

Обзор методов снижения размерности при обработке гиперспектральных оптических сигналов¹

А. ДЕЛЬ АГИЛА, Д.С. ЕФРЕМЕНКО¹, Т. ТРАУТМАНН

Институт методологии дистанционного зондирования Германского аэрокосмического центра (DLR), Оберфаффенхофен, Германия
E-mail: dmitry.efremenko@dlr.de

Аннотация

Гиперспектральные датчики проводят измерения в узких сопряжённых полосах спектра электромагнитного излучения. Целью при этом обычно является обнаружение определённого объекта или элемента среды, имеющих присущие только им спектральные характеристики. В частности, гиперспектральные измерения применяются при дистанционном зондировании атмосферы для выявления малых газовых компонент. Для улучшения эффективности алгоритма обработки гиперспектральных данных были использованы методы уменьшения количества данных. В статье описаны методы снижения размерности применительно к гиперспектральному дистанционному зондированию атмосферы. При снижении размерности происходит исключение из данных избыточной информации, и в настоящее время снижение размерности является неотъемлемой частью высокопроизводительных моделей переноса излучения. В обзоре описано, как можно использовать метод главных компонент² для моделирования спектрального распределения энергетической яркости и определения составляющих атмосферы, ускоряя тем самым на порядки скорость обработки данных. Представленные методы являются обобщёнными, и их можно непосредственно использовать для решения как атмосферных задач, так и задач из других связанных с материаловедением областей знаний.

Ключевые слова: пассивное дистанционное зондирование, гиперспектральные данные, метод главных компонент, физически полное машин-

ное самообучение, определение малых газовых компонент.

1. Введение

Гиперспектральные датчики регистрируют прошедшее или отражённое излучение в узких сопряжённых полосах спектра электромагнитного излучения. Целью при этом обычно является обнаружение определённого объекта или элемента среды, имеющих присущие только им спектральные характеристики, т.е. свои особенные «отпечатки пальцев». Гиперспектральная визуализация результатов зондирования стала одним из наиболее действенных методов в различных областях, включая астрономию, минералогию, сельское хозяйство, медицину и химию. Например, гиперспектральные данные (которые иногда называют гиперкубическими данными (*hypercube data*) или кубом изображения (*image cube*)), используются в астрономии и при дистанционном зондировании Земли для создания пространственно-разрешённого спектрального изображения, обеспечивающего возможность более точного отождествления и классификации объектов в поле зрения прибора. В то же время, для обработки гиперспектральной информации и получения определённой характеристики рассеивающей среды требуются существенные объёмы памяти и вычислительные возможности.

В этом обзоре мы сконцентрировали внимание на гиперспектральном дистанционном зондировании атмосферы. Пассивные датчики состава атмосферы регистрируют и записывают энергетическую яркость отражённого атмосферой Земли излучения в ультрафиолетовой, видимой и инфракрасной областях спектра. Информацию об атмосфере затем извлекают из этих спектральных данных при помощи так называемых атмосферных процессо-

ров (*atmospheric processors*), то есть алгоритмов, специально предназначенных для преобразования результатов произведённых этими датчиками измерений [1]. Извлечение информации о геофизических характеристиках (данные 2-го уровня) из спектральных распределений энергетической яркости (данные 1-го уровня) представляет собой большую вычислительную проблему и требует использования высокоскоростных компьютеров [2].

Последние достижения в области оптики, конструкции датчиков и методов измерений привели к существенному улучшению характеристик датчиков состава атмосферы, таких как пространственное разрешение и отношение сигнала к шуму. В таблице проведено сравнение приборов предыдущего поколения, таких как оборудование для глобального мониторинга озона (*Global Ozone Monitoring Experiment*) *GOME* и *GOME-2*, и новейшего прибора для мониторинга тропосферы (*TROPOspheric Monitoring Instrument*) *TROPOMI*, установленного на борту спутника *Copernicus Sentinel 5 Precursor (S5P)*. Пространственное разрешение прибора *TROPOMI* на два порядка выше, и он позволяет получить в день 21 млн спектров 1-го уровня, т.е. почти 8 млрд спектральных точек, тогда как отношение сигнала к шуму в ультрафиолетовом и видимом каналах достигает у него значений, примерно равных 1500. На рис. 1 приведён пример полученной с помощью этого прибора карты распределения тропосферного диоксида азота (*NO₂*). Обратите внимание на то, что на рисунке хорошо видны загрязнения атмосферы, обусловленные большими городами и морскими путями сообщения. Данные высокого разрешения позволяют не только зарегистрировать загрязнение воздуха над отдельными городами, но и определить источник загрязнения и, тем самым, выявить области, наиболее опасные с точки зрения загрязнения атмосферы. Такое имеющее высокую разрешающую способность дистанционное спутниковое зондирование атмосферы очень полезно для определения состава атмосферы в мировом масштабе, в частности, для выявления мелкомасштабных источников загрязнения и увеличения доли наблюдений, осуществляемых при отсутствии облачности. В то же время, высокая разрешающая способ-

¹ Перевод с англ. Е.И. Розовского

² В рассматриваемой области принято использовать имя существительное женского рода «компонента». – Прим. пер.

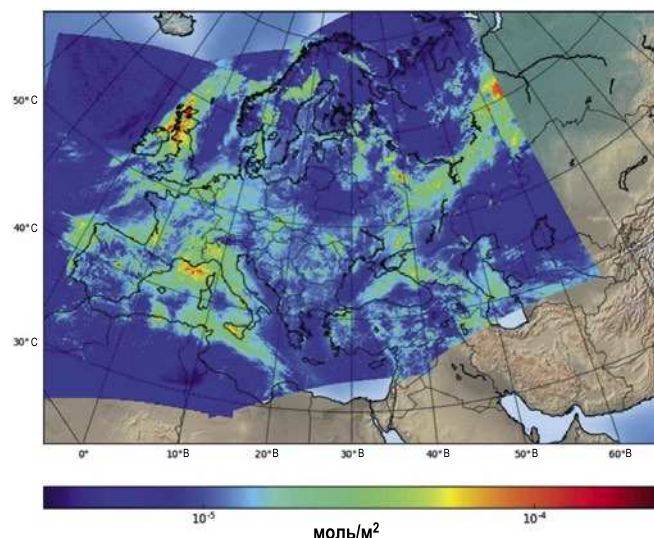
ность современных датчиков состава атмосферы выливается в чрезвычайно большой объём подлежащих обработке данных – 240 Тбайт в год для данных 1-го уровня.

Фактически, объём спутниковых данных возрастает быстрее, чем быстроедействие компьютеров [4]. Так как данные, получаемые при дистанционном зондировании, удовлетворяют 3V-критерию Дуга Лани (*Doug Laney*) (существенный рост в части объёма, скорости и многообразия (*volume, velocity, variety*)), то их относят к большим данным. Чтобы удовлетворить требования к быстройдействию, для атмосферных процессоров следует разработать новые эффективные методы обработки данных.

Модель переноса излучения (МПИ) является ключевым элементом и основным узким местом атмосферных процессоров. Кроме того, гиперспектральная МПИ включает в себя иерархию вложенных циклов вычислений [6], приведённую в виде псевдокода на рис. 2. В опубликованных недавно обзорах (см., например, [7–9]) показано, что существенное увеличение быстройдействия может быть обеспечено, скорее, благодаря оптимизации рамочной программы, в которой осуществляется вызов программы для расчёта переноса излучения (*Radiative transfer solver*), а не за счёт ускорения работы самой МПИ. Фактически, возможность улучшения быстройдействия программ для расчёта переноса излучения представляется маловероятной [10]. Было предпринято несколько попыток оптимизации циклов в части наземных пикселей и геометрии (см., например, [11–14] и приведённые в них ссылки на литературу). В настоящее время наиболее затратной с точки зрения гиперспектральных расчётов является цикл, связанный с длинами волн. Важной частью анализа больших данных является снижение размерности, которое в случае МПИ относится к циклу, связанному с длинами волн. Наиболее известным представителем способов снижения размерности является метод главных компонент (*principal component analysis – PCA*), который был предложен в 1901 г. К. Пирсоном [15] и в настоящее время является неотъемлемой частью гиперспектральных МПИ.

Были предприняты многочисленные попытки разработать гиперспектральные МПИ, в которых в полной

Рис. 1. Пример результатов измерений тропосферного диоксида азота (NO_2), выполненных 01.04.2019 при помощи прибора *Sentinel-5P* (данные находятся в свободном доступе на <https://s5phub.copernicus.eu/dhus>)



```

для каждого наземного пикселя
  для каждой длины волны: #<-- цикл гиперспектральной обработки
    для каждой геометрии:
      вызов программы для решения задачи переноса излучения (Radiative transfer solver);
  
```

Рис. 2. Иерархия циклов вычислений, приводящих к вызову программы для решения задачи переноса излучения (*Radiative transfer solver*)

мере учитывались бы взаимозависимость и статистические соотношения между данными 1-го и 2-го уровня [16]. Целью этого обзора является систематизированное описание основных концепций снижения размерности применительно к созданию атмосферных процессоров и сведения воедино последних достижений в этой области.

