

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 Workshop
10-12 May Nicosia, Cyprus

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 for 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)

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 financial resources.

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 and subdaily observations)

Goals for 2009-2011 : 6 million observations

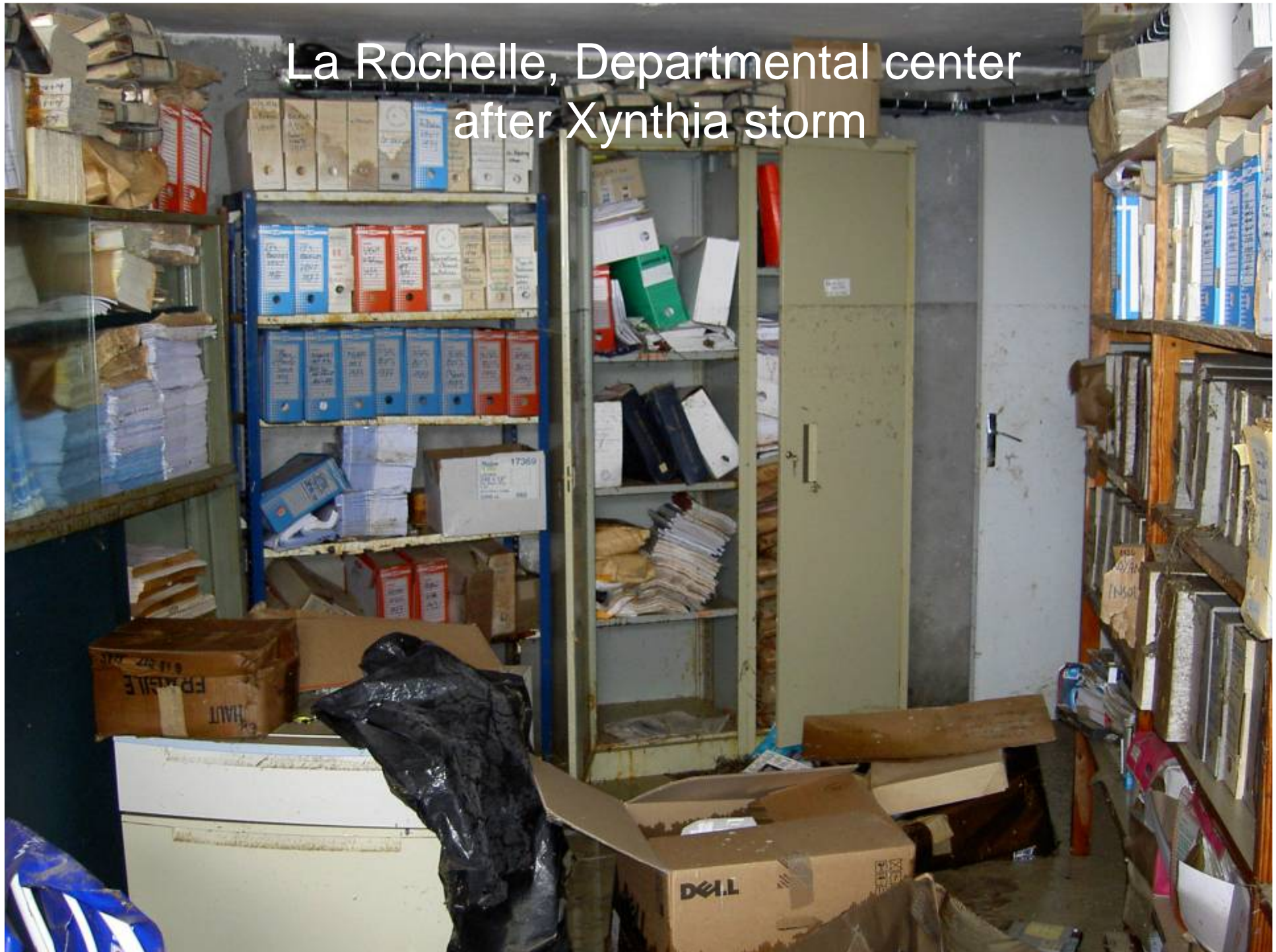
Preservation

- 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

La Rochelle, Departmental center
after Xynthia storm



Preservation

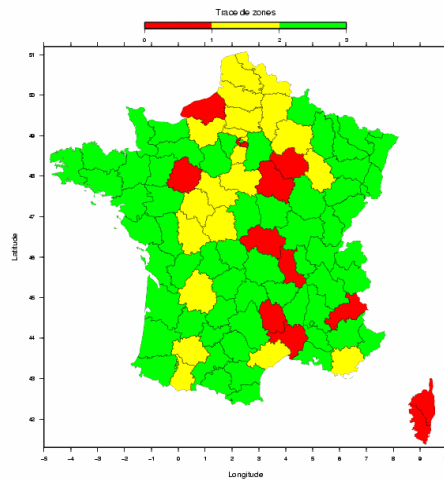
- Preservation of documents Documents are always in danger
 - Use acid free boxes to store records
 - digital imaging or scanning
 - Microfilms
 - Store documents in good air conditions ! Mould problems
 - Store in safe buildings
 - Flood risk prevention



Case of Paris-Montsouris archives last year

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 Climatometry 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
Should be finished in 2011

Inventory

- Gather all relevant metadata to trace a time series :
 - Information about location of the observations (name of the site, coordinates, ...)
 - time period, of the observations,
 - observed variables,
 - observation frequency and hours of observations,
 - information about how the data are available (digital, hardcopy) 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 and 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 complicated in France because material is scattered across a huge number of archives in France mainland, overseas and in other countries (for example DWD in Germany)
