

УДК 514.18

Л. М. Білокриницька¹, Г. М. Крученицький², І. Г. Міщишина¹

¹ Київський національний університет імені Тараса Шевченка, Київ

² Центральна аерологічна обсерваторія Росгідромету, Долгопрудний, Росія

БАЗА ДАНИХ КОЕФІЦІЄНТІВ АЕРОЗОЛЬНОГО ПОСЛАБЛЕННЯ НА ЧОТИРЬОХ ДОВЖИНАХ ХВИЛЬ ЗА СПОСТЕРЕЖЕННЯМИ SAGE II

Описано структуру та функціональні можливості об'єднаної бази даних, створеної на основі вимірювань приладу SAGE II на борту КА ERBS з жовтня 1984 р. по квітень 2000 р. на довжинах хвиль $\lambda\lambda$ 386, 452, 525 і 1020 нм. Розроблено зручний для користувача інтерфейс. На нульовому рівні об'єднана база даних містить первинні дані: вертикальні профілі концентрації озону, водяної пари, двоокису азоту і аерозолію на чотирьох довжинах хвиль. На першому рівні забезпечується візуальний перегляд даних, їхня фільтрація та пошук, можливість критичного перегляду даних за певними критеріями селекції. На другому рівні зберігається інформація про кліматичні моделі і нормальні образи параметрів (значення спектральних характеристик сезонної мінливості, межі відхилення від сезонного ходу і характеристики довготривалої мінливості). Крім того, є можливість поповнення об'єднаної бази даних значеннями аерозольних індексів, отриманих супутниковою апаратурою класу TOMS-OMI та внесками національних центрів даних. Продемонстровано приклади роботи з базою.

ВСТУП

З 24 жовтня 1984 р. по 22 серпня 2005 р. провадились дослідження стратосферних аерозолів і газів апаратурою SAGE II (Stratospheric Aerosol and Gas Experiment), яка була встановлена на борту супутника ERBS (Earth Radiation Budget Satellite, NASA). Тривалість спостережень була унікальною, кількість вимірювань становить 168389. Отримано понад 300 тисяч висотних профілів у чотирьох спектральних каналах ($\lambda\lambda = 386, 452, 525$ і 1020 нм) з просторовим розділенням по висоті 0.5 км, що становить загалом понад 50 млн відліків. Цей масив потрібно не лише зберегти, але й забезпечити можливість проведення детального огляду даних та їхніх перетинів, аналізу стандартними статистичними та аналітичними методами, а також візуалізації як самих даних, так і результатів аналізу.

ВИБІР СИСТЕМИ КЕРУВАННЯ ОБ'ЄДНАНОЮ БАЗОЮ ДАНИХ

Незважаючи на наявність в даний час значної кількості потужних систем керування базами даних (СКБД) з різноманітними користувацькими зручностями вибір СКБД для вирішення конкретного завдання нерозривно пов'язаний з її особливостями. Об'єднана база даних (ОБД) «Аерозольний індекс і довжина аерозольної екстинкції (коефіцієнт аерозольного послаблення) на довжинах хвиль $386, 452, 525$ і 1020 нм» має такі особливості:

- значні обсяги інформації, яка архівується ($10\text{—}100$ Гб),
- потреба в розвинених обчислювальних засобах, особливо математичної статистики,
- необхідність в ефективних засобах візуалізації одновимірних та двовимірних даних,
- доступність засобів інформаційного забезпечення для непідготовленого користувача,
- забезпечення можливості зміни СКБД.

© Л. М. БІЛОКРИНИЦЬКА, Г. М. КРУЧЕНИЦЬКИЙ,
І. Г. МІЩИШИНА, 2012

Тому питання про вибір СКБД доцільно вирішувати після аналізу складу і структури рівнів архівування даних з урахуванням результатів такого аналізу.

ВИБІР І ОБГРУНТУВАННЯ СКЛАДУ І СТРУКТУРИ РІВНІВ АРХІВУВАННЯ ОБ'ЄДНАНОЇ БАЗИ ДАНИХ

Загальні принципи побудови об'єднаної бази даних припускають трирівневу структуру рівнів архівування.

Рівень 0. Зберігаються первинні дані, отримані виконавцем спостережень у форматі, заданому ним же. Розробники та користувачі бази даних не мають права вносити зміни ні у формат архівації первинних даних, ні у значення самих даних. Навіть якщо архівні дані видаються абсолютно безглуздими, вони не можуть бути змінені ніким, крім тих виконавців, які переводили апаратні відліки у значення геофізичних параметрів. Розробники та користувачі ОБД можуть виключити дані, що не викликають їхньої довіри, з наступних рівнів архівації або з процесу обробки даних, але на «Рівні 0» їх необхідно зберегти. Програмне забезпечення «Рівня 0» повинно складатися виключно із засобів поповнення рівня новими даними.

Рівень 1. На цьому рівні дублюється інформація «Рівня 0», але при цьому забезпечуються такі можливості [2]:

- включення до складу програмного забезпечення «Рівня 1» засобів критичного перегляду даних і доведення критеріїв селекції даних до відома користувача;
- виключення чи оснащення спеціальними позначками тих даних, які не відповідають критеріям критичного перегляду, і доведення до відома користувача відповідності між змістом критеріїв та форматом позначок;
- включення до складу програмного забезпечення «Рівня 1» програмних засобів вичерпного візуального перегляду даних, а також фільтрації та пошуку.

