

SISCLIMA : capitalisation de données agroclimatiques au Brésil

Daniel CARDON ⁽¹⁾, Patrick SECHET ⁽²⁾
Tatiana D. de A. S. DINIZ ⁽³⁾, Maria Gil M. MALTEZ ⁽⁴⁾

RÉSUMÉ

Dans le cadre de son système d'informations géographiques sur l'environnement en milieu rural (SISGEO), l'entreprise Brésilienne de Recherches Agronomiques (EMBRAPA) a mis au point un système de capitalisation d'informations agroclimatiques (SISCLIMA). Celui-ci a comme objectif la fourniture automatique des produits les plus classiques dans ce domaine (bulletin météorologique, calculs d'évapotranspiration, etc.), à partir des données brutes collectées sur une station météorologique. Le module comprend une vérification systématique détaillée des données saisies et leur correction, et fournit un fichier opérationnel de données élaborées au niveau journalier. Ce travail rend compte des critères climatologiques adoptés et des méthodes informatiques utilisées pour sa réalisation.

MOTS-CLÉS : Climatologie – Banque de données – Logiciel de traitement – Vérification.

ABSTRACT

SISCLIMA : PROCESSING OF AGROCLIMATOLOGICAL DATA IN BRAZIL

Within the framework of its geographical environmental system (SISGEO), the Brazilian Agency for Agricultural Research (EMBRAPA), has developed a subsystem (SISCLIMA), for the storage and processing of agroclimatological data. This work has been performed jointly by climatologists of the Humid Tropics Centre of Agricultural Research (CPATU) or research centre of the EMBRAPA in the Amazon region, analysts of the Quantitative Methods Department of the EMBRAPA (DMQ) and with the support and participation of a climatologist and an analyst of the French Institute of Research and Development in Cooperation (ORSTOM). SISCLIMA permits the usual operations of agroclimatology (printing of agroclimatological bulletins, evapotranspiration calculation, etc.) from raw data collected at meteorological stations. After detailed quality control and possible corrections, these data are stored on a daily basis which is used for processing. In the version presently available SISCLIMA is only used with data collected by the CPATU in the Amazon region, but another is being prepared to allow processing of EMBRAPA data for the whole of Brazil. This paper presents the climatological requests and the processing techniques used for the implementation of the system.

KEY WORDS : Climate – Data base – Processing system – Quality control.

RESUMO

SISCLIMA : ARMAZENAMENTO E RECUPERAÇÃO DE DADOS AGROCLIMATOLÓGICOS NO BRASIL

No âmbito do seu sistema de informações geoambientais (SISGEO), a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) desenvolveu um módulo de armazenamento de dados agroclimatológicos (SISCLIMA). Este trabalho foi realizado em conjunto pelos climatologistas do Centro de Pesquisa Agropecuária do Tropicó Umido (CPATU) ou centro de pesquisa agrônoma da EMBRAPA em região amazônica, os analistas do Departamento dos Métodos Quantitativos

(1) ORSTOM, 213, rue de la Fayette, 75480 Paris.

(2) EMBRAPA (sede) Supercenter Venâncio 2000 – 70333 Brasília Brésil.

(3) CPATU CP 48 – 66000 Belém Brésil.

(4) UFPa campus Universitario do Guama – 66000 Belém Brésil.

(DMQ) e a participação de um climatologista e de um analista do Instituto Francês de Pesquisa e Desenvolvimento em Cooperação (ORSTOM). O sistema tem como finalidade a realização automática das operações mais comuns em agroclimatologia (impressão do boletim meteorológico, cálculo da evapotranspiração, etc.) a partir dos dados brutos coletados em estações meteorológicas. Tal sistema permite uma verificação pormenorizada da consistência dos dados captados, assim como as devidas correções, e produz um arquivo operacional de dados elaborados a nível diário. Na versão atual o SISCLIMA está sendo utilizado para a administração dos dados coletados em algumas estações mantidas pela EMBRAPA na região amazônica, estando em preparação uma outra possível de ser generalizada para todos os centros da EMBRAPA no território brasileiro. O trabalho a seguir apresenta as diretrizes climatológicas de realização, e as técnicas de processamento adotadas para a implementação do sistema.

PALAVRAS CHAVAS : Climatologica – dados – processamento – consistência.

