Missing Data

[Missing Data 1](#_Toc256000000)

[1. Principles ? 3](#_Toc256000002)

[1.1. Which Data ? (OSCAR Requirements , GCOS ECVs) 3](#_Toc256000003)

[1.2. Data completeness 3](#_Toc256000004)

[1.3. Traceability & Provenance 3](#_Toc256000005)

[1.3.1. May a comment on infilled data: As data processing centre it is important for us to have flags on such data as infilled. Depending on the application, infiled data could/should be removed from the time series as it could suggest a higher confidence of gridded data than the data set have. (Markus Z.) 3](#_Toc256000006)

[1.3.2. I'd like to second on the importance of flagging infilled data as mentioned by Robert and Markus (Ge Peng) 3](#_Toc256000007)

[1.3.3. Row, Computed, Estimated, Interpolated, etc. 3](#_Toc256000008)

[1.4. Data Control and Correction 3](#_Toc256000009)

[1.5. Avoid non-scientific approach 3](#_Toc256000010)

[1.6. Data Rescue : long-term strategy (William) 3](#_Toc256000011)

[2. Impacts of missing data 3](#_Toc256000012)

[3. How to prevent data gaps 3](#_Toc256000013)

[3.1. Knowing the sensors, the Automatic Weather Stations, the data flow 3](#_Toc256000014)

[3.2. Real time data acquisition monitoring over all acquisition steps (end-to-end data flow) 3](#_Toc256000015)

[3.3. Instrument/Network Maintenance 3](#_Toc256000016)

[3.4. instrument/network redundancy (GSRN) 3](#_Toc256000017)

[3.5. Security (hardware, software, document, human ressource) 4](#_Toc256000018)

[4. How to identify Data Gap ? 4](#_Toc256000019)

[4.1. Gap Monitoring & Inventory 4](#_Toc256000020)

[4.2. Real Time 4](#_Toc256000021)

[4.3. Archive / Historical 4](#_Toc256000022)

[4.4. Cause of the gap ? 4](#_Toc256000023)

[5. Where to find missing data 4](#_Toc256000024)

[5.1. Global and Regional Data Centers 4](#_Toc256000025)

[5.1.1. NOAA NCEI 4](#_Toc256000026)

[5.1.2. Lead Centers : DWD JMA etc. 4](#_Toc256000027)

[5.1.3. ECMWF - Copernicus 4](#_Toc256000028)

[5.1.4. RCCs 4](#_Toc256000029)

[5.1.5. Universities 4](#_Toc256000030)

[5.1.6. NMHSs 4](#_Toc256000031)

[5.1.7. ... 4](#_Toc256000032)

[5.2. Others 4](#_Toc256000033)

[5.2.1. ACRE 5](#_Toc256000034)

[5.2.2. ... 5](#_Toc256000035)

[5.3. Document 5](#_Toc256000036)

[5.3.1. Data Rescue 5](#_Toc256000037)

[5.3.2. GCOS 5](#_Toc256000038)

[6. What to do ? 5](#_Toc256000039)

[6.1. William 5](#_Toc256000040)

[6.1.1. a clear statement of what the issue is, and why it needs to be addressed; 5](#_Toc256000041)

[6.1.2. scope: at this stage I'd confine it to short to medium-term climate time-series (days up to a couple of months); 5](#_Toc256000042)

[6.1.3. a clear statement of the deliverable(s), such as a recommended method (or methods), backed by an extensive literature review; 5](#_Toc256000043)

[6.1.4. a review process, including by WMO Members. 6](#_Toc256000044)

[7. state-of-the-art on technics (infilling and disagregation) 6](#_Toc256000045)

[7.1. Data computation with missing Data 6](#_Toc256000046)

[7.1.1. WMO computing rules with missing data 6](#_Toc256000047)

[7.1.2. NMHSs practices 10](#_Toc256000048)

[7.1.3. Global Data Centres practices 12](#_Toc256000049)