Крім того, бажано:

- включення в доступну користувачеві інформацію метаданих, що полегшують доступ до даних «Рівні 1» можливостями його власних програмних засобів;

- включення до складу користувацького інтерфейсу «Рівня 1» розвинених обчислювальних засобів і скрипта для автоматизації обчислювальних та пошукових операцій на програмному рівні.

Допускається включення до складу даних на цьому рівні допоміжної інформації (наприклад, від світових центрів дії (СЦД) про стан сонячної активності, про квазидворічні коливання зональної складової екваторіального вітру та індекс субглобальних коливань тиску та ін.), сумісної з основною інформацією, зі зручним інтерфейсом та програмними засобами поповнення даних.

Рівень 2. На цьому рівні повинна зберігатися виключно інформація про кліматичні моделі і нормальні образи параметрів. Під нормальним образом слід розуміти значення спектральних характеристик сезонної мінливості, межі відхилення від сезонного ходу і характеристики довготривалої мінливості (довготривалі коливання, тренди та ін.).

На «Рівні 2» розробники ОБД повинні забезпечити виконання таких вимог:

- доступність для користувача алгоритмів побудови моделей і формування нормальних образів та їхньої програмної реалізації;
- наявність користувацького інтерфейсу для візуалізації побудованих моделей і сформованих образів;
- доступність користувачеві інформації про те, які з модельних даних отримані безпосередньо з інформації, що міститься на «Рівні 1», а які методами інтер- та екстраполяції;
- доступність користувачеві алгоритмів інтер- та екстраполяції, а також програмної реалізації цих алгоритмів;
- можливість для користувача модифікувати чи розробляти заново алгоритми і програми побудови моделей, здійснювати формування нормальних образів, а також інтер- та екстраполяцію модельних параметрів.

Програмне забезпечення всіх рівнів повинне бути доступним користувачеві, забезпеченим вичерпними коментарями і максимально можливою мірою відповідати вимогам інкапсуляції.

ВИБІР ТА ОБГРУНТУВАННЯ ВИМОГ ДО СИСТЕМ КЕРУВАННЯ БАЗАМИ ДАНИХ

Бази даних, створені на сучасному рівні вимог до користувацьких зручностей, повинні забезпечувати [2]:

- легкість доступу до наявних даних і простоту поповнення новими;
- можливість використання структурованих запитів із застосуванням мови QSL;
- можливість візуалізації (двовимірної і тривимірної) інформації, що міститься у базі даних та/або отриманої в результаті аналізу;
- можливість застосовувати для обробки інформації як стандартні аналітичні засоби, так і власні алгоритми користувача.

Всі ці вимоги легко здійснити шляхом застосування стандартних пакетів програм та засобів керування базами даних, таких як Access, FoxPro, MathCad і т. д. Проте освоєння цих пакетів і СКБД, тим більше на рівні, достатньому для реалізації власних алгоритмів, часто вимагає від користувача завеликих витрат часу. При цьому виконання всіх перерахованих вище вимог робить для користувача необхідним оволодіння принаймні однією СКБД і одним пакетом аналітичних програм. Для того щоб мінімізувати зусилля користувача, оптимальним є створення проміжного рівня представлення інформації у вигляді робочої книги Excel. Це забезпечує користувачам можливість застосування потужних систем керування базами даних, до FoxPro включно, вбудованих у Microsoft Visual Studio, і ефективних засобів аналізу, що включені у аналітичні пакети Maple, MathCad, MathLab тощо.

Наявність проміжного рівня надає користувачеві такі переваги:

- для вирішення більшості пошукових і дослідницьких завдань досить оволодіти навичками роботи з електронними таблицями Excel, які на сьогодні є одним з найбільш масових програмних продуктів;
- обробка стандартних запитів, візуалізація даних і використання більшості математичних та логічних функцій може здійснюватися звичайними засобами Excel;

- у пакеті користувацьких засобів аналізу та підготовки можна використовувати «швидкі» можливості Excel з обчислювальної та сортувальної роботи з масивами;

- для реалізації власних алгоритмів обробки даних та для автоматичного формування потоків вхідної інформації за власними критеріями користувачеві досить опанувати діалект VBA (Visual Basic for Applications) [3]. Розробку можна вести також на будь-якій з мов Microsoft Visual Studio;

- оформлені у вигляді робочих книг Excel, результати аналізу геофізичних процесів легко контролювати, тобто автоматично виконується вимога повторюваності результатів, яка часто ігнорується у сучасних геофізичних дослідженнях.

Як СКБД було вибрано електронні таблиці Excel 2007, інформаційні можливості яких істотно розширено порівняно з попередніми версіями: робочий аркуш має 16384 стовпці (замість 256) і 1048576 рядків (замість 65536).

ОПИС КЕРУВАННЯ ДАНИМИ ТА ЇХНЬОЇ ОБРОБКИ ЗАСОБАМИ ОБРАНОЇ СИСТЕМИ КЕРУВАННЯ БАЗАМИ ДАНИХ

Електронні таблиці Excel 2007 містять значний набір засобів керування даними та їхнього аналізу: перевірка даних, сортування і фільтрація даних, можливість використання поряд з автофільтрами користувацьких фільтрів, перегляд і пошук значень, візуалізація даних, зв'язування і консолідація даних, аналіз даних за допомогою введених таблиць.

До складу електронних таблиць Excel 2007 входить також розвинений і ефективний інструментарій для обробки даних, який включає:

- формули робочого аркуша, що містять функції елементарної математики, математичної статистики, лінійної алгебри, логіки, систем числення та ін.;
- пакет аналізу, що включає розширення засобів математичної статистики, спеціальні функції, засоби вирішення рівнянь, аналіз Фур'є.
- засоби аналізу даних за допомогою сценаріїв «що — якщо»;
- засоби аналізу даних за підбором параметрів і вибором рішення.

