Lessons learnt from the Météo-France data rescue program :

problems identified when locating-transferring-homogenising-archiving-disseminating the data

Sylvie Jourdain Météo-France sylvie.jourdain@meteo.fr

WMO/MEDARE International Wokshop
10-12 May Nicosia, Cyprusy

METEO FRANCE
Toujours un temps d'avance

Data Rescue

Definition of the data rescue (WMO/TMD-No 1128)

Two step process :

- Ongoing process of preserving all data at risk of being lost due to deterioration of the medium
- Digitization of the data into computer compatible form foeor easy access

Data Rescue involves rescuing both the data and metadata

- Chapitre 3.6 **Guide to climatological practices** third edition

Each NMHS should establish and maintain a data rescue programme

All National Meteorological Services have digitized most of their recent records, with many having most of the periode from 1950s digitized

(From 1961 for climatological stations at Météo-France and from the beginning for synoptical stations)

METEO FRANCE
Toujours un temps d'avance

Data Rescue Process

- Search and locate the data
- Make a detailed inventory
  - Prioritise records based on NMHS need and risk of loss
- Preservation
- Recovering
- Digitisation
- Ongoing maintenance
- Dissemination

Data Rescue at Météo-France

Keeping the memory of past climate is one of the key mission given to Météo-France, the French NMS

Météo-France has renewed its data rescue activity since 2007 with more human and finantial ressources.

Data rescue and long-term series are included in the Météo-France goals since several years.

The action is nationwide and includes the 4 overseas territories managed at Météo-France by the Climatology department.

Data rescue activities are undertaken by national, regional and departmental services (near 100 different services)

Efforts have been made to improve methodologies associated to the search of documents, inventories, digitization and monitoring tools.

From 2007 to 2009 M-F digitized 2 million observations (daily an subdaily observations

Goals for 2009-2011 : 6 million observations

METEO FRANCE
Toujours un temps d'avance

Preservation

- Preservation of documents Documents are always in danger
  - Use acid free boxes to store records
  - digital imaging or scanning
  - Microfilms
  - Store documents in good air conditions ! Mould problems
  - Store in safe buildings
  - Flood risk prevention

Case of Paris-Montsouris archives last year

METEO FRANCE
Toujours un temps d'avance


La Rochelle, Departmental center
after Xynthia storm
DELL

Preservation

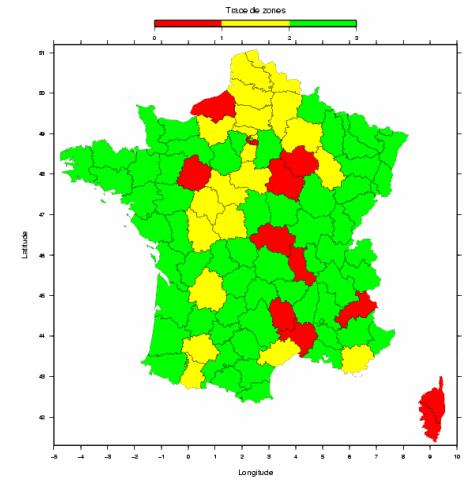
- Preservation of documents Documents are always in danger
  - Use acid free boxes to store records
  - digital imaging or scanning
  - Microfilms
  - Store documents in good air conditions ! Mould problems
  - Store in safe buildings
  - Flood risk prevention

Case of Paris-Montsouris archives last year

METEO FRANCE
Toujours un temps d'avance

Inventorying

- The Inventory of the climatological data available in the archives is essential . Inventory has a old fashioned and dusty image but is fundamental.
- A guide for inventorying climatological data has been written by the Climatomogy department to explain the aim of the action and to give recommendations to colleagues in the M-F centers



Monitoring of the national inventorying action

Inventory of the departmental centers is in process
Shoub be finished in 2011

METEO FRANCE
Toujours un temps d'avance

Inventory

- Gather all relevant metadata to trace a time series :
  - Information about location of the observations (name of the site, ccordinates, …)
  - time period, of the observations,
  - observed variables,
  - observation frequency and hours of observations,
  - information about how the data are available (digital, harcopy) and where the material is stored