[7.1.4. Manufacturers (AWS, Data Concentrators, Instruments, etc.) 12](#_Toc256000050)

[7.1.5. Other Organizations : ISO, OGC, Universities, etc. 12](#_Toc256000051)

[7.2. Data Estimation/Analysis 12](#_Toc256000052)

[7.2.1. Technics 12](#_Toc256000053)

[7.2.2. Strengh and weakness of each technic 13](#_Toc256000054)

[7.2.3. Lack of technics comparison 13](#_Toc256000055)

[7.2.4. Technics difficulties (which level of expertise is needed?) 13](#_Toc256000056)

[7.3. Reference 13](#_Toc256000057)

[7.3.1. WMO 13](#_Toc256000058)

[7.3.2. Other 19](#_Toc256000059)

[8. How to support members 20](#_Toc256000060)

[8.1. Training on principles 20](#_Toc256000061)

[8.2. Maintain up-to-date a document on the state-of-the art on this subject 20](#_Toc256000062)

[8.2.1. Does not exist yet ? 20](#_Toc256000063)

[8.2.2. Currentt practices 20](#_Toc256000064)

[8.2.3. "The" list 20](#_Toc256000065)

[8.3. Who? 20](#_Toc256000066)

[8.3.1. WMO Secretariat 20](#_Toc256000067)

[8.3.2. Expert Team(s) INFCOM & SERCOM & HMEI 20](#_Toc256000068)

[8.3.3. Research project 21](#_Toc256000069)

[8.3.4. GCOS 21](#_Toc256000070)

# Principles ?

## Which Data ? (OSCAR Requirements , GCOS ECVs)

## Data completeness

## Traceability & Provenance

### May a comment on infilled data: As data processing centre it is important for us to have flags on such data as infilled. Depending on the application, infiled data could/should be removed from the time series as it could suggest a higher confidence of gridded data than the data set have. (Markus Z.)

### I'd like to second on the importance of flagging infilled data as mentioned by Robert and Markus (Ge Peng)

### Row, Computed, Estimated, Interpolated, etc.

## Data Control and Correction

## Avoid non-scientific approach

## Data Rescue : long-term strategy (William)

# Impacts of missing data

# How to prevent data gaps

## Knowing the sensors, the Automatic Weather Stations, the data flow

## Real time data acquisition monitoring over all acquisition steps (end-to-end data flow)

## Instrument/Network Maintenance

## instrument/network redundancy (GSRN)

## Security (hardware, software, document, human ressource)

# How to identify Data Gap ?

## Gap Monitoring & Inventory

## Real Time

## Archive / Historical

## Cause of the gap ?

# Where to find missing data

## Global and Regional Data Centers

### NOAA NCEI

### Lead Centers : DWD JMA etc.

#### Buoy

#### Marine

#### Surface

#### ...

### ECMWF - Copernicus

### RCCs

### Universities

### NMHSs

### ...

## Others

### ACRE

### ...

## Document

### Data Rescue

#### WMO

##### I-Dare portal

##### ...

#### Copernicus

##### Data Rescue Portal

##### ...

#### ...

### GCOS

#### The 2022 GCOS Implementation Plan

# What to do ?

## William

### a clear statement of what the issue is, and why it needs to be addressed;

### scope: at this stage I'd confine it to short to medium-term climate time-series (days up to a couple of months);

### a clear statement of the deliverable(s), such as a recommended method (or methods), backed by an extensive literature review;

### a review process, including by WMO Members.

# state-of-the-art on technics (infilling and disagregation)

## Data computation with missing Data

### WMO computing rules with missing data

#### Normals

##### Mean

###### WMO-No. 1203, 2017 No more than 6 months missing over the 30 years (at least 80 % of the years)

###### WMO-No. 100, 2018 No more than 6 months missing over the 30 years (at least 80 % of the years) + No more than 3 consecutive missing months over the 30 years