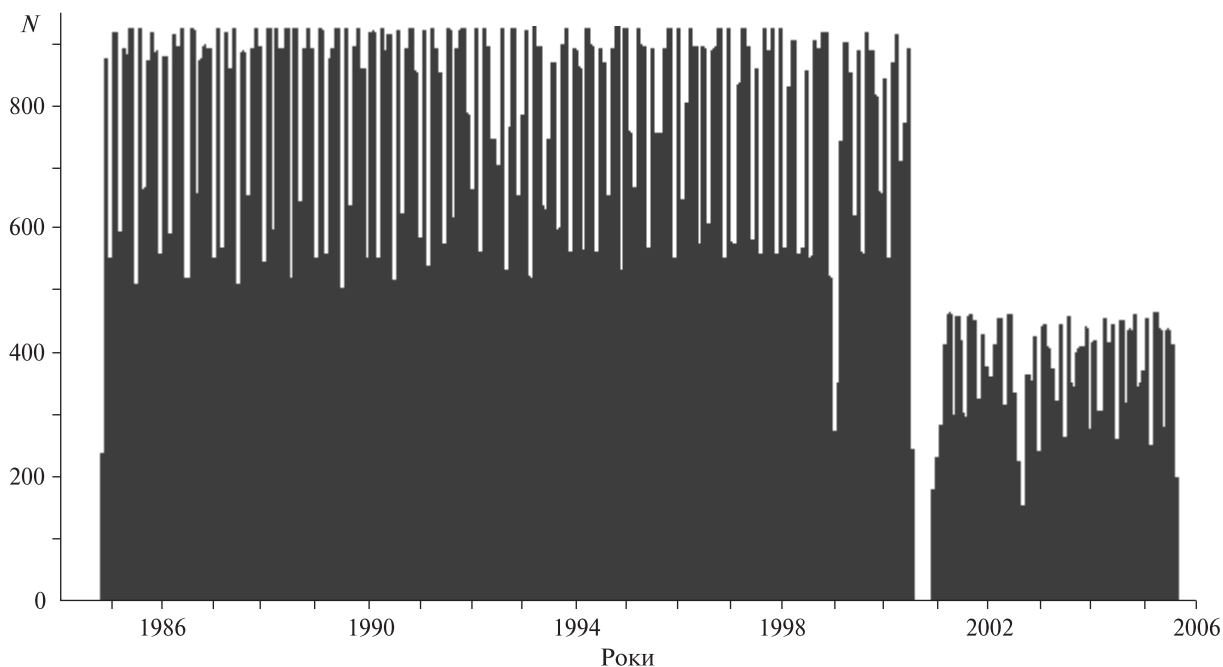


Рис. 1. Діаграма щомісячних вимірювань SAGE II

Засоби керування даними і їхнього аналізу доступні як безпосередньо з робочого аркуша, так і з коду VBA, що дозволяє автоматизувати операції керування та аналізу.

РОЗРОБКА БАЗИ ДАНИХ, ПІДГОТОВКА ДОКУМЕНТАЦІЇ

Короткий опис апаратури, режимів вимірювання та інші обставини спостережень, на основі яких будується ОБД наведені нижче. Вихідними даними про стратосферний аерозоль є коефіцієнт аерозольного послаблення на вказаних вище чотирьох довжинах хвиль в інтервалі висот від 0.5 до 40 км з кроком 0.5 км. Крім коефіцієнта аерозольного послаблення, в роботах з дослідження аерозолів часто використовується зворотна величина — довжина аерозольного послаблення.

Коротко структура ОБД відображена у таблиці.

У процесі роботи апаратура SAGE II вертикально сканує лімб атмосфери зі супутника під час заходу і сходу Сонця (15 заходів і 15 сходів кожного дня). Захід чи схід, під час якого виконуються вимірювання, називається подією. Нахилена на 57° орбіта корабля ERBS рівномірно розподіляє вимірювання SAGE II по широті. Детально схему

Структура об'єднаної бази даних

Рівень	Зміст
0	Вихідні файли з інформацією про всі події за місяць кожного року
1	Файли з вільним доступом
1_1	Файли метаданих про події
1_2	Файли візуалізації профілів і їхні прив'язки
1_3	Файли метаданих про географічний розподіл подій
2	Інформація про кліматичні моделі і нормальні обра-зи параметрів
2_1	Файли з реалізаціями, сезонним ходом і трендом для кожного параметру стратосфери у довільній забезпеченій 3D-комірці (з візуалізацією)
2_2	Файли з даними про розподіл параметрів часової мінливості за забезпеченими 3D-комірками
2_3	Файли з глобальною інтерполяцією даних рівня 2_2 за допомогою сферичних функцій

лімбових вимірювань потоків радіації спектрометром супутника наведено в роботі [1].

На рис. 1 показана діаграма кількості N щомісячних вимірювань, проведених апаратурою SAGE II за час здійснення проекту.

SAGE II провадив вимірювання у семи спектральних каналах від 385 до 1020 нм. За час своєї роботи SAGE II забезпечив великий об'єм корисних даних з хімії та динаміки зміни геофізичних параметрів у верхніх шарах тропосфери і стратосфери Землі (10—40 км).

