### **Data Access and NOMADS Update**

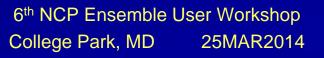
# Ensembles on NOMADS with Aggregation of the Data Set



Jordan C. Alpert

jordan.alpert@noaa.gov







# Address the Unprecedented Increase in Ensemble Model Data Production and Information Content

- A TB/day for the SREF and as much for GENS and a large number of images are generated to show Reliability, Probability, and Possibilities.
- A one year Archive would be a Peta byte. I have not mentioned Climate ensembles or Downscaling or increases in the current ensemble horizontal and vertical resolution.
- Can users continue to be asked to download peta-bytes of data?
- How can NWP and GCM data providers and associated data centers and archives begin to provide not only distributed access, but distributed web services capabilities from a semantic-web standpoint?

Prefixes

—
Kilo – 10<sup>3</sup>
Mega – 10<sup>6</sup>
Giga – 10<sup>9</sup>
Tera – 10<sup>12</sup>
Peta – 10<sup>15</sup>

 $Zetta - 10^{21}$ 



### Address the unprecedented increase in Ensemble model data production and information content (con't)

- Development of cross-discipline science studies linking data across massive databases to provide products for the attainment of knowledge and not only access to the raw data
- Allow researchers more time to spend on analyzing results and less time coding and worrying about file formats and data transfers.
- Improving services is a technology problem not a scientific one

#### Prefixes

Cilo – 10

 $Mega - 10^6$ 

Giga  $-10^9$ 

 $Tera - 10^{12}$ 

Peta  $-10^{15}$ 

 $Exa - 10^{18}$  $Zetta - 10^{21}$ 

# NOAA Operational Model Archive and Distribution System (NOMADS)

Designed to provide real-time and retrospective format independent access to climate, ocean and weather model data, and advance the integration of real time model data and applications:

- ✓ A digital archive of NOAA's operational weather models, and an "innovative data access philosophy to promote interoperable access across the geosciences" (BAMS, Rutledge et. al., 2006) and an integrator of common web services infrastructure to support the discovery, access and transport of data (NOAA GEO-IDE Concept of Operations Rept. to the DMC, 2005).
- "Completing the Forecast" (NRC, 2006) Recommendation 3.4: "NOMADS should be maintained and extended to include: (a) long-term archives of the global and regional <a href="Ensemble">Ensemble</a> forecasting systems at their native resolution, and (b) re-forecast datasets to facilitate post-processing."
- ✓ The real time data service is now at the (new) NCWCP Data Center
- ✓ Archives at National Climate Data Center
- ✓ Development continues at with servers and non-operational data sets for new applications and parallel testing of datasets.

## NOMADS: High availability servers from distributed locations using the same protocol to distribute model data.

- Other agencies use similar protocol to distribute environmental data, eg., GFDL, PMEL, NCAR, NASA... and collaborate through organizations like DMIT, GO-ESSP,...
- NOMADS open source services use:
  - http or fast/partial ftp access: inventory & client script
  - <u>ftp2/4u</u>/("GRIB filter"), slice, dice and area subset files delivering repackaged GRIB files
  - GDS/OPeNDAP(DODS), and TDS for values from http queries
  - pdisp ("Great Displays") development display program example, and other application examples.

#### **National Weather Service**

OTA SELL | S DILCCOLL | S LIGOTIGI ... | S LIGOTIGI ... | 🗻 OTAGO D... | 🔝 HOLILII

### **NCEP Central Operations**



Home News Organization Search Search

cast by

Go

Go

lazards s/Warnings

Conditions ations Images nagery Rivers Veather urface Analys

urface Analysis n sphere :e sis

Loops nental Models Info Status nalysis dance

ne

al Marine ather t Maps

Prediction Archives Salety teady

Library ibrary Affairs ment on urces

#### **NOMADS**

#### NOAA Operational Model Archive and Distribution System

Description of NOAA's NOMADS servers hosting NCEP model data

Users of NOMADS are reminded that they should use the URL http://nomads.ncep.noaa.gov to access the system and they will always be placed on the current active server. As of Tuesday October 7, 2009, users that have been using direct IP addresses to access NOMADS systems may no longer be able to access the system.

Help Desk: Questions or problems please use the link to submit a service ticket.

Background Background documents about the NOMADS project.

Service Description: OCWWS Service Description Document

Click on link in the Data Set field for description and availability info.

Click on the column headings for description of each data access method.

