

Speedwell Weather



**Introduction to Weather
Data Cleaning**



Speedwell Weather Limited

An Introduction

- Providing weather services since 1999
- Largest private-sector database of world-wide historic weather data
- Major provider of Settlement Data for weather risk contracts.
- Products:
 - Weather data
 - Weather forecasts
 - Weather derivative pricing and risk-management software
 - Weather risk contract settlement services
 - Consultancy/Weather station installation
- Emphasis on quality. We see weather data as a form of financial market data
- We serve clients in insurance, weather derivatives, banking, energy and agriculture sectors world-wide
- Offices in the United States and the United Kingdom



Introduction to Weather Data Cleaning

- Weather data quality control – “cleaning” - is fundamental to our provision of weather data.
- The availability of reliable “ground truth” is key to the pricing and settling of parametric weather risk contracts
- It is also important in the production of our site-specific downscaled forecasts.
- This document shows how we approach the cleaning task.
- We clean over 5,000 sites around the world every day with dedicated teams in the UK and USA
- At any one time we provide Settlement Data for dozens of sites to a specific Settlement Data Specification for those sites where clients have open weather risk. Please see separate documentation relating to the provision of Settlement Data



Data Cleaning

Problem: Weather data is not perfect

- Missing values
- Erroneous observations
- Consistency problems
- Multiple data sets claiming to be for the same weather station

~~**Solution (1):** Ignore the problem~~

- ~~➤ Erroneous observations will lead to inaccurate pricing~~

~~**Solution (2):** Use only “good” stations~~

- ~~➤ Difficult to determine what is good without cleaning it~~
- ~~➤ Greatly limits your ability to trade~~

Solution (3): Clean / fill the data

- Fill missing values
- Detect and replace erroneous observations
- Confirm the consistency of the data

Never purchase data from anyone unless they can provide satisfactory answers to these questions:

- What is the original source of the data?
- What is the observation convention?
- What are the attributes (lat, lon, elevation)?
- What has been done to the data
- Which data points have been QC'd, and what were they pre-QC?



Data Cleaning

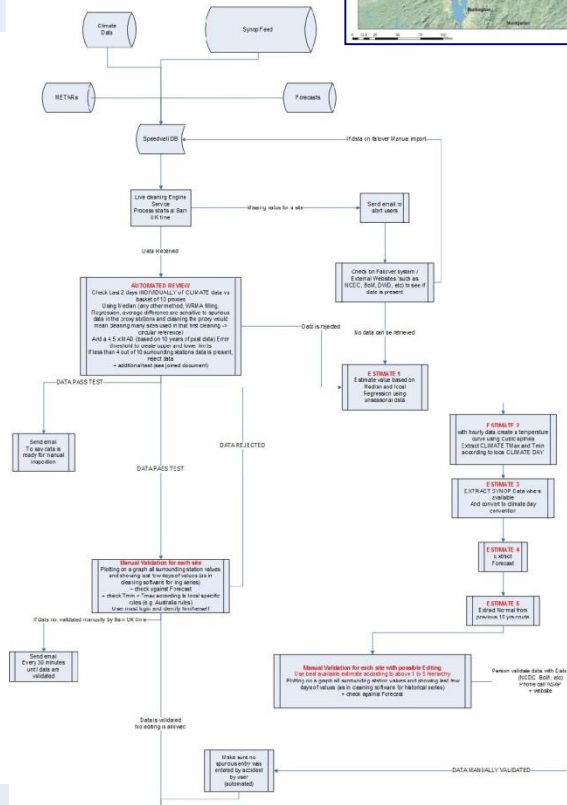
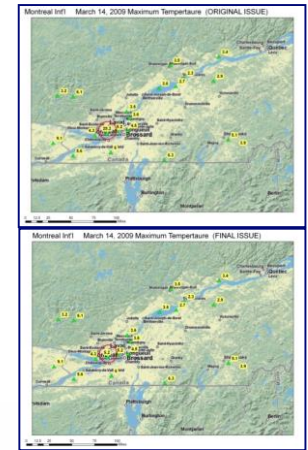
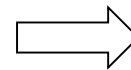
- The quality of Meteorological observations varies significantly even from G20 national met services
- Missing / erroneous observation are common place
- A lot of weather data available in public archives is stored in an inconsistent manner and is of low quality

Fundamentals of a proper data cleaning

- (1) Organization
- (2) Redundancy
- (3) Flexibility
- (4) Human interaction
- (5) Transparency

Fundamental to satisfying the above is the implementation of software systems infrastructure. ..but data cleaning cannot and **SHOULD not** be FULLY automated (see 4)