За даними вимірювань випромінювання у зазначених вище чотирьох спектральних каналах, розсіяного і поглиненого газами та аерозолями на різних висотах, відновлювалися вертикальні профілі концентрації озону, водяної пари, двоокису азоту і аерозолію. Описання методики обчислення функції пропускання атмосфери і аерозольної екстинкції можна знайти в роботі [1].

Детальний опис рівнів розробленої бази даних наведено нижче.

Опис «Рівня 0». Після обробки фахівцями NASA телеметрії, що була прийнята зі супутника, дані SAGE II поширюються у подвійному форматі. Архів, що містить результати вимірів апаратурою SAGE II, має 374 файли загальним обсягом 1.25 Гб і складається з двох видів файлів для кожного місяця спостережень: «*index*» і «*species*».

Кожен окремий запис файлу *index* містить просторово-часову інформацію для кожної події за весь місяць (до 930 подій). Файли *species* — це файли прямого доступу, що містять до 930 записів. Кожен запис містить профілі густини частинок і додаткову специфічну інформацію про подію. Кожні два окремих файли «*index*» і «*species*» об'єднують місячний ряд даних за всіма компонентами.

Дані, представлені у подвійному форматі, вимагають наявності спеціального програмного забезпечення для їхнього перегляду і використання. Програмні коди, що супроводжують такі дані, як правило, придатні для роботи з окремими файлами і не підходять для роботи з великим обсягом інформації. Тому для роботи з архівом SAGE II було розроблено програмне забезпечення для розпакування бінарних файлів у текстовий формат.

Після розпакування архів SAGE II складається з текстових файлів одного формату, які об'єднують місячний ряд спостережень. Імена

файлів (наприклад SAGE_II_SPEC_198410.6.00.txt) містять вказівку на:

- прилад (SAGE_II);
- рік і місяць проведення вимірювань у форматі YYYYMM (198410);
- номер версії програмного забезпечення для обробки телеметрії (6.00).

Оскільки телеметрія, отримана при супутниковому зондуванні, — це послідовність окремих подій, кожен файл має заголовок, який містить інформацію щодо всіх подій і блоки з інформацією за кожною окремою подією.

Заголовок містить:

- число вимірювань, проведених у даному місяці [Number of profiles (records) in these files],
- номер версії програмного забезпечення для обробки телеметрії [LaRC Driver version, LaRC Transmission version, LaRC Inversion version, LaRC Spectroscopy version],
- мережу висотних рівнів (км) [Geometric Altitudes] і її крок [Altitude Grid spacing],
- нижні і верхні межі за висотою для профілів всіх представлених параметрів атмосфери [Min & Max altitudes],
- перерахований перелік всіх подій за місяць із зазначенням для кожної дати порядкового номера у розглянутих добі, часу, координат прицільної точки, типу події як щодо космічного апарата, так і щодо землі [Event Specific Info useful for data subsetting].

Блок інформації про кожну окрему подію має заголовок із зазначенням порядкового номера події в даному місяці і включає:

- профілі тиску (мб), температури (К) і концентрації (молекул/см³) від 0.5 до 70 км у точках спостережень, отримані за результатами модельних розрахунків NMC (National Meteorological Center) на основі експериментальних даних в інших пунктах спостережень [Pressure, Temperature, Density];
- висоту тропопаузи (км) [Tropopause height];
- значення спектральних каналів вимірювань (мікрометри) [Channel wavelengths];
- профілі концентрації озону від 0.5 до 70 км (см⁻³) [O3 number density];
- профілі концентрації двоокису азоту від 0.5 до 50 км (см⁻³) [NO2 number density];

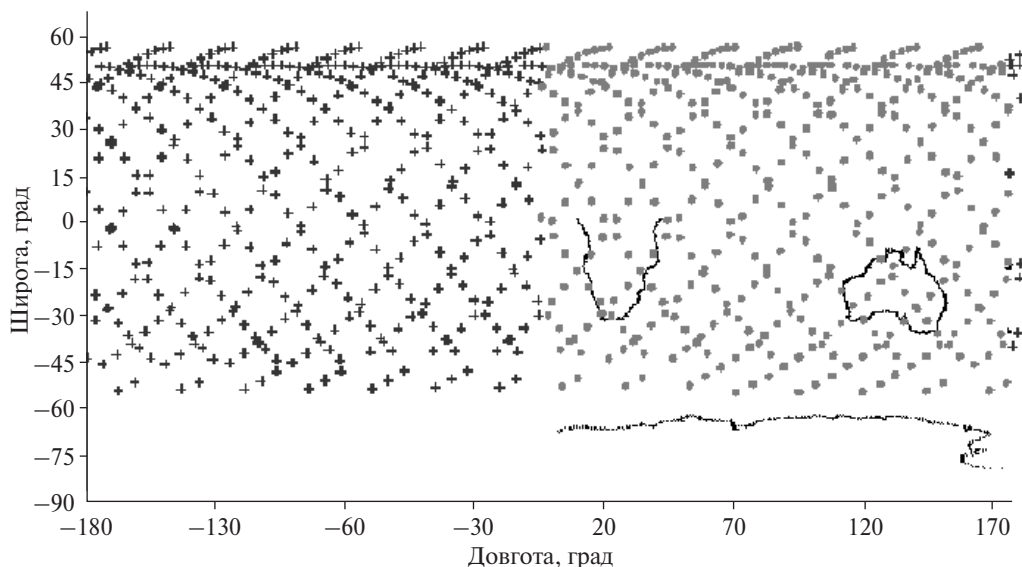


Рис. 2. Розташування подій у перший місяць роботи супутника. Хрестики — захід, плямки — схід Сонця

- профілі відносної концентрації водяної пари від 0.5 до 50 км (parts per part — ppp) [mixing ratio];
- профілі екстинкції аерозолі для чотирьох спектральних каналів: 1020, 525, 453, і 385 нм від 0.5 до 40 км (через 1 км) [386-nm, 452-nm, 525-nm, 1020-nm aerosol extinction];

Всі профілі мають просторову роздільну здатність вздовж вертикалі 0.5 км. За географією дані SAGE II охоплюють $\pm 80^\circ$.