- French climatological archives in national archives : no access since 2004 because asbestos!
- 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 copy 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 publications,
 - 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 coordinated 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 digitization 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



The screenshot shows the 'Climatologie' section of a website. On the left, there is a 'Climathèque' sidebar with a list of links: 'Accueil', 'Catalogue des produits' (highlighted), 'Mon panier', 'Ma boîte de réception', and 'Tout sur les stations météo'. Below this is an 'Infos pratiques' section with links for 'Mode d'emploi' and 'Contact en cas de problèmes'. The main content area features a 'Catalogue 'CLIMATOLOGIE'' tree menu with the following structure:

- Données climatologiques de base
 - Données horaires
 - Données quotidiennes
 - Données décadaires
 - Données mensuelles
- Climatologie de référence
 - Fiches climatologiques
- Secteurs professionnels
 - Chauffagiste-Climaticien
 - Agriculteur
 - BTP (Bâtiments Travaux Publics)
- Hydrométéorologue

On the right side of the interface, there is a blue box with the text 'Catalogue des données et produits climatologiques'. Below it, a link reads 'Accès aux formulaires de " commande en ligne " ...'. Further down, there is a text instruction: 'Choisissez une catégorie de produits dans l'arbre sur la gauche de l'écran,' accompanied by three small images: a tree, a blue sky with clouds, and a close-up of a tree branch.

