментов. При этом имитируется ситуация проведения эксперимента по оценке дисперсии ошибки измерения в точке с меньшей точностью (большей дисперсией), чем в эксперименте для проверки адекватности модели. Результаты этих расчетов приведены в табл. 2.

Таблица 2

Рассчитанные значения критерия χ^2 при вариации параметров ε_1 и ε_2

ε_1	χ^2	ε_2	χ^2
–1,0	63,66	0,0	7,25
–0,8	50,75	0,1	2,21
–0,6	37,83	0,2	6,58
–0,4	20,12	0,3	11,79
–0,2	6,58	0,4	15,33
0,0	7,25	0,5	16,77
0,2	16,58	0,6	24,71
0,4	40,54	0,7	29,29
0,6	77,62	0,8	32,62
0,8	109,58	0,9	33,46
1,0	123,87	1,0	36,79

Видно, что при наличии существенных ошибок в теоретических или экспериментальных данных сформулированная выше нуль-гипотеза не подтверждается и вопрос об адекватности используемой математической модели остается открытым.

Заключение

Предложенный метод оценки адекватности математических моделей распространения примесей в атмосфере позволяет объединять в одну выборку данные отдельных экспериментов, проведенных в различных метеорологических условиях. Он может быть использован даже в тех случаях, когда имеются только однократные измерения концентрации, в том числе с различными законами распределения в каждой точке пространства.

1. *Берлянд М.Е.* Прогноз и регулирование загрязнения атмосферы. Л.: Гидрометеоиздат, 1985. 272 с.
2. *Марчук Г.И.* Математическое моделирование в проблеме окружающей среды. М.: Наука, 1982. 320 с.

3. *Пененко В.В., Алоян А.Е.* Модели и методы для задач охраны окружающей среды. Новосибирск: Наука, 1985. 256 с.
4. *Джонсон Н., Лион Ф.* Статистика и планирование эксперимента в технике и науке. М.: Мир, 1980. 336 с.
5. *Худсон Д.* Статистика для физиков. М.: Мир, 1967. 298 с.
6. *Десятков Б.М., Котлярова С.С., Никонов А.М., Павленко Ю.С.* Сравнение теоретических и экспериментальных значений интегральных концентраций и плотности отложений при обработке полей БСЗР: Материалы отраслевого совещания // Биологические и технологические проблемы создания вирусных препаратов для интегрированной защиты растений. Пос. Кольцово, Новосибирская обл., 26–29 сентября 1989 г. С. 72.
7. *Десятков Б.М., Сарманаев С.Р., Бородулин А.И.* Численно-аналитическая модель переноса аэрозоля в термически стратифицированном пограничном слое // Оптика атмосф. и океана. 1996. Т. 9, № 6. С. 815–820.
8. *Бородулин А.И., Десятков Б.М.* Моделирование распространения примесей в атмосферном пограничном слое. Новосибирск: НГУ, 2007. 376 с.

V.A. Zhukov, B.M. Desyatkov, N.A. Lapteva. **A new statistical method for assessment of adequacy of models of admixture distribution in the boundary air layer.**

An important stage of creating mathematical models of aerosol distribution in the boundary air layer is to check their validity using the data of experimental measurements. However, the impossibility of repeating the experiments under the invariable weather conditions (with the resulting small sample size of the measurements) does not allow one to estimate model adequacy correctly.

A new method of the assessment of adequacy of mathematical models based on an integral transform is proposed in this work. The method can be used even when there are only one-time measurements of concentration, including those with various laws of distribution at each point of space. It is quite possible to combine these separate experiments made under various weather conditions. The efficiency of the method is illustrated by an example.