INTRODUCTION

Dès 1983 le CPATU (Centro de Pesquisa Agropecuária do Tropicó Umido), centre de recherches de l'EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) en Amazonie brésilienne, avait exprimé la nécessité de créer localement un système de données agroclimatiques informatisées. Celui-ci devait permettre à un non-informaticien d'effectuer de façon automatique le stockage et la vérification des données ainsi que leur restitution sous une forme élaborée : bulletin météorologique, annuaire, calculs d'évapotranspiration, etc. Dans le même temps, à l'échelle brésilienne cette fois, était en cours de réalisation à l'EMBRAPA un Système d'Informations Géographiques (SISGEO), dont un sous-système SISOLOS (Système d'Informations Pédologiques du Brésil), était déjà opérationnel (MENEGUELLI & *al.* 1983 – MENEGUELLI & SECHET 1984). Le stockage de l'information agroclimatique était une des prochaines étapes prévues dans la réalisation de SISGEO. SISCLIMA est la résultante de ces deux directions de recherche. Il constitue un sous-système de SISGEO donc, sur le plan géographique, son application est à l'échelle régionale. Il a été créé et actuellement fonctionne en Amazonie brésilienne ce qui marque fortement sa réalisation. En effet, dans cette région, les données sont parfois sujettes à caution, or elles sont trop rares pour être rejetées facilement, SISCLIMA a donc été axé en priorité sur leur vérification.

Le stockage et le traitement informatique des données agroclimatiques est un problème auquel de nombreux utilisateurs ont été confrontés. Ce fut le cas de MILDE (1985) au Brésil, dans le Sud de l'État de Bahia, avec le système CADE. Ces logiciels répondent le plus souvent à des besoins particuliers, de plus ils sont en général adaptés aux moyens informatiques locaux. Leur généralisation apparaît donc difficile à réaliser. Parfois aussi ils sont placés en aval de une ou plusieurs stations météorologiques automatiques, or, actuellement au Brésil, de telles stations sont expérimentales et encore, dans le sud du pays. Il existe également des logiciels d'emploi plus généraux tels CLICOM créé par l'Organisation Météorologique Mondiale (1986). Dans ce cas, la limitation se situe au niveau de la configuration informatique minimum indispensable, très au-delà des possibilités des centres de l'EMBRAPA en général. Ainsi CLICOM requiert un micro-ordinateur type IBM PC de 512 kbytes de mémoire interne et équipé d'un disque dur de 20 Mbytes. Dans ces conditions la création d'un système propre à la région est apparu comme une nécessité et, par analogie avec SISOLOS, on l'a appelé SISCLIMA.

Par l'intermédiaire de SISCLIMA on se propose de rationaliser la collecte, la saisie et l'exploitation des données agroclimatiques. Or la composition d'une station météorologique, tout comme le type de capteur utilisé, sont susceptibles de variations importantes dans la région considérée. C'est là une des grandes difficultés à surmonter dans la réalisation d'un tel système : créer une information homogène à partir de données de base hétérogènes. Afin de mieux cerner ce problème, tout le premier paragraphe est consacré à la description des données. Ensuite SISCLIMA a été présenté suivant deux points de vue : celui du climatologue et celui de l'informaticien. Le point de vue du climatologue c'est en quelque sorte les directives qu'il donne à l'informaticien après analyse de son problème. Deux paragraphes lui sont consacrés. Dans le paragraphe 2 on décrit les produits attendus de SISCLIMA et, compte tenu de l'importance qui lui a été accordée, le paragraphe 3 se rapporte en entier au contrôle de la qualité des données. Dans le dernier paragraphe, l'informaticien expose les solutions techniques adoptées pour résoudre ce problème.

1. LES DONNÉES

Dans une station météorologique sont effectuées des mesures permettant de connaître un certain nombre de paramètres climatiques. Ce sont ces mesures ou données brutes qui nous intéressent ici. Le but de ce paragraphe n'est pas la description des données brutes en soi mais la mise en évidence des problèmes posés par leur saisie

et leur traitement en informatique. Dans cet esprit on a décrit le fonctionnement d'une station météorologique et les paramètres qu'on peut y obtenir. Les mesures ont ensuite été classées suivant le type de relation les unissant aux paramètres recherchés.

