

Mutli-dimensional Presentation and Analysis of Stationary Land Observation Data of Atmospheric Pollutant Concentrations for Verification of Hydrodynamic Forecast Models

Valery M. Tokarev¹, Marina Y. Zdereva¹

¹FSBI “Siberian Regional Hydrometeorological Research Institute”

Abstract: The paper describes a possibility of assessing the influence of weather conditions on the concentration of anthropogenic impurities using the parameters of hydrodynamic models of the atmosphere. The sample included official data of observations of eleven impurities at ten observational stations of the city of Novosibirsk over the period of four years, and synchronous parameters obtained from the output fields of global hydrodynamic models of the atmosphere. An algorithm for factor analysis for statistically non-uniform samples is proposed. The possibility of using binary trees to estimate the dependence of impurities on atmospheric parameters is demonstrated.

Keywords: dynamics of atmospheric impurity concentration, factor analysis, dispersion decomposition, binary trees.

МНОГОМЕРНОЕ ПРЕДСТАВЛЕНИЕ И АНАЛИЗ ДАННЫХ СТАЦИОНАРНЫХ НАЗЕМНЫХ НАБЛЮДЕНИЙ ЗА КОНЦЕНТРАЦИЯМИ АТМОСФЕРНЫХ ПРИМЕСЕЙ ДЛЯ ВЕРИФИКАЦИИ ГИДРОДИНАМИЧЕСКИХ МОДЕЛЕЙ ПРОГНОЗА

Токарев В.М.¹ Здерева М.Я.¹

¹ФГБУ «Сибирский региональный научно-исследовательский гидрометеорологический институт»

В статье описана возможность оценки влияния погодных условий на концентрацию антропогенных примесей с помощью параметров гидродинамических моделей атмосферы. В качестве выборки использованы официальные данные наблюдений за одиннадцатью примесями на десяти постах города Новосибирска за четыре года и синхронные параметры, полученные из выходных полей глобальных гидродинамических моделей атмосферы. Предложен алгоритм факторного анализа для статистически неоднородных выборок. Продемонстрирована возможность использования бинарных деревьев для оценки зависимостей примесей от параметров атмосферы.

Ключевые слова: динамика концентраций атмосферных примесей, факторный анализ, разложение дисперсий, бинарные деревья

Динамика концентраций антропогенных атмосферных примесей, несомненно, зависит от погодных условий. Однако эмиссии таких примесей, за редким исключением, имеют мало общего с атмосферными процессами. Таким образом, задача моделирования и прогнозирования аэрозольного и газового состава атмосферного воздуха сильно осложняется. На входе модели надо иметь не только пространственные данные метеорологических наблюдений, но и оценки эмиссий от множества разнородных источников. К сожалению, системы мониторинга качества воздуха пока не могут предоставить данные наблюдений достаточного для моделирования и верификации объема и детальности. В случае же отсутствия подробной карты эмиссий, задача представляется невыполнимой. Попробуем все-таки извлечь полезную информацию из доступной с помощью не совсем традиционных подходов и алгоритмов.

Для исследования были использованы архивные данные наблюдений за концентрациями атмосферных примесей Новосибирского ЦМС (4 года, 11 контролируемых примесей на 10 постах) и синхронные архивные данные глобальных прогностических моделей Exeter и ECMWF.

На первом этапе из данных наблюдений примесей формировался многомерный массив: 11 примесей * 10 постов * 3 суточных срока * 12 месяцев * 4 года. Таким образом, далее можно было оценивать влияние таких факторов, как местоположение на территории мегаполиса, суточный и сезонный ход, и межгодовой тренд. Однако от оценок суточного хода пришлось отказаться из-за отсутствия ночных наблюдений (в выходные также были пропуски).

Качественный характер динамики примесей оценивался по перекрестным сравнительным графикам с разным временным масштабом (от суток до годов). И первые выводы были неутешительные: в рядах присутствует очевидная сильная нестационарность в масштабах от часов до нескольких суток, а тренды и сезонность трудноразличимы и немонотонны. Это означало, что придется иметь дело с выделением очень слабого полезного сигнала (влияния погодных условий) на фоне преобладающего шума от неизвестной динамики эмиссий и, возможно, технологических особенностей мониторинга.

