1.4

OBSERVATIONAL DATA USED FOR ASSIMILATION IN THE NCEP NORTH AMERICAN REGIONAL REANALYSIS

Perry C. Shafran*, Jack Woollen
SAIC/GSO and NCEP/EMC, Camp Springs, MD

Wesley Ebisuzaki
NCEP/CPC, Camp Springs, MD

Wei Shi, Yun Fan
RSIS and NCEP/CPC, Camp Springs, MD

Robert W. Grumbine
NCEP/EMC, Camp Springs, MD

Michael Fennessy
COLA, Calverton, MD

1. INTRODUCTION

The National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR, Mesinger et al, this volume) assimilated a great deal of observational data. The data used was assimilated into the analysis, used as boundary conditions, or directly used during the execution of the Eta model. Here, the data usage and preparation are described in depth.

Table 1: Data Used in Global Reanalysis and Regional Reanalysis

Dataset	Observed variable	Source
Rawinsondes	Temperatures, wind, moisture	NCEP/NCAR Global Reanalysis (GR)
Dropsondes	Same as above	GR
Pibals	Wind	GR
Aircraft	Temperature and wind	GR
Surface	Pressure	GR
Cloud drift	Winds from geostat.sat.	GR

2. DATA USED IN GLOBAL REANALYSIS AND REGIONAL REANALYSIS

Much of the NARR input dataset was created during the preparation period of the NCAR/NCEP Reanalysis, also known as the Global Reanalysis (GR). Most, but not all, of that data was also used in the NARR.

Temperature, winds, and moisture from radiosondes were used in the NARR. Figure 1 shows a 00Z radiosonde distribution for a typical day. About 100-130 radiosondes were available for assimilation. Also included were dropsondes, instruments dropped from airplanes, that also measured temperature, winds, and moisture. Wind data was used from pibals. Commercial aircraft measured temperature and winds. While NCEP surface data was available in the GR, it only used the sea-level pressure in their assimilation, and thus it was available for usage in the NARR. Also, wind profilers have been available since 1990 and are assimilated from that time on. Vertical Azimuth Display (VAD) winds have also been available since about 1990 as well. Finally, cloud drift winds from geostationary satellites also were used during the creation of the analyses.

The dataset was basically the same as what was used in the GR. Therefore, it required only a small amount of preparation. To make the data useful for the NARR, though, the 6-hourly files had to be split into 3-hourly files. Also, data outside the Eta domain was cut from the file for easy file management.

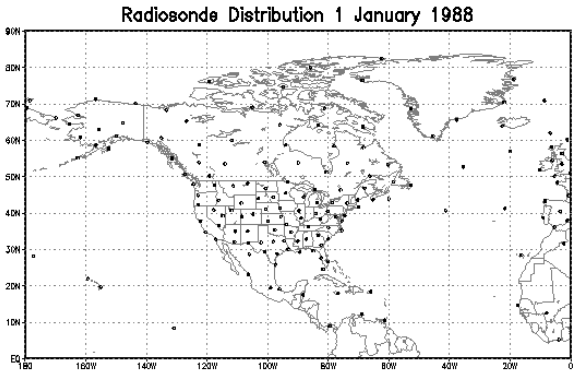

Figure 1: Distribution of radiosondes assimilated in the NARR 1 January 1988.

Corresponding author address: Perry C. Shafran, SAIC/GSO and NCEP/EMC, 5200 Auth Rd. Rm. 207, Camp Springs, MD 20746; email: Perry.Shafran@noaa.gov

Table 2: Data Added or Improved Upon for Regional Reanalysis (*, not assimilated)