Recovering

- More and more french meteorological publications on Internet
  - Gallica : numerical french library :memories, annals, journals
  - Googlebooks : french meteorological annals
  - Site of Astronomical annals :yearbooks published in annals
- Locating the data is easier because more an more catalogs on the web from universities and public archives (SUDOC) .
- But the recovering is not so easy and takes a long time
- Most of archives are used to duplicate the document with microfilms to preserve the documents but not to disseminate the copies

Recovering

- A vast amount of french climate data is not stored in Météo-France archives but in public archives, in universities, in observatories … A guide for M-F colleagues has been written by the Climatology department about the different archives

- It is often impossible to lend the data for digitization in public archives. The solution depends on the rules of the archives !
- Several possibilities :
  - imaging (scanning or photos),
  - Microfilms
  - Photocopies
  - Digitization in situ
- Fortunately, it is often possible to lend documents from universities and observatories after agreement about dissemination of images.

French climate data recovering

- Situation is complicate in France because material is scattered accross a huge number of archives in France mainland , overseas and in other countries (for example DWD inGermany)
- French climatological archives in national archives : no access since 2004 because abestos!
- Météo-France , old colonial power stores a lot of old African climate data (microfilms and microfiches)

- Close cooperation between NHMS is vital for data recovering
- New M-F and DWD collaboration : recovery of german observations in France during the World War II. Handwritten observations stored in DWD in hard harcopy form recently inventoried by DWD.

North Africa data recovering

- Cooperation M-F with National Meteorological Institutes
  - Inventory of the documents stored at M-F (microfilms, books) in Toulouse and Paris
  - Selection of non digitized reliable long –term series
  - Selection and imaging books with useful data
  - Imaging microfilms
  - Digitization of daily and monthly data
  - Maroc example temperature, rainfall and pressure series :
    - Casablanca, (192-1962) Fez, (1924-1962), Kasb-Tadla (1949-1962), Meknes 1(924-1962) Rabat-Salé, (1947-1961), Rabat-Ville(1924-1961)

Station identification

- Identification is not easy because plenty stations are in the same city at the same time.

- Need to use all information :
  - metadata gathered in the inventory (name of the station, coordinates , station description), period of observations, observed variables)
  - meteorological publcations,
  - data and metadata in the database for comparison

- Difficulties to identify the station in french annals : hundreds rainfall data without station code
- Digitization can be decided after identifying the station

Digitization

- Digitization is coordonnated by the Climatology department at M-F :

Recovering and Digitization Priorities (period, station, variables), training , monitoring

The digitization is financed by M-F

M-F uses a french private company for several years (good experience of meteorological documents) and M-F agents in regional and departmental centers. The M-F agents can key directly in the database with the climatological operational tool Climsol available in all centers or key in spreadsheets sent to Toulouse

The digitization must be seriously prepared , an expert in data rescue has to check the data source before having it digitized

Digitization

- At M-F , a digitization document describing the documents and the data is written for each action

source of the data, parameters, hours of observation, formats, units, missing data, and at the end data we decide to digitize

Another document is written after that for the private society to give precise directives for the keying

The preparation of digization is task consuming because a lot of things can change

The next step is quality control to correct typing errors , changes of columns….

Digitized data Dissemination

- Controlled digitized data are rapidly inserted in the national database BDCLIM. The operational infrastructure of M-F for centrally archiving is used.