Следующим шагом было получение количественных оценок факторов. Для этого использовались условные статистические оценки из известного разложения дисперсии [1]:

$$DY = DM(Y|x) + MD(Y|x), \quad (1)$$

где: $DM(Y|x)$ определяет долю дисперсии Y за счет влияния X ,

$MD(Y|x)$ – неопределенная (остаточная) дисперсия Y .

Тогда:

$DM(Y|x)/DY$ - дисперсионное отношение, мера (нелинейной) связи.

Дополнительную полезную информацию можно получить, рассматривая отдельно выборки из минимальных и максимальных по городу наблюдений. При этом минимум можно условно интерпретировать, как "фон", т.е. пункт, достаточно удаленный от максимума, т.е. источника "выброса".

Все расчетные характеристики сведены в таблицу (Таб.1)

Таблица 1. Компоненты дисперсионной идентификации концентрации примесей по пространственно-временным осям матрицы данных

Компоненты дисперсионной идентификации концентрации примесей по пространственно-временным осям матрицы данных

	по городу				N	посты		тренд		сезонность	
	макс	мин	сред	дисп		DM(Y/X)	DM(Y/X)/DY	DM(Y/X)/DY		DM(Y/X)/DY	
	MY	MY	MY	DY				выброс	фон	выброс	фон
							%	%	%	%	%
Пыль	0.55405	0.12934	0.30519	0.04998	10	0.00293	5.86	2.71	2.24	6.95	13.96
SO2	0.01860	0.00136	0.00782	0.00026	6	0.00001	6.18	8.42	5.07	22.53	13.32
CO	4.78524	1.28747	2.58200	4.07565	10	0.27580	6.76	11.38	0.25	2.46	19.45
NO2	0.16659	0.03972	0.09193	0.00533	10	0.00030	5.66	12.62	53.07	2.62	0.61
NO	0.08238	0.02959	0.05714	0.00274	7	0.00066	24.32	37.35	51.67	1.80	0.85
H2S	0.00046	0.00046	0.00046	0.00000	1	0.00000	0.00	5.05	5.05	0.46	0.46
Фенол	0.00740	0.00124	0.00415	0.00004	3	0.00000	0.47	11.80	10.21	0.42	0.37
Сажа	0.09714	0.00164	0.03330	0.00527	10	0.00038	7.38	0.67	4.26	7.49	9.87
HF	0.00763	0.00763	0.00760	0.00008	1	0.00000	0.00	1.49	1.49	3.21	3.21
Аммиак	0.13945	0.02303	0.07739	0.01305	3	0.00067	5.17	5.27	1.27	2.20	1.90
Форм	0.02357	0.00734	0.01525	0.00018	6	0.00001	10.65	21.87	52.55	0.47	0.86

Полученная таблица позволяет выполнить довольно тонкий статистический анализ. Так, средний максимум по городу для диоксида серы превышает средний минимум более, чем в 10 раз (больше всех таблице). То есть, на первый взгляд, для этой примеси местоположение поста является ключевым фактором. Однако, дисперсионное отношение

оставляет этому фактору лишь 6% всей дисперсии, а сезонной компоненте - от 13% ("фон") до 23% ("выброс"). Так что все весьма относительно. И преобладание сезонных колебаний выбросов SO₂ за счет сжигания угля котельными и частным сектором вполне логично и объяснимо, впрочем, этим же объясняются такие различия в местоположении (частный сектор в Новосибирске занимает отдельные компактные территории).

Самый большой трендовый вклад - у оксидов азота и формальдегида, и это никак не радует, поскольку по графикам - это почти монотонный рост по годам. И, если оксиды азота объясняются ростом парка автотранспорта, то тренд формальдегида, скорее всего, связан с неконтролируемым использованием пластиков и эпоксидных смол.

Подобный анализ можно продолжить, но нам важно, что примененный алгоритм позволяет извлекать полезную количественную информацию из очень статистически непростых для обработки данных.

Остается вопрос, а где же здесь погода? Да где-то в очень значительной остаточной дисперсии (для фенола это почти 90%), как трудноразличимый фактор на фоне высокоамплитудных аварийных или "планово-незаконных" выбросов, которые хорошо видны на графиках. Анализируемую выборку сложно разбить на отрезки с "хорошей" и "плохой" погодой для концентраций всех примесей. Нет такого одного модельного параметра атмосферы. А чтобы это проверить и оценить используем один из алгоритмов распознавания - построение логического бинарного дерева решений. Собственно он может сразу строить прогностические решающие правила по обучающей выборке, что и делали авторы в другой работе.