<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

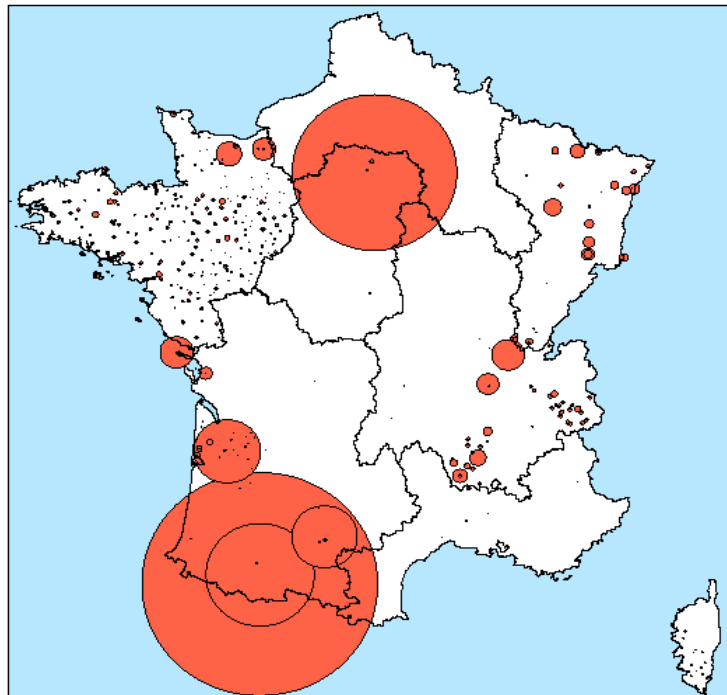
- station code : 67124001 City : Entzheim
- Périod : 01/10/1941 au 31/10/1944
- Table : Q (daily data)
- Parameters : RR – TN – TX
- Digitizer : Marlyse Colobert climatology department Entzheim
- Insertion mod : keying in CLIMSOL

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

Monitoring : map of inserted data (2009)



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



Homogénéisation : Why? La Rochelle example

- 1910 Primary school



- 1999 departmental center
Le Bout Blanc



Homogenization: Why?

