

Příprava fenologické databáze v ORACLE

The preparation of the phenological database in ORACLE

Coufal Lubomír – Nekovář Jiří

Český hydrometeorologický ústav Praha

Abstract

In this article are described the principles for the creation of the database of phenological data in CHMI. The attention is devoted especially to metadata. The rules are mentioned, according which are described stations, observed plants and observations performed in individual locations. It belongs here the declination and exposure of the slopes, pedological conditions, irrigation, pests and diseases. Data tables are designed according to the kinds of stations. Data input is provided by the virtual electronic form, data checking repose upon either the time interval among particular phenophases or in the area checking, based on the percentage of the probability of occurrence. Normal values are derived data as well as products for the operative needs of the service. For the data retrieving serve standard products of ORACLE. It is assumed that the database will be fully functional in the year 2005.

Key words

Phenology, ORACLE, database

Úvod

V posledních letech se stále více ukazuje potřeba používat moderní databázové metody pro archivaci dat, které získávají meteorologické / hydrometeorologické služby celého světa. Tato tendence se začala nejdříve projevovat v meteorologii a klimatologii, postupně však začíná pronikat i do jiných vědeckých oborů. V našem příspěvku se budeme věnovat aplikaci ORACLE ve fenologii tak, jak je v našem ústavu připravována.

V tom budeme vycházet ze zkušeností, které jsme získali při tvorbě naší klimatologické aplikace CLIDATA, která je dnes plně funkční a je distribuována v rámci mezinárodní pomoci do zahraničí. Před tím byla v roce 2002 úspěšně testována odbornou komisí, kterou ustavila WMO a jejímž cílem bylo nalézt optimální aplikace, které by v mezinárodním měřítku mohly nahradit systém CLIDATA, který byl vyvinut před více než 20 lety a který již nestačí požadavkům, které jsou na databázové systémy kladeny.

Podotýkáme, že se jedná o první verzi návrhu, která může během programování nebo později v období zkušebního provozu doznat řady změn, někdy i dosti podstatných.

Databázová struktura

Relační databázové struktury jsou velmi citlivé na navrženou strukturu; od ní se odvíjí jak programování, tak samotné využívání databáze. V našem pojetí rozdělíme systém do následujících šesti základních skupin, kterými jsou metadata (popisy stanic a pozorování), vlastní (pozorovaná) data, pořizování dat a jejich kontrola, odvozená data, produkty a pomocné tabulky, sloužící k zajištění

všech požadovaných funkcí databáze. Kromě toho si musíme být rovněž vědomi, že aplikace musí vyjít vstříc nejen současnému způsobu pozorování, ale musí být schopna akceptovat data, získaná v souladu s předchozími návody pro pozorovatele a, pokud budeme uvažovat v mezinárodních souvislostech, i metodám pozorování v zahraničí.

Moderní databázové prostředí je nezbytné a bude velmi dobře využitelné pro připravovaný nový projekt vědecko-technické spolupráce (COST) Evropské Unie nazvaný The Use of Phenological Data for Climatological Purposes, který bude řešen i aktivitou ČHMÚ v letech 2004 - 2007.

Metadata

Je zřejmé, že ve fenologii nevystačíme s jednoduchým popisem stanice tak, jak je známe z meteorologie. Pozorování se provádí na rozsáhlém územním celku, který může být místě značně rozdílný co do jeho svažitosti a orientace svahů, pedologických podmínek, zavlažování kulturních plodin a celé řady dalších faktorů. Nezbyvá tedy nic jiného, než tyto poměry u každého místa, v němž pozorování probíhá, standardními metodami popsat. Kromě toho není od věci doprovodit slovní popis jednak mapou, jednak i náčrtkem, který informace, vložené do databáze, vhodným způsobem rozvine nebo doplní. Přitom informace, kterými např. popisujeme pedologické podmínky, jsou do databáze vkládány předem jako konstanty (výčet všech možných hodnot), což umožňuje klást aplikaci kvalifikovaně dotazy (query). Jedině na základě těchto informací si můžeme ověřit, zda vyskytnuvší se anomálie od běžných podmínek jsou skutečnými chybami nebo naopak zda jde o rozdíly, vyvolané rozdílnými podmínkami stanoviště.