2. Снижение размерности

2.1. Наследство метода k -распределения

В науке об атмосфере и астрофизике методам снижения размерности предшествовали методы, которые хотя и не в полной мере использовали *PCA*, но всё же были основаны на сильной взаимозависимости гиперспектральных данных. В [17] отмечалось, что перенос излучения в спектральном интервале связан не с *line-by-line (LBL)* зависимостью коэффициента поглощения k от длины волны λ , а, скорее, с распределением коэффициента поглощения в пределах этого спектрального интервала. Этот подход называют методом k -распределения. Он основан на кумулятивном распределении частоты k , а именно, на $G(k)$, и обрат-

ная функция распределения $k(G)$ как раз и является k -распределением. Так как $G(k)$ представляет собой гладкую функцию, а $k(\lambda)$ сильно зависит от длины волны, то с точки зрения эффективности расчётов полезно заменить интегрирование в пространстве λ на интегрирование в пространстве G (в этом случае для численного интегрирования потребуется меньшее количество точек). В [18] этот метод был распространён на неоднородную атмосферу в приближении наличия корреляции между k -распределениями при различных уровнях давления – это метод коррелированных k -распределений. В других методах, таких, как подбор суммы экспоненциальных функций (*exponential sum fitting*) [19], спектральное отображение (*spectral mapping*) [20], основанный на k -бинировании подход (*k-binning approach*) [21], метод выборки прозрачности (*opacity sampling method*) [22], метод многомерных k -распределений (*multi-dimensional k-distribution method*) [23] и быстрые модели k -распределений (*fast k-distribution models*) [24, 25], используются аналогичные, хотя и основанные на более сложной математике, подходы. В [26] была рассмотрена модификация метода k -распределений, в которой интегрирование про-

водилось в исходном пространстве λ . В [27] был предложен основанный на однородных спектральных сетках эмпирический метод выбора наиболее характерных спектральных точек при проведении *LBL* расчётов. В совокупности, эти связанные с *k*-распределением исследования говорят о том, что гиперспектральную энергетическую яркость можно рассчитывать с использованием гораздо меньшего количества монохроматических расчётов, чем требуется в рамках *LBL* программы, и тем самым эти работы формируют основу для снижения размерности рассматриваемой задачи.

2.2. Базовая концепция снижения размерности

Простому говоря, снижение размерности – это представление исходного набора данных при помощи меньшего, чем вначале, количества параметров. Его можно считать одним из примеров сжатия данных с потерями [28]. Снижение размерности имеет решающее значение для стабильной и высокопроизводительной обработки результатов спектральных измерений. Оно обеспечивает исключение несущественной информации из исходного набора данных, уменьшение количества независимых параметров и улучшение эффективности машинного самообучения.

Существует разница между линейными и нелинейными методами снижения размерности. Более подробный обзор можно найти в [29–30] и в приведённых в данной статье ссылках на литературу. Сравнение линейных и нелинейных методов было проведено в [33]. Результаты этих численных экспериментов говорят о том, что нелинейные методы хорошо работают в отдельных искусственно сформированных случаях. Однако при решении реальных задач они вряд ли превосходят *PCA*. К аналогичным выводам пришли в [34], где несколько методов снижения размерности сравнивались применительно к ускорению рассмотрения переноса излучения. Принимая во внимание тот факт, что в результате сравнительного анализа не был выявлен ни один метод, который можно было бы считать наилучшим (более сложные и требующие больших временных затрат методы снижения размерности приводят к получению более точных результатов, и наоборот),

наш анализ будет ограничен классическим *PCA*.

2.3. Метод главных компонент

Несмотря на то, что методы снижения размерности хорошо известны и входят во многие статистические библиотеки (например, библиотека *scikit-learn* [35] для языка программирования *Python*), мы приведём короткий математический анализ, который позволит лучше понять вышеизложенное. Для простоты обозначим размер матрицы как $\in \mathbb{R}^{\text{строки} \times \text{столбцы}}$, вектор-строку энергетической яркости атмосферы для *W* длин волн $\{\lambda_w\}_{w=1, \dots, W}$ – как

$$\mathbf{y} = (y(\lambda_1), y(\lambda_2), \dots, y(\lambda_W)), \quad \mathbf{y} \in \mathbb{R}^{1 \times W}.$$

Тогда совокупность *S* спектров образует матрицу $\mathbf{Y} \in \mathbb{R}^{S \times W}$ с *i*-ой строкой \mathbf{y}_i , а \mathbf{y}_i можно записать в новой системе координат, как:

$$\mathbf{y}_i = \bar{\mathbf{y}} + \sum_{k=1}^W t_{ik} \mathbf{f}_k,$$

где $\bar{\mathbf{y}} = \sum_{i=1}^S \mathbf{y}_i / S$, $\bar{\mathbf{y}} \in \mathbb{R}^{1 \times W}$ – выборочное среднее для спектра (усреднённый спектр), t_{ik} – *k*-я координата вектора \mathbf{y}_i в новой системе координат, а

$$\mathbf{f}_k = (f_k(\lambda_1), f_k(\lambda_2), \dots, f_k(\lambda_W)) \in \mathbb{R}^{1 \times W}$$

– *k*-ый базисный вектор новой системы координат. Учитывая то обстоятельство, что высокоразмерные реальные данные обычно образуют линейное многообразие меньшей размерности, то спектр \mathbf{y}_i может быть спроецирован на имеющее размерность *K* подпространство ($K < W$) следующим образом:

$$\mathbf{y}_i \approx \bar{\mathbf{y}} + \sum_{k=1}^K t_{ik} \mathbf{f}_k, \quad (1)$$

или в соответствующем исходному набору данных матричном виде:

$$\mathbf{Y} \approx \bar{\mathbf{Y}} + \mathbf{T}\mathbf{F},$$

где $\bar{\mathbf{Y}} = \{\bar{\mathbf{y}}_1, \dots, \bar{\mathbf{y}}_S\} \in \mathbb{R}^{S \times W}$,

$$\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K\}^T \in \mathbb{R}^{K \times W},$$

$\mathbf{T} \in \mathbb{R}^{S \times K}$ – матрица с элементами

$$\{t_{ik}\}_{i=1, \dots, S}^{k=1, \dots, K}. \text{ Здесь и далее верхний индекс } T \text{ означает «транспонированная»}.$$

Преобразование (1) можно произвести методами снижения размерности, такими, как *PCA* [15]. В последнем случае, входящие в (1) базисные векторы \mathbf{f}_k называют «главными компонентами» (ГК) или эмпирическими ортогональными функциями (ЭОФ) и используют их в качестве *K* собственных векторов, соответствующих наиболее важным собственным значениям ковариационной матрицы $\text{cov}(\mathbf{Y}, \mathbf{Y}) \in \mathbb{R}^{W \times W}$. Координаты t_{ik} в новой системе координат и соответствующую матрицу \mathbf{T} называют «координатами главных компонент» (*principal component scores*).

3. Основанная на главных компонентах модель переноса излучения для гиперспектральных сигналов

В самом простом по своей сути подходе используется тренировочный набор спектров, позволяющий при помощи *PCA* сформировать набор ЭОФ, а затем при помощи *K* монохроматических энергетических яркостей восстановить гиперспектральный сигнал в *W* спектральных точках. Естественно, $K < W$.

Теория *PCA*, кратко изложенная в предыдущем разделе, говорит о линейной зависимости между координатами ГК и монохроматическими энергетическими яркостями:

$$\mathbf{y}(\lambda) = \bar{\mathbf{y}}(\lambda) + t_1 \mathbf{f}_1(\lambda) + \dots + t_K \mathbf{f}_K(\lambda). \quad (2)$$

Так что для заданного набора из *K* ЭОФ и *K* спектральных точек можно получить замкнутую систему из *K* линейных уравнений:

$$\begin{cases} \mathbf{y}(\lambda_1) = \bar{\mathbf{y}}(\lambda_1) + t_1 \mathbf{f}_1(\lambda_1) + \dots + t_K \mathbf{f}_K(\lambda_1), \\ \mathbf{y}(\lambda_2) = \bar{\mathbf{y}}(\lambda_2) + t_1 \mathbf{f}_1(\lambda_2) + \dots + t_K \mathbf{f}_K(\lambda_2), \\ \dots \\ \mathbf{y}(\lambda_K) = \bar{\mathbf{y}}(\lambda_K) + t_1 \mathbf{f}_1(\lambda_K) + \dots + t_K \mathbf{f}_K(\lambda_K). \end{cases} \quad (3)$$

Важно, что значения энергетической яркости в *K* спектральных точках представлены теми же ЭОФ. Решив (3), мы получаем координаты ГК t_1, \dots, t_K , а при помощи уравнения (2) можно с лёгкостью восстановить полный спектр в *W* спектральных точках.