Exemple de La Rochelle

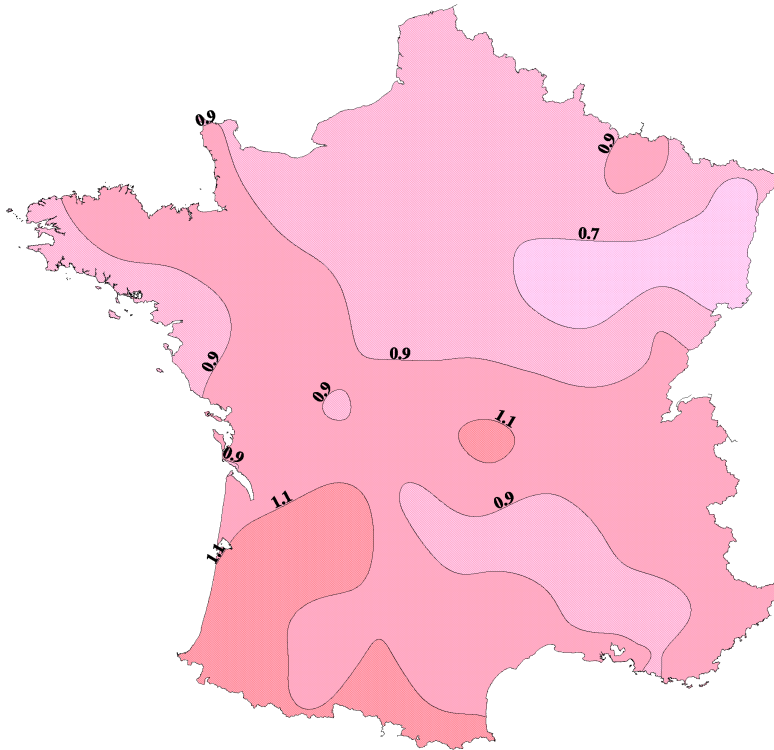
- La Rochelle Le Bout Blanc in 1995



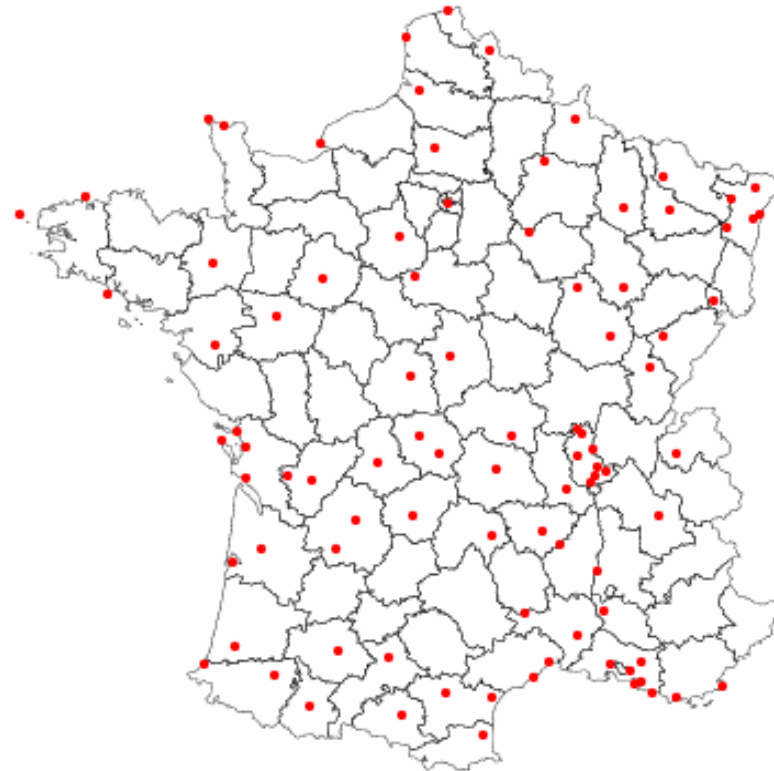
- Le Bout Blanc in 2004



Homogenized mean temperature series 1901-2000 Mestre(2004)