Dataset	Details	Source
Precipitation, disaggregated into hours	CONUS (with PRISM), Mexico, Canada, CMAP Over oceans (<35N)	NCEP/CPC, Canada, Mexico
TOVS-1B radiances	Temperature, precipitable water over oceans	NESDIS
Surface	Temperature*, wind, moisture	GR
TDL surface	Pressure, temperature*, wind, moisture	NCAR
COADS	Ship and buoy obs.	NCEP/EMC
Air Force snow	Snow depth	COLA and NCEP/EMC
SST	1-deg. Reynolds, with Great Lakes surface temp.	NCEP/EMC, GR
Sea and lake ice	Includes data on Canadian, Great Lakes	NCEP/EMC,GLERL, Canadian Ice Center
Tropical cyclones	Locations used for blocking CMAP precipitation	LLNL

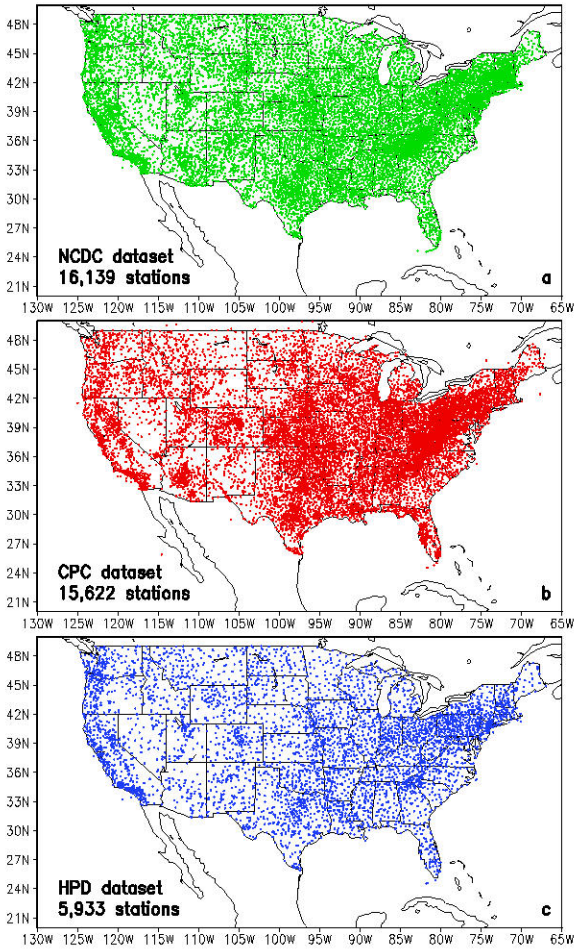

Figure 2: Total number of stations ever reported in a) NCDC cooperative dataset, b) RFC (CPC) cooperative data set, and c) HPD dataset. Figure from Higgins et al (2000).

3. PRECIPITATION

Precipitation is used to take advantage of the Eta's precipitation assimilation (Lin et al, 2001). Precipitation comes from several sources. Oceanic data comes from the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP). Land data comes from different sources for the Continental United States (CONUS), Mexico, and Canada. CMAP (Xie and Arkin, 1997) is a merged dataset of satellite-based data and rain gauges (Wei Shi, personal communication). It is a global 2.5-degree dataset that was interpolated onto the Eta grid. This particular dataset has reliable data only up to about 50 deg N, so no CMAP data is used there. To make sure that there is no sharp discontinuity, a blending was added to the Eta code to ease the influence of the CMAP data within a 15-degree zone centered at 35 degrees N.

The data is packaged in a pentad, and is disaggregated to an hourly dataset using hourly precipitation weighting based on GR precipitation. The CMAP dataset was available starting January 1979.

CMAP is not reliable in areas of very heavy precipitation (< 100 mm/day) or near the centers of tropical storms, so in those locations CMAP is filtered out and the EDAS is allowed to produce its own precipitation.