Data Set	freq	grib filter	http	gds	
Global Models					
FNL	6 hours	grib filter	http	OpenDAP	
GFS 1.0x1.0 Degree	6 hours	grib filter	http	OpenDAP	
GFS 0.5x0.5 Degree	6 hours	grib filter	http	OpenDAP	
GFS 2.5x2.5 Degree	12 hours	grib filter	http	OpenDAP	
GFS Ensemble high resolution	6 hours	grib filter	http	OpenDAP	
GFS Ensemble Precip Bias-Corrected	daily	grib filter	http	OpenDAP	
GFS Ensemble high-resolution Bias-Corrected	6 hours	grib filter	http	OpenDAP	
GFS Ensemble NDGD resolution Bias-Corrected	6 hours	grib filter	http	OpenDAP	
NAEFS high resolution Bias-Corrected	6 hours	grib filter	http	OpenDAP	



#### All NOMADS Holdings have a MetaData Description

#### Man or Machine readable

GrADS Data Server - info for /gens\_bc/gens20110511/gep\_all\_00z : dds das

OPeNDAP/DODS Data URL: http://nomads.ncep.noaa.gov:9090/dods/gens\_bc/gens20110511/gep\_all\_00z

Description:

bias corrected GEFS member fcsts starting from 00Z11may2011, downloaded May 11 05:40 UTC

Documentation:

Longitude:0.0°E to 359.0°E (360 points, avg. res. 1.0°)

Latitude:-90.0°N to 90.0°N (181 points, avg. res. 1.0°)

Altitude:1000.0 to 10.0 (10 points, avg. res. 110.0)

Time:00Z11MAY2011 to 00Z27MAY2011 (65 points, avg. res. 0.25 days)

Ensemble:1 to 21 (21 points) Variables:

(total of 13)

hgtprs

\*\* (1000 925 850 700 500.. 250 200 100 50 10) geopotential height [gpm]

Data location, Description, Extent, Variables, Units..., Geo-spatial,

# Using http queries (URLs) to extract data from the GDS server Example: Aggregated Global Ensemble data set

#### ...DODS/OPenDAP/GDS constrained query:

http://nomads.ncep.noaa.gov:9090/dods/gens/gens20090501/gep\_all\_00z.ascii?tmpprs[0:19][0:21][1:1][129:129][243:243]

http://nomads.ncep.noaa.gov:9090/dods/gens/gens20110506/gep\_all\_00z.ascii?tmin2m[0:20][0:21][130:130][255:255]

#### Notice the 5-Dimensional query for each variable (6-D data cube!), e.g, temperature (tmpprs):

<i>i</i> <sup>th</sup> Ensemble component	[0:20]	Lists all 20 Ensemble components at,
Forecast times	[0:21]	IC and every 6-hour interval to 5-days is indicated,
Vertical levels	[1:1]	975Mb indicated, and [0:0] would mean 1000mb,
Latitude,	[129:129]	is measured from SP (0) to NP (for a 1 degree grid)
Longitude,	[243:243]	Beginning at the 0 meridian - we show Baltimore Intl

The ordering of the square bracketed values:

[Ens1:Ens2][Fcst1:Fcst2][lev1:lev2][lat1:lat2][lon1:lon2]

Units and other information represented in the metadata descriptor file and a stride is also possible [start:stride:finish] using colon separated values

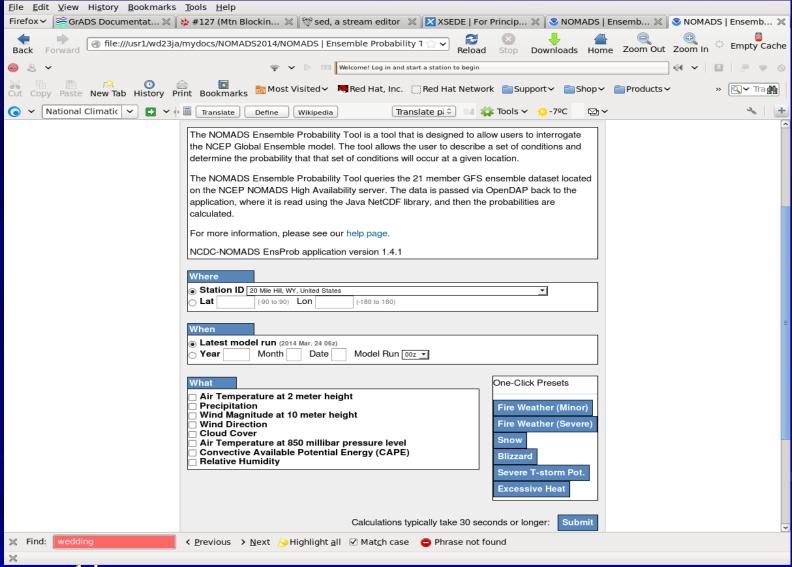
Use a non-interactive web download program like "wget" or cURL, and place the URL in cron scheduler and a cgi-bin script to provide user interaction or obtain information for customized data flow.