Логически алгоритм [2] очень прост. Он последовательно разделяет исходную выборку на две с разнесением вероятностей бинарного предиктанта в каждой подвыборке. Так получается дерево с дwoящимися ветвями, в узлах которых фиксируется выбранный лучший параметр разделения (предиктор) и значение этого порога. Таким образом, выходное дерево содержит список выбранных лучших параметров с лучшими пороговыми значениями распознавания двух образов предиктанта ("да" или "нет") и со знаковой информацией о связи предиктора и предиктанта. Эта связь определяется тем, совпадают ли условия деления по порогу предиктора (больше/меньше порога) с увеличением/уменьшением соответствующей вероятности образа предиктанта ("да"/"нет"). Совпадение означает прямую связь: чем больше значение предиктора, тем больше вероятность 1-го образа предиктанта ("да") и наоборот. Не вдаваясь в подробности, по каким критериям выбираются лучшие варианты разделения ветвей и правило останова (они важнее для целей прогноза), перейдем к получению статистических связей для нашей задачи. Сведение концентраций примесей к бинарной форме выполнялось по индивидуальному пороговому условию: больше или меньше ПДК. Матрица предикторов формировалась из списка различных выходных параметров атмосферы глобальных моделей Exeter и ECMWF в ближайших узлах сетки ГРИБ у земли, на верхней границе пограничного слоя и в свободной атмосфере.

Полученные результаты обучения сведены в таблицу (Таб.2) Для фенола и аммиака к обычной обучающей выборке фактов превышения ПДК - "хотя бы на одном посту" добавлено обучение для двух и более постов.

Таблица 2. Прямые и обратные связи превышения ПДК примесей и параметров атмосферы

Прямые(+) и обратные(-) связи превышения ПДК примесей и параметров атмосферы.

Параметры атмосферы	Примеси								
	Пыль	СО	NO2	фенол2	фенол1	сажа	аммиак2	аммиак1	форм
r0	-	-	0	-	0	-	0	+	-
VV	-	0	0	0	0	0	0	0	0
P0	0	-	0	+	0	0	0	0	0
gradP0	0	0	+	+	0	0	0	0	0
lapP0(+цикл,-анти)	0	0	0	0	+	0	0	0	0
dP0(t-24)	0	0	0	0	0	0	0	0	0
Tz-T850	-	0	-	0	-	-	+	0	0
lapT850(+холод,-тепло)	0	0	0	0	0	0	-	0	-
gradT850	0	0	-	0	0	0	0	0	0
dT850(t-24)	0	0	0	0	0	0	0	0	0
H500	+	+	-	0	0	0	+	0	0
lapH500(+цикл,-анти)	0	-	0	-	+	0	0	0	0
gradH500	0	0	0	0	0	0	0	+	0
dH500(t-24)	0	0	0	0	0	0	0	+	0

Первое, что можно отметить, это немногочисленность найденных связей, но это ожидаемо для таких зашумленных данных. Также ожидаема отрицательная связь относительной влажности у земли (r0), но для аммиака это не так, причем, и другие связи у аммиака чисто индивидуальные. Логично и рассеяние примесей с изменением вертикального градиента температуры в сторону неустойчивости (Tz-T850). Следует также отметить вероятность такого сценария рассеяния примеси, когда слабый ветер и устойчивая стратификация способствуют накоплению концентрации непосредственно около точечного источника, а неустойчивая погода будет рассеивать примесь, донося ее до постов наблюдений.

Еще одно методическое замечание относится к выбору критического порога для разбиения предиктанта на образы. ПДК - не физическая характеристика, а биологическая, так что выбор других порогов, например, единых для всех примесей частотных квантилей, должен обеспечить более четкое разделение градаций концентраций по условиям погоды.

Полученные в работе результаты не совсем очевидны и могут помочь в настройках численных моделей на прогноз уровней загрязнения атмосферы антропогенными примесями и последующей валидации и анализа результатов.

Список литературы:

1. Дисперсионная идентификация/Под ред. Н.С.Райбмана.- М.: Наука,1981-336с.
2. Манохин А.Н. Алгоритм DW для распознавания образов: Пакет прикладных программ ОТЭКС.- Новосибирск: изд-во Новосибирского государственного университета, 1981.-С.3-30.