#### Monthly

##### Mean

###### WMO-No. 1203, 2017 it is therefore recommended that, where a monthly value is the mean of that month’s daily values, it should not be calculated if either of the following criteria are satisfied: – Observations are missing for 11 or more days during the month; – Observations are missing for a period of 5 or more consecutive days during the month. However, it should be noted that countries which decided in the past to follow the 5/3 rule (accepting only 5 days of missing observations per month and not more than 3 consecutive missing days as per Calculation of Monthly and Annual 30-year Standard Normals (WMO, 1989)), or other missing data criteria that are stricter than those above, may wish to continue applying these stricter rules to ensure consistency of their national climate record. (In principle, this note applies also to the below paragraphs of section 4.4.1.

###### WMO-No 100, 2023 No more recommendation

###### WMO-No 100, 2018 It is recommended that a monthly value should not be calculated if more than ten daily values or five or more consecutive daily values are missing.

###### WMO-TD-No 1188 (Handbook on CLIMAT and CLIMAT TEMP Reporting), 2009 If observations on more than ten days are missing, or if there is a period of five consecutive days without any observations, the symbolic letters for the respective parameter shall be encoded as slashes (“/”). It is, though, common practice at many national meteorological services (USA, Russia, …) to define three days as an acceptable limit of missing days from the record for a parameter during a month for the majority of the parameters included in CLIMAT Reports, and zero days for parameters such as R1 (total precipitation or snow water equivalent for the month) and S1 (total sunshine duration for the month) to avoid possible significant observational errors for monthly values.

###### WMO-TD-No. 341 Calculation of monthly and annual 30-year standard normals When arithmetic means are to be calculated for each month of each year from daily data the following rule (hereafter referred to as the "3/5 rule") applies. If more than 3 consecutive daily values are missing or more than 5 daily values in total in a given month are missing, the monthly mean should not be computed and the year-month mean should be considered as missing.

a. Precipitation Total—Totals shall be calculated for each  
month of each year from daily data. Monthly totals should be  
based on a full month's data. However, accumulated amounts  
during the month are acceptable in lieu of individual daily  
totals provided that each accumulation is for 3 or less  
days. If accumulated data are used, the monthly total should  
be identified with an "accumulation" indicator. If any daily  
totals are missing and the corresponding accumulated totals  
are also missing, the monthly total should not be computed  
and the year-month total should be considered as missing.  
  
b. Days With Precipitation Greater Than or Equal to 1mm—Totals  
should be calculated for each month of each year from daily  
data. Monthly totals should be based on a full month's data,  
that is, no missing daily counts are permitted.  
  
c. Temperature—Calculate average monthly maximum (Tx), minimum  
(Tn), and mean (Tm) temperature from the daily values Tx, Tn,  
Tm as follows:   
  
etc.

##### Count

###### WMO-No. 1203, 2017 For count of parameter: "A ratio or percentage should not be calculated for the month if there are 11 or more days with missing observations, or 5 or more consecutive missing days. In these cases, the monthly value should be considered to be missing."

##### Extreme

###### WMO-No. 1203, 2017 In the context of calculation of normals, the purpose of calculating an extreme value during an individual month is as an intermediate step of calculating extreme values over the entire period under consideration. At this point in the process, extreme values should be calculated for a month, regardless of the amount of available data during that month. (Whether sufficient data exist for the reliable reporting of an extreme value over the entire period is considered in section 4.4.2.