1.1. DESCRIPTION GÉNÉRALE ET FONCTIONNEMENT D'UNE STATION MÉTÉOROLOGIQUE

Les mesures sont effectuées dans un parc météorologique engazonné, suivant un ordre déterminé. Il est cependant arrivé que certaines séries de données aient été obtenues en dehors du parc. Par exemple les données de rayonnement fournies par le CPATU à Belem de 1978 à 1982 provenaient en fait d'une station voisine (2 km). On peut éviter ce genre de choses pour l'avenir, mais pour les données historiques il ne reste qu'à le mentionner.

Un parc météorologique est composé d'un abri météorologique et d'un certain nombre de capteurs disposés à l'air libre ou dans le sol. L'influence de l'abri sur les mesures est très faible mais non nulle. Il importe donc de signaler le type d'abri en service. Enfin, pour des raisons pratiques, certaines mesures sont parfois effectuées sans couverture végétale (température dans le sol, évaporation du bac classe A). Ceci est surtout vrai dans la zone sèche du Brésil, mais dans tous les cas l'utilisateur potentiel doit être informé de cette particularité.

Si on excepte les diagrammes de rayonnement dépouillés journallement, les mesures dans une station météorologique s'effectuent trois fois par jour. Dans la plus grande partie du Brésil l'horaire de ces mesures est : 9, 15, et 21 heures. Pour l'état d'Amazonie par contre c'est 8, 14, et 20 heures. Si on ajoute que les différentes heures légales sont appliquées suivant un découpage politique du pays sensiblement différent des fuseaux horaires, on se rend compte que l'heure réelle de collecte des données varie passablement. Ces informations doivent également figurer dans le système.

1.2. PARAMÈTRES CONSIDÉRÉS

La liste du tableau I correspond à une situation idéale, les stations météorologiques étant généralement incomplètes. Certaines grandeurs, pourtant importantes, telles l'évapotranspiration réelle par lysimètre pesable, n'y figurent cependant pas. C'est parce qu'on ne les rencontre jamais mesurées de façon continue dans la région considérée.

TABLEAU I
Paramètres obtenus dans une station

TEMPERATURES	- maximum (*) - minimum (*) - sèche (psychromètre) (*) - humide (psychromètre) (*) - dans le sol (5, 10, 20, 30, 50, 100 cm) - enregistrée (*)
HYGROMETRIE	- enregistrée (*) - (voir psychromètre) *
RAYONNEMENT	- enregistré (actinographe) - mesuré (solarimètre) - durée d'insolation
PLUVIOMETRIE	
EVAPORATION	- évaporomètre Piche (*) - bac classe A
ANEMOMETRIE	- vitesse instantanée du vent - vitesse moyenne du vent - direction du vent

(*) Paramètres obtenus à partir de mesures sous abri.

1.3. TYPES DE MESURES

Les paramètres recherchés ne sont pas tous mesurés directement et, pour les obtenir, il faut faire subir aux données brutes un certain traitement. Parfois ce traitement fait appel à des données complémentaires (coefficient d'étalonnage, tables, etc.), qu'il faut avoir introduit dans la machine. Il arrive également que le type de capteur associé à une même donnée brute, et donc le traitement à faire subir à cette donnée, varient avec la station (anémométrie).

2. PRODUITS ATTENDUS DE SISCLIMA

Suivant que l'on se place du point de vue de celui qui collecte les données, ou de celui qui les utilise, les résultats attendus sont différents. Le premier désire constituer un fichier facilement utilisable de données sûres (fichier opérationnel), tandis que le second est intéressé par les produits issus de l'utilisation de ce fichier.

2.1. CONSTITUTION DU FICHER OPÉRATIONNEL

Pour créer ce fichier il faut passer par différentes étapes. La première est évidemment la collecte des données sur le terrain ou données brutes, suivie par l'informatisation avec une phase éventuelle de retranscription. Ces données doivent ensuite être vérifiées, parfois corrigées et alors seulement elles seront traitées en informatique pour constituer le fichier opérationnel.

Ce fichier est constitué de données journalières. Elles sont de trois types :

- des paramètres climatiques obtenus en traitant les données brutes,
- des paramètres astronomiques,
- des informations complémentaires concernant la station météorologique.