- Data is available through the internet by the so-called Climatheque, the Météo-France climate data and products access service but not freely.
- Remark : Long-term series are freely provided to researchers and students

Data dissemination

http://climatheque.meteo.fr/

Data Digitization and Insertion Monitoring

spreadsheet with metadata (parameter, period, digitalizer, QC actor)

and

Data Insertion monitoring : dedicated sql table in the climatological database

Insertion report for each insertion

177	Metz Ecluse de l'Esplanade (57463001) 1892-1914 Q.RR	
178	Frayol (07319001) Q.RR,TN,TX + H.T,PSTAT,PMER 1935-1939	Arch
179	Bourg en Bresse (01053001) Q.RR,TN,TX 1890-1898	
180	Paris Impérial 1860 H.PMER, T	
181	Bourg en Bresse (01053001) 1890-1902+1909-1922 H.PSTAT,T et Q.TN,TX	
182	Sarreguemines (57631003) 1893-1914 Q.RR	
183	Gex (01173011) 1890-1898 Q.RR	
184	Belley (01034001) 1890-1898 Q.RR	
185	Bull. Obs. Imp. Paris Lyon Télégraphe(69123024) 1857-1860 H.PSTAT,T	Scan
186	Remelfing (57568002) 1920-1932 Q.RR	
187	Cap Ferret (33236002) 1887-1899 Q.RR,TN,TX	Arch
188	StPierre(9750200X) Q.RR,TN,TXINS,NEIG,BROU H.PSTAT,T 1866-1948	Scan
189	Annales du B.C.M. 1878-1900 Q.RR,TN,TX H.PSTAT,T	Arch
190	Savoie Q.RR	
191	Isère Q.RR	
192	Saint-Michel-sur-Meurthe (88428001) 01/10/1921-31/12/1934 Q.RR	
193	Saint-Michel-sur-Meurthe Gare (88428002) 01/10/1939-31/07/1944 Q.RR	
194	Commission Mété. Gironde plusieurs postes 06/1894-12/1910 Q.TN,TX	Arch
195	Saint-Nicolas-de-Port (54483001) 1912-1920 Q.RR,TN,TX	
196	Sewen-Lac d'Alfeld (68307001) 1893-1914+1920-1934 Q.RR	

DIRNE Marlyse Colobert	DIRNE	11/09/2009
CAA 20090921/20091005	Alco	18/01/2010
DIRCE	Alco	23/09/2009
Stagiaire DEV	Alco	24/09/2009
Stagiaire DEV	Alco	07/10/2009
DIRNE/DET/MC	DIRNE	24/09/2009
Stagiaire DEV	DEV	01/10/2009
Stagiaire DEV	DEV	01/10/2009
CAA 20081210/200901	Alco	30/09/2009
DIRNE Marlyse Colobert	DIRNE	01/10/2009
CAA 20091006/20091015	Alco	
Chry20091105/20091222	Alco	
Chry……../……		
DIRCE Annick Auffray	DIRCE	02/12/2009
DIRCE Annick Auffray	DIRCE	15/01/2010
DIRNE Marlyse Colobert	DIRNE	20/11/2009
DIRNE Marlyse Colobert	DIRNE	24/11/2009
CAA 20091215/…….		
DIRNE Marlyse Colobert	DIRNE	30/12/2009
DIRNE Marlyse Colobert	DIRNE	11/01/2010

Data Insertion Report

date of the insertionn : 05/15/2008

- station code : 67124001 City : Entzheim
- Période : 01/10/1941 au 31/10/1944
- Table : Q (daily data)
- Parameters : RR – TN – TX
- Digitizer : Marlyse Colobert climatology department Entzheim
- Insertion mod : keying in CLIMSOL