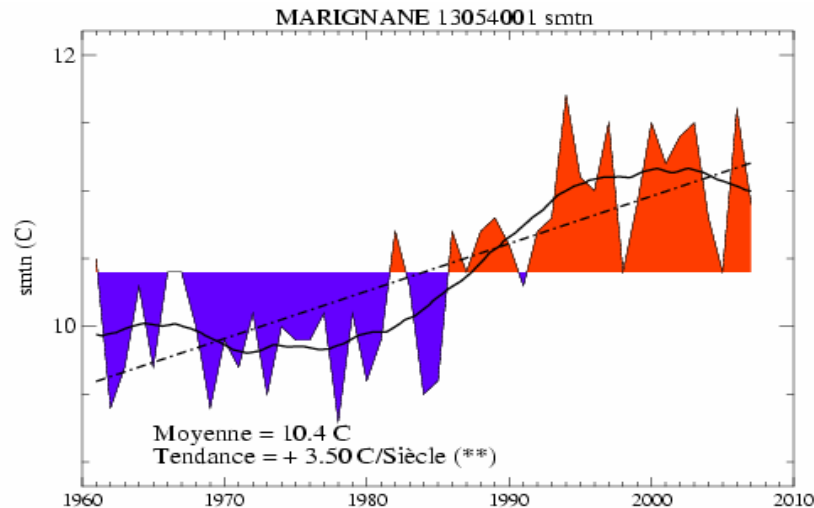
Trend in °C/century



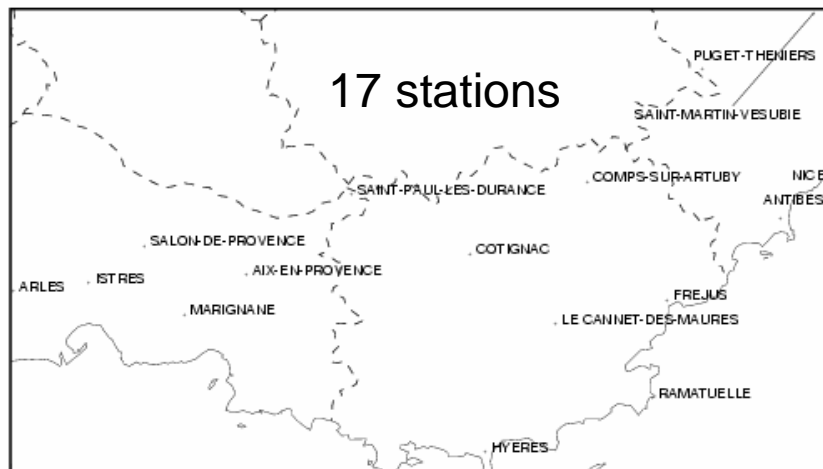
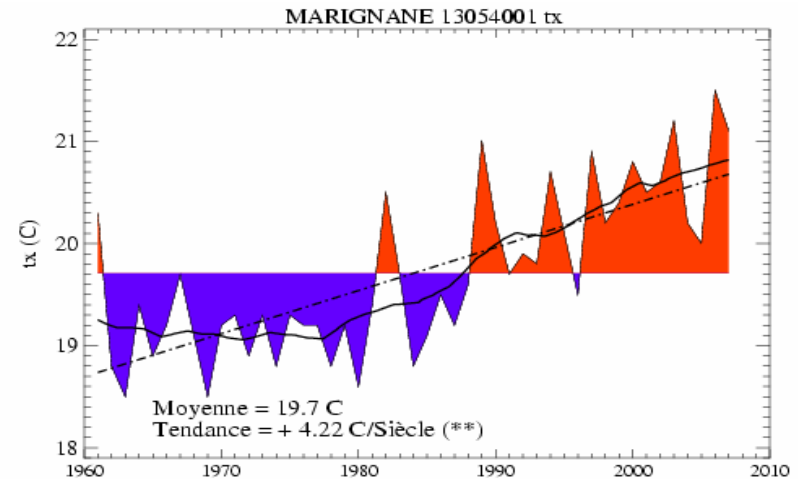
stations

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°C due to
Shelter change
and measurement automation

Homogenizing lessons

- Most of long-term series are heterogeneous, many methods to homogenize long-term series (see COST HOME action)
- Metadata collection 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

08

The screenshot shows the Météo-France web interface. At the top, there are radio buttons for "Métropole" (selected) and "Outremer". Below this is a map of France with a legend on the left showing "DIR-CE", "DIR-IC", "DIR-SE", "DIR-SO", and "France". A pop-up window titled "Département" is open, showing "13 - Bouches-du-Rhône" selected. Below the map, a dark blue bar contains the text "Étape 2: Choisir un paramètre, lancer la recherche, puis sélectionner une ou plusieurs séries". Underneath, there is a dropdown menu set to "Précipitations" and a "Rechercher" button. A table lists the available series for the selected department. To the right of the table, there are buttons for "Tout sélectionner", "Inverser la sélection", "Ne rien sélectionner", "Séries avant 1940", and "Séries après 1940".

Étape 2: Choisir un paramètre, lancer la recherche, puis sélectionner une ou plusieurs séries

Précipitations [v] Rechercher

Liste des séries (Précipitations) disponibles pour le département (13) :

Numéro	Nom de la station	Code évaluation	Début de série	Fin de série
13004001	ARLES-VILLE	0	01/1895	12/2005
13004002	ARLES-SALIN	0	01/1895	12/2005
13005003	AUBAGNE	0	01/1895	12/2005
13022003	CASSIS	0	01/1895	12/2005
13030001	CUGES-LES-PINS	0	01/1895	12/2005
13037001	LA FARE LES OLIVIERS	0	01/1895	12/2005
13047001	ISTRES	0	01/1895	12/2005
13050001	LAMBESC	0	01/1895	12/2005
13054001	MARIGNANE	0	01/1895	12/2005
13055001	MARSEILLE-OBS	0	01/1895	12/2005
13056002	CAP COURONNE	0	01/1895	12/2005
13060002	MEYRETH	0	01/1895	12/2005

Code évaluation (0 : qualité moyenne, 1 : bonne qualité)

Trier la liste des longues séries :
Par numéro [v]

Tout sélectionner
Inverser la sélection
Ne rien sélectionner
Séries avant 1940
Séries après 1940

Monthly homogenized long-term series product

Product description

Annual and monthly 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-Franczal 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

Année 1878.