Le passage des données brutes aux paramètres climatiques journaliers se fait en deux étapes. Tout d'abord on passe des grandeurs mesurées aux valeurs instantanées des paramètres climatiques. Suivant le type de mesure considéré dans cette phase interviennent les caractéristiques des capteurs et leurs coefficients d'étalonnage. Dans le cas de la pluviométrie, du rayonnement, du vent moyen et des mesures d'évaporation, cette phase préliminaire se rapporte à la période séparant deux heures de collecte successives. Ensuite on calcule les moyennes journalières ou les sommes dans le cas des quatre derniers paramètres cités. Dans toute cette partie les calculs sont effectués en suivant les normes de l'Institut Météorologique Brésilien (INEMET) ; ainsi le « jour » en cours est la période séparant la dernière heure d'observation de la veille à la dernière heure d'observation du jour même. Font également partie de cette catégorie de données certaines grandeurs climatiques plus complexes à calculer, mais d'emploi courant dans les formules d'évapotranspiration, telles la tension moyenne de vapeur d'eau.

Dans les applications agroclimatiques, la durée du jour et la quantité de rayonnement arrivant au sommet de l'atmosphère interviennent constamment : évapotranspiration suivant PENMAN (1948), estimation du rayonnement global à partir de la durée d'insolation d'après ANGSTRÖM (1924), etc. C'est pourquoi ces deux paramètres ont été annexés au fichier opérationnel. Leur calcul fait intervenir dans des formules très complexes la latitude du lieu et la date. De nombreux auteurs ont expliqué comment les obtenir. Dans SISCLIMA on a utilisé les formules présentées par BAILLE & al. (1982), en appliquant la valeur de la constante solaire préconisée par l'Organisation Météorologique Mondiale à sa réunion de 1981.

Au paragraphe 1.1. il a été vu que dans certains cas les données informatisées pouvaient avoir été collectées en dehors de la station. Il est arrivé que, pour une période, des capteurs ne soient pas installés exactement suivant les normes. Ceci peut entraîner des anomalies au niveau des applications et, pour les interpréter, il faut disposer de ces informations. Ces données complémentaires seront introduites sous forme d'un code littéral associé à chaque donnée d'un autre type. Les situations rencontrées sont très diverses et on ne peut prétendre en dresser la liste exhaustive. L'utilisateur doit donc pouvoir établir cette liste lui-même, en la modifiant chaque fois qu'un nouveau cas se présente.

Dans certaines applications (bilan hydrique) on a besoin de séries de données complètes. Les séries de mesures comportent souvent des valeurs manquantes (panne de capteur, absence de l'observateur, etc.). Dans un certain nombre de cas le système permet la reconstitution de données faisant défaut dans le fichier opérationnel. Lorsque les paramètres considérés ont été obtenus de deux façons différentes (température, humidité relative), on substitue la valeur enregistrée à la valeur calculée et vice versa. Si deux paramètres sont étroitement corrélés, on reconstitue le premier à partir du second. C'est ce qui se produit entre la durée d'insolation (S) et le rayonnement global (G) lorsqu'on applique la formule d'ANGSTRÖM (1924) : $G/G_0 = a * S/S_0 + b$, où G_0 et S_0 sont respectivement la durée du jour et le rayonnement au sommet de l'atmosphère, et a et b deux coefficients empiriques spécifiques de la station considérée. Il faut pouvoir différencier les données reconstituées des autres, ceci est effectué automatiquement en donnant la valeur « R » au code associé à la donnée reconstituée.

2.2. UTILISATION DU FICHER OPÉRATIONNEL

L'utilisation la plus simple, mais aussi la plus recherchée, est la constitution de bulletins météorologiques mensuels comportant la somme, la moyenne, l'écart-type et les valeurs extrêmes d'un certain nombre de paramètres. Deux types de bulletin sont édités : un bulletin « standard » comportant les paramètres les plus usuels, et un

bulletin « à la carte » où l'utilisateur choisit lui-même les paramètres qu'il désire voir figurer. Enfin il a également été donné la latitude d'exclure des calculs certaines données en fonction des valeurs prises par les codes littéraux leur correspondant. Se rattache à ce type de produit l'édition d'annuaires, autrement dit l'ensemble des valeurs annuelles d'un paramètre avec les mêmes calculs et les mêmes facilités que précédemment.