##### Sum

###### WMO-No. 1203, 2017 A monthly value for a sum parameter (for example, total precipitation) can only be calculated if there are complete data over the month. This means that, in general, a sum parameter cannot be calculated if there are any missing observations during the month

#### 10 Day

#### 5 Day

#### Daily

##### Mean

###### WMO-No. 1188, 2009 Mean daily values shall be calculated as an average of observation values at the UTC standard times for surface synoptic observations which correspond to a given day j in local time (0000 - 2359 local time) for all days of the respective month. All four or eight observations shall be used for daily averaging. If any value necessary for the calculation of a mean daily value is missing, the missing value, if possible, should be taken from appropriate autographic records. If this cannot be done, and if it was intended to calculate the mean daily value on the base of eight standard times for surface synoptic observations, then only the four main or intermediate standard times for surface synoptic observations shall be used for calculation. If this cannot be done, the respective daily mean value shall be marked as missing. It is not allowable to use less than four either main or intermediate standard times for surface synoptic observations for the calculation of a mean daily value.

###### WMO-TD-No. 341 from 24 hourly observations per day. If 24 hourly values are not available each day, then 8 (00,03,06,12,15,18,21Z) observations each day should be used. The number of observations each day should be identified with a frequency indicator.

#### Hourly

#### 6 Minute

#### 5 Minute

#### 1 Minute

#### Infra-Minute

### NMHSs practices

#### Météo-France

##### General Principles Missing data could be automatically reconstituted, or estimated by human or corrected by human or left as missing, with the main following rules: \* Daily time series of TX, TN and RR should be complete Monthly RR complete \* Never correct (human action) more than 2 hourly consecutive observations (except for rainfall in case of dry day) \* Never exceed ten consecutive days of estimation or correction of daily data \* It is forbidden to estimate any hourly data and 6 minutes, except in the case of dry day "Les Consignes pour assurer l’expertise climatologique Référence : MF\_GT\_CLIMAT\_Consignes\_Expertise\_Clim" (Météo-France)

###### Automatic Reconstruction

Parameters

Temperature

Priority on reconstruction  
1. use of minute data if available  
2. use of he previous hour and the following hour if available (interpolation)  
3. Spatial Analysis tool

Wind

Precipitation

Humidity

Radiation

Pressure

Snow

Principles

\* Only for certain station types (AWS)  
\* Automatic process to retrieve observed data not received  
\* Automatic reconstitution of permanently missing data  
\*Automatic checks: automatically reconstructed data pass the automatic checks in the same way as other observation data. On the other hand,   
automatically reconstructed data are not used as a reference for spatial checks (i.e. the checked data are not not compared to the data of a neighboring station if this one has been automatically reconstructed)

###### Human Estimation or Correction

###### Computing with missing data

### Global Data Centres practices

### Manufacturers (AWS, Data Concentrators, Instruments, etc.)

#### HMEI

### Other Organizations : ISO, OGC, Universities, etc.

## Data Estimation/Analysis

### Technics

#### Statistical methods

##### Spatial

###### neighbors technics

###### ...

##### Temporal

##### Spatial/Temporal (e.g. Marine data)

##### ...

#### Homogenisation

#### Machine learning

#### Numerical Models

##### Reanalyses

##### ...

#### Remote sensing measurements

#### Alternative measurements

#### Data fusion

##### Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)

##### ...

### Strengh and weakness of each technic

### Lack of technics comparison

### Technics difficulties (which level of expertise is needed?)

## Reference

### WMO

#### WMO-No. 100 Guide to Climatological Practices, Edition 4

##### 3.5.8 Data estimation

###### Statistic methods

Interpolation

Extrapolation

Inverse distance weighting

Curve fitting (Spline)

###### Using other parameters (e.g. Sunshine duration for global radiation)

###### Proxy data

###### Deterministic methods (e.g.regression)

###### Spatial interpolation

Stochastic methods

Kriging

Climatologically Aided Interpolation (CAI)

###### Machine learning

###### Analysis Using Relief for Hydrometeorology (AURELHY)

###### Parameter-elevation Regressions on Independent Slopes Model (PRISM)

###### Numerical models

#### WMO-No. 1203 Calculation of Climate Normals

##### 4.6 - Estimation of data

###### Spatial interpolation

###### Temporal interpolation

###### Use of alternative elements

using cloud amounts to estimate a missing daily  
sunshine value.

