

# BANCO DE DADOS METEOROLÓGICOS PARA PREVISÃO DO TEMPO E ESTUDOS CLIMÁTICOS

Sérgio Henrique S. Ferreira  
Centro de Previsão de Tempo e Estudos Climáticos – INPE  
Rod. Pres. Dutra, km 40, Cachoeira Paulista-SP, 12630-000  
[henrique@cptec.inpe.br](mailto:henrique@cptec.inpe.br)

Luciana Santos Machado Carvalho  
Centro de Previsão de Tempo e Estudos Climáticos – INPE  
Rod. Pres. Dutra, km 40, Cachoeira Paulista-SP, 12630-000  
[lmachado@cptec.inpe.br](mailto:lmachado@cptec.inpe.br)

Enivaldo Freire do O' Filho  
Centro de Previsão de Tempo e Estudos Climáticos – INPE  
Rod. Pres. Dutra, km 40, Cachoeira Paulista-SP, 12630-000  
[enivaldo@cptec.inpe.br](mailto:enivaldo@cptec.inpe.br)

## **ABSTRACT**

At CPTEC several processes related to weather forecast and climate studies are routinely performed, which manipulate data of different kinds, like observations, satellite imagery and numerical prediction products. Here we discuss the present status of implementation of the Meteorological Database, whose purpose is to be a reliable, efficient and suitable mechanism for the storage and retrieval of these meteorological data.

## **1. INTRODUÇÃO**

Para implementar sistemas de armazenamento de dados meteorológicos que sejam confiáveis, flexíveis e eficientes, tem sido comum nos grandes centros meteorológicos do mundo o uso da tecnologia de Banco de Dados (Pottier, 1995), (Nyfors, 1995), (Sanders, 1997), (Raoult, 1997) e (Stanek, 1999). Observa-se entretanto que as implementações variam em função das particularidades de cada centro.

No CPTEC iniciou-se a implementação do Banco de Dados Meteorológicos com o subsistema relativo aos dados observacionais, que se encontra operacional desde julho de 1999. Outros subsistemas, como de dados climatológicos, de imagens de satélite e de produtos numéricos de modelos de previsão estão sendo implementados.

Na primeira parte deste trabalho descreve-se o subsistema de dados observacionais. O subsistema de dados climatológicos é abordado de forma resumida a seguir e ao final são colocados os planos futuros do CPTEC relativos a Banco de Dados.

## **2. SUBSISTEMA DE DADOS OBSERVACIONAIS**

Este subsistema é responsável pelo armazenamento e recuperação das observações meteorológicas que chegam regularmente ao CPTEC pela rede GTS da OMM através do INMET. O controle de qualidade é feito por um outro sistema, denominado Pré-processamento (Dereczynski, 1996), que agrega às observações índices de confiabilidade dos valores reportados para as variáveis meteorológicas. Como resultado, o Pré-processamento codifica cada observação e sua respectiva informação de qualidade em BUFR, o formato padrão para observações adotado pela OMM (WMO, 1994)

Existem atualmente 10 tipos principais de observações cujas características principais são mostradas na tabela 1.

Tabela 1 - Principais tipos de observações

TIPO	DESCRIÇÃO	HORÁRIO (GMT)	QUANT. MÉDIA DIÁRIA	
			REGIONAL <sup>(*)</sup>	GLOBAL
SYNOP	superfície (terra)	00/06/12/18	2.400	15.600
SHIP	superfície (navio)	00/06/12/18	120	2.400
BUOY	superfície (bóia)	a todo instante	950	8.500
METAR	superfície (aeroporto)	a cada hora	1.750	1.750
AIREP	altitude, 1 nível, avião	a todo instante	200	15.000
SATOB	altitude, 1 nível, satélite	a cada hora	1.000	7.000
SATEM	sondagem vertical, satélite	a cada hora	800	1.000
TOVS	sondagem vertical, satélite	a todo instante	2.000	45.000
TEMP	sondagem vertical, não satélite	00/06/12/18	70	1.150
PILOT	sondagem vertical, não satélite	00/06/12/18	40	850

(\*) Área geográfica com latitude entre 90°S e 20°N e longitude entre 100°O e 20°O

## 2.1 CARACTERÍSTICAS TÉCNICAS

Este subsistema é uma adaptação do BDM da Météo-France (Pottier, 1995), o qual por sua vez utiliza a arquitetura básica no NEONS, desenvolvido pela Marinha dos EUA (Tsui, 1991). Neste caso contemplou-se apenas o tipo genérico de dados do NEONS denominado "observação". Em função das diferenças nas necessidades locais e infraestrutura computacional disponível, sua implementação para fins operacionais em outro ambiente terá que sofrer as devidas adaptações.