Этот подход требует наличия заранее рассчитанных ЭОФ, которые могут быть получены при помощи тренировочного набора смоделированных или измеренных спектров. На рис. 3

приведён пример работы *PCA*, применённого к набору спектров, рассчитанному для полосы Хартли-Хаггинса, используемой при определении количества озона. Этот набор данных состоит из 10^5 спектров. При формировании спектра отражения изменяли следующие параметры: зенитный угол солнца, зенитный угол наблюдения, относительный азимутальный угол, альbedo поверхности, приведённую толщину слоя озона, высоту расположения поверхности и температуру. В правой части рис. 3 видно, что почти 99,9 % изменений данных можно объяснить влиянием только 5-ти ГК.

Чтобы добиться максимальной эффективности, параметры этой схемы, такие, как K и набор выбранных длин волн, следует подобрать опытным путём. Число K зависит от желательного уровня изменений, который должен быть замечен ГК. Для определения оптимального количества ГК было предложено несколько полумпирических правил (например, модель сломанной трости [36]). Однако нет никакого универсального правила выбора K – этот выбор зависит от области применения. Если же говорить о выборе длин волн, то в [37] был предложен основанный на использовании корреляционной функции метод выбора длин волн монохроматического излучения. Этот метод включает в себя следующие этапы:

1. Расчёт коэффициентов корреляции для значений энергетической яркости и последующее преобразование их в векторные углы при помощи функции «арккосинус».

2. Перестановка спектральных данных в соответствии с величиной коэффициента корреляции.

3. Отбор монохроматических энергетических яркостей, основанный на выборе предикторов с равным расстоянием между значениями коэффициентов корреляции.

Схема основанной на *PCA* модели гиперспектрального переноса излучения приведена на рис. 4. Входные данные, включающие в себя оптические параметры атмосферы, соответствующие набору длин волн, передаются в программу для расчёта переноса монохроматического излучения. Для получения наиболее репрезентативного набора данных рекомендуется использовать метод умной выборки [38], основанный на последовательностях Халтона [39]. В результате фор-

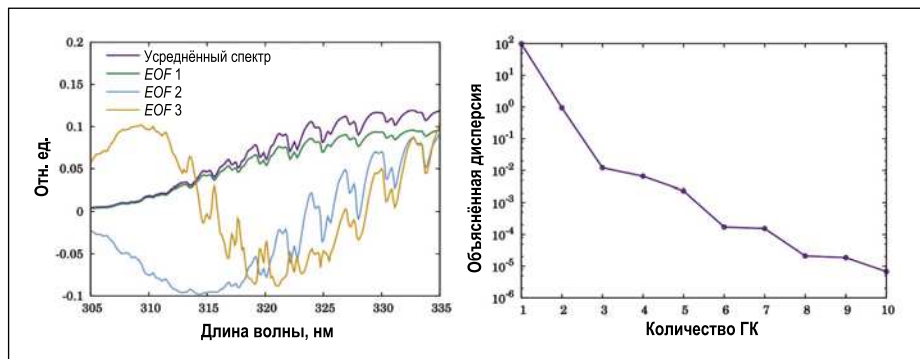


Рис. 3. Усреднённый спектр и три первые ЭОФ, рассчитанные в полосе Хаггинса (слева), и зависимость объяснённой дисперсии от количества ГК (справа)

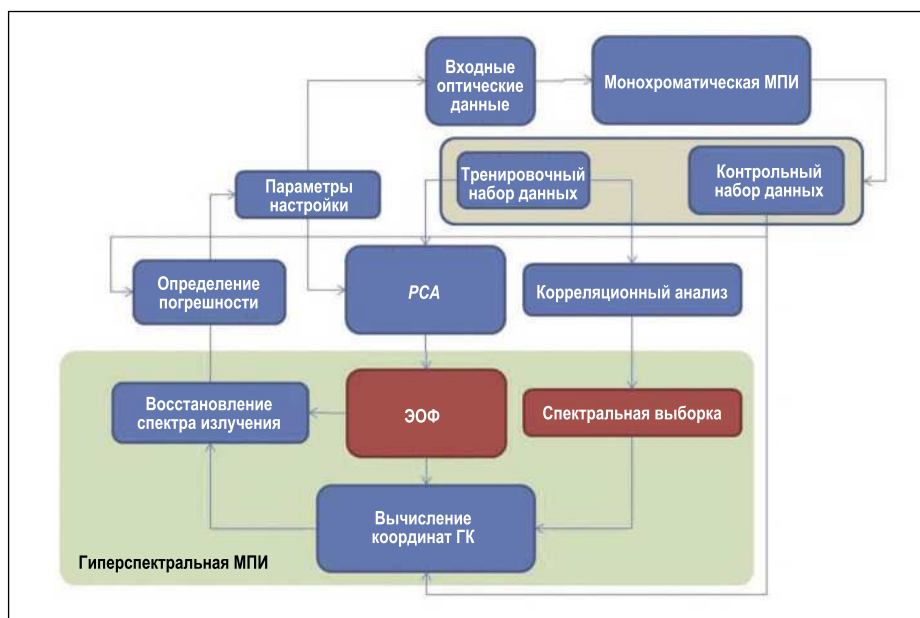


Рис. 4. Схема основанной на *PCA* модели переноса излучения с использованием предварительно рассчитанных ЭОФ – стадия «офлайн»

Рис. 5. То же, что и на рис. 4, но на стадии «онлайн»



мируется набор спектров, который делится на тренировочный и контрольный наборы данных. Применение к тренировочному набору данных *PCA* и корреляционного анализа позволяет, соответственно, получить систему ЭОФ и выбрать подмножество спектральных точек (спектральная выборка). Эти два результата сохраняются и используются для определения координат ГК для контрольного набора данных. Спектры во всём интервале длин волн восстанавливают при помощи уравнения (2), после чего можно определить погрешность этого восстановления. Если погрешность превышает допустимую, то увеличивают количество сгенерированных спектров и ГК. Основным результатом тре-

нировочной стадии является полученные ЭОФ и спектральной выборки (на рис. 4 эти результаты выделены красным фоном), которые позволяют обрабатывать новые данные в режиме онлайн (см. рис. 5).

В [37] отмечалась линейность оператора свёртки аппаратной функции и *PCA*. Поэтому и координаты ГК свёрнутых спектров являются линейными функциями монохроматических энергетических яркостей. Так что на тренировочной стадии соответствующие весовые коэффициенты линейной зависимости сохраняют совместно с системой ЭОФ и спектральной выборкой. Авторы [37] утверждают, что для ИК области, охватывающей промежуток между 650 и 3500 см^{-1} ,

использование свёрнутых спектров приводит к уменьшению количества монохроматических вычислений с нескольких тысяч до нескольких сот.

Описанный подход оказался эффективным и был использован в нескольких пакетах программ (например, в *PCTRM* [37], *RTTOV* [40] и других [41, 42]). Основным недостатком этого подхода является то, что он требует проведения длительных вычислений для тренировочного набора данных.

4. Метод главных компонент в случае дифференциальной спектроскопии оптического поглощения

В этом разделе мы рассмотрим использование *PCA* применительно к дифференциальной спектроскопии оптического поглощения (ДСОП) [43]. ДСОП широко используется для выявления малых газовых концентраций. Основными достоинствами этого метода являются его простота и устойчивость, обусловленные его способностью отфильтровывать влияние тех факторов, которые не рассматриваются должным образом в рамках *PCA*. Основная идея ДСОП состоит в том, что обусловленное газами поглощение приводит к появлению характерных особенностей в спектре энергетической яркости, тогда как влияние многократного рассеяния и иных факторов в пространстве длин волн обычно оказывается плавным. Согласно [44], энергетическую яркость излучения, отражённого атмосферой, состоящей из N_g газообразных веществ с коэффициентами поглощения $\sigma_{abs, g}$, можно в соответствии с законом слабого поглощения Бугера-Ламберта-Бера представить в виде:

$$\ln y(\lambda) = -\sum_{g=1}^{N_g} S_g \sigma_{abs, g}(\lambda) - P(\lambda) - RRS(\lambda), \quad (4)$$

где S_g – численная плотность газа g вдоль оптической траектории (которую называют ещё и плотностью наклонного столба (*slant column density*)), $P(\lambda)$ – полином, учитывающий влияние как рэлеевского рассеяния и рассеяния на аэрозолях и в облаках, так и отражения от поверхности, а RRS учитывает вращательное комбинационное рассеяние [45, 46].