Une autre application est l'estimation de l'évapotranspiration sur une période donnée. Onze méthodes, parmi les plus utilisées, ont été retenues. Leur justification et leur description ont été tirées de l'ouvrage présenté par DOORENBOOS & PRUITT (1975) ou des publications de leurs auteurs respectifs. Il s'agit de la formule de PENMAN (1948), de la méthode de radiation, de celle du bac classe A, de THORNTHWAITE, de JENSEN & HAISE, de PRIESTLEY & TAYLOR, de LINACRE, de BLANEY & CRIDDLE (1962), de BÉNAVIDES & LOPEZ (1970), de HARGREAVES (1973) et de la formule de TURC (1961). Dans un premier temps le système interroge l'utilisateur pour savoir la période de travail (jour, pentade, semaine, décade ou mois). Il sélectionne ensuite les méthodes compatibles avec la réponse, les fait apparaître sur la vidéo du micro-ordinateur et demande celle qui a été choisie. Si nécessaire le système réclame quelques informations complémentaires telles que l'albedo dans le cas de la formule de Penman. Si aucune précision n'est apportée le système prend alors une valeur par défaut (0,25 dans le cas de l'albedo).

Les applications présentées sont actuellement disponibles avec le système. Leur liste est loin d'être close et, en collaboration avec l'utilisateur, de nombreuses autres possibilités restent à définir.

3. CONTRÔLE DES DONNÉES

La constitution du fichier opérationnel s'est faite par un certain nombre d'étapes. Au cours de chacune d'elles a été instauré un contrôle de la qualité des données, ce contrôle a pris deux formes : la réduction des causes d'erreur et leur détection et correction.

3.1. RÉDUCTION DES CAUSES D'ERREURS

Chaque intervention humaine est génératrice de son lot d'erreurs. Il convient d'en réduire le nombre et de les rendre moins graves. Ces interventions sont les suivantes : collecte des données, retranscription et prétraitement éventuels, numérisation. On ne peut échapper ni à la collecte ni à la numérisation. Par contre, l'emploi de formulaires adaptés au traitement informatique facilitera grandement la saisie qui doit pouvoir être effectuée par l'observateur météorologique lui-même. Pour faciliter son travail, l'image apparaissant sur l'écran du micro-ordinateur est la réplique la plus fidèle possible du formulaire. D'autre part l'ordre de saisie est identique à celui de la collecte. Enfin, la fraction décimale étant placée par programme dans les données, l'opérateur ne devra saisir que des données entières ou des données littérales. Une partie du bordereau réalisé pour SISCLIMA est présentée fig. 1.

EMBRAPA
CPATU-DMQ
SISCLIMA

BOLETIM DE OBSERVAÇÕES DIÁRIAS

NOME DA ESTAÇÃO _____ DATA _____
Dia _____ Mes _____ Ano _____

OBSERVAÇÃO :

LISÍMETROS (1) (2) (3) 9h (3) MEDIDAS DE RADIAÇÃO (Insolacão) (Actinog) RESPONSÁVEIS 9h 15h 21h

OBSERVAÇÕES DE 09:00H

ABRIGO METEOROLÓGICO

TS _____ TU _____
TX _____ TN _____
REGISTRADORES (UR) _____
NÍVEIS EV PICHE (NV1) (NV2) _____

VENTO (VL) _____ DIR _____

TANQUE CLASSE A (VT) _____
NÍVEIS DO TANQUE (NV1) (NV2) _____
TEMPERATURAS (TX) (TN) _____

CHUVAS (CHU) _____ RADIAÇÃO (RG) _____

TEMPERATURAS NO SOLO (T2) (T5) (T10) _____
(T20) (T30) _____
(T50) (T100) _____

Fig. 1. – Bordereau de saisie des observations journalières (détail)

3.2. DÉTECTION ET CORRECTION DES ERREURS

Elle se fait à trois niveaux : la saisie, la vérification et la constitution du fichier opérationnel.

En cours de saisie, seule la nature informatique de la donnée sera analysée. Si l'opérateur tente de placer un caractère autre que numérique (une virgule par exemple) dans un entier, un signal sonore retentit et la saisie ne pourra reprendre que si un caractère numérique est entré en remplacement.