A infraestrutura computacional usada no CPTEC é composta pelo sistema de gerenciamento de banco de dados (SGBD) Oracle Server (versão 7.3) e pelas ferramentas de desenvolvimento Pro\*C (pré-compilador) e Developer (pacote com Forms, Reports e Graphics), rodando em ambiente Unix da Compaq (Tru64 Unix).

## 2.2 ORGANIZAÇÃO DO BANCO DE DADOS

Um conceito fundamental para o entendimento da organização deste banco de dados é o conceito de **dataset**, ou seja, um conjunto de observações que são manipuladas como uma unidade, sendo armazenadas fisicamente juntas e removidas ou arquivadas ao mesmo tempo. Como exemplo de dataset, pode-se citar o conjunto de observações regionais do tipo Synop relativas a um único dia.

Os dados podem ser classificados em 3 categorias ou domínios:

- primários**: são os dados propriamente ditos, ou seja, as observações, que correspondem a mais de 95% do espaço físico total do banco de dados. Estes dados são agrupados em datasets, o que permite tratar de forma diferenciada as observações dependendo do seu tipo e região geográfica;
- associativos**: contêm informações relativas aos dados primários, usadas como índice para localizá-los. Todo dataset do domínio primário é referenciado por um elemento do domínio associativo, que descreve suas coberturas temporal e geográfica, local de armazenamento e demais características.
- descritivos**: são dados relativamente estáticos por natureza, como tabelas de estações meteorológicas, países, etc. Além destes, os dados descritivos também incluem os metadados, ou seja, informações sobre os dados propriamente ditos. São exemplos de metadados os tipos de observações e respectivas características, como destino final, tempo de retenção antes de ir ao destino final, espaço físico estimado por dia, índices suplementares, dentre outras.

Como os dados estão sob a supervisão do SGBD Oracle, eles são armazenados em estruturas denominadas tabelas. No caso dos dados primários, elas são denominadas tabelas primárias, cuja estrutura será detalhada mais adiante. A tabela 2 relaciona as tabelas mais importantes relativas aos domínios descritivo e associativo.

Tabela 2 - Principais tabelas do banco de dados (domínios associativo e descritivo)

TABELA	CONTEÚDO	DOMÍNIO
AS_LLTBUFR	apontadores para seleção e acesso aos datasets	Associativo
CODE_FREQUENCE	códigos com que observações podem ser feitas	Descritivo
FAIXA_OMM	faixas de indicativos OMM alocadas aos vários países	Descritivo
FREQUENCE_OBS	frequências com que observações são feitas nas várias estações	Descritivo
LIST_STATION	listas de estações	Descritivo
LIST_STATION_ELEM	estações que fazem parte de cada lista	Descritivo
LLTBUFR_DSET	tipos de datasets associados a cada tipo de observação	Descritivo
LLTBUFR_GRUPO_DSET	grupos de datasets	Descritivo
LLTBUFR_IND	índices suplementares associados às tabelas primárias	Descritivo
LLTBUFR_IND_COL	colunas que compõem a chave de cada índice suplementar	Descritivo
LLTBUFR_PREFIXO_TBS	prefixos de tablespaces para tabelas primárias e/ou índices	Descritivo
LLTBUFR_SEQ	tipos de observações	Descritivo
LLTBUFR_SEQ_COL	colunas desnormalizadas de cada tipo de observação	Descritivo
LLTBUFR_SEQ_GDSET	associações entre tipos de observações e grupos de datasets	Descritivo
PAYS	países	Descritivo
STATION	estações meteorológicas	Descritivo

As tabelas com o prefixo "LLTBUFR\_" possuem dados descritivos do tipo metadado, citado anteriormente. Elas são usadas efetivamente pelos processos de criação das tabelas primárias e respectivos índices. Estes processos ocorrem automaticamente a cada final de mês, preparando o banco de dados para armazenar as observações que serão recebidas no(s) mês(es) seguinte(s).

A tabela AS\_LLTBUFR é a única tabela que possui dados do domínio associativo. Ela é usada sempre que for efetuada uma recuperação de dados, uma vez que ela contém os elementos necessários para localizar todos os datasets que possuem observações que podem atender àquela recuperação. Uma estrutura simplificada desta tabela é mostrada na tabela 3.