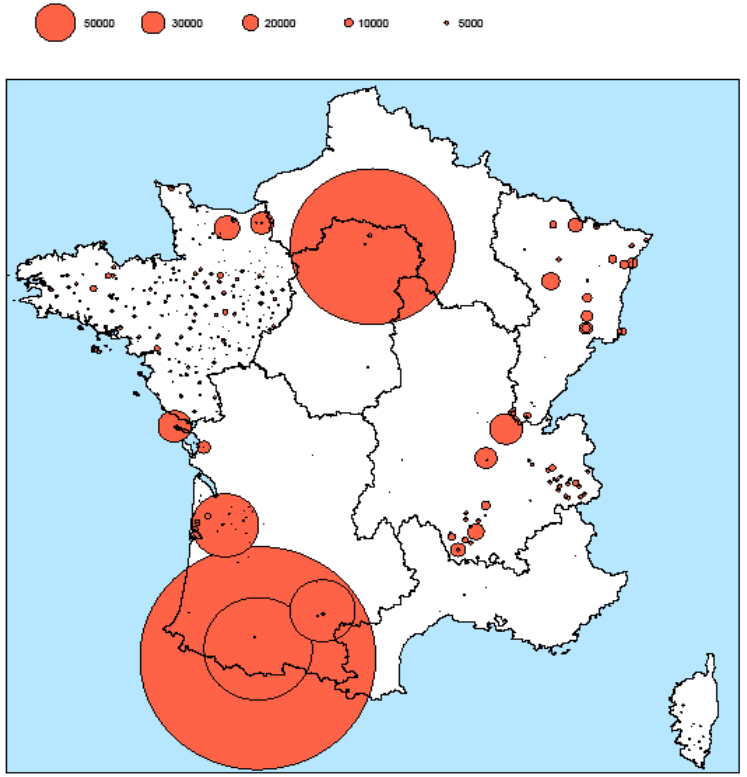
- Type of documents : original copies
- Documents : Reichswetterdienst Monatstabelle für meteorologische Stationen II. Ordnung

METEO FRANCE
Toujours un temps d'avance

Monitoring : map of inserted data (2009)

METEO FRANCE
Toujours un temps d'avance

Nombre d'Insertions dans la BDCLIM
Du 01/01/2009 au 31/12/2009
Dans le Cadre de l'Action DATA RESCUE



METEO FRANCE
Toujours un temps d'avance

Homogénéisation : Why? La Rochelle example

- 1910 Primary school

- 1999 departmental center Le Bout Blanc

METEO FRANCE

METEO FRANCE
Toujours un temps d'avance

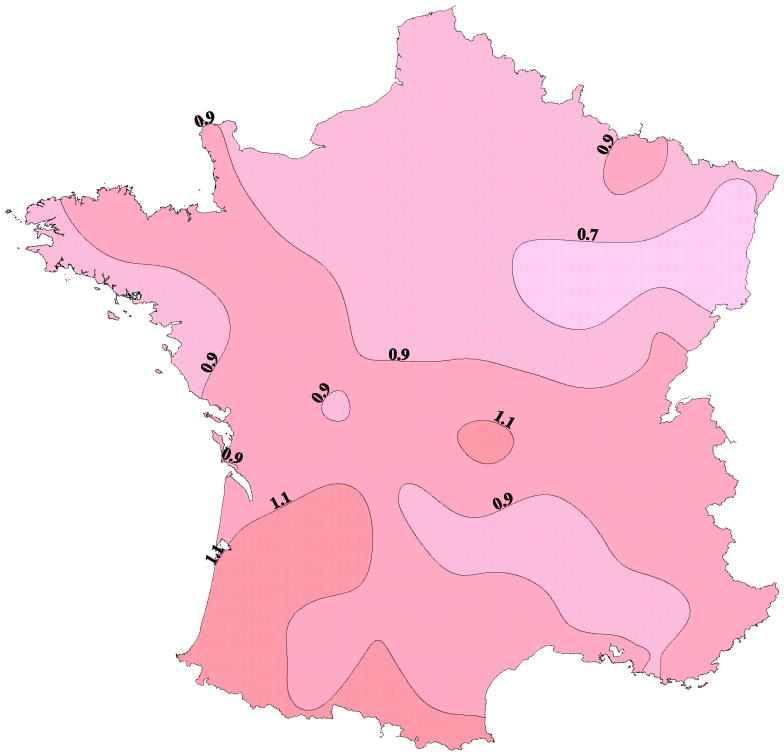
Homogenization: Why?
Exemple de La Rochelle