Опис «Рівня 1». «Рівень 1» містить сім файлів. З них чотири файли є текстовими з вільним доступом, в інших трьох розташовано шість робочих книг електронних таблиць Excel 2007.

Файли з вільним доступом мають імена PTD_profiles.dat, SCA_profiles.dat, AEL_profiles.dat і RelErr_profiles.dat. У всіх файлах запис (рядок) відповідає події які мають хронологічну послідовність. Записи містять такі поля (стовпці):

- гринвіцький час події;
- тип події (схід, захід);
- номер широтного поясу, в якому сталася подія (широтні пояси мають ширину 5° , і їхній рахунок починається з Півдня);
- номер довготної частки, в якій сталася подія (довготні частки мають ширину 10° , і їхній рахунок починається з центрованої на -175°);

- номер комірки географічної сітки (комірки першого широтного поясу нумеруються від 1 до 36 з заходу на схід, комірки другого широтного поясу від номера 37 до 72 і т. д.);
- далі йдуть записи з відліками або кодом відсутності даних (-999).

Файл PTD_ має 420 полів з відліками, в яких записані послідовно тиск, температура і концентрація по 140 висотних рівнях. Файл SCA_ має 340 полів з відліками, послідовно озон — по 140 висотних рівнях, двоокис азоту і водяна пара по 100 висотних рівнях. Файл AEL_ має 320 полів з відліками по 80 висотних рівнів на кожен довжину хвилі в порядку зростання довжин хвиль. І нарешті, файл RelErr_ має 1080 полів з відліками, у яких записані значення відносних похибок або код відсутності. Послідовність полів: 420 як в PTD_, за ними 340 як в SCA_ і за ними 320 як в AEL_.

Три робочі книги PTD_profiles.xlsx, SCA_profiles.xlsx, AEL_profiles.xlsx повторюють структуру текстових файлів, але у форматі Excel 2007 і з заголовками полів. Це дублювання пов'язано з необхідністю забезпечити, з одного боку, користувачеві зручності і можливості в рамках обраної СКБД, а з іншого — можливість зміни СКБД. Робоча книга Step_1.xlsx містить детальний опис

