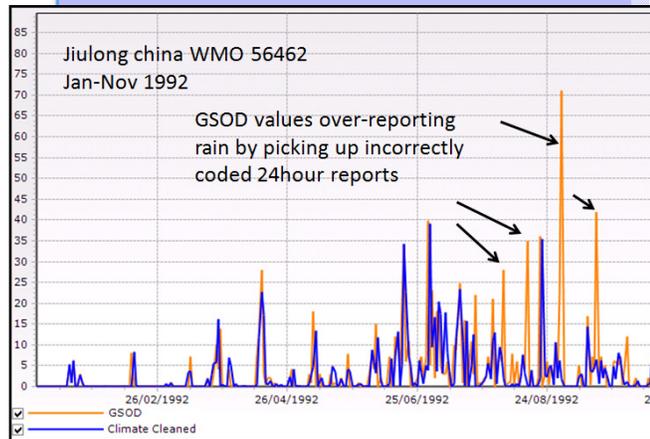




Understanding the limitations of the GSOD data set

Users of weather data are likely familiar with a global data set available from NOAA known as GSOD or “Global Summary of Day”. This is an important data resource and is freely available. The usability of this data varies depending upon the intended analysis to be performed. In certain circumstances the data can be very useful while in others it needs to be used with caution. The difficulty is not knowing when the data *can* be trusted. We caution that this data **should never be relied upon** to price weather risk contracts and should be avoided where possible when analyzing weather risk for commodity transactions, crop yield models, as well as other sensitive analyses.

China: Cleaned Climate Data vs. GSOD



Shown above is daily precipitation for Jiulong, China for 1992. The blue line is Speedwell Cleaned Climate data which is overlaid on top of GSOD data in orange. **In this period GSOD over estimates rainfall by 24%.**

Anqing WMO 58424 Annual Precip Totals

| Year | GSOD (mm) | Cleaned Climate (mm) | GSOD Error % |
|------|-----------|----------------------|--------------|
| 2000 | 1,189.23 | 1,119.50 | 6% |
| 2001 | 1,190.75 | 1,211.60 | -2% |
| 2002 | 1,443.48 | 1,319.80 | 9% |
| 2003 | 1,502.66 | 1,405.80 | 7% |
| 2004 | 1,227.58 | 1,133.70 | 8% |
| 2005 | 1,412.75 | 1,290.90 | 9% |
| 2006 | 1,106.68 | 1,060.50 | 4% |
| 2007 | 1,433.58 | 1,310.90 | 9% |
| 2008 | 1,146.56 | 1,079.90 | 6% |
| 2009 | 1,629.16 | 1,422.90 | 14% |
| 2010 | 2,289.05 | 2,044.10 | 12% |
| 2011 | 1,254.25 | 1,117.50 | 12% |
| 2012 | 1,907.29 | 1,566.10 | 22% |

Looking at another site in China, Anqing 58424 we see that these problems are found in recent years as well as older history with the overestimate (GSOD less Climate) in 2012 of 22%.

What is GSOD?

GSOD is an important and wide-ranging archive of global daily weather variables such as daily maximum and minimum temperature and rainfall. This dataset is a derived dataset that is computed from archived daily & sub-daily data reported on the global telecommunications system (and other sources) and archived by NOAA. Daily observations are estimated based on what is known about a given day. For example, daily rain may consist of one 24 hour report, or the sum of 12 hour reports or the sum of available 6 hour reports. Likewise, daily temperatures will be based on available hourly reports.

Why is this important?

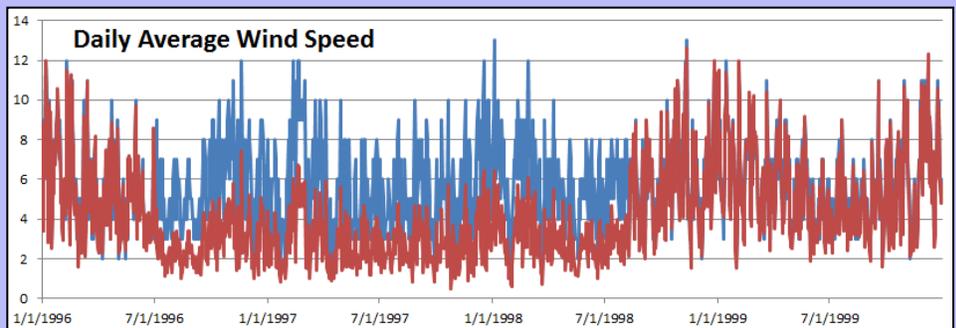
- The underlying data is frequently incomplete. In this case the data point can only be a “best estimate” based on the available data.
- The data used may have been mis-reported in the original synoptic report
- Sometimes the reported data is archived incorrectly before being processed into the GSOD archive

About Speedwell...

Founded in 1999 and with offices in the USA and UK, Speedwell supplies historical and real-time weather data for tens of thousands of weather stations around the world. We have direct supply agreements with a wide range of national meteorological services and provide a single-point of contact for the provision of high quality weather data and forecasts to the financial markets.

Ireland: Comparison of Climate Data

Shown below is the daily wind for Dublin Airport, Ireland. The blue line is official quality controlled data from the Irish Met Office. The red is GSOD. We can see that there is a two-year period where the GSOD data is incorrect. Minor differences are found in other years.



A few things to keep in mind

- Much of the data supplied freely on the internet (and by some vendors) is frequently based upon GSOD data. 
- No weather data set is perfect and *sometimes* in the absence of all other possibilities imperfect data **may** be better than none. Our approach is that **if we do not understand the provenance of a data set we do not deliver it**. For the reasons shown, we at Speedwell do not deliver GSOD data to our clients unless it is specifically requested.

Doing a better job

- It is necessary to have a complete understanding of weather data, how it is reported, archived, and how it should be interpreted.
- Source as much data as possible from national met offices (or from vendors that do).
- **Use Quality controlled data**

Info@SpeedwellWeather.com