### TWO Applications

Global Ensemble Probability and the Event Probability Tool

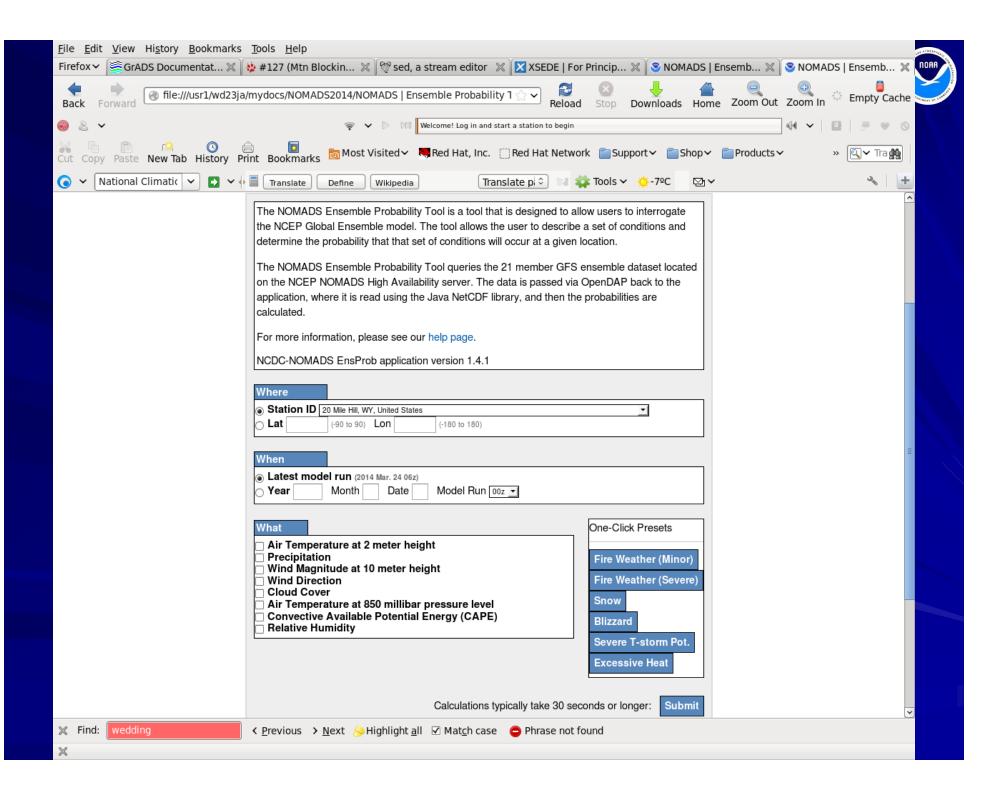
## Screen shot of a web page containing user prompts for defining the probability of a weather event from Global Ensembles.

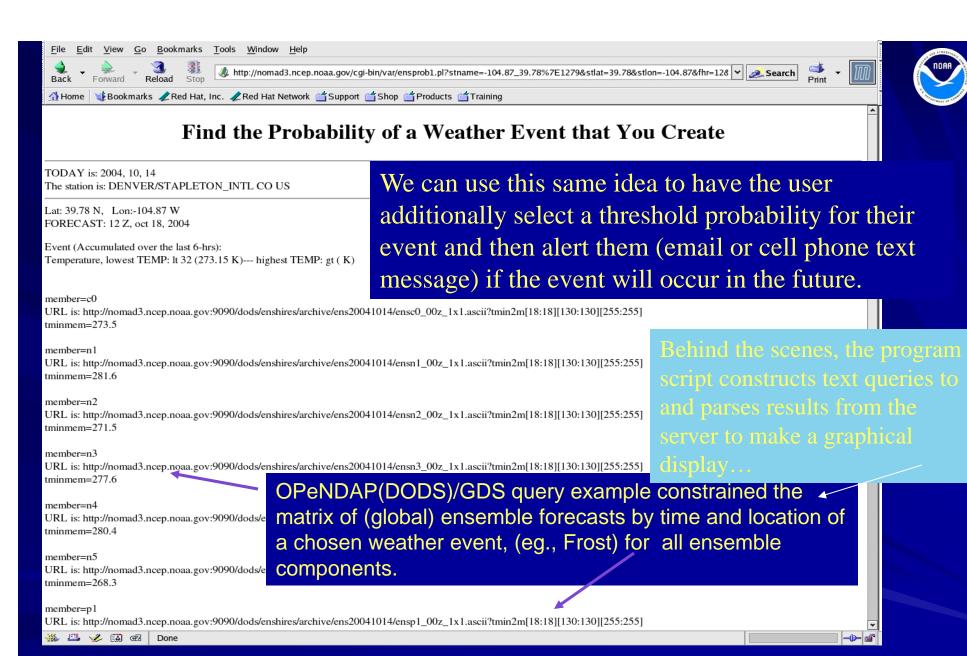




One can try this at

(http://nomads.ncdc.noaa.gov/EnsProb)





The Ensprob tool can be used to hone the probability threshold value for the users chosen event with a display of the above informatin on the next slide.... NOMADS