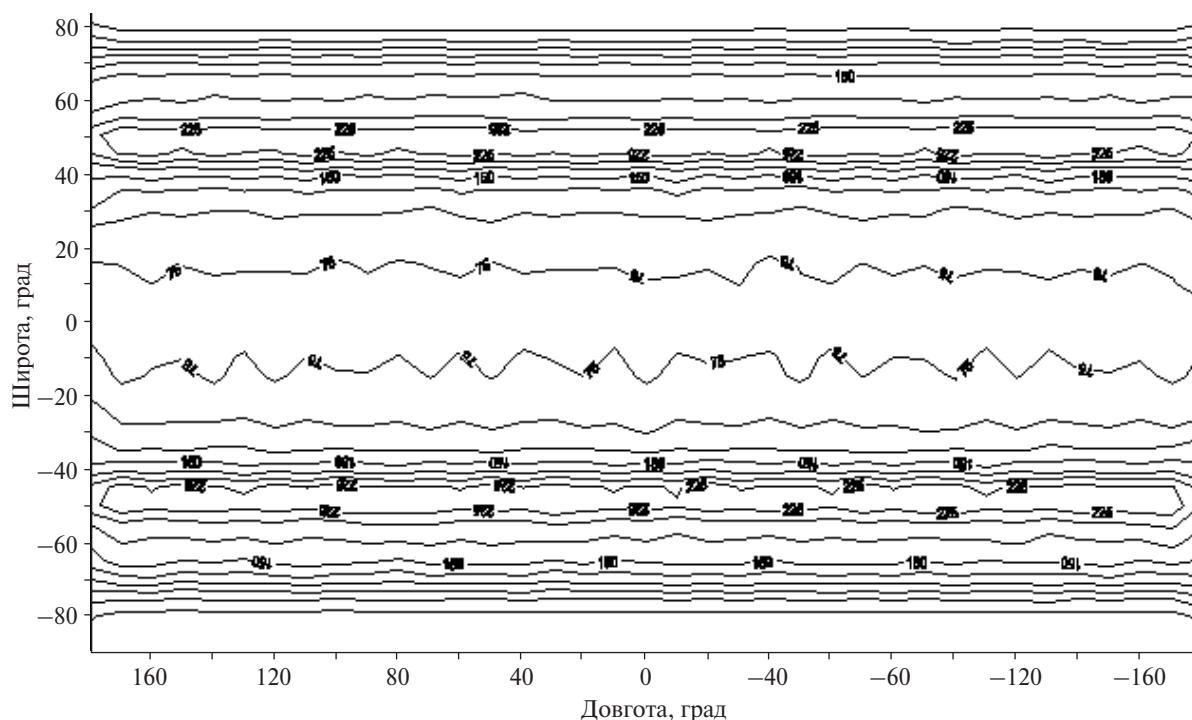


Рис. 3. Географічний розподіл подій

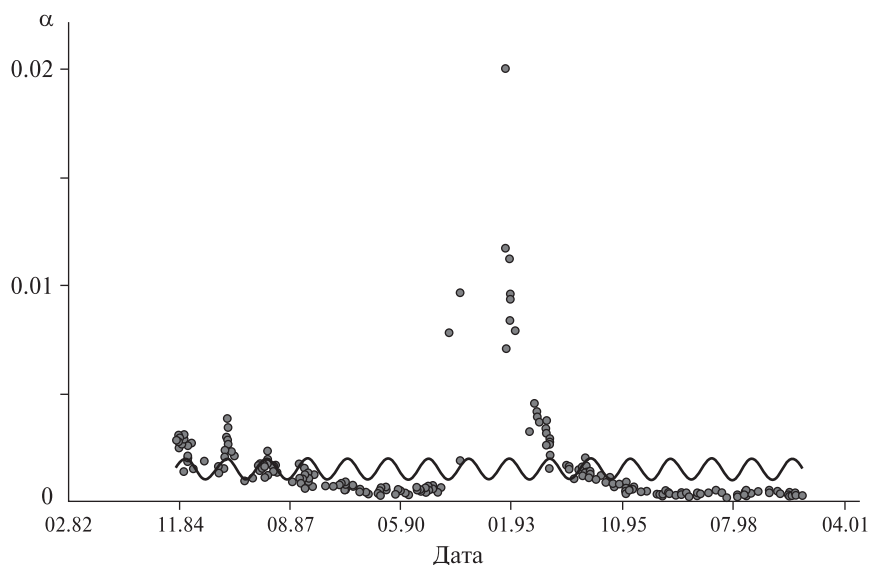


Рис. 4. Сезонний хід і реалізація коефіцієнта α аерозольної екстинкції на довжині хвилі λ 452 нм

усіх 143309 подій, відображених у БД. Кожній події відповідає запис з 11 полів, що містять точну географічну і часову прив'язку, умови вимірювання і висоту тропопаузи під час вимірювань.

Робоча книга Step_2.xlxb призначена для перегляду географічної прив'язки подій за обра-

ним користувачем часом і будь-якої пари профілів параметрів з вимірюваних у вибраній користувачем події. Робоча книга Step_3.xlxb містить дані про глобальний розподіл подій. Приклад доступних для перегляду діаграм наведені на рис. 2 та 3.

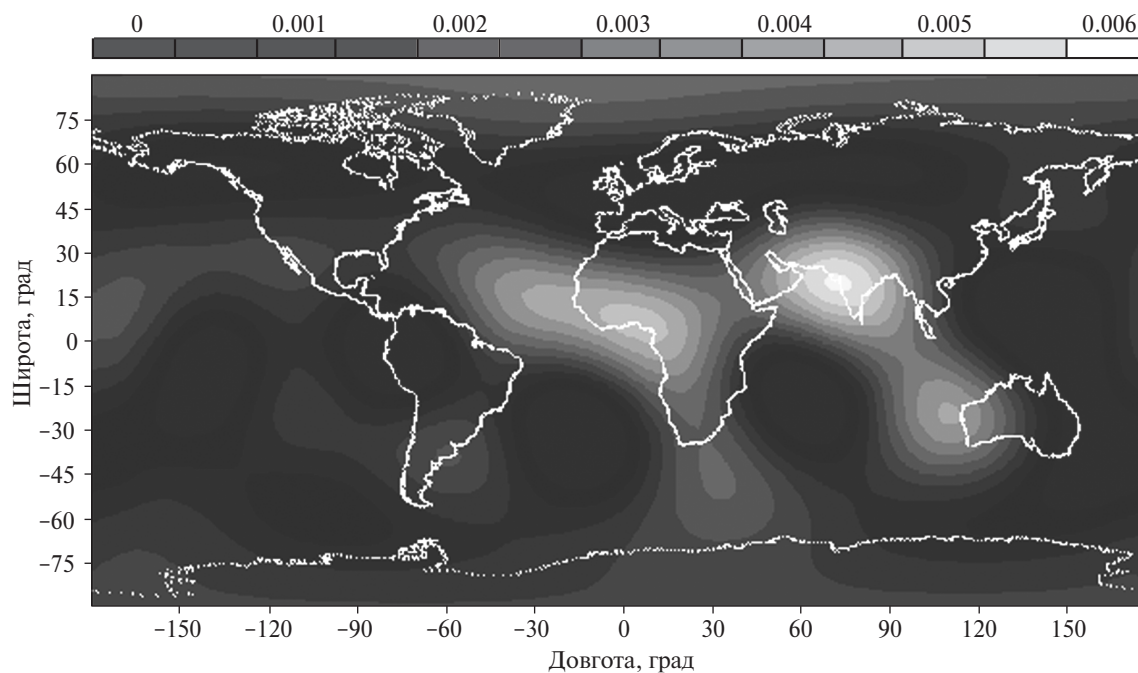


Рис. 5. Глобальний розподіл постійної складової коефіцієнта аерозольної екстинкції на довжині хвилі λ 452 нм для $H = 14.5$ км

