Accuracy of Weather Forecasting

Question: What is the accuracy of a weather forecast as the forecasting date is further into the future.

Data: Historical Weather forecasts for 2013 and 2014 for various cities in the continental U.S. Data will be predicted daily high and low temperatures and predicted probability of precipitation. Data will vary by how many days ahead is the forecast for. Data will also include the actual weather data for each forecasted date.

Availability of Data: It is easy to historical actual weather conditions data, and it is easy to find historical weather model data (forecasts but different – more technical – numbers, more about that later). But after much searching, I found: U.S. DOC[1] > NOAA > NWS > NCEP > WPC has archived weather forecasts (www.wpc.ncep.noaa.gov and then click Product Archive). The most promising seems to be the medium range forecasts (3 – 7 days ahead). Unfortunately, the data is in the form of a graphic (.gif) with the numbers embedded in the graphic.

Maximum temperature for 6/17/15 predicted 7 days before
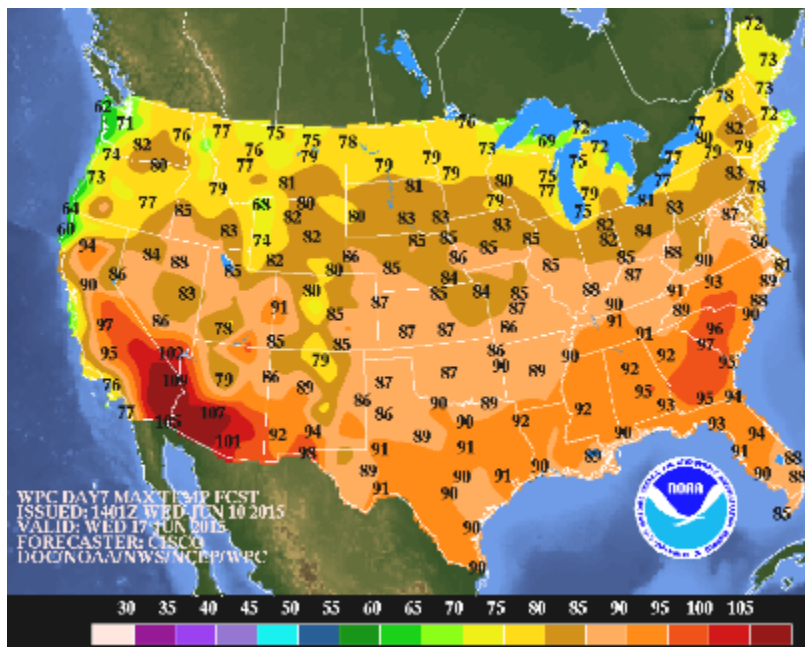
DAY7_MAX_2015061012_filled.gif

Probability of Precipitation for 6/17/15 predicted 7 days before
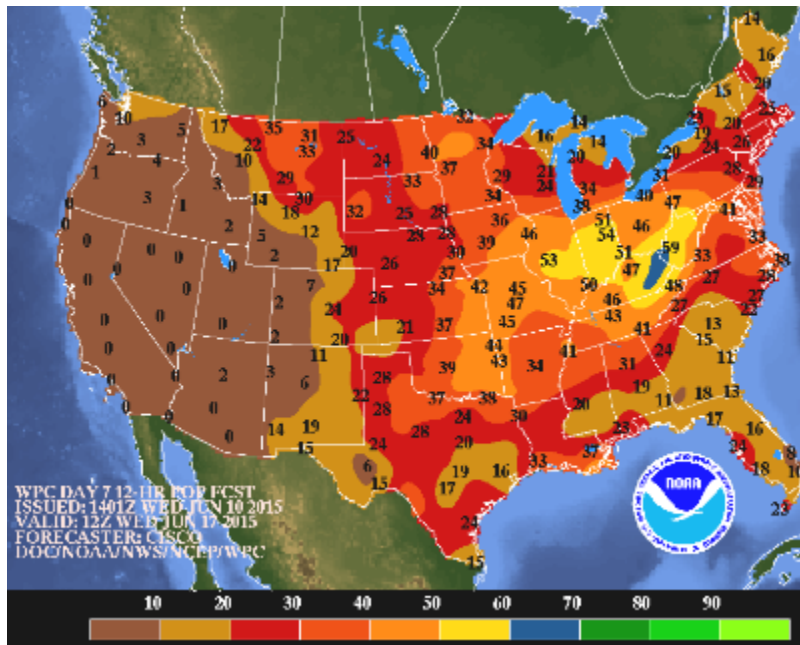
DAY7_POP1_2015061012_filled.gif