Tabela 3 - Estrutura da tabela associativa (AS\_LLTBUFR)

COLUNA	DESCRIÇÃO
lltbufr_id	identificador do dataset
seq_type	tipo de observação
dset_name	nome do dataset
tbl_name	nome da tabela primária onde estão as observações do dataset
db_name	nome do banco de dados que contém a tabela <i>tbl_name</i>
min_lat	latitude mínima existente
max_lat	latitude máxima existente
min_lon	longitude mínima existente
max_lon	longitude máxima existente
min_dat	data mínima existente
max_dat	data máxima existente
min_val	lltbufr_val mínimo existente
max_val	lltbufr_val máximo existente
rec_cnt	número de observações do dataset
stamp_time	data de criação do dataset
stor_per	período de retenção (dias)
fate	destino do dataset ao final da retenção
status	status atual ("load" ou "rdonly")

As colunas cujos nomes começam com "min\_" ou "max\_" são estatísticas mantidas pelo subsistema, acerca das observações que fazem parte de cada dataset. Toda vez que observações são inseridas no banco de dados estas estatísticas são devidamente atualizadas. Assim, quando for feito um pedido para recuperar, por exemplo, as observações de um determinado tipo, área geográfica e período de tempo, a tabela AS\_LLTBUFR será primeiramente consultada para se determinar quais os possíveis datasets que devem ser pesquisados. De posse da lista destes

datasets, é feita uma busca detalhada em cada um deles usando-se para isso as colunas *db\_name* e *tbl\_name*, que contêm os nomes do banco de dados e da tabela primária onde suas observações se encontram.

Esta organização permite que o banco de dados possa ser distribuído em bancos de dados menores, inclusive em máquinas distintas, em função de critérios como tipo de observação, idade dos datasets, etc. Desde que a tabela AS\_LLBUFR do banco principal continue referenciando os datasets residentes nos bancos de dados menores, o conjunto completo de observações pode ser recuperado de forma transparente. Na implementação do CPTEC esta característica não foi utilizada, ou seja, todos os datasets residem num único banco de dados e possuem o atributo *db\_name* nulo.

As tabelas primárias, usadas para armazenar as observações meteorológicas, possuem uma estrutura comum, detalhada na tabela 4. Este fato permite que o conjunto de rotinas de manipulação de dados deste subsistema seja genérico, tratando de forma independente e absolutamente neutra os diversos tipos de observação existentes. As particularidades associadas a cada tipo de observação estão cadastradas nas tabelas descritivas com prefixo "LLBUFR\_" (vide a tabela 2), o que torna este subsistema flexível quanto à inclusão de novos tipos de observações.

Tabela 4 - Estrutura das tabelas primárias

COLUNA	DESCRIÇÃO
lltbufr_id	Identificador do dataset
latitude	Latitude
longitude	Longitude
dat	Data e hora da observação ( formato AAAAMMDDHHmmss)
lltbufr_val	Identificador do observador (indicativo OMM, ship, OACI etc.)
amendment	Indicador de versão (0, 1, ...)
date_insert	Data de inserção no banco de dados
flag_pretr	Flag do pré-processamento (menor índice de confiabilidade encontrado)
last_arrived	Indica a última versão da observação, quando dela se tiver recebido diversas
C1	Coluna desnormalizada 1 (nome na tabela LLBUFR_SEQ_COL)
...	.....
Cnn	Coluna desnormalizada <i>nm</i> <sup>(*)</sup> (nome na tabela LLBUFR_SEQ_COL)
bitstream	Observação codificada em BUFR

(\*) *nm* deve ser menor ou igual a 25

As colunas das tabelas primárias podem ser separadas em 3 grupos:

- colunas fixas, iguais para todos os tipos de observação; são as 9 primeiras colunas listadas (desde *lltbufr\_id* até *last\_arrived*);
- colunas desnormalizadas (C1 ... Cnn): correspondem aos valores observados para um conjunto selecionado de variáveis meteorológicas de maior interesse;
- *bitstream*: é a coluna que contém a observação codificada em BUFR.