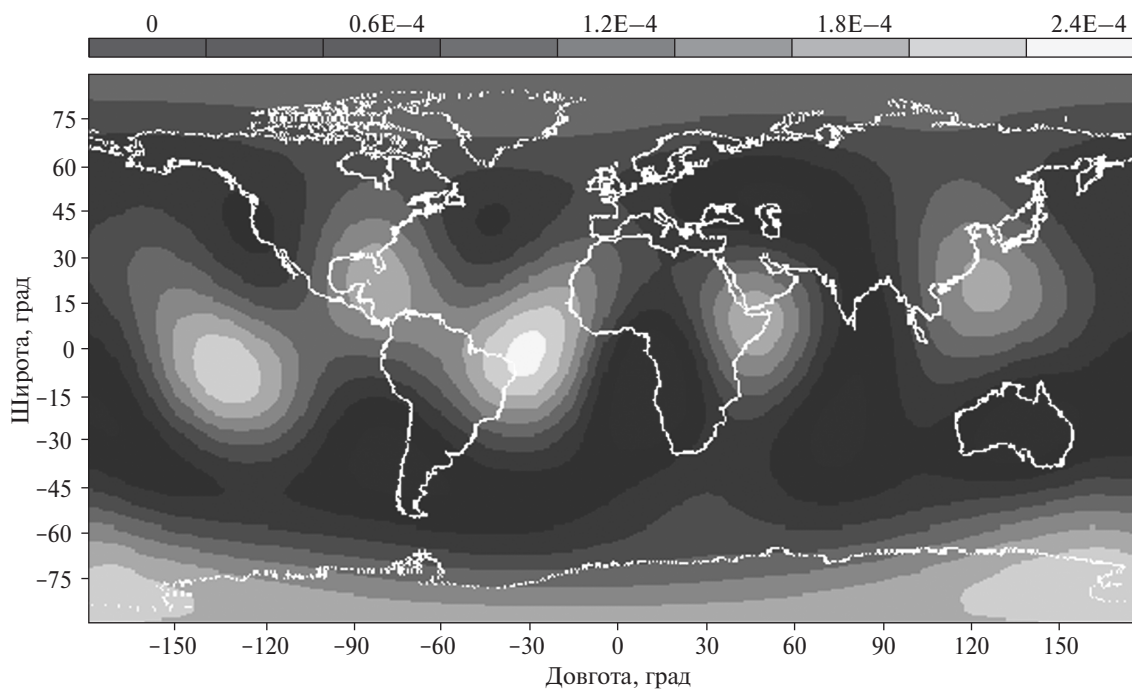


Рис. 6. Глобальний розподіл амплітуди I гармоніки коефіцієнта аерозольної екстинкції на довжині хвилі λ 452 нм для $H = 14.5$ км

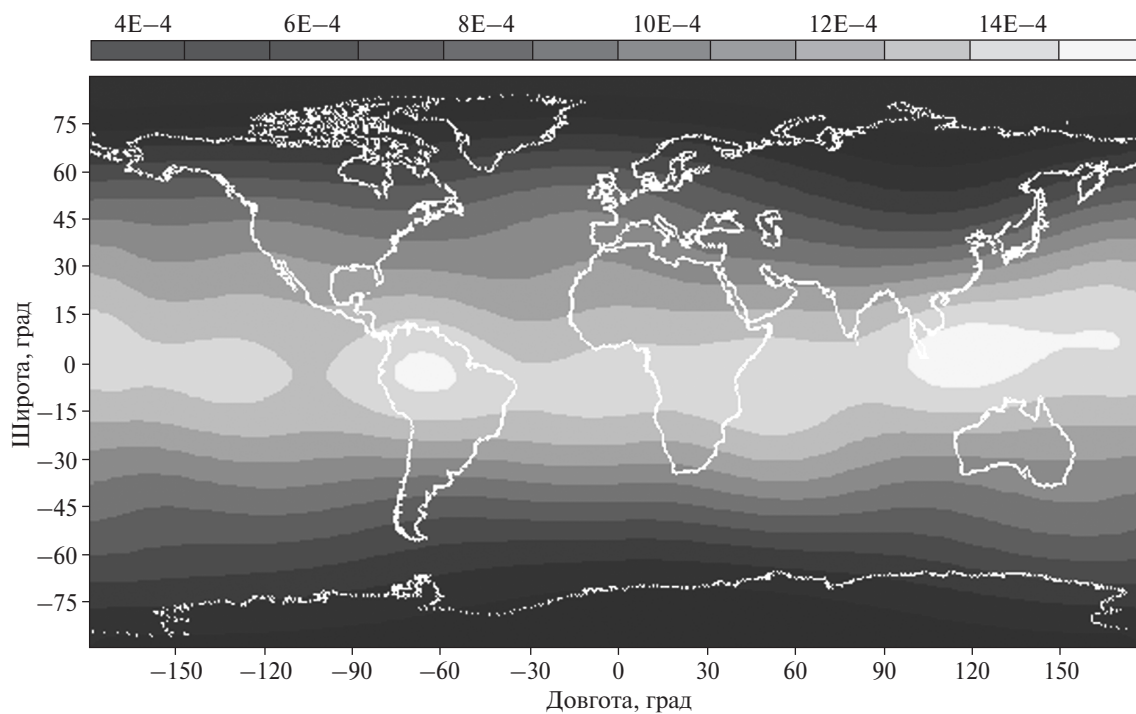


Рис. 7. Глобальний розподіл постійної складової коефіцієнта аерозольної екстинкції на довжині хвилі λ 452 нм для $H = 21$ км