The precipitation data for the CONUS is a gage-based daily precipitation dataset analyzed on a 1/8-degree grid. The CONUS precipitation comes from a variety of sources: NCDC daily cooperative stations, River Forecast Center stations, and daily accumulations of the Hourly Precipitation Data (HPD) set. Fig. 2 shows all the datasets that have reported. In a typical day, the number of daily reports is 8000 NCDC, 7000 RFC, and 2500 HPD stations (Higgins et al, 2000). This particular dataset is analyzed using an orographic dataset known as the Parameter-elevation Regressions on Independent Slopes Model (PRISM, Daly et al 1994). This allows the effects of mountainous terrain to be more accurately analyzed for the RR. Also, a least-squares weighting scheme was used as well. The daily precipitation was disaggregated using the HPD.

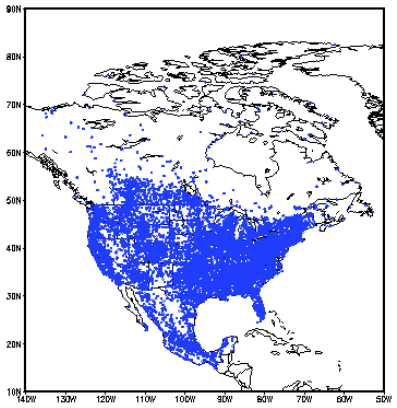

Figure 3: Distribution of precipitation observations for a typical day for CONUS, Canada, and Mexico.

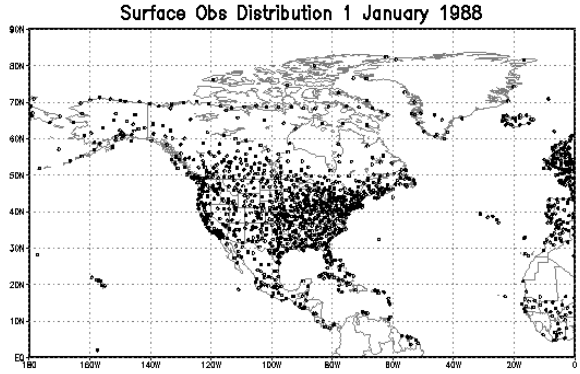

Figure 4: Distribution of surface observations assimilated in the NARR 1 January 1988.

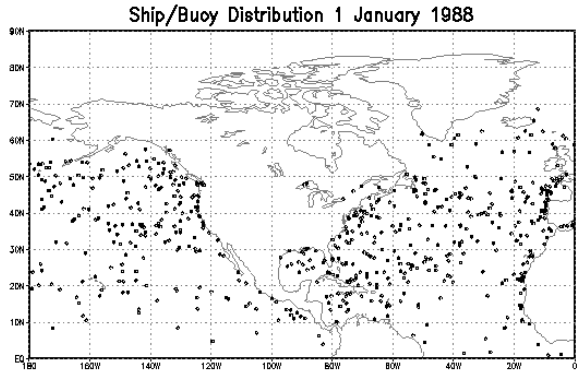

Figure 5: Distribution of ship and buoy observations assimilated in the NARR 1 January 1988.

Both the Mexico and Canada data were also daily gage-based datasets on 1-degree grids (E. Yarosh, personal communication). These datasets were disaggregated using the hourly precipitation from the GR. Then all four of the datasets were applied onto the Eta grid. The borders along Canada, the United States, and Mexico, were blended together to minimize the effect of boundaries.

4. ADDITIONAL DATA IMPROVED UPON FOR THE NARR

Several other data sources were included that were not used for the GR. Some of this data include radiances, additional surface data, ship and buoy data, snow, sea-ice, and sea-surface temperatures.

4.1 *Radiances*

Radiances, which provide winds and precipitable water data, mainly over oceans, comes from the National Oceanic and Atmospheric Administration's (NOAA) TIROS (Television Infrared Operational Satellite) Operational Vertical Sounder (TOVS, Kidwell 1995). The NARR utilized data from two instruments of this polar orbiter: the High Resolution Infrared Radiation Sounder/2 (HIRS/2) and the Microwave Sounding Unit (MSU). These satellites started receiving data in October 1978. The data for 1978-1997 was collected from the Satellite Active Archive (SAA) of the National Environmental Satellite Data and Information Services (NESDIS). These data were collected by scans and were converted into the format required for the NARR, in 3-hourly blocks. From 1998 to the present, the data comes from NCEP's operational run history tapes, already in the format required for usage in the NARR, except that the data had to be converted from a 6-hourly set to a 3-hourly set. This set was used instead of the TOVS retrievals that were assimilated in the GR.