Lat: 39.78 N, Lon:-104.87 W FORECAST: 18 Z, oct 15, 2004

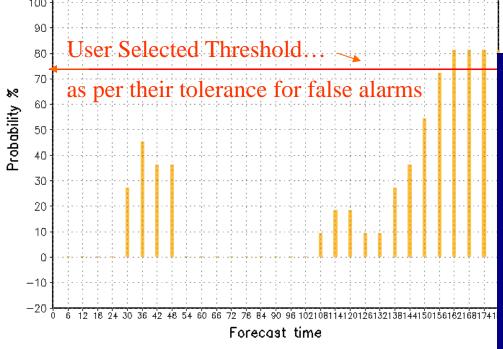
Event (Accumulated over the last 6-hrs): Temperature, lowest TEMP: lt 32 (273.15 K) -

errcnt=0

The application delivers ensprob program "code" in the form of the returned URL http://etc... address for the user to repeat the action by copy-pasting it into any browser or...

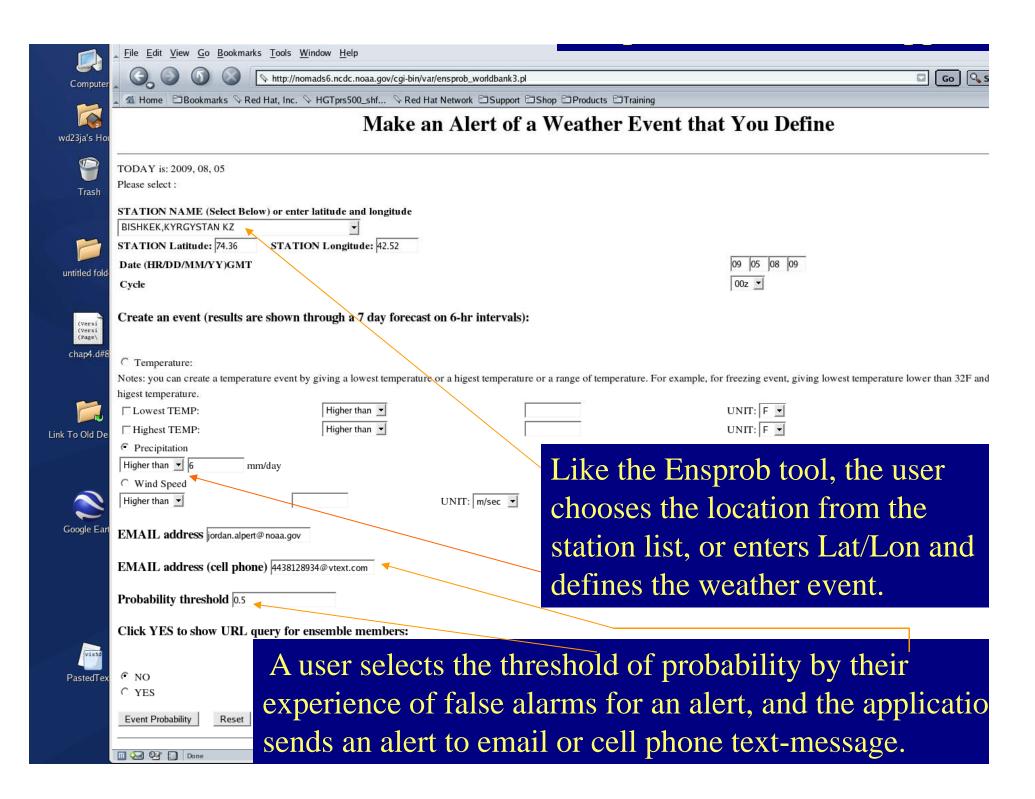
🚟 🕮 🅓 🖼 🗹 Done

Ensemble probability tool: A web based client application that can be used to provide threshold information to the user. The program obtains the Global Ensemble Forecast information matrix from the server, and returns the information to the user as a display (below) to allow the user to determine a proper threshold of their user defined weather event.



User determines their threshold (to tolerate false alarms) for an alert, the application sends an alert to email and cell phone textmessage when the threshold is met (next slide).

The user can re-issue this "check for an alert" automatically from a scheduler like cron using a non-interactive web download command like "wget" with the returned URL.



### Result from the World Bank Example App

NORA NORAN STATE OF THE PARTY O

----- Original Message ------

Subject: ALERT WEATHER EVENT:

Precip: gt 5 mm/day

Date: Fri, 31 Jul 2009 16:01:39 -0400From: Apache <apache@noaa.gov>

Precip: gt 5 mm/day > 50%, chance @ ft=

2009 aug 01 18Z

24 hr fcst

Information and communications technologies (ICT), in particular, high-speed internet connections, content servers, and mobile application services are transforming public service delivery of critical information and democratizing innovation (McNamara, K. S., "Workshop on Mobile Innovations for Social and Economic Transformation", World Bank, September 16, 2009).