###### Use of alternative observation methods

using radar or satellite observations to  
estimate precipitation data where the original observation is missing

###### Rule : no more than 10 missing day in a month

#### WMO-No. 1204 WMO Guidelines on Generating a Defined Set of National Climate Monitoring Products

##### 3.5 Interpolating data

###### Krigging method recommended for index

#### WMO-No. 1287 Developing the Climate Science Information for Climate Action

##### 5. DEVELOPING CLIMATE SCIENCE INFORMATION WITH LIMITED OR ABSENT DATA

###### Approaches for addressing limited or absent data

Remote-sensing observations + gridded data

Reanalysis products

##### Annex I: Guidance on Methods, Tools and Data

##### Annex II: Use cases

#### WMO/TD-No.1188, 2004, Handbook on CLIMAT and CLIMAT TEMP reporting

##### For daily means for pressure, temperature and vapour pressure it is written that the means are the arithmetic means of all 4 or 8 values at observation time (00, 03, 06, 09, 12, 15, 18, 21). The following rule on missing data should be followed: "where this cannot be done, then only either the four principal or the four intermediate observation times shall be used for calculation of the mean daily value. If this is also not possible, the mean daily value for the day in question shall be regarded as missing."

#### WMO/TD-No.341, 1989, Calculation of monthly and annual 30-year standard normals

##### It is recommended to use either the 24 hourly observations or the 8 synoptic observations plus one Flag to indicate the number of observations used for the calculation.

#### WMO-No. 1203, 2017, WMO Guidelines on the Calculation of Climate Normals

##### For temperature it is said that "Definition of observation day, and the way in which daily mean temperature is calculated, should be according to national standards and documented in metadata (see also section 4.9). Different methods are in operational use for the calculation of daily mean temperature." For pressure it is said that "Daily values should be calculated, if possible, as the mean of either eight evenly spaced 3-hourly observations or four evenly spaced 6-hourly observations. If this is not possible, they should use a set of observation times that is consistent over time at that station and documented in metadata."

#### WMO-No. 100, 2018\*, Guide to climatological practices

##### "There are many methods for calculating an average daily temperature. These include methods that use a daily maximum and daily minimum, 24 hourly observations, synoptic observations and observations at certain specified hours in the course of a day. The best statistical approximation of an average is based on the integration of continuous observations over a period of time; the higher the frequency of observations, the more accurate the average. Practical considerations generally preclude the calculation of a daily average from a large number of observations evenly distributed over a 24-hour period because many observing sites do not measure an element continuously. For comparative purposes, a standard processing methodology is desirable for all stations worldwide, with as many stations as possible. All ordinary climatological stations observe a daily maximum and minimum temperature (see 2.2.1). Hence, the recommended methodology for calculating average daily temperature is to take the mean of the daily maximum and minimum temperatures. Even though this method is not the best statistical approximation, its consistent use satisfies the comparative purpose of normals. An NMHS should also calculate daily averages using other methods if these calculations improve the understanding of the climate of the country." \*: be careful the new edition does not have advice on computing daily averages.

#### WMO-No. 1202, 2017, Challenges in the transition from conventional to automatic weather observing networks for long -term climate records

#### WCDMP No 62/WMO-TD No 1378 Guidelines for managing changes in climate observation programmes

#### Guide to Instruments and Methods of Observation (WMO-No. 8)

##### Volume I –Measurement of Meteorological Variables

###### Output Averaging time : in general 1 minute for most parameters, 1 & 10 minutes for wind speed and direction, 3 secondes for Gusts

##### Volume II – Mesure de variables de la cryosphère

##### Volume III – Observing Systems

###### 1.1.3 Meteorological requirements

The overlap time8 is dependent on the different measured variables and on the climatic region. In tropical regions and islands, the overlap time could be shorter than in extratropical and mountainous regions. The following general guidelines are suggested for a sufficient operational overlap between existing and new automated systems: (a) Wind speed and direction: 12 months; (b) Temperature, humidity, sunshine, evaporation: 24 months; (c) Precipitation: 60 months. A useful compromise would be an overlap period of 24 months (that is, two seasonal cycles).