В обычных ДСОП S_g определяют посредством минимизации методом наименьших квадратов невязки между измеренным (левая часть уравнения (4)) и смоделированным (правая часть уравнения (4)) спектрами энергетической яркости. Затем S_g преобразуют в плотность вертикального столба Ω_g с использованием приведённой толщины воздуха (*air mass factor*). Последнюю вычисляют для одной длины волны в предположении заданного вертикального распределения газа g [47, 48].

В [49] была предложена модификация метода ДСОП применительно к определению приведённой толщины слоя SO_2 , когда *PCA* применяют к спектрам, измеренным в районах с несущественным наличием SO_2 , например, в экваториальной зоне Тихого океана:

$$\ln y(\lambda) = \overline{\ln y}(\lambda) + \sum_{i=1}^K t_i \mathbf{f}_k(\lambda). \quad (5)$$

При этом ЭОФ охватывают всё многообразие данных, обусловленное протеканием физических процессов (например, рэлеевским рассеянием, вращательным комбинационным рассеянием и поглощением озоном). Кроме того, ЭОФ полностью учитывают особенности прибора (например, ухудшение его характеристик, аппаратную функцию и измерительные устройства). Это тренировочная стадия. Что касается районов, загрязнённых SO_2 , то в их случае уравнение (5) приведёт к невязке при определении содержания SO_2 . Так что для этих районов

$$\ln y(\lambda) = \overline{\ln y}(\lambda) + \sum_{i=1}^K t_i \mathbf{f}_k(\lambda) + \Omega_{SO_2} \frac{\partial \ln y(\lambda)}{\partial \Omega_{SO_2}}. \quad (6)$$

где Ω_{SO_2} – плотность вертикального столба SO_2 . Производную, которая входит в последний член этого уравнения, можно определить или методом конечных разностей, или при помощи линейаризованных моделей переноса излучения [50, 51]. Уравнение (6) позволяет с лёгкостью определить Ω_{SO_2} .

Этот метод был использован применительно к данным, полученным при помощи прибора для мониторинга озона (*Ozone Monitoring Instrument*) [52] в диапазоне длин волн 310–340 нм. Так как ГК высоких поряд-

ков отражают скорее шум, чем полезный сигнал, то усечение ГК выступает и в качестве фильтра. Для восстановления спектральных энергетических яркостей требовались, по меньшей мере, 20–30 ГК, тогда как при наличии обусловленных SO_2 относительно сильных сигналов это количество можно было уменьшить до 8. Авторы [52] утверждают, что шум в данных был уменьшен вдвое, тем самым обеспечив большую чувствительность к антропогенным источникам SO_2 .

На настоящий момент нет сообщений об аналогичном подходе к регистрации других малых газовых компонент. Одна из причин этого связана со сложностью получения системы ЭОФ для участков, в которых те или иные малые газовые компоненты отсутствуют. Вторая причина заключается в том, что, строго говоря, описание (6) является приблизительным. Для SO_2 эта формула работает нормально, и невязка непосредственно связана с обусловленным SO_2 сигналом. Для других малых газовых компонент уравнение (6) может оказаться несправедливым, и для них нужен более сложный подход.

5. Снижение размерности в МПИ с машинным самообучением

5.1. Общие соображения

Как и в [53], обратная задача решается посредством сведения её к задаче оптимизации. Основная идея этого метода заключается в поиске вектора состояния, обеспечивающего минимизацию невязки между результатами моделирования и измерений. Нелинейную обратную задачу решают методом последовательных приближений [54, 55]. Задав априорный вектор состояния \mathbf{x} , линейризуем около него нелинейную прямую модель. Затем можно с лёгкостью обратить эту линейризованную модель и сформировать новый вектор состояния. Этот итерационный подход широко используется как для выявления малых газовых концентраций, так и для определения характеристик аэрозолей и облаков [56]. Однако этот метод обращения требует очень больших временных затрат из-за повторяющегося обращения к моделирующим энергетическую яркость сложным прямыми моделям переноса излучения и матри-

цам Якоби (то есть к матрицам частных производных первого порядка спектральной энергетической яркости по x) и последующего обращения сравнительно больших матриц. Эти соображения служат побудительной причиной для разработки альтернативных методов обращения для осуществляемого в реальном времени дистанционного зондирования, в основу которых были бы положены методы машинного самообучения и которые иногда называют физически полными инверсивными самообучающимися машинами (*full-physics inverse learning machines – FP-ILM*) [57].

5.2. Концепция самообучающихся машин для исследований атмосферы

Алгоритмы машинного самообучения охватывают задачу оптимизации не полностью. Скорее, они обеспечивают *получение знания* от заданной совокупности данных и делают прогнозы относительно нужных параметров. В принципе, алгоритм машинного самообучения состоит из тренировочной стадии, в рамках которой на основе данных, смоделированных при помощи МПИ (которая генерирует «физический полный» набор компонент), получают оператор обращения, и операционной стадии, в рамках которой этот оператор применяют к результатам реальных измерений. Основным преимуществом этого подхода перед традиционным состоит в том, что моделирование осуществляется в режиме офлайн, а сам оператор обращения оказывается устойчивым и простым с вычислительной точки зрения.

На рис. 6 представлена схема возможной реализации самообучающейся машины. На тренировочной стадии при помощи физически полной прямой модели, которой в рассматриваемом случае является МПИ, производится расчёт тренировочного набора данных. Для выявления важных элементов смоделированного набора данных и во избежание «избыточной размерности» (так называемого эффекта Хьюза [58]) смоделированные спектральные данные сжимают при помощи подходящего метода снижения размерности. Преобразование смоделированных спектров уменьшенной размерности в нужный параметр осуществляется при помощи машинного самообучения.

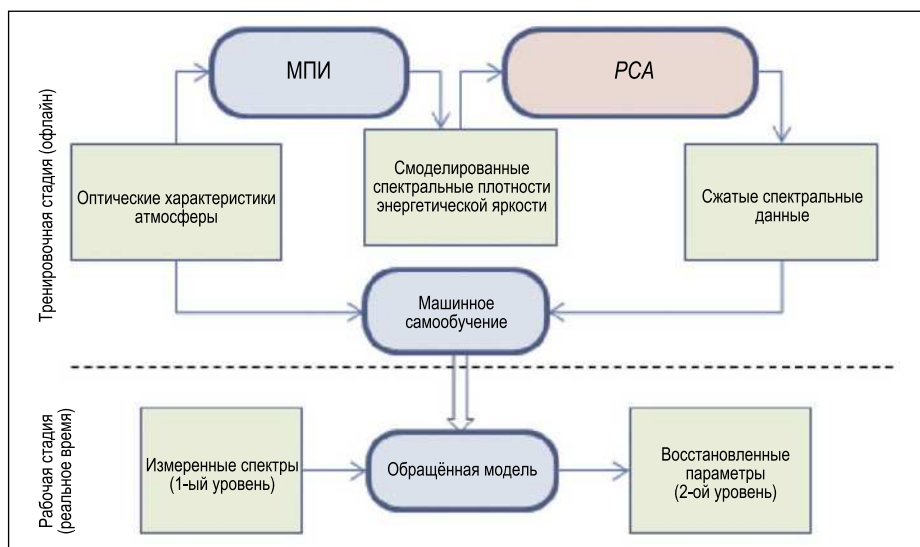


Рис. 6. Схема алгоритма поиска в рамках машинного самообучения, в котором используется снижение размерности спектрального распределения энергетической яркости

5.3. Машинное самообучение, основанное на линейной регрессии

В алгоритмах поиска, основанных на линейной регрессии, используется следующее представление искомого параметра x :

$$x = c + \sum_{w=1}^W l_w y(\lambda_w),$$

где c – линейное смещение, а l_w – коэффициенты регрессии. Метод регрессии главных компонент (РГК) основан на модели линейной регрессии, которая связывает x и координаты ГК спектральной энергетической яркости:

$$x = c + \sum_{k=1}^K l_k t_k;$$

Так как $K \ll W$, то размерность модели линейной регрессии (и соответствующей обратной задачи) уменьшается. Более того, так как аппаратные шумы не влияют на координаты ГК низких порядков, то вся схема обращения оказывается более стабильной.