4.2 *Surface observations*

Some land-surface observations were available for the Global Reanalysis but were not used by them. From the GR dataset the NARR assimilated 10-m winds and 2-m moisture values. The 2-m land temperatures from the GR dataset were not assimilated in the NARR because testing showed that the errors of 850-mb and 700-mb winds and temperature were worse when they were assimilated (Mesinger et al, 2004, this volume).

In addition, another surface dataset was provided by the National Centers for Atmospheric Research (NCAR). This dataset originated from the Model Development Laboratory (MDL) and was based on observations that have been taken since 1976 (R. Jenne, personal communication). The dataset was merged with the operational NCEP surface dataset. The obs were carefully checked against the existing surface dataset to ensure that there were no duplicates (J. Woollen, personal communication). When a duplicate was noted, the MDL observation was used and the NCEP observation was thrown out.

For the ocean, the NARR used an updated version of the Comprehensive Ocean-Atmosphere Data Set

(COADS, Woodruff et al, 1993). This dataset contains temperatures, winds, moisture, and pressure from ships. and buoys. The new dataset was finished up to 1997, so for the period 1998-present the GR existing ship and buoy dataset was used.

4.3 *Sea and lake ice*

Sea ice comes from a variety of sources. The main data over the oceans come from a satellite-based ice sensor that was interpolated onto the Eta 32-km grid (Grumbine 1996). This data is a sea-ice concentration between 0 and 100% of the gridpoint. For the NARR's purposes, if a gridpoint contained 50% or more sea ice, it was considered an ice point, otherwise, it was considered a water point.

For the Great Lakes, a digital ice dataset was available from the Great Lakes Environmental Research Laboratory (GLERL, R. Assel, personal communication) up to 1978-2000 and applied to the Eta grid. For 2001 and beyond, a 5-year climatology (1995-2000) was calculated and applied. Canadian ice data came from the Canadian Ice Service (CIS). The CIS provided data on a per-lake basis, not by gridpoint. A comparison to the Eta's land-sea mask and an atlas provided the gridpoints representative on each lake. This comparison listed the gridpoints for all the lakes in Canada, some which were not available by the CIS. For those available by the CIS, the value given for each lake was applied to every gridpoint on that lake. For those lakes not available by the CIS, the value of the nearest CIS-provided lake was applied to those gridpoints. For inland bodies of water that are near coastlines, the oceanic value was applied. The incorporation of ice from these three sources provided a complete history of ice for all water-based points.

4.4 *Sea-surface temperatures*

Sea-surface temperatures (SSTs) also come from a variety of sources. The main source is a 1-degree Reynolds dataset analyzed using a optimal interpolation algorithm, available from 1981 to the present (Reynolds et al 2002). Prior to 1981, the SSTs originate from a from a reconstructed SST dataset using COADS (Smith and Reynolds 2003). These datasets were interpolated between the Pacific and Atlantic Oceans. Inland lakes are not properly represented by this interpolation. Using the previously calculated ice data set, if a Canadian lake point is ice, it is given a value of 273.15 K, the freezing point of ice. If it is not ice, it is left alone. For the Great Lakes, the SSTs were provided by GLERL (R. Assel, personal communication) and applied to the Eta grid. Values for the Great Salt Lake and the Gulf of California were provided by climatologies. The use of climatological values for the Gulf of California was especially important in the southwestern United States to accurately analyze the monsoon cycle.

4.5 *Snow cover*

A snow cover data set comes from an Air Force snowfall set (Hall 1986). This dataset originated on a global 512x512 grid and was interpolated to the Eta grid over land. The snowfall from 1979-1999 was supplied by the Center for Ocean-Land-Atmosphere Studies (COLA). The data from 1999-present was available from NCEP's run history tapes.