O conjunto de colunas desnormalizadas de uma tabela primária é específico para cada tipo de observação. Seus valores foram previamente extraídos da codificação BUFR armazenada na coluna *bitstream*. Embora esta redundância represente mais espaço físico para armazenamento, a solução adotada apresenta as seguintes vantagens:

- a) é possível recuperar observações usando como critério de seleção valores observados, e não apenas dados de identificação como região geográfica e período de tempo;
- b) é mais rápido recuperar informações se elas já estiverem em colunas desnormalizadas, não sendo necessário recuperar nem decodificar o código BUFR;
- c) as aplicações que processam dados em BUFR podem usar diretamente a coluna *bitstream*; este é o caso do Metview (vide item 2.3);

Uma vantagem na solução adotada de guardar os dados em BUFR é permitir o armazenamento de determinados tipos de uma forma bastante efetiva. Este é o caso das sondagens verticais da atmosfera (tipos Temp, Pilot, Satem e Tovs), que reportam valores em dezenas e até centenas de níveis de altitude. Sem o BUFR seria

inviável o armazenamento destas observações usando-se uma estrutura única de tabela primária para todos os tipos de observações.

A coluna *bitstream*, embora existente em todas as tabelas primárias, pode ser, para um dado tipo de observação, colocada como não obrigatória, ou seja, nem todas as observações daquele tipo precisam ter valor armazenado para esta coluna. Isso significa que é possível a criação de um novo tipo de observações que não estejam codificadas em BUFR, possuindo apenas colunas desnormalizadas (além das fixas, que são obrigatórias).

A organização aqui apresentada permite estruturar o banco de dados em diversas tabelas. Esta estruturação é absolutamente indispensável, devido à grande quantidade de observações que chegam (aproximadamente 100 mil por dia, ou 40 milhões por ano) e precisam ser armazenadas. O critério para particionar este conjunto gigantesco em subconjuntos menores (datasets) usa, em primeiro lugar, o instante de tempo e, depois, a região geográfica. Com isto atende-se a uma série de processos que demandam dados especificando como critério um intervalo de tempo razoavelmente estreito. Este é o caso das aplicações relacionadas com previsão de tempo (e.g. análise objetiva, previsão de meteorologistas, plotagem de cartas sinóticas e verificação de modelos), que referenciam um período desde 6 horas até alguns dias.

Para aplicações que requerem, diretamente ou mediante funções estatísticas, dados de uma região relativos a períodos muito maiores de tempo (maiores que um ano), esta organização não é a mais apropriada. Isto ocorre porque os dados resultantes de tal demanda se encontram em várias tabelas, o que torna esta organização ineficiente e até inviável para atender estas aplicações. A inviabilidade deve ocorrer no momento em que for implementado no CPTEC um sistema de arquivamento automatizado que, integrado com o SGBD, permita a migração dos datasets mais antigos para dispositivos de armazenamento de outros tipos, como disco ótico ou biblioteca de fitas controlada por robôs.

As aplicações deste último tipo, que solicitam dados de um intervalo longo de tempo, deverão ser atendidas pelo subsistema de dados climatológicos, que está em fase de implementação no CPTEC e é tratado no item 3 deste trabalho.

### 2.3 INTERFACES DISPONÍVEIS

A figura 1 mostra as interfaces do subsistema de dados observacionais hoje existentes. Como se pode notar, a única forma de alimentar ou alterar as observações é através do sistema de Pré-processamento. Os dados descritivos (cadastros de apoio e metadados) podem ser alterados e consultados usando a interface interativa. As demais interfaces são usadas apenas para consulta e recuperação de dados. A seguir descreve-se cada uma destas interfaces.