Mobile platforms (cell phones) are emerging as the single most powerful way to extend ICT opportunities and key services to millions of people. Poor access to information and communication is an intangible dimension of rural poverty.

Our goal is to increase smallholder productivity and incomes by reducing uncertainty for variables of interest of agricultural importance. In this case we use forecasts from National Weather Service (NWS) operational Global Forecast System (GFS) global spectral model ensemble to alert agriculture users if and when their selected weather events will occur. The opportunity and challenge is to frame the response as useful and understandable answers to the questions that need to be answered. Verification to identify practical lead times for the economically important variables is a needed first step. (See AGU Poster U21A-0007).



### Summary

- Real Time NOMADS now at the NOAA Center for Weather and Climate Prediction (NCWCP) Data Center
- GRIB(2) filtering (ftp4/2u), Http (Fast or partial ftp) Scalable and distributive....
- OPeNDAP services using GrADS Data Server and TDS for NCEP suite of model operational data in real time or archives at NCDC are available so the data can appear like a local file.
- Global Ensemble data matrix, is available across space, time including *i*<sup>th</sup> component for all forecast times, variables, levels and location with one server query, virtually a 6-D data cube
- Aggregation of data sets is the key for users to obtain the data they need to make time critical decisions for their own projects
- Application examples are shown for users to construct probabilities from NCEP GFS (global) Ensemble matrix of data to obtain, eg., threshold information for alerts through email and cell phone SMS-messages



### Future NOMADS

- High resolution operational model datasets: Ensembles that are calibrated (bias corrected) and aggregation of Global and Regional data sets
- BUFR/NETcdf files from Operational Conventional and Non-conventional Observations, an NCEP value added product, to be served in a similar way to grid data sets
- NCEP catalog/aggregation servers, THREEDS/TDS
- Continue development with NCDC archive and NCEP real time model data and observations for seamless access of data and to promote useful applications. (CFS and ensemble archives)



# NOMADS Servers for NCEP Model Data

(High availability Severs)

- http://nomads.ncep.noaa.gov (Data)
- http://nomads.ncdc.noaa.gov (Archives)

### (Development)

- http://nomad1.ncep.noaa.gov (Data/Applications) http://nomad3.ncep.noaa.gov (Data/Applications)
- http://nomad5.ncep.noaa.gov (Data/Applications)

Global ensemble probability threshold alert to a cell phone text message: <a href="http://nomads-cloud.net/cgi-bin/ensprob\_worldbank4.pl">http://nomads-cloud.net/cgi-bin/ensprob\_worldbank4.pl</a>

Global Ensemble probability event display: http://nomads.ncdc.noaa.gov/Ensprob

## Background Slides

### Model Forecast Ensemble as a Tool for Decision Support Additional slides for Poster IN33B-1550

Example Script program:

To Get Data from NOMADS GDS/Opendap service

### Compose a Query for the ensemble database

- Each Global Ensembles file is a separate file for each component and future forecast time
- The model output is aggregated over forecast time and ensemble component so there is one entry.
- Each Grib2 file contains hundreds of variables over vertical levels and global latitude/longitude

# Assumptions using Global Ensembles on a 1x1 degree global grid for this example:

- We start with the temperature, tmpprs, from aggregated file entry, gep\_all\_00z:
  - Of the 21 ensemble components, to start we will get the 1st (Control), [0:0], of 20 components
  - Of the 64 forecast times on 6-hour intervals, out to 16 days, we will get the initial condition, f00, or [0:0]
  - Of the 26 vertical levels, to start we choose the 2<sup>nd</sup>, 975mb
     [1:1]
  - One or more points are possible for this example but we will pick one point, so the printout fits on a slide
  - We choose SFO airport at (38N,-122W) = 128 degrees
     North of the South Pole counting from zero, and 238 degrees counting from 0 Meridian so our point is [128:128][238:238]

### **Constrained Query**

http://nomads.ncep.noaa.gov:9090/dods/gens/gens20131208/gep\_all\_00z.ascii ?tmpprs[0:0][0:0][1:1][128:128][238:238]

If typed into any browser this will return:

tmpprs, [1][1][1][1][1] [0][0][0][0], 277.2

ens, [1] 1.0 time, [1] 735211.0 lev, [1] 975.0 lat, [1] 38.0 lon, [1] 238.0

\_\_

First the variable name is returned along with its array structure, the first element of the 5D single point array has a value of 277.2 Kelvin. The first (1.0) ens is found. The time indication is in Rata Die, the number of days from the year 1, counting from zero. For 20131208 00Z it is 735211.0 The lat/lon of SFO are returned and confirmed.