4.6 *Tropical cyclones*

Finally, the locations of tropical cyclones were useful (Fiorino 2002). The CMAP precipitation dataset is not reliable near tropical cyclones, and the Eta model produces better precipitation near the cyclones. The only use of the tropical cyclones locations was to determine where to block the CMAP in their vicinity.

Table 3: Fixed Fields and Initial/Boundary Conditions

Dataset	Use	Source
Green vegetation fraction	Initialization of Vegetation	GR
Baseline snow-free albedo	Initialization of Albedo	GR
Sigma-level data	Lateral boundary conditions	GR
Surface binary files	Initialization of land states; option exists to initialize with NARR data	GR

5. FIXED FIELDS AND INITIAL AND BOUNDARY CONDITIONS

The baseline snow-free albedo and the vegetative fraction are fixed fields were updated daily during the NARR run. These were the same fixed fields that were used by the GR.

The lateral boundary conditions were supplied by the sigma-level data from the GR. These boundary conditions were interpolated to the Eta grid every 6 hours. The surface-level files were available with the option to be used to initialize the land-surface states.

6. REFERENCES

Daly, C., R.P. Neilson, and D.L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140-158.

Fiorino, M., 2002: Analyses and forecasts of tropical cyclones in the ECMWF 40-year Reanalysis (ERA-40). *25th Conf. on Hurricanes and Tropical Meteorology*. San Diego, CA, Amer. Meteor. Soc., 261-265.

Grumbine, R. W., 1996: Automated passive microwave sea ice concentration analysis at NCEP. DOC/NOAA/NWS/NCEP/EMC/OMB Technical Note 120, 13 pp.

Hall, S.J., 1986. AFGWC Snow Analysis Model, AFGWC/TN-86/001, AFGWC, Offutt AFB, NE, 23 pp.

Higgins, R.W., W. Shi, E. Yarosh, and R. Joyce, 2000. Improved US Precipitation Quality Control System and Analysis, NCEP/Climate Prediction Center Atlas No. 7.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S-K. Yang, J. J. Hnilo, M. Fiorino, and G.L. Potter, 2002: NCEP/DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc*., **83**, 1631-1643.

Kidwell, K.B, 1995: NOAA Polar Orbiter Data User's Guide, U.S. Dept. of Commerce, NOAA/NESDIS, Washington, DC, 410 pp.

Lin, Y., M.E. Baldwin, K.E. Mitchell, E. Rogers, and G.J. DiMego, 2001: Spring 2001 changes to the NCEP Eta Analysis and Forecast System: Assimilation of observed precipitation data. Preprints, *18th Conference on Weather Analysis and Forecasting* and *14th Conference on Numerical Weather Prediction*, Fort Lauderdale, FL, Amer. Meteor. Soc., J92-95.

Mesinger, F, G. DiMego, E. Kalnay, P. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, M. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, K. Mitchell, D. Parrish, E. Rogers, W. Shi, 2004: NCEP North America Regional Reanalysis. *15th Symp. On Global Change and Climate Variations*, Poster 1.1, Combined Preprints CD-ROM, 84th AMS Annual Meeting, Seattle, WA, Jan. 2004.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W.Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609-1625.

Smith, T.M., and R.W. Reynolds, 2003: Extended Reconstruction of Global Sea Surface Temperatures Based on COADS Data (1854-1997), *J. Climate*, **16**, 1495-1510.

Woodruff, S.D., S.J. Lubker, K. Wolter, S.J. Worley, and J.D. Elms, 1993: Comprehensive Ocean-Atmosphere Data Set (COADS) Release 1a: 1980-92. *Earth System Monitor*, **4**, 1-8.

Xie, P., and P.A. Arkin, 1997: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs, *Bull. Amer. Meteor. Soc*., **78**, 2539-2558.