Péče o stanici

O stanici rovněž existuje celá řada informací, které nemají ryze odborný charakter. Máme zde především pozorovatele (včetně jeho adresy a ostatních náležitostí, jakými je např. rodné číslo, číslo občanského průkazu nebo evidenční číslo smlouvy a částky, na kterou je uzavřena). Bez významu rovněž nejsou návštěvy na stanici, sloužící k udržování kontaktu s pozorovatelem a ke zlepšení práce. Všechny tyto údaje, dnes často rozptýlené na celé řadě míst, poslouží ke zlepšení práce lidí, zodpovědných za staniční síť, např. k plánování návštěv nebo automatických výplat odměn.

Rostliny, fenofáze, odrůdy

Zde máme na mysli nikoliv vlastní pozorování, ale jeho systém. Dnes jsou v Česku prováděna pozorování odděleně pro polní plodiny, ovocné dřeviny a divoce rostoucí (lesní) rostliny. Tyto stanice se dnes prakticky nepřekrývají. Naproti tomu dříve (před rokem 1982) fenologické stanice prováděly pozorování všeobecná. Tomu musí struktura databáze vyhovovat, především tím, že data jsou ukládána jednotně.

K tomu slouží speciální tabulky (součást metadat), obsahující jednak výčet všech pozorovaných druhů, jednak, v závislosti na tomto výčtu, rovněž fenofáze, které jsou u jednotlivých

druhů pozorovány. Tím je znemožněno uložit do databáze rostlinu či fázi, která není v tomto výčtu obsažena. K tomu přistupují u polních plodin a ovocných dřevin rovněž seznamy pozorovaných odrůd, které se v souvislosti s odrůdovou skladbou postupně doplňují a mění.

Pozorování

Z výše uvedených tabulek se vyberou ty případy, které se na stanici pozorují. Tím je jednoznačně spojena definice všech možných pozorování a jeho podmnožinou, které bude na stanici věnována pozornost. Sloupce „počátek pozorování“ a „konec pozorování“ pak vymezují roky, v nichž je nebo byla rostlina pozorována a dále je pozorování vymezeno shodným způsobem pro jednotlivé odrůdy. Stanice tady může v rámci jednoho období platného pro druh pozorovat více jeho odrůd. Kromě toho se pěstební plochy střídají v rámci tzv. rotace plodin. Proto jsou plochy definovány v popisu stanice (opět s počátkem a koncem pozorování na stanovišti). Tím je naprosto přesně zajištěn vztah mezi lokalitou a zjištěnými údaji pro jednotlivé rostliny a jejich odrůdy. Tomuto faktoru přikládáme obzvláštní důraz, neboť může zásadním způsobem ovlivňovat získané údaje.

Choroby a škůdci

K metadatům patří i seznam chorob a škůdců (jen významných), které se mohou na rostlinách vyskytovat. Jsou definovány pro jednotlivé polní plodiny a ovocné dřeviny ve zvláštní tabulce a jsou pak připojena k datům. Mohou sloužit pro analýzu závislosti rychlosti vývoje a výnosu plodin na počasí.

Data

Fenologická data jsou rozvržena do tří základních tabulek podle jejich povahy (polní - ovocné - lesní). Řádek tabulky obsahuje data z jednoho roku, z jedné stanice a pro jednu rostlinu. Jednotlivé fenologické fáze mají vyhrazeny sloupce, do nichž se zapisují kalendářní data, v nichž se nástup dané fenofáze vyskytl. Databáze obsahuje nástroje pro práci s datovými formáty a převody jsou tudíž automatizovány. U polních plodin se sledují také fenometrické údaje (počet listů, délka rostlin) a ukládají se i údaje o výnosu; tato data jsou v tabulkách definována jako numerické hodnoty.

Uvedený návrh bude třeba ještě dále optimalizovat. To se týká především případného výběru ze spektra pozorovaných fenologických fází. Pak by se naskýtala možnost vytvořit více základních tabulek, dělených právě podle tohoto schématu.

Vstup (pořizování) dat

Budeme zde hovořit o standardním způsobu pořizování dat manuálním způsobem. Historická data pořizovaná předchozími metodami v letech 1984 - 2003 a nacházející se nyní v elektronické podobě budou importována přímo do databáze. K tomu se jako mezistupeň použijí textové (ASCII) soubory, které jsou standardním exportem z jakéhokoliv dříve užívaného formátu (XLS, DBF a podobně).