- La Rochelle Le Bout Blanc in 1995

- Le Bout Blanc in 2004

METEO FRANCE
Toujours un temps d'avance

Homogenized mean temperature series
1901-2000 Mestre(2004)



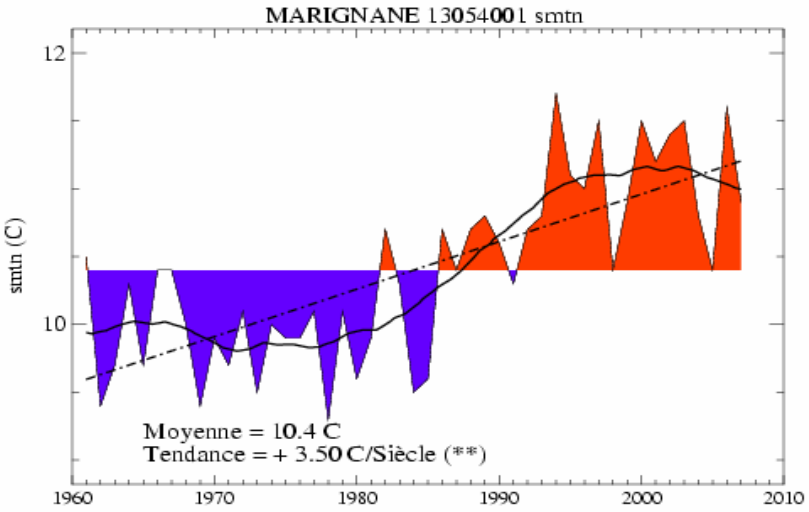
Trend in ℃/century

stations

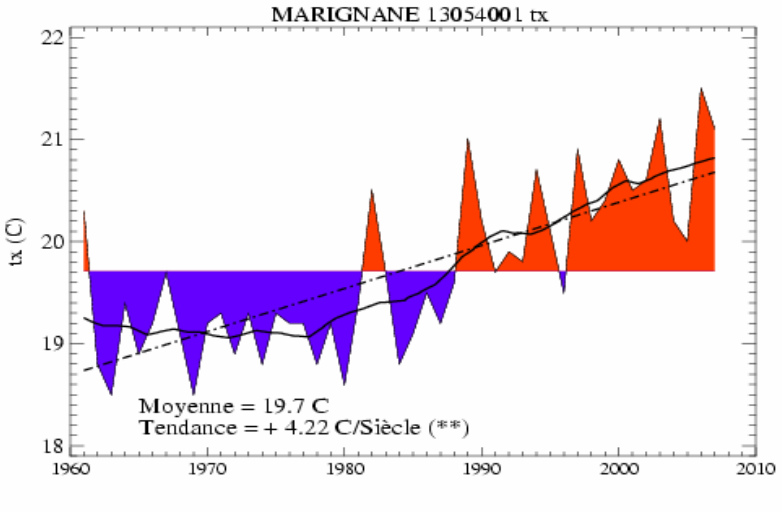
METEO FRANCE
Toujours un temps d'avance

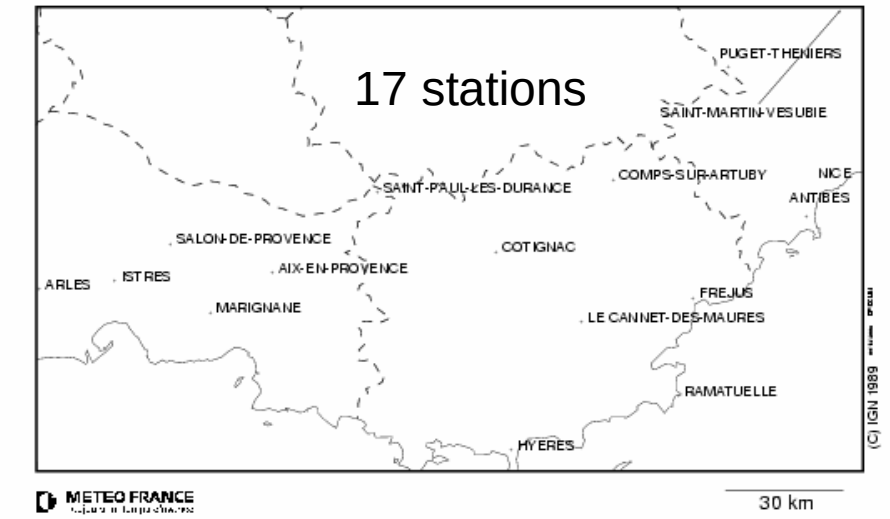
South-East of France temperature Homogenization 1961-2007