### **Constrained Query Script**

A Unix script can be made to obtain the temperature value using this URL and the non-interactive web download command from above:

http://nomads.ncep.noaa.gov:9090/dods/gens/gens20131208/gep\_all\_00z.ascii ?tmpprs[0:0][0:0][1:1][128:128][238:238]

```
# --- program to get a value
ANSWER=`wget -O -
    "http://nomads.ncep.noaa.gov:9090/dods/gens/gens20131208/gep_all_00z.ascii?tmpprs[0:0][0:0][
    1:1][128:128][238:238]"`
VAR=`echo $ANSWER|cut -f1 -d" "`
echo " ***** The variable is $VAR "
TEMP=`echo $ANSWER|cut -f4 -d" "`
ENS=`echo $ANSWER|cut -f7 -d" "`
TIME=`echo $ANSWER|cut -f10 -d" "`
LEV=`echo $ANSWER|cut -f13 -d" "`
LAT=`echo $ANSWER|cut -f16 -d" "`
LON=`echo $ANSWER|cut -f19 -d" "`
echo " ***** TEMP=$TEMP Kelvin "
echo " ***** ENS=$ENS "
echo " ***** TIME=$TIME which is 20131208 00Z "
echo " ***** LEV=$LEV mb "
echo " ***** LAT=$LAT degrees North "
echo " **** LON=$LON degrees East "
exit
```

#### Executing the script gives...

```
./AGUtest1.sh

***** The variable is tmpprs,

***** TEMP=277.2 Kelvin

***** ENS=1.0

***** TIME=735211.0 which is 20131208 00Z

***** LEV=975.0 mb

***** LAT=38.0 degrees North

***** LON=238.0 degrees East
```

**--**

Note that only one query to the server is needed, all the information is in "ANSWER". The values are then "cut" out of "ANSWER" to be used in subsequent calculations. One could also use sed and awk commands or other languages such as Perl and Python. Fortran can also be used with a call to system. We constrained this example to one temperature value but all the dimensions may be varied using the square bracketed template with colon separated values, each dimension as noted in the metadata description:

[Start\_value:stride:End\_value]

### Query Generation of a 2D field

Instead of one temperature value we could have downloaded the entire set of ensembles at SFO Intl airport. A URL for the model output variable, minimum surface (2m) temperature, tmin2m, (over 6-hour interval) for all the ensembles [0:20] and a week of forecasts on 6-h intervals [0:28]:

http://nomads.ncep.noaa.gov:9090/dods/gens/gens201312 09/gep\_all\_00z.ascii?tmin2m[0:20][0:28][128:128][238:2 38]

Returns a 21 X 29 = 609 values from as many separate packed binary GRIB2 files, in this case for the model forecast of minimum temperature, a 2D field (no vertical layers are specified), all the information for probabilities.

### The Alert Threshold

The URL was generated by the

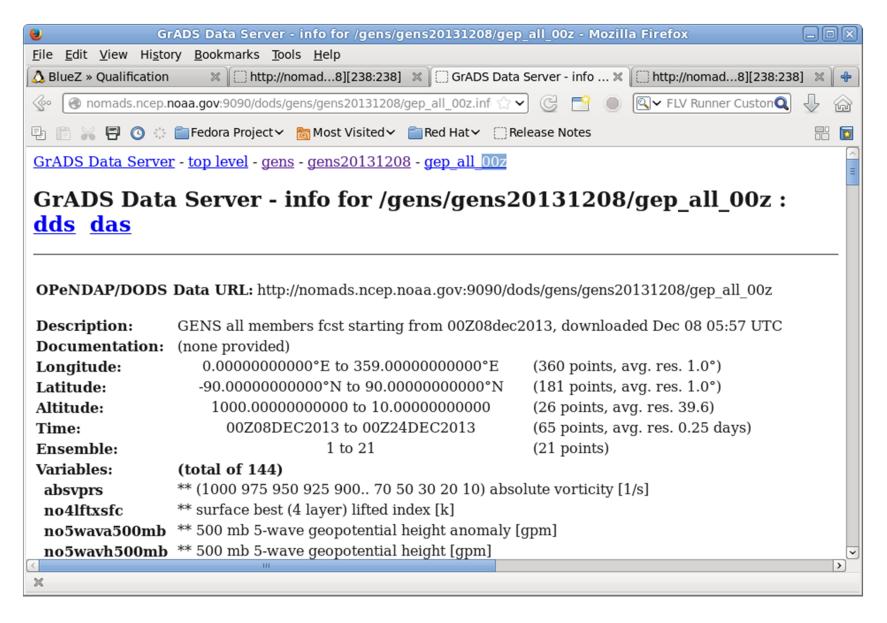
http://nomads-cloud.net/cgi-bin/ensprob\_wb4.pl

program which calculated the probability of the SFO temperature falling below say 35F (274.8K).