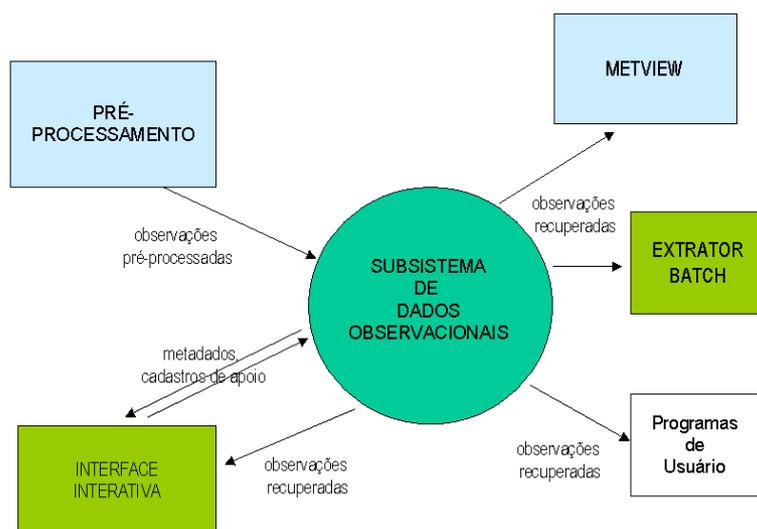


Figura 1 : Interfaces do subsistema de dados observacionais

- Pré-processamento: gera observações em BUFR com o controle de qualidade já efetuado, que são submetidas ao processo de ingestão no banco de dados; este ocorre na mesma frequência do Pré-processamento, ou seja, a cada 30 minutos;
- Interface Interativa: é usada para consulta às observações meteorológicas e sua eventual exportação, consulta e alteração dos dados descritivos e geração de gráficos com estatísticas de quantidade de observações alimentadas; as observações são mostradas no formato de formulários, ou seja, linhas e colunas; existem dois tipos de usuários: o normal, habilitado apenas para consulta, e o administrador, que pode modificar os dados descritivos;
- Metview: é uma ferramenta para acesso, manipulação e visualização de dados meteorológicos (Almeida et al., 2000); ela acessa o banco de dados fazendo uso de um conjunto de rotinas disponibilizadas a todos os usuários, que recuperam e retornam as observações em blocos de 30; neste caso, apenas a coluna *bitstream* (que contém o BUFR) é retornada, uma vez que o Metview sabe tratar este formato;
- Extrator *batch*: um utilitário de extração de observações em arquivos texto e/ou BUFR, a partir de um conjunto de diretivas como critério de seleção, colunas desejadas e respectivos formatos, etc.; ele é usado em scripts operacionais para gerar arquivos usados em aplicações anteriores à implementação do banco de dados;
- Programas de usuário: outros programas que acessam as observações através do mesmo conjunto de rotinas usadas pelo Metview e pelo extrator batch; eles podem rodar em qualquer máquina da rede interna do CPTEC, acessando remotamente o servidor onde reside o banco de dados.

## 2.4 RESULTADOS OBTIDOS

A implantação do subsistema de dados observacionais trouxe os seguintes benefícios para o CPTEC:

- Compartilhamento de dados: eliminaram-se as cópias redundantes dos dados observados, criadas em função de necessidades específicas dos grupos de trabalho; os dados passaram a residir, de forma definitiva, apenas no banco de dados;
- Integração: devido ao compartilhamento de dados, vários processos passaram a trocar informações através do banco de dados, que passou a servir como elemento de integração entre eles;
- Eficiência na recuperação dos dados: o tempo para recuperação dos dados passou a ser substancialmente menor do que anteriormente, quando do uso de varredura sequencial em arquivos convencionais contendo observações;
- Sofisticação na consulta aos dados: foram viabilizadas consultas mais elaboradas como, por exemplo, aonde e quando uma variável meteorológica ultrapassou um determinado valor (se existir uma coluna desnormalizada correspondente);

Em termos de desempenho, os tempos citados abaixo são tempos totais decorridos, ou seja, englobam os processos do subsistema, do SGBD, do sistema operacional e até dos demais processos em execução no momento da medição. A operação de inserção de 800 novas observações consome por volta de 30 segundos. Este valor pode variar em função do número de colunas desnormalizadas e da existência ou não de duplicatas entre as observações (o que leva a rotina de inserção a sair do modo vetorial, em blocos de 30, para o modo individual). A recuperação de 800 observações do tipo Synop no formato BUFR e sua gravação em arquivo convencional leva em torno de 4 segundos.

## 3. SUBSISTEMA DE DADOS CLIMATOLÓGICOS

O subsistema de Dados Climatológicos, atualmente em desenvolvimento no CPTEC, corresponde a uma complementação do subsistema de dados observacionais, decorrente da necessidade de acesso mais eficiente a dados relativos a longos períodos de tempo, porém restritos a localidades e a variáveis meteorológicas de interesse.

Portanto, a ênfase deste subsistema, é o acesso as informações climatológicas para áreas geográficas determinadas, auxiliando assim, diversos tipos de pesquisas. As consultas a estes dados poderão ser relativas as estações meteorológicas, cidades, bacia e sub-bacias hidrográficas, que poderão ser realizadas de forma interativa, com saídas para tela e também arquivos.

Também são integrados diferentes tipos de informações climatológicas, pertinentes à América do Sul e procedentes de diversas fontes, tais como rede de estações sinópticas, Plataformas de Coleta de Dados (PCDs) e de

outros bancos de dados. Desta forma, uma outra característica deste subsistema é a flexibilidade na forma de inclusão dos dados, integrando processos de digitação, processamento batch, ou ainda interação direta entre bancos de dados.