Parc Saint Maur.

Dates	JUILLET					AOUT					SEPTEMBRE					OCTOBRE					NOVEMBRE					DÉCEMBRE				
	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol
1	16,2	24,4	17,30	0,0		14,5	25,5	20,05	-		12,0	20,2	16,00	3,5		9,0	15,7	12,35	5,0		0,9	6,5	3,35	-		0,8	8,5	2,20	3,8	
2	15,1	22,2	18,65	2,1		12,7	23,5	18,10			13,1	21,2	17,10	-		8,0	16,1	12,20	-		3,1	9,5	5,80	0,6		-0,4	4,5	2,02	6,5	
3	11,5	17,7	14,60	7,0		11,0	20,0	15,50	0,9		9,0	22,0	15,50	-		3,2	15,1	9,15	-		2,3	8,2	5,25	1,0		2,5	2,2	0,15	-	
4	7,7	20,0	14,05	-		10,5	25,0	17,95	-		10,5	23,5	17,00	-		3,9	17,6	10,65	-		0,0	6,7	3,35	5,1		-0,3	5,3	2,60	1,8	
5	8,0	24,7	16,35	-		12,5	27,0	20,20	0,4		12,1	25,5	19,55	0,1		6,0	18,5	13,10	-		1,0	10,0	6,00	0,0		1,0	10,0	6,00	0,0	