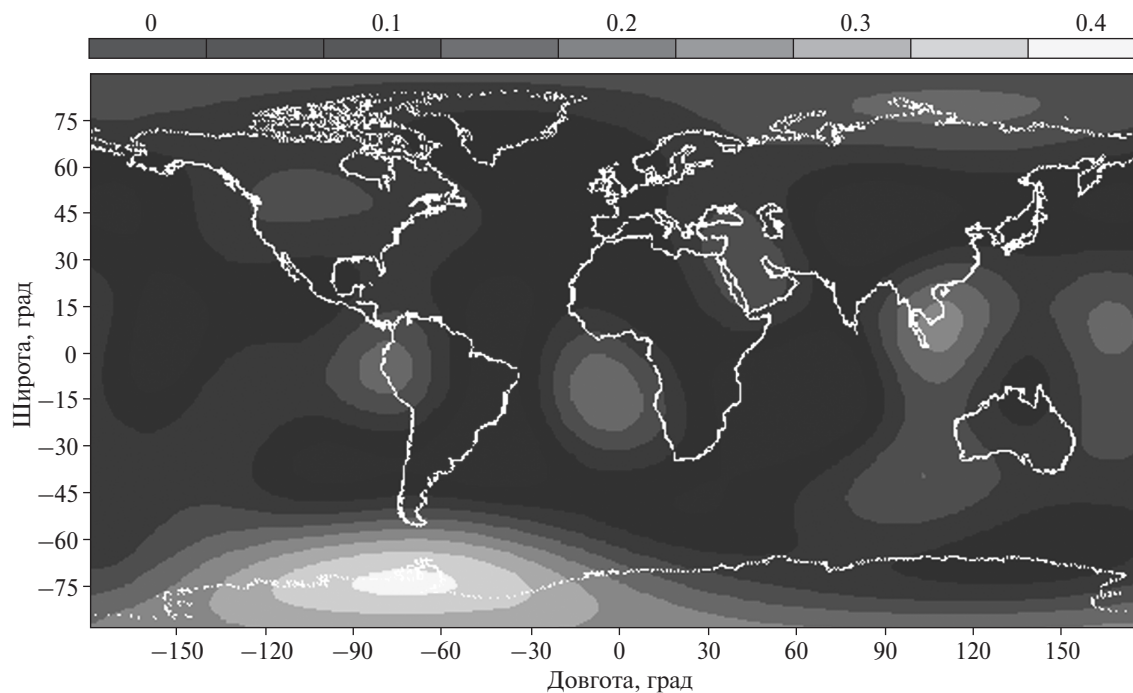


Рис. 8. Глобальний розподіл амплітуди I гармоніки коефіцієнта аерозольної екстинкції на довжині хвилі λ 452 нм для $H = 21$ км

Опис «Рівня 2». Об'єднана база даних містить 12 файлів «Рівня 2», по три файли на кожну з чотирьох довжин хвиль. Приклад розрахованого сезонного ходу та реалізації для коефіцієнта аерозольної екстинкції на довжині хвилі 452 нм наведено на рис. 4. Робоча книга Clim_T_1 призначена для тестування статистичної значимості параметрів сезонного ходу і трендів та організації статистично значимих даних у вигляді таблиць, що відображають структуру географічної сітки на кожному висотному рівні. І нарешті, робоча книга Clim_T_2 призначена для інтерполяції статистично значимих даних регресією за сферичними функціями на всю географічну сітку для тих висотних рівнів, число статистично значимих відліків на яких не менше ніж у чотири рази перевищує число базисних сферичних функцій.

ПРИКЛАДИ РОБОТИ

З ПРЕДСТАВЛЕНОЮ БАЗОЮ ДАНИХ

Для ілюстрації найпростіших можливостей бази розраховано кліматичні норми для коефіцієнтів аерозольного послаблення, визначено межі його природної мінливості на чотирьох довжинах хвиль і оцінено похибки для цих норм. Деякі приклади результатів наведено на рис. 4—8.

1. Білокриницька Л. М., Міщишина І. Г., Крученицький Г. М. Вибір та обґрунтування структури інформаційного та алгоритмічного забезпечення даними про оптичну активність стратосферного аерозолі бази даних коефіцієнтів аерозольного послаблення. — Київ, 2011. —

44 с. — (Рукопис деп. у ГНТБ України 25.08.2011; № 66 Ук-Д2011).

2. Галкина И. Л., Зуев В. В., Крученицкий Г. М. Использование электронных таблиц Excel для изучения озоносферных процессов: архивация, анализ, моделирование // Оптика атмосферы и океана. — 2003. — 16, № 10. — С. 943—951.
3. Уокенбах Д. Професійне програмування на VBA в Excel 2007. — М.—СПБ—Київ: Діалектика, 2008. — 784 с.

Надійшла до редакції 19.09.11

L. M. Belokrinitskaya, G. M. Kruchenitsky,
I. I. Mischishina

THE DATABASE OF AEROSOL EXTINCTION COEFFICIENTS FROM MEASUREMENTS BY SAGE II MISSION INSTRUMENTS AT FOUR WAVELENGTHS

We describe the structure and resources of the combined database developed from the measurements of the SAGE II mission from October 1984 till April 2000 at four wavelengths: 386, 452, 525, and 1020 nm. For the database, human-friendly interface is elaborated and three levels are created. The zeroth level of the database contains some data on vertical profiles of concentrations for ozone, water vapour, nitrogen dioxide and aerosol at the four wavelengths mentioned above. The first level provides a possibility of data searching, filtrating and visual analyzing, sorting the data by certain criteria and describing these criteria. The second level keeps some information on climate models and normal (standard) parameters, namely, spectral behaviour of season variations, variations of season trend and long-term variations. Besides, the possibility exists of attaching additional data on aerosol indexes which are obtained from measurements of TOMS-OMI satellites and of adding new data from Data Centres. We give some examples demonstrating possibilities of the database.