La critique véritable s'effectue à l'aide d'un programme spécialisé, toute une série de tests étant appliqués aux données brutes. Ces tests sont de quatre types :

type 1 : on vérifie que la valeur de toutes les données est bien comprise entre deux valeurs extrêmes,

type 2 : lorsqu'il existe une relation entre deux ou plusieurs données, on utilise cette relation pour vérifier que les valeurs de ces données sont compatibles,

type 3 : dans le cas de données différentielles (niveaux du Piche et du bac, enregistreurs de vent ou de rayonnement) on compare les valeurs de données chronologiques,

type 4 : lorsque la même grandeur est obtenue de deux façons différentes (mesure et enregistrement) on compare les résultats obtenus.

Ces différents types de tests n'ont pas la même puissance. Les tests de type 1 sont, et de loin, les plus imprécis. Si une donnée est infirmée par un test de type 1, dans les tests de type 2, 3, 4 où elle interviendra, elle risque de l'être également, ainsi que toutes les autres données auxquelles elle sera comparée. Pour éviter ce problème le programme de vérification n'applique les tests de type 2, 3, 4 que si aucune des données comparées n'a été infirmée auparavant dans un test de type 1.

Pour affiner les tests, les valeurs de comparaison doivent être adaptées à chaque station. Une température maximum de 40 degrés C à Rio de Janeiro est parfaitement admissible, à Belém elle est impensable. Tant qu'une station est peu connue on doit accepter une amplitude de variation très grande pour les données. Mais ce n'est plus le cas lorsqu'on est mieux renseigné. Le système doit donc permettre à l'utilisateur de fixer et de modifier les limites de variation correspondant aux différents tests.

Dans le tableau II on donne la liste des tests de type 1 appliqués, le symbole et l'unité des données traitées et les limites extrêmes entre lesquelles ces données sont autorisées à varier. On peut ramener tous les tests à des tests de type 1, soit à vérifier si une grandeur est comprise entre deux valeurs extrêmes. La seule différence est que, dans le cas de tests de type 2, 3, ou 4, la grandeur testée n'est plus une donnée brute mais une expression plus complexe, où interviennent différentes mesures. Dans le tableau III on considère tous les tests de type différent de 1. Sont indiqués le type du test, le symbole des données y intervenant, la grandeur testée et les limites entre lesquelles cette grandeur peut varier. Dans le cas de tests de type 3 les symboles correspondant aux données brutes ont été affectés de l'indice « *i* » pour l'observation de l'heure considérée, et de l'indice « *i-1* » pour l'observation antérieure. Enfin on utilise dans les tableaux II et III le symbole « *x* » chaque fois que l'on désigne une limite ou un paramètre que l'utilisateur peut choisir.

TABLEAU II
Description des tests de type 1

Nom de la donnée testée	Symbole	Unité	Limite	
			Mini.	Maxi.
Température sèche (m)	T _{s_m}	°C	XXXX	XXXX
Température sèche (e)	T _{s_e}	°C	XXXX	XXXX
Température humide	Th	°C	XXXX	XXXX
Température maximale	Tx	°C	XXXX	XXXX
Température minimale	Tn	°C	XXXX	XXXX
Températures dans le sol	T _{nnn}	°C	XXXX	XXXX
Humidité relative (m)	Hr _m	%	XXXX	XXXX
Humidité relative (e)	Hr _e	%	XXXX	XXXX
Totalisateur de vent	Chvt	Km	0	XXXX
Vitesse du vent	V	Km/h	0	XXXX
Niveau du Piche	P	mm	0	400
Niveau du bac classe A	ClA	mm	0	300
Durée d'insolation	S	h	0	So
Radiation globale *	G*	ly	0.1*Go	0.8*Go
Radiation globale **	G**	ly	0.1*Go	0.8*Go
Pluviométrie	PP	mm	0	XXXX

(m) : mesuré

(e) : enregistré

nnn : profondeur de la mesure

(002cm, 005cm, 010cm, 020cm, 030cm, 050cm, 100cm)

* : solarimètre

** : actinographe

TABLEAU III
Description des tests de type 2, 3 et 4
(mêmes notations que pour le tableau II)