Année 1878

PARC S^t MAUR

Dates	JANVIER					FÉVRIER					MARS					AVRIL					MAI					JUIN				
	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol	Min	Max	Moy	Pluie	Insol
1	1,3	6,4	2,55			0,0	1,9	0,35			11,0	14,0	12,50	1,1		4,0	8,5	6,25	9,2		3,0	21,2	13,10	1,6		8,5	21,5	15,00		
2	2,0	1,0	-0,10			0,9	5,1	2,10	0,4		10,6	15,9	12,35	3,0		-1,6	11,9	5,45			10,8	19,5	13,30			8,5	19,1	13,80	3,7	
3	1,5	3,7	1,10			-1,0	5,8	2,40	0,3		5,1	13,8	8,45			1,2	12,5	7,10	4,8		7,8	18,8	13,30			13,1	22,5	17,70	2,1	
4	0,9	5,6	3,20	0,8		2,8	6,5	3,65			0,0	15,3	7,65			5,1	15,0	9,05	8,2		12,0	19,7	15,85			13,4	23,5	18,45	0,4	
5	-1,0	3,0	1,00			-2,0	2,5	0,25			0,6	11,1	5,25			2,5	14,2	8,55	0,7		6,0	21,5	14,25			15,0	20,0	16,60	1,1	
6	0,2	4,5	2,15	0,3		-2,4	4,8	1,20			2,3	11,9	7,10	0,6		-0,6	14,0	6,20			13,0	21,4	17,20	0,5		9,2	20,8	15,00		
7	1,5	7,4	5,00	0,8		-1,1	0,2	-0,45			5,7	13,2	10,95	0,9		1,1	13,3	8,45			13,0	15,6	14,10	11,7		8,5	24,1	15,55		
8	0,3	3,4	1,85	0,3		-2,9	1,1	-2,00			8,1	10,8	9,65	0,2		3,9	18,7	11,30			10,0	19,3	14,65	0,7		11,9	26,6	19,25	1,7	
9	0,0	2,0	1,00			-3,6	-1,5	-2,55			4,9	9,5	7,10			4,8	14,9	9,85			5,5	20,8	12,85			13,5	21,5	17,40	0,4	
10	-5,6	2,1	-1,70	1,8		-5,0	3,2	1,5			4,2	10,5	7,35	1,4		12,0	12,50	3,5			7,0	21,1	15,80			10,6	20,0	15,30	3,7	
11	-7,0	3,0	-2,00			1,6	9,2	5,40			7,1	12,1	9,60			1,7	14,1	11,50			10,0	22,7	16,05	3,8		9,4	22,7	16,05	3,8	
12	-9,1	-1,6	-5,35			1,0	7,7	4,35			7,0	12,6	9,80			1,7	14,1	11,50			12,6	19,4	16,00	5,0		12,6	19,4	16,00	5,0	
13	-2,3	7,9	4,20			0,0	12,6	6,30			0,6	7,0	5,30	1,0		5,4	19,1	10,75			11,5	20,8	16,15	1,6		10,3	19,3	15,00	3,2	
14	0,5	6,9	3,10			7,0	11,8	9,40	1,5		-0,9	8,8	5,35	0,6		5,9	22,7	14,30			10,4	20,8	15,60	5,6		11,6	17,7	14,65	3,6	
15	4,7	10,0	7,35			7,8	13,9	10,55			2,2	7,2	4,70	1,3		9,7	21,0	16,35			8,5	20,0	14,15	5,7		10,2	16,2	13,20		
16	7,1	9,3	8,10			7,0	12,0	9,50	1,3		-2,1	5,4	1,65			12,3	20,0	14,45	1,7		5,3	19,8	14,55			8,1	19,2	13,65	0,2	
17	6,5	8,3	7,35	0,8		5,5	12,0	11,25			-3,0	6,0	1,25			9,1	16,0	12,55	7,8		11,8	20,2	13,00			7,9	20,2	14,05	0,4	
18	3,1	8,2	5,65			7,3	12,2	10,05	2,1		1,0	11,9	6,45	0,9		3,5	16,7	12,10			15,2	27,5	11,35			9,0	20,9	14,35	0,5	
19	5,4	6,2	4,30			1,1	6,0	3,75			3,7	11,5	9,60	0,6		9,6	18,0	14,30	9,4		11,2	20,9	15,95			12,6	16,2	14,40	5,0	
20	-2,8	5,4	1,30			3,6	5,7	4,65			4,8	13,3	9,05			16,1	17,7	14,40	3,3		8,0	18,0	13,00	0,6		7,4	21,0	14,10		
21	-1,6	7,5	2,95			2,5	9,5	6,00			5,5	13,8	9,35			6,8	12,8	12,80	0,3		7,7	15,8	11,75			15,1	26,7	13,30		
22	7,5	11,2	9,35	0,8		1,9	10,5	6,10			5,8	13,9	9,85	5,6		5,7	19,1	11,40			3,1	18,0	10,90			15,1	26,7	13,30		
23	7,5	9,5	8,50	3,6		1,6	10,7	7,10			1,2	5,9	3,35	5,6		8,0	19,1	13,55	0,1		7,5	14,9	11,70	10,1		16,3	28,1	22,20	13,5	
24	2,4	5,5	3,95	10,1		1,5	6,3	5,90			-2,0	5,4	1,10	3,7		11,6	19,0	13,80	0,2		10,7	20,7	14,95			15,2	25,5	19,50		
25	1,7	4,6	3,10	6,1		5,4	10,6	7,80	0,7		-1,0	5,2	2,10	1,9		9,2	12,5	13,25			8,8	15,5	12,15	3,3		13,2	25,5	19,50		
26	-1,1	4,8	1,85			6,2	10,8	8,30	0,1		-0,8	5,6	3,90	0,1		8,0	13,9	14,95	2,0		11,9	19,6	11,05	1,8		16,1	28,7	21,40		
27	-2,3	2,5	-0,45			3,6	12,1	9,35	2,0		-1,0	11,0	5,15			5,1	10,7	10,00	0,2		8,6	19,9	14,25			15,2	29,1	22,15		
28	-1,0	4,3	1,65	4,1		3,4	13,0	11,50	0,3		2,1	6,6	4,70	5,0		6,0	20,5	15,15			6,6	18,3	10,95	11,1		16,0	27,9	11,95		

Thank you for attention

rescue@meteo.fr

Sylvie.Jourdain@meteo.fr