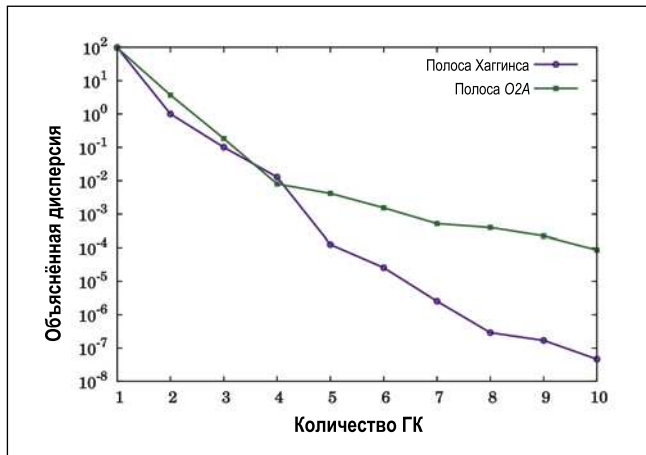
Для данных с большим уровнем шума собственные векторы F следует рассчитать для матрицы $C_Y + C_\epsilon$, где C_ϵ – ковариационная матрица для шума, а не для матрицы C_Y . В этом случае, координаты ГК для данных с большим уровнем шума коррелированы, и поэтому их называют «спроецированными ГК» (*projected principal components*) [59]. Если статистические характеристики шума неизвестны, то ковариационную матрицу можно получить, сделав ряд допущений (например, гауссовский шум) или воспользо-

вавшись аппроксимацией $C_\epsilon \approx \alpha I$, где I – единичная матрица, а α – параметр регуляризации. Эта процедура уменьшает влияние ГК высоких порядков.

Обобщением метода РГК является ядерная гребневая регрессия (ЯГР) (*kernel ridge regression*), которая используется для прогнозирования атмосферных профилей на основе данных, полученных при помощи инфракрасного интерферометра для зондирования атмосферы (*LAS*) [61]. Недостатком основанных на РГК и ЯГР моделей является то, что для представления результатов измерений используются базисные векторы F , тогда как содержащаяся в X информация во внимание не принимается. Альтернативной моделью, в которой этот недостаток обходится, является регрессия частных наименьших квадратов [62]. В [63, 64] было показано, что регрессия частных наименьших квадратов позволяет подобрать модель с использованием меньшего, чем в случае РГК, количеством ГК. В свою очередь, регрессию частных наименьших квадратов можно обобщить на случай, когда требуется получить не одну переменную x , а набор коррелированных параметров (например, температурный профиль). Соответствующий метод называют методом канонических корреляций [65]. Использование канонической корреляции в науке об атмосфере рассмотрено в [66].

Основанный на РГК подход был успешно применён к решению задач определения высот выброса пе-

Рис. 7. Зависимость объяснённой дисперсии от количества ГК в полосе Хаггинса и полосе O2A



пла вулканами на основе результатов измерений, сделанных приборами GOME-2 [67] и TROPOMI [68], а также спутником GOSAT [69, 70].

6. Снижение размерности входных оптических данных

6.1. Моделирование спектров

В [71] был предложен эффективный метод, основанный на снижении размерности входных оптических данных. В этом методе используется конечно-разностная локальная линейризация МПИ применительно к входным параметрам. Для уменьшения обращений МПИ к расчёту конечных разностей, линейризация осуществляется в уменьшенном пространстве данных. Этот метод можно кратко описать следующим образом.

Сначала вводится поправочная функция

$$Q(\lambda_w) = \ln[y(\lambda_w) / y_a(\lambda_w)]. \quad (7)$$

где y – энергетическая яркость, рассчитанная при помощи полной МПИ, а y_a – энергетическая яркость, рассчитанная при помощи приближённой модели (например, двухпоточной модели (*two-stream model*)). Затем для атмосферы, состоящей из L слоёв, вводится вектор состояния $\mathbf{x}_w \in \mathbb{R}^{2L+1}$, содержащий оптические параметры всех слоёв, например,

$$\mathbf{x}_w^T = \left[\begin{array}{c} \sigma_{\text{abs},1}(\lambda_w), \dots, \sigma_{\text{abs},L}(\lambda_w), \\ \sigma_{\text{sct},1}(\lambda_w), \dots, \sigma_{\text{sct},L}(\lambda_w), \rho(\lambda_w) \end{array} \right],$$

где $\sigma_{\text{sct},i}$ и $\sigma_{\text{abs},i}$ – соответственно, коэффициенты рассеяния и поглощения для i -го слоя, а ρ – альбедо поверхности. В результате, всё обусловленное зависимостями от длины волны мно-

гообразии оптических параметров, представляющих собой входные данные программы для расчёта переноса излучения, оказывается заключённым в векторе \mathbf{x}_w . Обратите внимание на то, что фазовая функция считается постоянной в пределах заданного спектрального интервала, вследствие чего она не вошла в вектор \mathbf{x}_w . Применив PCA к $\{\mathbf{x}_w\}_{w=1}^W$, получаем:

$$\mathbf{x}_w \approx \bar{\mathbf{x}} + \sum_{k=1}^K t_{wk} \mathbf{f}_k, \quad \bar{\mathbf{x}} = (1/W) \sum_{w=1}^W \mathbf{x}_w.$$

На рис. 7 приведены результаты PCA, осуществлённого применительно к входным оптическим данным, относящимся к полосе Хаггинса (315–335 нм) и полосе O2A (755–775 нм). Оптические данные были взяты из [72]. Обратите внимание на то, что для описания 99,9 % изменчивости набора данных достаточно четырёх ГК.

Теперь предположим, что $Q(\mathbf{x}_w)$ можно достаточно хорошо аппроксимировать рядом Тейлора относительно $\bar{\mathbf{x}}$:

$$Q(\mathbf{x}_w) \approx Q(\bar{\mathbf{x}} + \Delta \mathbf{x}_w) \approx Q(\bar{\mathbf{x}}) + \Delta \mathbf{x}_w^T \nabla Q(\bar{\mathbf{x}}) + \frac{1}{2} \Delta \mathbf{x}_w^T \nabla^2 Q(\bar{\mathbf{x}}) \Delta \mathbf{x}_w, \quad (8)$$

где ∇Q и $\nabla^2 Q$ – градиент и гессиан Q соответственно. Воспользовавшись центральными разностями для аппроксимации входящих в (8) производных по направлению первого и второго порядков, получаем:

$$Q(\mathbf{x}_w) \approx Q(\bar{\mathbf{x}}) + \frac{1}{2} \sum_{k=1}^K \left[Q(\bar{\mathbf{x}} + \mathbf{f}_k) - Q(\bar{\mathbf{x}} - \mathbf{f}_k) \right] t_{wk} + \frac{1}{2} \sum_{k=1}^K \left[Q(\bar{\mathbf{x}} + \mathbf{f}_k) - 2Q(\bar{\mathbf{x}}) + Q(\bar{\mathbf{x}} - \mathbf{f}_k) \right] t_{wk}^2. \quad (9)$$

Из (9) и (7) ясно видно, что для расчёта поправочного коэффициента требуется $2K+1$ раз вызвать полную и двухпоточную модели. Обратите внимание на то, что если бы поправочную функцию получали при помощи конечных разностей в исходном пространстве данных, то полную и двухпоточную модели потребовалось бы вызвать $2L+1$ раз. А так как обычно $K \ll L$, то в результате было обеспечено существенное сокращение затрачиваемого на вычисления времени.

Этот подход был использован применительно к моделированию спектров в полосе O2A [71], полосе Хаггинса [34] и полосах CO2 [73, 74]. В [75] аналогичный подход был использован применительно к моделированию энергетической яркости в области спектра, охватывающей УФ, видимый и ближний ИК диапазоны (0,3–3000 нм). Во всех случаях авторы сообщали, что среднеквадратическая погрешность расчётных значений энергетической яркости была порядка 0,01 %, и при этом было достигнуто почти 10-кратное увеличение скорости расчётов. Большим достоинством этого метода является то, что, в отличие от рассмотренных ранее методов, он не требует наличия баз данных, содержащих рассчитанные предварительно спектры.

В [76] эффективность методов снижения размерности входных и выходных данных была рассмотрена применительно к моделированию в полосе Хартли-Хаггинса. Было предложено комбинированное применение этих методов. Методы снижения размерности пространства выходных данных и спектральной выборки были использованы применительно к двухпоточным решениям при помощи соответствующих просмотрных таблиц, тогда как многопоточные вычисления выполнялись применительно к снижению размерности входных данных, которая описана в этом разделе. Было установлено, что комбинированное применение этих методов обеспечивает точность, превышающую 0,05 %, причём скорость выполнения расчётов возрастает примерно в 20 раз.

6.2. Расчёты в случае сокращённого пространства входных данных

Так как задача определения параметров атмосферы является чрезвычайно

Характеристики датчиков состава атмосферы

Прибор	GOME	GOME-2	TROPOMI
Носитель	ERS-2	MetOp (A, B, C)	Sentinel 5 Precursor
Пространственное разрешение, км ²	320×40	80×40	7×3,5
Количество данных 1-го уровня (Тбайт/год)	0,8	4,2	240
Применялся, годы	1995–2011	2006-настоящее время	2017- настоящее время

Рис. 8. Схематическое представление основанного на PCA алгоритма классификации, используемого применительно к гиперспектральным данным



некорректной, то физически правильные результаты можно получить при помощи регуляризации, в рамках которой учитывается ряд априорных сведений. В этом случае снижение размерности пространства входных данных можно рассматривать как особый вид регуляризации, то есть получаемые параметры должны подчиняться определённой зависимости, описанной выбранным набором ЭОФ.