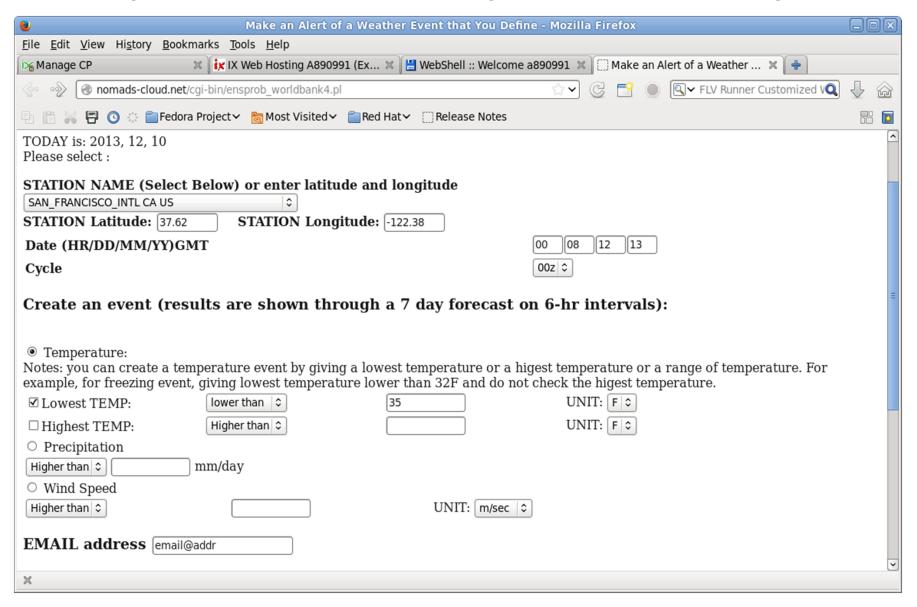
The program "cuts" or counts the number of ensemble forecasts that exceed the criteria, that is, temperature falling below 35F, and counts the number of ensemble components, the hits, dividing by the total number of forecasts to obtain a probability of below 35F weather event.

A user supplied threshold of what probability will constitute an alert is checked and if and only if the alert threshold criteria is met then the user is signaled with email and a SMS text message to the users cell phone number.

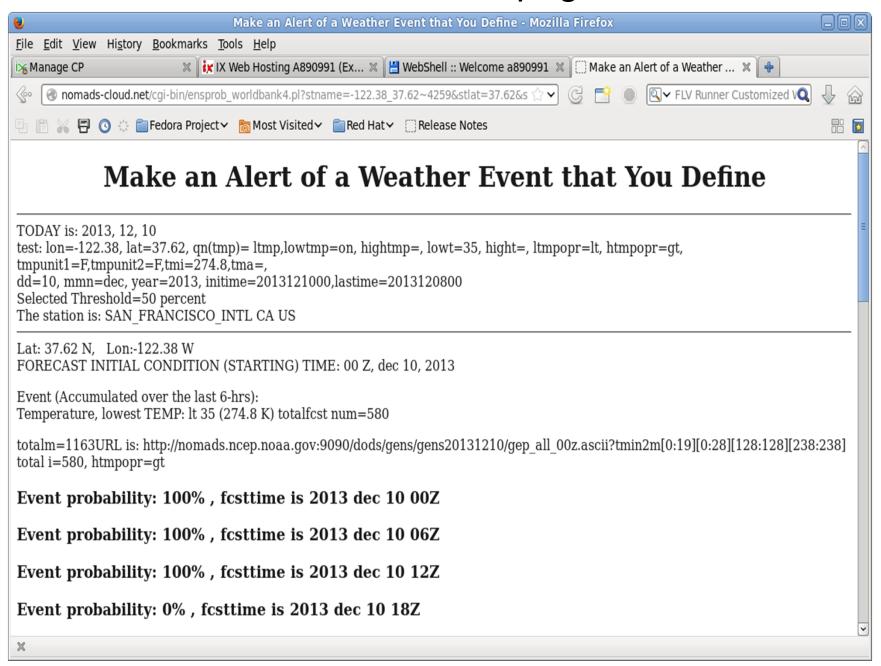
### NOMADS GDS OpeNDAP Global Ensemble Metadata Excerpt from 20131208



### http://nomads-cloud.ensprob.worldbank4.pl



### Result from alert page



### Returned SMS text message command The message received by cell phone in Red

mail -s "ALERT WEATHER EVENT:" Temperature, lowest TEMP: It 35 (274.8 K) email@addr < Temperature, lowest TEMP: It 35 (274.8 K) > 50%, chance @ ft= 2013 dec 10 00Z 0 hr fcst 2013 dec 10 06Z 6 hr fcst 2013 dec 10 12Z 12 hr fcst 2013 dec 11 12Z 36 hr fcst (PLEASE DO NOT REPLY TO THIS EMAIL ADDRESS), snd=1

### The URL Generated by the Alert program

http://nomads.ncep.noaa.gov:9090/dods/gens/gens20131210/gep\_all\_00z.ascii?tmin2m[0:19][0:28][128:128][238:238]

\_\_

tmin2m, [20][29][1][1]

[0][0][0], 9.999E20

[0][1][0], 273.59

[0][2][0], 273.12

[0][3][0], 273.15

.