Température minimale



Température maximale





Marignane (Marseille Airport)
Synoptical and professional station

3 breaks
Breaks around 0.4℃ due to
Shelter change
and measurementautomation

METEO FRANCE
Toujours un temps d'avance

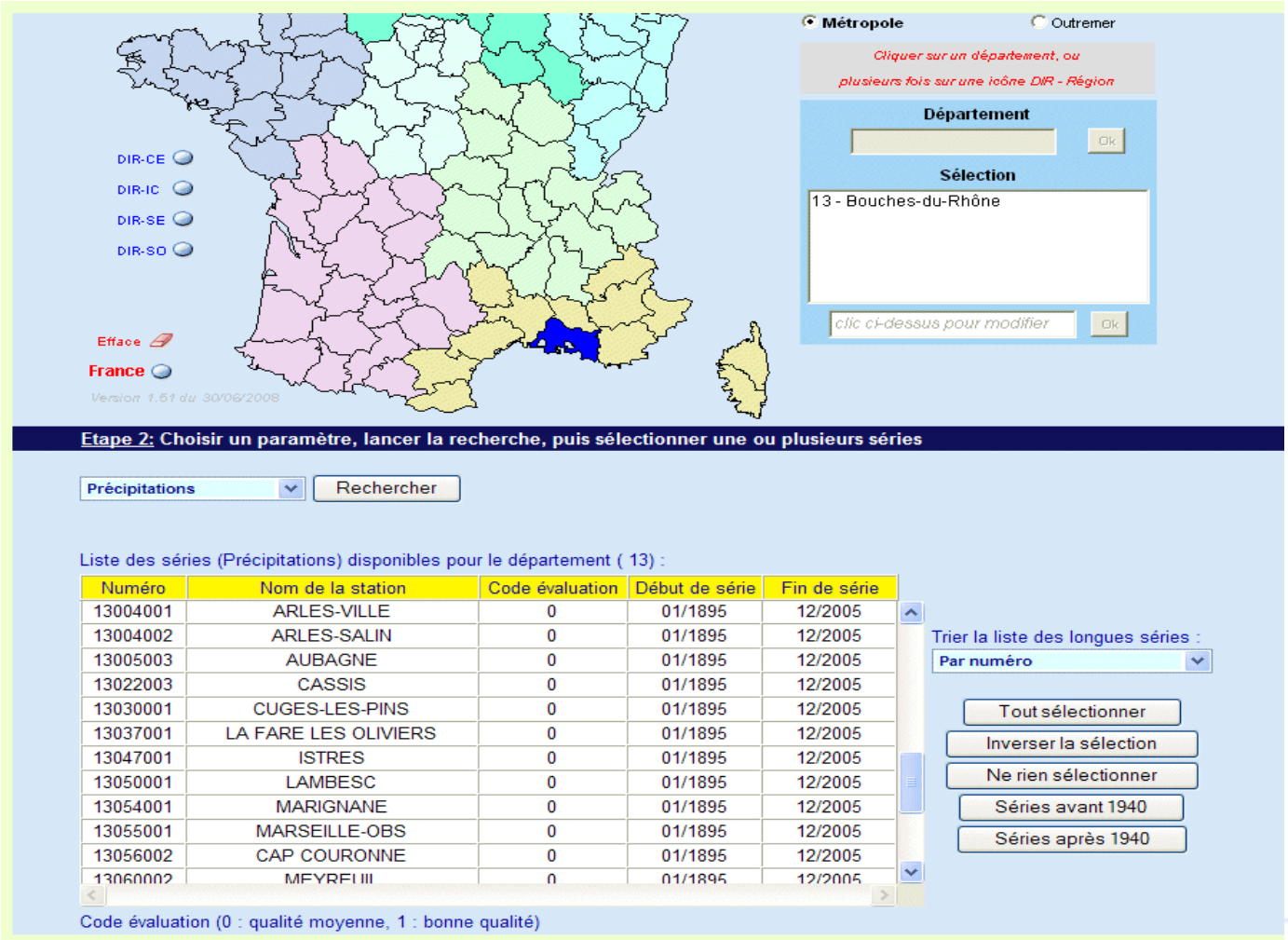
Homogenizing lessons

- Most of long-term series are heterogeneous, many methods to homogenize long-term series (see COST HOME action)
- Metadata collectionare must be performed before homogenization metadata are essential to validate the date of the breaks
- Quality control and selection of reliable series are essential too
- Homogenizing must be carried by professional

- M-F uses Caussinus Mestre's technique (free available)
- Problems when changes are simultaneous (shelter changes and automation in professional stations)
- Non expert people put too many breaks
- For french temperature long-series, France is divided in 20 areas : problems for stations in the border because corrections depend on the data set

French homogenized series dissemination

Météo-France Product



METEO FRANCE
Toujours un temps d'avance

Monthly homogenized long-term series product

Product description

Annual and monthlly homogenized series archived in the database in specific tables

- Series metadata