Part of the Speedwell Data cleaning process diagram





Data Cleaning: Organization

Fundamentals of a proper data cleaning

(1) Organization

- logical flow
- data management

(2) Redundancy

(3) Flexibility

(4) Human interaction

(5) Transparency

Some of the 50+ Speedwell data quality types

- **SYNOP U - NOAA U:** Unedited synop/NOAA data
- **SYNOP E - NOAA E:** Edited synop/NOAA data
- **SYNOP Cleaned:** Cleaned Synop data (a Speedwell Product only)
- **Climate U - NCDC U:** Unedited climate/NCDC data
- **CLIMATE E - NCDC E:** Edited climate/NCDC data
- **CLIMATE Cleaned:** Cleaned Climate data (a Speedwell Product only)
- **Exchange Initial:** Exchange Initial / Preliminary Settlement data (a Speedwell Product only)
- **Exchange Final:** Exchange Final Settlement data (a Speedwell Product only)
- **RECONS U -:** Unedited reconstructed data
- **RECONS E:** Edited reconstructed data
- **RECONS E2:** Edited reconstructed data series

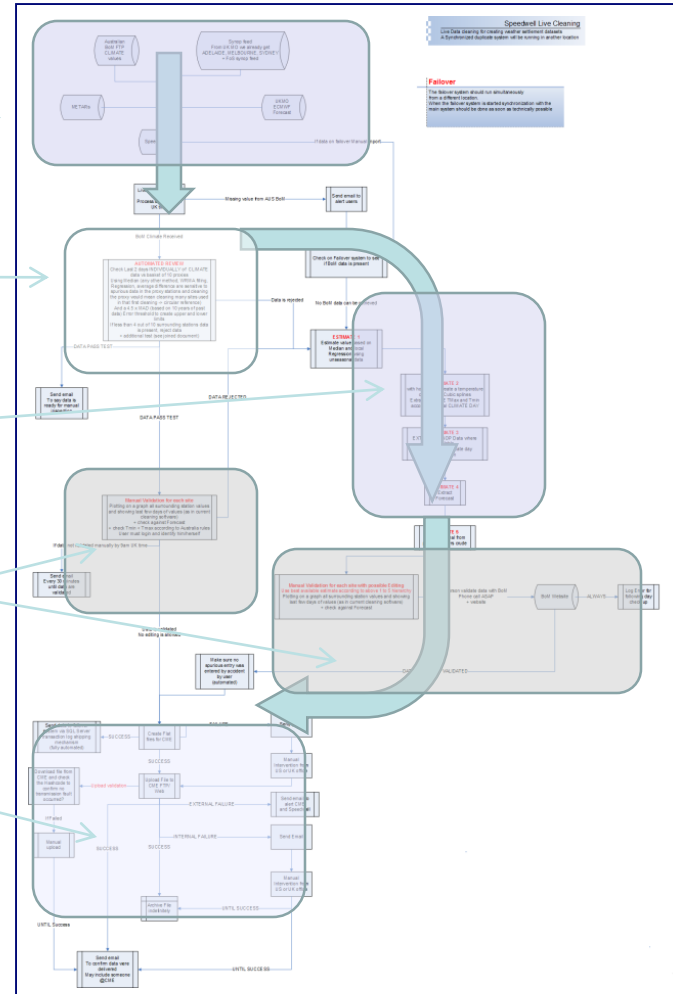
Data preparation

Initial Review

In-depth analysis / data filling

Manual Review

Data delivery





Data Cleaning: Flexibility

Fundamentals of a proper data cleaning

(1) Organization

(2) Redundancy

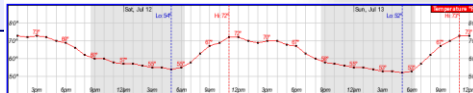
(3) Flexibility

- consider the situation
- appropriateness of tests

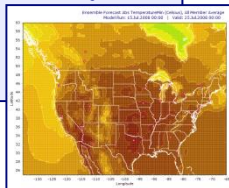
(4) Human interaction

(5) Transparency

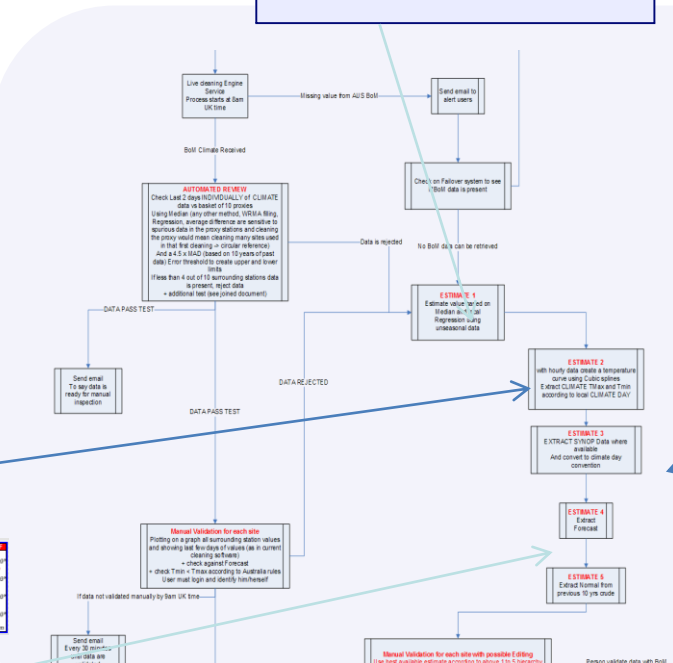
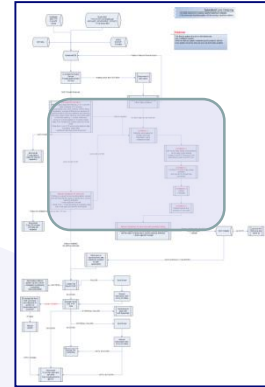
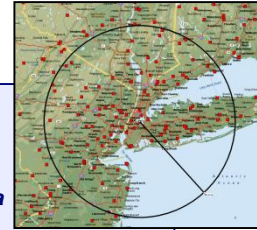
Estimate #2
Estimates of daily observations from hourly observations (curve fitting)