###### 1.2.3 Instrument Double or triple sensors for security reason

It can be desirable to double (or even triple) some instruments. This approach can minimize the probability of missing values in case of instrument failure and/or introduce measurement redundancy in the system to detect possible instrument drift. The difference between two instruments indicates a drift of at least one of them; if three instruments are used, it becomes possible to identify automatically which instrument is drifting and choose to exclude its values. This procedure of using multiple sensing elements is used within some instruments. Several commercial models of barometers are available with one, two or three cells.

###### 1.4.2.2

Volume I, Chapter 1, Annex 1.A of the present Guide recommends that "instantaneous” values of most of the meteorological variables be a one-minute average (except for wind and visibility).

###### 1.4.2.5

The algorithms used to derive values in the AWS are as important as the choice of instrument. Small and subtle changes can be introduced over time that can have significant impacts afterwards. A register of algorithms as well as software versions should be kept as a part of the metadata for the AWS. Where the algorithm is created in-house, it is wise to document the process and to develop data test sets so that changes to software can be consistently checked in the future.  
CIMO is involved in a regular programme to survey and standardize algorithms for all variables. The results are published in WMO (2003). See also the corresponding chapters of Volume I of the present Guide for details on the meteorological variables.

### Other

#### Jose

##### BEN SALEM K, 1999|Design and analysis of an iterative algorithm for incomplete data estimation|International Journal of Computer Mathematics, 71:71-82, DOI: 10.1080/00207169908804793

##### BENNET RJ, HAINING RP, GRIFFITH DA, 1984|The problem of missing data on spatial surfaces|Annals Assoc. Amer. Geogr., 74:138-156.

##### KANG H, 2013|The prevention and handling of the missing data|Korean J. Anesthesiol., 64:402-406

##### KASHANI MH, DINPASHOH Y, 2012|Evaluation of efficiency of different estimation methods for missing climatological data|Stoch. Environ. Res. Risk Assess., 26:59-71.

##### PAPPAS C, PAPALEXIOU SM, KOUTSOYIANNIS D, 2014|A quick gap filling of missing hydrometeorological data|Jour. Geophys. Res. Atmos., 10.1002/2014JD021633, 11 pp.

##### SIMOLO C, BRUNETTI M, MAUGERI M, NANNI T, 2010|Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach|Int. J. Climatol., 30:1564­1576.

##### TEEGAVARAPU RSV, ALY A, PATHAK CS, AHLQUIST J, FUELBERG H, HOOD J, 2018|Infilling missing precipitation records using variants of spatial interpolation and data-driven methods: use of optimal weighting parameters and nearest neighbour-based corrections|Int. J. Climatol., 38:776-793.

##### XIA Y, FABIAN P, STOHL A, WINTERHALTER M, 1999|Forest climatology: estimation of missing values for Bavaria, Germany|Agric. and Forest Meteor., 96:131-144.

#### Denis

##### "Estimating daily climatological normals in a changing climate": https://hal.science/hal-01980565

##### " An updated assessment of past and future warming over France based on a regional observational constraint" : https://esd.copernicus.org/articles/13/1397/2022/

# How to support members

## Training on principles

## Maintain up-to-date a document on the state-of-the art on this subject

### Does not exist yet ?

### Currentt practices

#### WMO rules

#### NMHSs practices

#### Global Data Centres practices

#### Manufacturers

### "The" list

## Who?

### WMO Secretariat

### Expert Team(s) INFCOM & SERCOM & HMEI

### Research project

#### University partnership?

#### subject for a post-doctoral fellow ?

#### [14:45] Ge Peng We have a ML team who may be able to do the job systematically and efficiently if there is a funding opportunity (Ge Peng)

### GCOS