  - Stations with periods used to build the long-term series
  - Dat of breaks
  - Residual breaks amplitude
  - Quality of the homogenized series

Toulouse 1878-2000
metadata of homogenized series

- **Définition du paramètre extrait**

  + *Afficher les informations*

- **Liste des séries**

  - **Série TOULOUSE-BLAGNAC (MTN031069001)**

    - *Masquer les informations*

Description générale				
Numéro Météo-France	**Nom usuel**	**Date début**	**Date fin**	**Qualité**
31069001	TOULOUSE-BLAGNAC	01/1878	12/2000	bonne qualité

Liste des stations de la série			
Numéro Météo-France	**Nom usuel de la station à défaut celui de la commune**	**Date début**	**Date fin**
31555016	TOULOUSE	01/1878	12/1920
31157001	TOULOUSE-FRANCAZAL	01/1921	05/1927
31157001	TOULOUSE-FRANCAZAL	06/1927	12/1946
31069001	TOULOUSE-BLAGNAC	01/1947	12/2000

Liste des ruptures de la série	
Date des ruptures	**Amplitude minimale détectable de la série**
12/1888	0,25
12/1920	
12/1946	
12/1970	
12/1987	

Toulouse long-term series

Exemple de la série de Toulouse

1. **Observatoire de Toulouse** 1733 –1984

3 sites : en ville sur une tour des remparts, puis observatoire sur colline de Jolimont en 1841 puis campus de Toulouse Rangueil en 1981

2. **Toulouse-Francazal** 1922- aujourd'hui**.** Fermeture annoncée de la Base aérienne