V soustavě importních tabulek se předem popíše struktura importované věty; import se pak provádí automaticky bez jakéhokoliv dalšího programování.

Pro vlastní pořizování dat bude sloužit virtuální tiskopis (3 typy), který je elektronickou kopií papírového dokladu (výkazu pozorování nebo dílčího hlášení). Ten se vytváří v administrátorské sekci aplikace tak, že se (podle předlohy) vybírají pouze rostliny. Jejich fenofáze (případně i numerické hodnoty) jsou pak přiřazeny automaticky. V záhlaví tohoto elektronického tiskopisu je stanice a rok.

Pokud si při vlastní práci zvolíme libovolnou stanici, pak podle přehledu pozorování se umožní vstup jen do těch sloupců, které mají význam, ostatní jsou nepřístupné a pro lepší přehled vybarveny šedě. Tato metoda zabrání pořizování dat pro plodiny nepozorované na stanici nebo pro neexistující fenofáze.

Data se (po uložení tiskopisu) okamžitě vkládají do databáze, kde jsou připravena k další kontrole. Není třeba zvláště zdůrazňovat, že práci lze v libovolném bodě přerušit a po kratší či delší době se k nim zase vrátit. Uložení v databázi umožní používat v operativní činnosti i neprověřená data, validační příznak ukazuje na skutečnost, že data dosud nejsou zkontrolována.

Odvozená data

Kromě základních dat, získaných pozorováním v sítích, obsahuje databáze i data odvozená. Jde o normálové (průměrné) hodnoty, soubory dlouhodobých charakteristik, které databáze průběžně aktualizuje. Metody výpočtu budou udržovány v souladu s doporučeními WMO. V tabulce jsou dále zahrnuty empirické pravděpodobnosti výskytu, které slouží kromě jiného ke kontrole získaných dat. Pochopitelně je nutnou podmínkou mít dostatečně dlouhé prověřené řady pozorování. Tato podmínka je snáze splnitelná pro divoce rostoucí rostliny. U polních plodin a ovocných dřevin je dlouhé řady možno sestavovat jen za předpokladu nepřehlížení k odrůdě.

Kontrola dat

Je samozřejmé, že data získaná pozorováním mohou obsahovat (a také obsahují) různé druhy chyb. Povaha takových chyb může být rozličná, od chybějících údajů přes chybně stanovený údaj (den výskytu) až po nepochopení metodiky pozorování.

Při kontrole dat v rámci jedné stanice slouží ke kontrole především posloupnost nástupu jednotlivých fenofází, časové meze přijetí pro fenofázi a také pro interfázi - jsou tak stanoveny maximálně možné odchylky fenofáze od normálu a časové odstupy mezi nimi. Při dalším rozvoji kontrolních metod pak počítáme s využitím klimatologických dat (především teploty a srážek), které umožní upřesnění (věrohodnost) nalezených rozdílů.

Při územní kontrole získaných dat použijeme normované hodnoty (podle informací o pravděpodobnosti výskytu z normálových hodnot), tedy procentuální hodnoty pravděpodobností. Tím se anulují vlivy fyzicko-geografických podmínek, nadmořských výšek a jiné. Na mapce obdržíme

barevné body podle pravděpodobnosti výskytu (odpovídající stanici, pozorované rostlině a fenofázi), přičemž analýza rozptylu těchto údajů umožní vysledovat, která data jsou pravděpodobně chybná.

Při kontrole dat jsou k údajům přiřazovány verifikační příznaky, označující jejich kvalitu. Údaj, který považujeme za podezřelý, bude označen jiným příznakem než údaj, u něž jsme hodnotu změnili nebo doplnili (v případě chybějícího pozorování). Nově získané datum můžeme odhadnout nebo vypočítat, což rozlišíme opět příznakem. Nakonec je i celý řádek (stanice, rok, rostlina) označen speciálním příznakem, který označí, že data byla validována a jsou (pravděpodobně) bez chyb. Tím končí cyklus staničních a územních kontrol.