.

.

Continue on next slide ...

#### Returned values from URL (excerpt continued)

# NOAA Web Operations Center (WOC) NOMADS Commitment (Beginning Feb 15, 2009)

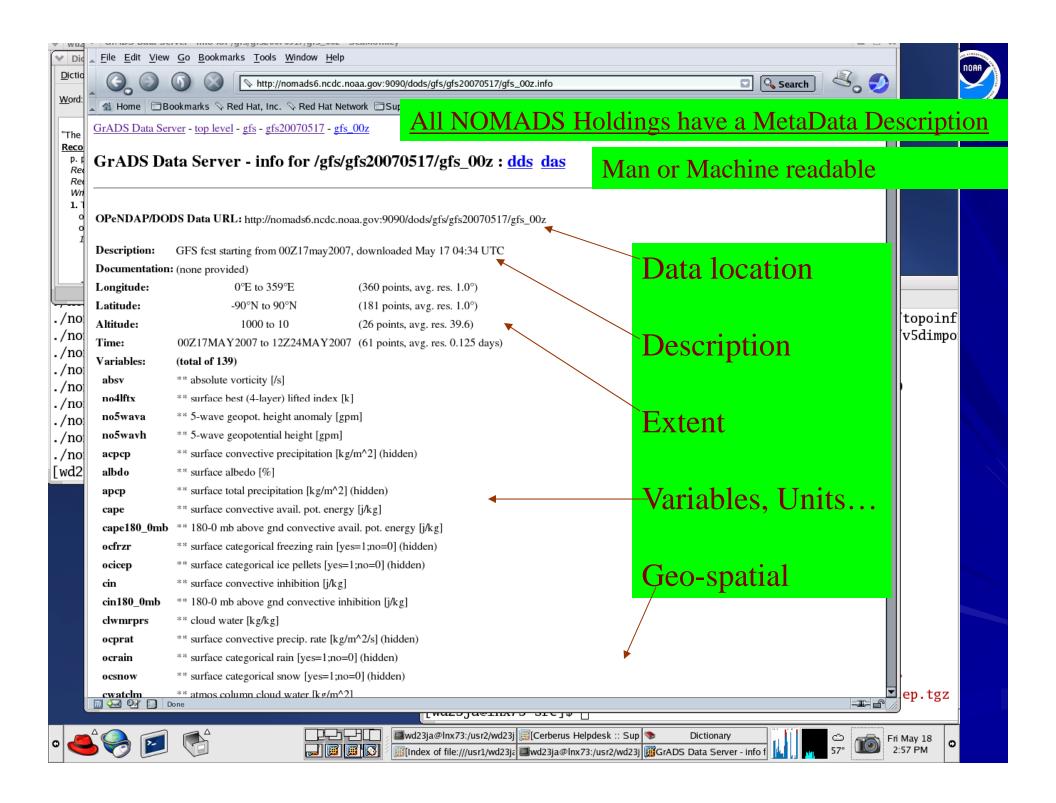
- ✓ NOAA has committed that the Web Operations Center (WOC) is high availability, 24/7 operations
- ✓ NCEP Central Operations (NCO) is committed to the data flow aspect of NOMADS for data to be present and on time from their Operational super computers.
- ✓ NOAA WOC has committed to maintain NOMADS servers now and into the future, as well as day to day operations with costs shared by NOAA and NCEP base
- ✓ A development commitment continues at NCEP Environmental Modeling Center (EMC), a development division, to keep up with new data sets and create applications
- ✓ Data review groups, official committees and procedures for moving new data sets and applications from development to operations follows the existing NCO framework.



### The GrADS-Data Server (GDS) OPeNDAP(DODS):

#### Open Source

- NOMADS participants serve their data sets through a client-server relationship. The data sets have machine and man readable metadata descriptions.
- Display is done by the client.
- GDS combines both GrADS, a freeware client (from COLA) and DODS (OPeN-DAP) server to unpack, cache and exchange data from many formats using http in response to user queries.
- This means that server data can appear to the user or client application as a <u>local file!</u>
- DODS requests are made by many freeware and commercial high level language clients like GrADS and MATLAB.
- http queries to the DODS server can create value added products in addition to the basis of scientific work for the public, federal agiences, etc.
- The aggregation of separate grided binary (GRIB2) files over forecast time and ensemble component is part of NOMADS



# Global Ensembles and Event Probability Tool (see Alpert & Wang, 2005, AMS 21 IIPS 17.5)

- NCEP ensembles are constructed from many (20) model integrations by slightly changing the initial conditions.
- Ensembles attempt to span the space of possible forecasts and ensemble members are equally probable forecasts if not, we can apply a suitable normalization.
- Probability estimates can be defined simply as the percentage of forecasts that satisfy a specified weather event over the total sample space (total number of components).
- We apply this to weather elements like surface temperature, wind speed, or precipitation at a location, at future model forecast times.