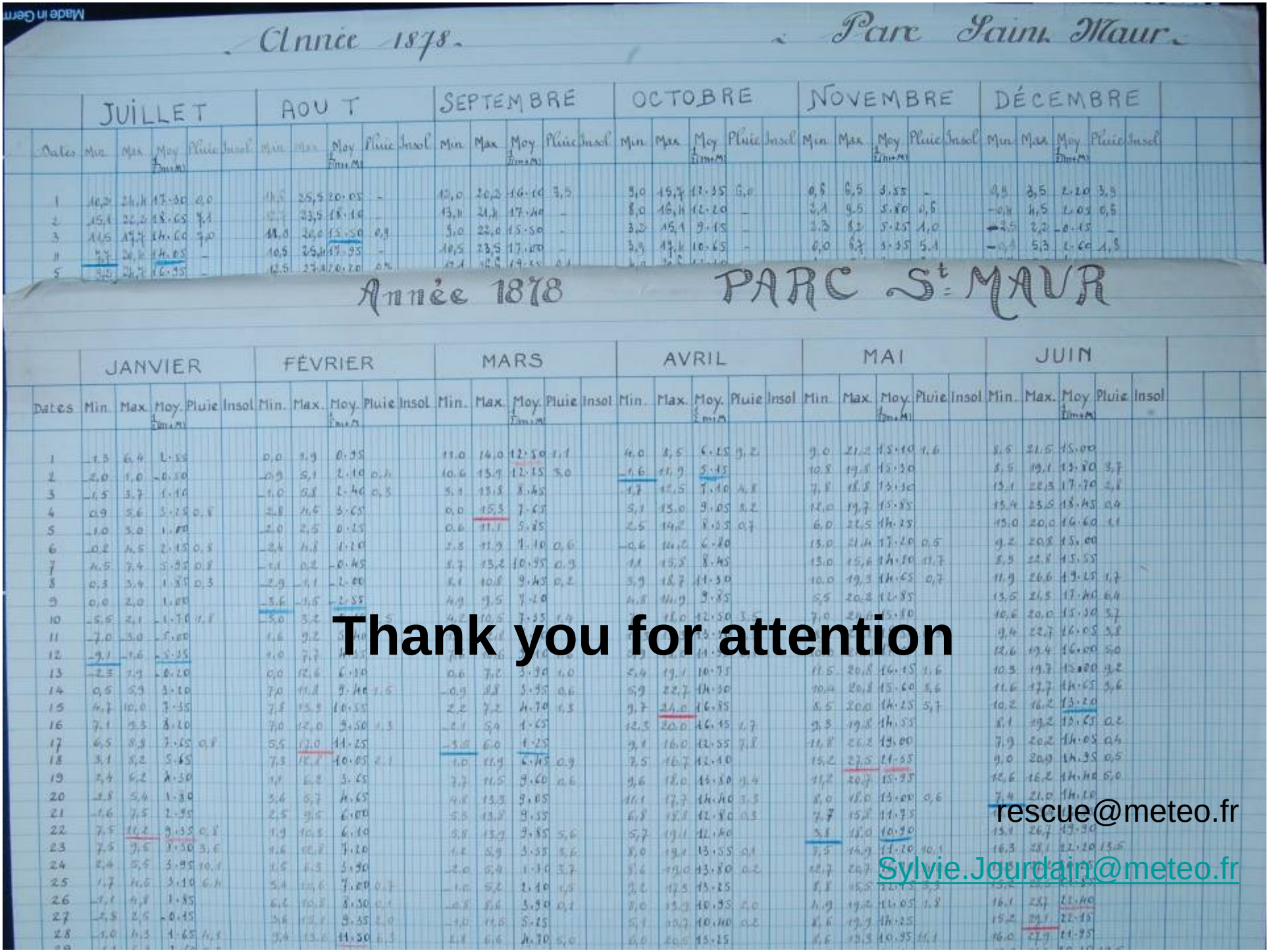
3 sites différents

3. **Toulouse-Blagnac** 1947-aujourd'hui

Toulouse Blagnac en 1999

Observatoire de Jolimont

METEO FRANCE
Toujours un temps d'avance


Année 1878
Parc Saint Maur
JUILLET
AOUT
SEPTEMBRE
OCTOBRE
NOVEMBRE
DÉCEMBRE
Année 1878
PARC St MAUR
JANVIER
FÉVRIER
MARS
AVRIL
MAI
JUIN
Thank you for attention
rescue@meteo.fr
Sylvie.Jourdain@meteo.fr