Produkty a pomocné tabulky

Databáze počítá definované produkty automaticky podle předem připravených schémat a časových rozvrhů. Jde především o ty úlohy, které slouží pro operativní zpracování dat. Jejich členění i definice odpovídají konkrétním požadavkům služby a mohou se v průběhu času měnit nebo doplňovat.

K produktům patří do jisté míry i inventář, poskytující přehled dat, která dosud v databázi chybí. Ten je užitečný především v období, kdy jsou do databáze importována nebo pořizována data.

Pomocné tabulky nejsou pro uživatele obvykle přístupné. Slouží k zajištění bezproblémového chodu databáze.

Vybavování dat, spolupráce s jinými softwarovými produkty

K vybavování dat z databáze slouží jednak interaktivní programový produkt DISCOVERER, poměrně jednoduchý uživatelský nástroj, který na základě zadaných vyhledávacích podmínek poskytne data ve formě tabulek, podobných Excelu. Kromě toho může uživatel použít programovací jazyk SQL+ v jeho poslední verzi 8.0. Výstupem jsou pak textové soubory.

Instance ORACLE velmi jednoduše poskytují data, o která má uživatel zájem. V našem případě půjde o data fenologická a klimatologická. Ostatní softwarové produkty mohou využívat data z ORACLE pomocí ODBC driverů (např. GIS, Statistica apod.).

Výhledy

Uvažujeme, že databáze bude uvedena do zkušebního provozu během příštího roku. Pak budou následovat importy starších dat z let 1923 -1984, opravy a dodatečné programy včetně replikací dat, která odpovídá distribuovanému způsobu získávání dat v ČHMÚ. Z toho vyplývá, že v plném provozu bude databáze v roce 2005.

Shrnutí

V příspěvku jsou popsány zásady, podle nichž vzniká v ČHMÚ databáze fenologických dat. Pozornost je věnována především metadatům. Jsou zmíněna pravidla, podle nichž budou popsány

stanice, sledované rostliny a pozorování prováděná na jednotlivých lokalitách. Patří sem sklonitost a orientace svahů, pedologické podmínky, zavlažování, choroby a škůdci. Podle druhů stanic jsou navrženy datové tabulky. Vstup dat je zajištěn virtuálním elektronickým tiskopisem, kontrola dat pak spočívá jednak v časovém intervalu mezi jednotlivými fenologickými fázemi, jednak v územní kontrole vycházející z procentuální pravděpodobnosti výskytu. Normálové hodnoty jsou odvozená data stejně jako produkty, sloužící k operativní potřebě služby. K vybavování dat slouží standardní produkty ORACLE. Předpokládá se, že databáze bude plně funkční v roce 2005.

Literatura

VALTER, J., 1982. Návod pro činnost fenologických stanic - polní plodiny (Metodický předpis č.2). 1.vyd. Praha: Český hydrometeorologický ústav. 157 s.

VALTER, J., 1982. Návod pro činnost fenologických stanic - ovocné dřeviny (Metodický předpis č.3). 1.vyd. Praha: Český hydrometeorologický ústav. 148 s.

Návod pro činnost fenologických stanic - lesní rostliny (Metodický předpis č.10). Český hydrometeorologický ústav Praha 1988. 111 s.

TOLASZ, R. - COUFAL, L., 1998. The Static Data Quality Process as Part of CLIDATA Application. In: *Oesterreichische Beiträge zu Meteorologie und Geophysik*, Heft 19 Proceedings of the 2nd European Conference on Applied Climatology. Wien: Zentralanstalt für Meteorologie und Geodynamik. S. 155.

NEKOVÁŘ, J., 2000. The Structure of the Czech Phenological Database. In: *International Conference Progress in Phenology - Monitoring, Data Analysis and Global Change Impacts*. Freising, Germany: Technical University Munich, October 4-6, 2000. 3 p.

COUFAL, L. - TOLASZ, R., 2001. Klimatologická databáze CLIDATA - datový model a jeho aplikace. *Meteorologické zprávy*, roč. 54, č. 3, s.83-93. ISSN 0026-1176.

Kontaktní adresa

RNDr Lubomír COUFAL, Český hydrometeorologický ústav, CZ-14306 Praha 4-Komořany, Na Šabatce 17. Tel: 420-24403-2203, fax: 420-24403-2276, e: coufal@chmi.cz