Strategy to obtain the data: The image files are actually twice as big as shown here. The files have systematic naming convention. So write a unix script using the curl utility to download ~14,000 files (~100KB each, 1.4GB total for 2 years).

Strategy to obtain the numerical data: The image files (at least the ones I looked at) are all the same format. The numbers always appear in the same place and are always the same font, color and size for each city. The (black) numbers overwrite all other colors. But gif format is compressed so use a utility to convert to bmp format. Then write python script to look at the pixel data in specific locations in the bmp file for patterns of black color for digits 0 – 9, and minus sign. (Poor man's OCR).

Data Integrity: The technique would be to extract the numbers from the image files only if what is found is exactly what is expected, leaving a null value otherwise. Then interrogate the numbers for missing values and also for obviously erroneous values. Try to fill in the missing values.

Questions Possibly Answered: I think comparing the forecasts with the actual weather conditions will be able to answer the question of how forecast accuracy varies as the forecast date is further into the future. Other questions possibly answered: Is the inaccuracy random or systematic? Does it vary by city? Does it vary by city's surrounding geography (mountainous, flat plains, coastal)? Does it vary over the course of a year or by season? Does it vary by type of forecast (rainy or sunny day, unusually hot or cold day – I would need to obtain average

max/min temperature per city/day of year)? Is it wrong only temporally (e.g. heat wave came but a day late)?

Modelling: *KNN* - for a given forecast time period, try to find predictors (city, season of the year, type of forecast) that reliably classify the forecast accuracy as good, OK, or bad. *Linear Regression* – Fit a line or curve to forecast accuracy verses forecast time period. *Visualization* – For example, for the day 7 forecast, a box plot of forecast error of daily maximum temperature verses city.

Background Weather Forecast Information: Various weather models (e.g. GFS) produce technical weather forecast data. Complicated data science is done (e.g. MOS) to produce weather forecast data in a form understandable to "the person on the street". (I've learned just a very little bit about weather forecasting; my data science is looks at the understandable forecast data).

[1]Acronyms:

```
CDO    Climate Data Online
ConUS  Continental U.S.
DOC    Dept. of Commerce
DWM    Daily Weather Map
DWML   Digital Weather Markup Language
EMC    Environmental Modeling Center
GCM    Global Climate Model
GFS    Global Forecast System
GIS    Geographic Information System
KML    KeyHole Markup Language
MDL    Meteorological Development Laboratory
MOS    Model Output Statistics (model data -> weather forecast)
MPD    Mesoscale Precipitation Discussion
NCDC   National Climate Data Center
NCEI   National Centers for Environmental Information (née NCDC)
NCEP   National Centers for Environmental Prediction
NDFD   National Digital Forecast Database
NESDIS Nat'l Environ'l, Satellite, Data and Information Service
NHC    National Hurricane Center
NMC    National Meteorological Center
NOAA   National Oceanic and Atmospheric Administration
NOMADS National Operational Model Archive and Distribution System
NWP    Numerical Weather Prediction
NWS    National Weather Service
OCR    Optical Character Recognition
PQPF   Probabilistic QPF
QPF    Quantitative Precipitation Forecast
SPC    Storm Prediction Center
SRRS   Service Records Retention System
WFO    Weather Forecasting Office
WPC    Weather Prediction Center
WRH    Western Regional HQ
```