O controle de qualidade dos dados, abrange vários níveis e difere conforme cada tipo de dado meteorológico e conforme sua origem. Neste processo, os dados com erros serão descartados automaticamente, as informações consideradas duvidosas, dentro dos critérios de consistência deste subsistema, serão incluídas como suspeitas e passam a aguardar validação por meteorologista responsável antes de ser armazenado em definitivo.

Serão armazenados dados processados em forma de valores médios, acumulados, extremos, entre outros, que serão processados e organizados de forma a facilitar as pesquisas por períodos diários, mensais, anuais e normais. Demais períodos, como por exemplo o sazonal, serão obtidos indiretamente através do agrupamento dos períodos pré - definidos.

Cada variável climatológica poderá ter vários valores relacionados a ela, em todos os períodos. Em primeira fase de operacionalização deste subsistema estarão disponíveis os valores relativos a precipitação, temperatura e umidade, conforme tabela 5

Tabela 5 - Principais variáveis do Subsistema Dados Climatológicos

Variáveis	Valores Relacionados
Precipitação	Valor acumulado
	Valor máximo
	Número de dias consecutivos com chuva
	Número de dias consecutivos sem chuva
	Número de dias com chuva não consecutivos
	Número de dias sem chuva não consecutivos
Temperatura Máxima, Mínima e Médias	Valores médios
	Valores máximos
	Valores mínimos
Umidade Relativa	Valores médios
	Valores mínimos

Na segunda fase de implementação, serão inclusos dados de vento, evaporação, pressão, radiação e nebulosidade, de forma análoga aos dados da primeira fase.

### 3.1 CARACTERÍSTICAS TÉCNICAS

O subsistema de dados climatológicos será implantado usando a estrutura computacional do CPTEC, utilizando o SGBD Oracle Sever versão 8i, ferramenta de desenvolvimento Pro\*C (pré-compilador) e Oracle Developer versão 6.0 (pacote com Forms, Reports e Graphics) e ferramentas de desenvolvimento para Web Oracle WebDB.

### 3.2 RESULTADOS ESPERADOS

Este subsistema deverá ser implantado no CPTEC no segundo semestre deste ano. A versão inicial irá ser disponibilizada com uma interface interativa baseada em formulários na forma de tabelas e gráficos apropriados, para Unix. E uma segunda versão será disponibilizada posteriormente para Web, utilizando sistema geo-referenciado.

## 4. PRÓXIMOS PASSOS

A implementação do Banco de Dados Meteorológicos do CPTEC deverá ser feita em etapas. A primeira delas foi concluída em 1999, quando o subsistema de dados observacionais entrou em regime operacional. Outro

subsistema, de dados climatológicos, se encontra em desenvolvimento, devendo ser operacionalizado ao final de 2000. Os subsistemas de produtos numéricos de modelos e de imagens de satélite ainda se encontram em fase de concepção, devendo ser liberados para uso operacional ao final de 2001. Estes dois últimos subsistemas exigem um sistema de armazenamento hierárquico, uma vez que são extremamente elevados os requisitos de espaço para armazenar estes tipos de dados.

## **5. REFERÊNCIAS**

Almeida, E.S.; Mello, F.O. Metview 2.0 - Um desktop meteorológico para tratamento e visualização de dados meteorológicos. Submetido ao XI Congresso Brasileiro de Meteorologia, Rio de Janeiro, 2000.

Dereczynski, C.; Cintra, R. Monitoramento de Dados Meteorológicos Observacionais. IX Congresso Brasileiro de Meteorologia, Campos do Jordão, 1996.

Nyfors, V. NEONS at the Finnish Meteorological Institute. Fifth Workshop on Meteorological Operational Systems, ECMWF, 1995.

Pottier, M. Operational Databases at Météo-France. Fifth Workshop on Meteorological Operational Systems, ECMWF, 1995.

Raoult, B. Architecture of the new MARS. Sixth Workshop on Meteorological Operational Systems, ECMWF, 1997.

Sanders, C. Data Management in the Australian NMC. Sixth Workshop on Meteorological Operational Systems, ECMWF, 1997.

Stanek, R. The Operational Database System of the Deutscher Wetterdienst (DWD) based on a commercial RDBMS. Seventh Workshop on Meteorological Operational Systems, ECMWF, 1999.

Tsui, T. Naval Environmental Operational Nowcasting System (NEONS). Third Workshop on Meteorological Operational Systems, ECMWF, 1991.

WMO. Manual on Codes, volume 1, International Codes, Part B - Binary Codes, com Suppl. no 6B (V.1994). World Meteorological Organization No. 306, Geneva.