Estimate #4
Day +1 forecasts can actually be very good...



Estimate #1
surrounding station regression using deseasonalized data



Estimate #3
Estimates of daily observations by manipulating other data types (Synoptic, METAR, 1/2 hourly)

Estimate #6, #7, #8...
Flexibility allows you to add any appropriate estimates. The possibilities are unlimited.
- satellite derived values
- installed stations
- reanalysis

Estimate #5
Climatology – worst case scenario



Data Cleaning: The Human element and Transparency

Fundamentals of a proper data cleaning

(1) Organization

(2) Redundancy

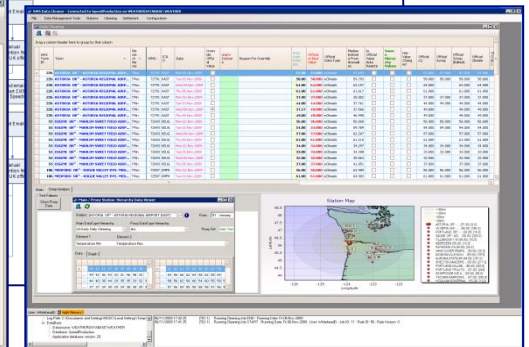
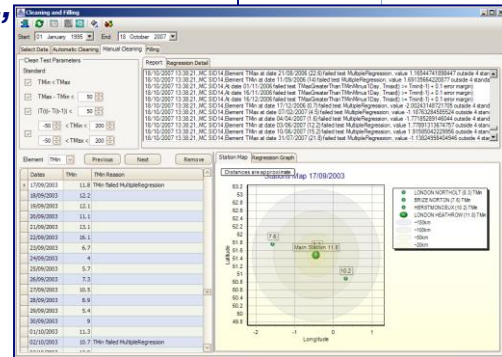
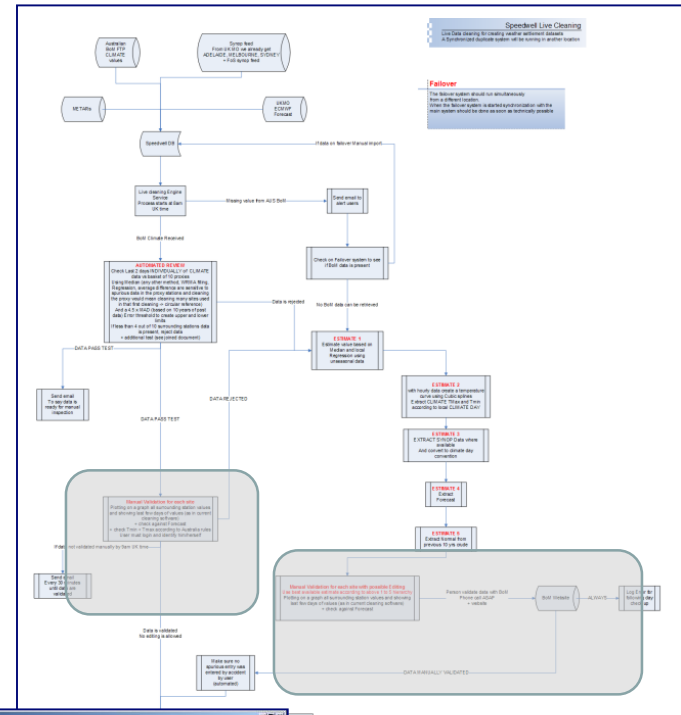
(3) Flexibility

(4) Human interaction

- meteorology is complicated
- introduction of non-automated information

(5) Transparency

- explanation of the process
- share what has been cleaned
- no-one likes “black boxes”



T: +44 (0) 1582 465 551
E: datateam@speedwellweather.com
Harpenden, UK | Charleston, SC, USA

www.speedwellweather.com

Speedwell
Weather