В [77] метод снижения размерности был использован для параметризации коэффициента аэрозольного ослабления с целью его последующего использования в алгоритме обращения, предназначенном для получения соответствующих координат ГК, а не для описания аэрозольного ослабления. На основе теории Ми [78] и алгоритмов, разработанных для ансамблей частиц, была получена система ЭОФ для набора коэффициентов аэрозольного ослабления. И наконец, в [78] снижение размерности было выполнено применительно к пространствам входных (профили температуры и влажности) и выходных (спектральные энергетические яркости) данных, тогда как искусственные нейронные сети использовались для установления взаимосвязей между координатами ГК. Благодаря уменьшению количества независимых параметров, такая схема оказалась более устойчивой и эффективной, чем традиционная.

7. Снижение размерности гиперспектральных данных в задачах классификации

Снижение размерности играет важную роль в алгоритмах классифика-

ции, используемых применительно к гиперспектральным данным. Онлайновое обнаружение аномалий и распознавание объектов на изображениях, получаемых с помощью дистанционного зондирования, имеет чрезвычайно большое значение для мониторинга лесных пожаров и вулканической деятельности. В подобных областях применения приходится решать задачи классификации. В сокращённом пространстве данных количество участвующих в классификации переменных уменьшается, хотя их количество и остаётся большим, чем в исходном пространстве. Поэтому алгоритмы классификации (например, алгоритм « k ближайших соседей») оказываются более устойчивыми и эффективными [80, 81]. PCA можно использовать для визуализации гиперспектральных данных в виде двумерных изображений, идентифицируя при этом участки с определёнными особенностями. Концепция совместного использования PCA и классификаторов проиллюстрирована на рис. 8. Подобный подход используется не только при обработке данных, полученных с борта космического аппарата, но и в других областях применения, таких как материаловедение [82, 83], табачная промышленность [84] и производство продуктов питания [85].

8. Заключение

В обзоре было рассмотрено несколько методов обработки гиперспектральных данных, причём в основу всех этих методов положено снижение размерности. Было показано, что применительно к гиперспектрально-

му моделированию PCA может быть реализован несколькими способами. Поэтому термин «основанная на PCA модель переноса излучения», по-видимому, не годится, так как он не описывает особенности этого алгоритма (например, в [40, 49, 75] представлены абсолютно разные модели, и при этом все они «основаны на PCA»).

Было показано, что концепция снижения размерности может быть положена в основу разработки гиперспектральной МПИ, в которой непосредственно учитывается сильная взаимозависимость спектральных данных. Для более тщательного рассмотрения возможности совмещения описанных в данной работе методов потребуется провести дополнительные исследования. Например, в [86] был описан гибридный подход, совмещающий метод коррелированных k -распределений и снижение размерности входных данных. Подобные модели имеют очень большое значение для обработки больших данных, полученных в результате дистанционного зондирования при реализации текущих проектов, и становятся господствующей тенденцией при разработке атмосферных процессоров нового поколения.

Описанные основы снижения количества гиперспектральных данных общеизвестны и могут использоваться в разных областях применения, включая материаловедение и электронную спектроскопию, что обусловлено схожестью используемых в этих областях методик [87]. Кроме того, PCA является перспективным инструментом для анализа гиперспектральных оптических данных в медицине. В частности, недавние исследования показали, что снижение размерности кубов данных позволяет улучшить алгоритмы распознавания и классификации, что было бы чрезвычайно полезно

для ранней оптической диагностики болезней [88].

Авторы благодарны главному редактору журналов «Светотехника» и «Light & Engineering» проф. В.П. Будаку за приглашение написать этот обзор, а также анонимному рецензенту за внимательное ознакомление с рукописью и ценные предложения.

СПИСОК ЛИТЕРАТУРЫ

[1]. D.G. Loyola, S. Gimeno Garcia, R. Lutz, A. Argyrouli, F. Romahn, R.J.D. Spurr, M. Pedernana, A. Doicu, V. Molina Garcá, and O. Schüssler. The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 precursor. *Atmospheric Measurement Techniques*, 11(1):409–427, 2018.

[2]. C.A. Lee, S.D. Gasster, A. Plaza, C. Chang, and B. Huang. Recent developments in high performance computing for remote sensing: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(3):508–527, 2011.

[3]. J.P. Veefkind, I. Aben, K. McMullan, H. Forster, J. de Vries, G. Otter, J. Claas, H.J. Eskes, J.F. de Haan, Q. Kleipool, and et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, 120:70–83, 2012.

[4]. Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, and W. Jie. Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51:47–60, 2015.

[5]. P. Liu. A survey of remote-sensing big data. *Frontiers in Environmental Science*, 3:45, 2015.

[6]. V.M. Roozendael, R. Spurr, D. Loyola, C. Lerot, D. Balis, J.C. Lambert, W. Zimmer, J. Gent, J. Van Geffen, M.E. Koukoulis, J. Granville, A. Doicu, C. Fayt, and C. Zehner. Sixteen years of GOME/ERS2 total ozone data: the new direct-fitting GOME Data Processor (GDP) Version 5: I. algorithm description. *J Geophys Res: Atmospheres*, 117: D03305(1–18), 2012.

[7]. V. Natraj. A review of fast radiative transfer techniques. In A.A. Kokhanovsky, editor, *Light scattering reviews*, volume 8, pages 475–504. Springer Berlin Heidelberg, 2013.

[8]. D. Efremenko, A. Doicu, D. Loyola, and T. Trautmann. Acceleration techniques for the discrete ordinate method. *J Quant Spectrosc Radiat Transfer*, 114:73–81, 2013.

[9]. D.S. Efremenko, D. Loyola, A. Doicu, and T. Trautmann. Data-intensive computing in radiative transfer modelling. In P. Soille and P.G. Marchetti, editors, *Proc. of the 2016 conference on Big Data from Space (BiDS16)*, pages 188–191, Santa Cruz de Tenerife, Spain, 2016.

[10]. V.P. Budak, G.A. Kaloshin, O.V. Shagalov, and V.S. Zheltov. Numerical modeling of the radiative transfer in a turbid medium using the synthetic iteration. *Opt. Express*, 23(15): A829, 2015.

[11]. D.S. Efremenko, D.G. Loyola, A. Doicu, and R.J.D. Spurr. Multi-core-CPU and GPU-accelerated radiative transfer models based on the discrete ordinate method. *Computer Physics Communications*, 185(12):3079–3089, 2014.

[12]. D.S. Efremenko, D. Loyola, R.J.D. Spurr, and A. Doicu. Acceleration of radiative transfer model calculations for the retrieval of trace gases under cloudy conditions. *J Quant Spectrosc Radiat Transfer*, 135:58–65, 2014.

[13]. R.J.D. Spurr. VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. *J Quant Spectrosc Radiat Transfer*, 102(2):316–342, 2006.

[14]. R.J.D. Spurr, T.P. Kurosu, and K.V. Chance. A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval. *J Quant Spectrosc Radiat Transfer*, 68(6):689–735, 2001.

[15]. K. Pearson. On lines and planes of closest fit to systems of points in space. *Phil Mag*, 2(6):559–572, 1901.

[16]. R. D. Morris, A. Kottas, M. Taddy, R. Furfaro, and B.D. Ganapol. A statistical framework for the sensitivity analysis of radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12):4062–4074, 2008.

[17]. V.A. Ambartzumyan. The effect of the absorption lines on the radiative equilibrium of the outer layers of the stars. *Publ. Obs. Astron. Univ. Leningrad*, 6:7–18, 1936.

[18]. R.M. Goody, R. West, L. Chen, and D. Crisp. The correlated k-method for radiation calculations in nonhomogeneous atmosphere. *J Quant Spectrosc Radiat Transfer*, 42(6):539–550, 1989.

[19]. W.J. Wiscombe and J.W. Evans. Exponential-sum fitting of radiative transmission functions. *J Comput Phys*, 24(4):416–444, 1997.

[20]. R. West, D. Crisp, and L. Chen. Mapping transformations for broadband atmospheric radiation calculation. *J Quant Spectrosc Radiat Transfer*, 43(3):191–199, 1990.

[21]. E. Boesche, P. Stammes, R. Preusker, R. Bennartz, W. Knap, and J. Fischer. Polarization of skylight in the O2A band: effects of aerosol properties. *Applied Optics*, 47(19):3467, 2008.

[22]. Christiane Helling and Uffe Grae Jorgensen. Optimizing the opacity sampling method. *Astronomy and Astrophysics*, 337:477–486, 1998.

[23]. Frédéric André, Longfeng Hou, Maxime Roger, and Rodolphe Vaillon. The multispectral gas radiation modeling: A new theo-

retical framework based on a multidimensional approach to k-distribution methods. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 147:178–195, nov 2014.

[24]. Boris A. Fomin. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 1. FKDM, fast k-distribution model for the longwave. *Journal of Geophysical Research*, 109(D2), 2004.

[25]. Boris Fomin. A k-distribution technique for radiative transfer simulation in inhomogeneous atmosphere: 2. FKDM, fast k-distribution model for the shortwave. *Journal of Geophysical Research*, 110(D2), 2005.

[26]. V. A. Falaleeva and B.A. Fomin. Overcoming spectroscopic challenges in direct problems of satellite sounding of the atmosphere. *Atmospheric and Oceanic Optics*, 30(1):1–6, jan 2017.

[27]. B.A. Fomin. Effective interpolation technique for line-by-line calculations of radiation absorption in gases. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 53(6):663–669, jun 1995.

[28]. S. Najmabadi, P. Offenhäuser, M. Hamann, G. Jajnabalkya, F. Hempert, C. Glass, and S. Simon. Analyzing the effect and performance of lossy compression on aeroacoustic simulation of gas injector. *Computation*, 5(4):24, 2017.

[29]. S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(22):2323–2326, 2000.

[30]. M.A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991.

[31]. A.N. Gorban, B. Kégl, D.C. Wunsch, and A.Y. Zinovyev, editors. *Principal Manifolds for Data Visualization and Dimension Reduction*. Springer Berlin Heidelberg, 2008.

[32]. I.K. Fodor. *A Survey of Dimension Reduction Techniques*. Office of Scientific and Technical Information (OSTI), 2002.

[33]. L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik. *Dimensionality Reduction: A Comparative Review*. Tilburg University Technical Report, TiCC-TR2009–005, 2009.

[34]. D.S. Efremenko, A. Doicu, D. Loyola, and T. Trautmann. Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone retrievals. *J Quant Spectrosc Radiat Transfer*, 133:128–135, 2014.

[35]. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[36]. Robert H. MacArthur. On the relative abundance of bird species. *Proc Natl Acad Sci USA*, 43(3):293–295, 1957.

- [37]. X Liu, W.L. Smith, D.K. Zhou, and A. Larar. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. *Applied Optics*, 45(1):201–208, 2006.
- [38]. D. G. Loyola, M. Pedernana, and S. Gimeno Garcia. Smart sampling and incremental function learning for very large high dimensional data. *Neural Networks*, 78:75–87, 2016.
- [39]. J. H. Halton. Algorithm 247: Radical-inverse quasi-random point sequence. *Commun ACM*, 7(12):701–702, 1964.
- [40]. M. Matricardi. A principal component based version of the RTTOV fast radiative transfer model. *Quarterly Journal of the Royal Meteorological Society*, 136:1823–1835, 2010.
- [41]. P. D. Hurlley, S. Oliver, D. Farrah, L. Wang, and A. Efstathiou. Principal component analysis and radiative transfer modelling of Spitzer Infrared spectrograph spectra of ultraluminous infrared galaxies. *Monthly Notices of the Royal Astronomical Society*, 424(3):2069–2078, 2012.
- [42]. A. Hollstein and R. Lindstrot. Fast reconstruction of hyperspectral radiative transfer simulations by using small spectral subsets: application to the oxygen A band. *Atmospheric Measurement Techniques*, 7(2):599–607, 2014.
- [43]. U. Platt. Differential optical absorption spectroscopy (DOAS). *Chem Anal Series*, 127:27–83, 1994.
- [44]. U. Platt and J. Stutz. *Differential Optical Absorption Spectroscopy: Principles and Applications*. Springer-Verlag, Berlin, Heidelberg, 2008.
- [45]. G.W. Kattawar, A.T. Young, and T.J. Humphreys. Inelastic scattering in planetary atmospheres. I. The Ring effect, without aerosols. *Astrophys J*, 243:1049–1057, 1981.
- [46]. M. Vountas, V.V. Rozanov, and J.P. Burrows. Ring effect: Impact of rotational Raman scattering on radiative transfer in earth's atmosphere. *J Quant Spectrosc Radiat Transfer*, 60(6):943–961, 1998.
- [47]. James Slusser, Kyle Hammond, Arve Kylling, Knut Stamnes, Lori Perliski, Arne Dahlback, Donald Anderson, and Robert DeMajistre. Comparison of air mass computations. *Journal of Geophysical Research: Atmospheres*, 101(D5):9315–9321, 1996.
- [48]. Lori M. Perliski and Susan Solomon. On the evaluation of air mass factors for atmospheric near-ultraviolet and visible absorption spectroscopy. *Journal of Geophysical Research*, 98(D6):10363, 1993.
- [49]. Can Li, Joanna Joiner, Nickolay A. Krotkov, and Pawan K. Bhartia. A fast and sensitive new satellite SO₂ retrieval algorithm based on principal component analysis: Application to the ozone monitoring instrument. *Geophysical Research Letters*, 40(23):6314–6318, 2013.
- [50]. R.J.D. Spurr. LIDORT and VLIDORT. Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In A.A. Kokhanovsky, editor, *Light scattering reviews*, volume 3, pages 229–275. 2008.
- [51]. A. Doicu and T. Trautmann. Two linearization methods for atmospheric remote sensing. *J Quant Spectrosc Radiat Transfer*, 110(8):477–490, 2009.
- [52]. P.F. Levelt, G.H.J. van den Oord, M.R. Dobber, A. Malkki, Huib Visser, Johan de Vries, P. Stammes, J.O.V. Lundell, and H. Saari. The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, 44(5):1093–1101, 2006.
- [53]. A.N. Tikhonov and V.Y. Arsenin. *Solution of Ill-Posed Problems*. Winston, 1977.
- [54]. C.D. Rodgers. *Inverse methods for atmospheric sounding: Theory and Practice*. World Scientific Publishing, 2000.
- [55]. A. Doicu, T. Trautmann, and F. Schreier. *Numerical Regularization for Atmospheric Inverse Problems*. Springer, 2010.
- [56]. D. S. Efremenko, O. Schüssler, A. Doicu, and D. Loyola. A stochastic cloud model for cloud and ozone retrievals from UV measurements. *J Quant Spectrosc Radiat Transfer*, 184:167–179, November 2016.
- [57]. J. Xu, O. Schussler, D.G. Loyola Rodriguez, F. Romahn, and A. Doicu. A novel ozone profile shape retrieval using full-physics inverse learning machine (FP-ILM). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(12):5442–5457, 2017.
- [58]. G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inform. Theory*, 14(1):55–63, 1968.
- [59]. W.J. Blackwell and F.W. Chen. *Neural Networks in Atmospheric Remote Sensing*. Lexington, 2009.
- [60]. John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, NY, USA, 2004.
- [61]. G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, L. Guanter, and X. Calbet. Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5):1759–1769, May 2012.
- [62]. Roman Rosipal and Nicole Krämer. Overview and recent advances in partial least squares. In *Subspace, Latent Structure and Feature Selection*, pages 34–51. Springer Science + Business Media, 2006.
- [63]. Peter D. Wentzell and Lorenzo Vega Montoto. Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemometrics and Intelligent Laboratory Systems*, 65(2):257–279, 2003.
- [64]. Saikat Maitra and Jun Yan. Principal component analysis and partial least squares: Two dimension reduction techniques for regression. In *Discussion Papers: 2008 Discussion Paper Program – Applying Multivariate Statistical Models*, pages 79–90. Casualty actuarial society, 2008.
- [65]. H. Hotelling. The most predictable criterion. *Journal of Educational Psychology*, 26:139–142, 1935.
- [66]. D. Wilks. *Statistical Methods in the Atmospheric Sciences, 3rd Edition*. New York: Elsevier, 2011.
- [67]. D.S. Efremenko, D.G. Loyola R., P. Hedelt, and R.J.D. Spurr. Volcanic SO₂ plume height retrieval from UV sensors using a full-physics inverse learning machine algorithm. *International Journal of Remote Sensing*, 38(sup1):1–27, 2017.
- [68]. Pascal Hedelt, Dmitry S. Efremenko, Diego G. Loyola, Robert Spurr, and Lieven Clarisse. SO₂ layer height retrieval from Sentinel-5 Precursor/TROPOMI using FP_ILM. *Atmospheric Measurement Techniques Discussions*, pages 1–23, feb 2019.
- [69]. M.Yu. Kataev, A.K. Lukyanov, and A.A. Bekerov. Modification of the empirical orthogonal functions method for solving the inverse task of retrieving of the CO₂ total content from satellite data. *Journal of Siberian Federal University. Engineering & Technologies*, 11(1):77–85, 2018.
- [70]. M.Y. Kataev and A.K. Lukyanov. Empirical orthogonal functions and its modification in the task of retrieving of the total amount CO₂ and CH₄ with help of satellite Fourier transform spectrometer GOSAT (TANSO-FTS). In G.G. Matvienko and O.A. Romanovskii, editors, *22nd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*. SPIE-Intl Soc Optical Eng, 2016.
- [71]. V. Natraj, X. Jiang, R.L. Shia, X. Huang, J.S. Margolis, and Y.L. Yung. Application of the principal component analysis to high spectral resolution radiative transfer: A case study of the O₂A-band. *J Quant Spectrosc Radiat Transfer*, 95(4):539–556, 2005.
- [72]. I.E. Gordon, L.S. Rothman, C. Hill, R.V. Kochanov, Y. Tan, P.F. Bernath, M. Birk, V. Boudon, A. Campargue, K.V. Chance, B.J. Drouin, J.-M. Flaud, R.R. Gamache, J.T. Hodges, D. Jacquemart, V.I. Perevalov, A. Perrin, K.P. Shine, M.-A.H. Smith, J. Tenynson, G.C. Toon, H. Tran, V.G. Tyuterev, A. Barbe, A.G. Császár, V.M. Devi, T. Furtenbacher, J.J. Harrison, J.-M. Hartmann, A. Jolly, T.J. Johnson, T. Karman, I. Kleiner, A.A. Kyuberis, J. Loos, O.M. Lyulin, S.T. Massie, S.N. Mikhailenko, N. Moazzen-Ahmadi, H.S.P. Müller, O.V. Naumenko, A.V. Nikitin, O.L. Polyansky, M. Rey, M. Rotger, S.W. Sharpe, K. Sung, E. Starikova, S.A. Tashkun, J. Vander Auwera, G. Wagner, J. Wilzewski, P. Weiso, S. Yu, and E.J. Zak. The HITRAN2016 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 203:3–69, 2017.

[73]. P. Somkuti, H. Boesch, V. Natraj, and P. Kopparla. Application of a PCA-based fast radiative transfer model to XCO₂ retrievals in the shortwave infrared. *Journal of Geophysical Research: Atmospheres*, 122(19):10,477–10,496.

[74]. V. Natraj, R.L. Shia, and Y.L. Yung. On the use of principal component analysis to speed up radiative transfer calculations. *J Quant Spectrosc Radiat Transfer*, 111(5):810–816, 2010.

[75]. P. Kopparla, V. Natraj, R. Spurr, R. Shia, D. Crisp, and Y.L. Yung. A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 173:65–71, 2016.

[76]. Ana del Águila, Dmitry Efremenko, Víctor Molina García, and Jian Xu. Analysis of two dimensionality reduction techniques for fast simulation of the spectral radiances in the Hartley-Huggins band. *Atmosphere*, 10(3):142, mar 2019.

[77]. Y. Timofeyev, A. Polyakov, H. Steele, and M. Newchurch. Optimal eigenanalysis for the treatment of aerosols in the retrieval of atmospheric composition from transmission measurements. *Appl. Opt.*, 42(15):2635, July 2003.

[78]. G. Mie. Beitrage zur optik trueber medien, speziell kolloidaler metalloesungen. *Annalen der Physik*, 330(3):377–445, 1908.

[79]. A. V. Polyakov, Yu.M. Timofeev, and Ya.A. Virolainen. Using artificial neural networks in the temperature and humidity sounding of the atmosphere. *Izvestiya, Atmospheric and Oceanic Physics*, 50(3):330–336, 2014.

[80]. Michael T. Eismann, Joseph Meola, and Russell C. Hardie. Hyperspectral change detection in the presence of diurnal and seasonal variations. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1):237–249, 2008.

[81]. Dandan Ma, Yuan Yuan, and Qi Wang. Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation. *Remote Sensing*, 10(5):745, may 2018.

[82]. David B Brough, Daniel Wheeler, and Surya R. Kalidindi. Materials knowledge systems in python – a data science framework for accelerated development of hierarchical materials. *Integrating Materials and Manufacturing Innovation*, 6(1):36–53, mar 2017.

[83]. Sai Kiranmayee Samudrala, Prasanna Venkataraman Balachandran, Jaroslaw Zola, Krishna Rajan, and Baskar Ganapathy-subramanian. A software framework for data dimensionality reduction: application to chemical crystallography. *Integrating Materials and Manufacturing Innovation*, 3(1), jun 2014.

[84]. Pilar B. Garcá-Allende, Olga M. Conde, Ana M. Cubillas, César Jáuregui, and José M. López-Higuera. New raw material discrimination system based on a spatial optical spectroscopy technique. *Sensors and Actuators A: Physical*, 135(2):605–612, apr 2007.

[85]. Juan Xing, Cédric Bravo, Pál T. Jancsó, Herman Ramon, and Josse De Baerdemaeker. Detecting bruises on ‘golden delicious’ apples using hyperspectral imaging with multiple wavebands. *Biosystems Engineering*, 90(1):27–36, jan 2005.

[86]. V. Molina García, S. Sasi, D.S. Efremenko, A. Doicu, and D. Loyola. Radiative transfer models for retrieval of cloud parameters from EPIC/DSCOVR measurements. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 213:228–240, 2018.

[87]. Афанасьев В.П., Будак В.П., Ефременко Д.С., Капля П.С. Применение фотометрической теории светового поля в задачах рассеяния электронов // Светотехника.— 2018.— № 5. — С. 44–50.

[88]. F. Vasefi, N. MacKinnon, and D.L. Farkas. Hyperspectral and multispectral imaging in dermatology. In *Imaging in Dermatology*, pages 187–201. Elsevier, 2016.

Открыта регистрация на конференцию Lux Pacifica 2020

Светотехническая конференция стран Тихоокеанского региона *Lux Pacifica* пройдёт 20–22 апреля 2020 года в Санкт-Петербурге. Организатором конференции выступит ВНИСИ им. С.И. Вавилова при поддержке журнала «Светотехника», Научно-технического совета светотехнической отрасли РФ и Государственного Эрмитажа, в стенах которого и пройдёт мероприятие.

Регистрация на конференцию открыта на сайте www.luxpacifica2020.vnisi.ru. В фокусе конференции — актуальные проблемы светотехники, такие как музейное освещение, незрительные эффекты света, интеллектуальное освещение, фотометрия и колориметрия, и другие. Тезисы на конференцию принимаются до 1 ноября 2019 года, регистрация участников продлится до 15 февраля 2020 года.

www.vnisi.ru

LITE-ON запускает самый мощный в мире УФ-С диод для обеззараживания

Компания *LITE-ON Technology Corp.* объявила о выпуске новой серии мощных УФ-С диодов «W35», у которых единственный кристалл, излучающий в диапазоне 270–280 нм, имеет рекордный поток излучения 130 мВт при токе диода 350 мА и напряжении на нём 6 В (т.е. при мощности 2 Вт). Диоды «W35» имеют размеры корпуса 3,3×3,5 мм и позволяют повышение их тока до 600 мА – для получения потока излучения 220 мВт (при мощности диода около 4 Вт). Не полагаясь только на успехи в области УФ-С кристаллов, *LITE-ON* разработала прогрессивную ультрасовременную технологию корпусирования, позволяющую на 40 % повысить эффективность выведения УФ-С излучения по сравнению с ныне используемой технологией.

Два выпускаемых в настоящее время компанией УФ-С диода серии «G35» (80 мВт при 600 мА) обеспечивают работу потребительской системы очистки воды с производительностью 2 л/мин. (Диоды «G35» дают эффективность обеззараживания свыше 99,99 %.) Более совершенные диоды «W35» производят столь же надёжное обеззараживание при меньшем на 46 % выделении тепла. Они станут очередной вехой на пути развития УФ-С диодов для обеззараживания проточной воды.

LITE-ON предлагает широкий ассортимент УФ диодов, охватывающий диапазоны длин волн от 275 (УФ-С) до 430 нм (УФ-А), разной мощности. Все они имеют степень защиты IP58 и испытаны на обеспечение достаточного для конечного изделия ресурса.

ledinside.com
28.06.2019



Ана дель Агила (Ana del Águila).

Окончила Гранадский университет, Испания (2015 г.). Аспирант (PhD) Германского аэрокосмического центра. Область научных интересов:

атмосферные аэрозоли, лидары, дистанционное зондирование, перенос излучения, анализ больших данных



Дмитрий С. Ефременко (Dmitry S. Efremenko),

PhD, доктор технических наук. Окончил Московский энергетический институт (2009 г.). Научный сотрудник Германского аэрокосмического центра. Область научных интересов:

перенос излучения, дистанционное зондирование, анализ больших данных



Томас Траутманн (Thomas Trautmann), Dr.,

профессор. Руководитель отдела Германского аэрокосмического центра. Область научных интересов:

перенос излучения, рассеяние электромагнитных волн, дистанционное зондирование атмосферы.