Tests de type numéro 2			
Données testées	Grandeur testée	Limite inf.	Limite sup.
Température maximale (TX) Température minimale (TN)	TM - Tm	0	XXXXXX
Température sèche mesurée (Tsm) Température maximale (TX) Température minimale (TN)	Tsm	TN	TX
Température sèche enreg. (Tse) Température maximale (TX) Température minimale (TN)	Tse	TN	TX
Radiation globale * (G*) Durée d'insolation (S)	a+b+S/So - G*	-0.2*Go	0.2*Go
Tests de type numéro 3			
Données testées	Grandeur testée	Limite inf.	Limite sup.
Totalisateur de vent (ChVt)	ChVt _i - ChVt _{i-1}	0	XXXXXX
Niveau du Piche (P)	P _i - P _{i-1}	0	XXXXXX
Niveau du bac classe A (ClA) Pluviométrie (PP)	ClA _{i-1} - ClA _i + PP	0	XXXXXX
Radiation globale * (G*)	G* _i - G* _{i-1}	0	XXXXXX
Tests de type numéro 4			
Données testées	Grandeur testée	Limité inf.	Limite sup.
Température sèche mesurée (Ts _m) Température sèche enreg. (Ts _e)	Ts _m - Ts _e	-XXXXXX	XXXXXX
Humidité relative mesurée (Hr _m) Humidité relative enreg. (Hr _e)	Hr _m - Hr _e	-XXXXXX	XXXXXX
Radiation globale * (G*) Radiation globale ** (G**)	G* - G**	-XXXXXX	XXXXXX

* (mêmes notations que pour le tableau II)

Lorsqu'on analyse un fichier de données, un rapport est émis donnant la liste des valeurs à vérifier. Dans le cas de tests de type 1 une seule donnée est considérée, le problème est simple. Dans tous les autres cas il faut vérifier plusieurs données par anomalie rencontrée. Par exemple si une température est supérieure à la température maximum, il faut vérifier les deux données pour découvrir qu'en général une seule est erronée. Afin de faciliter le travail de critique le type de test a été annexé à la donnée qu'il désigne dans le rapport de sortie.

Une fois en possession de ce rapport on effectue la vérification de chaque donnée présumée douteuse. Quatre cas sont possibles :

- 1 - la donnée est absurde, l'erreur n'a pas été découverte,
- 2 - la donnée est erronée, l'erreur est décelée,
- 3 - la donnée est douteuse, aucune erreur n'a été découverte,
- 4 - la donnée est acceptable.

Dans le premier cas la donnée est supprimée, dans le second elle est corrigée et dans le dernier aucune intervention n'est nécessaire. Dans le cas numéro 3 (le plus fréquent) il faut signaler la donnée comme douteuse. Ceci est permis par l'association à la donnée considérée d'un code littéral approprié dans le fichier opérationnel.

La correction ou l'invalidation sont des opérations effectuées par un opérateur. Elles peuvent être omises ou mal faites. Pour éliminer le risque de laisser subsister une erreur grossière, une dernière vérification est effectuée automatiquement lors de la constitution du fichier opérationnel. A chaque donnée on applique un test de type 1 mais avec des limites suffisamment larges pour que seules les données absurdes soient distinguées. Elles sont alors éliminées.

4. RÉALISATION DU SYSTÈME

Conformément à la méthodologie de travail adoptée par l'équipe d'analystes de SISGEO, la réalisation du système informatique envisagé par les climatologues de Belém passe par un processus en quatre étapes distinctes : conception, spécification, programmation et implémentation.

4.1. MODÉLISATION DU SYSTÈME

L'objectif essentiel de la première phase consiste à produire un modèle conceptuel du système idéalisé par l'utilisateur, en définissant rigoureusement les fonctions incorporées, les mouvements et transformations de données, les personnes qui interviennent et le rôle respectif de chacune d'elles.

Cette étape est extrêmement délicate dans la mesure où elle repose sur un dialogue difficile entre l'analyste, ignorant de climatologie et le scientifique, peu au fait des techniques informatiques. Au cours de cette étape, il convient de plus d'éviter soigneusement de se préoccuper de détails liés à la réalisation physique, il est au contraire important de conserver une vision hiérarchique de l'ensemble des mécanismes qui composent le système à mettre au point.

La technique d'analyse structurée de systèmes (GANE & SARSON 1979), constitue un guide précieux pour la réalisation de cette étape du travail. Les diagrammes de flux de données (DFD), dessinés par l'analyste dès les premières réunions avec l'utilisateur, constituent très rapidement le support privilégié de ce dialogue. La simplicité de cet outil graphique, qui matérialise les flux de données à prévoir dans le système, repose sur le faible nombre de symboles utilisés (un carré pour une entité externe, un rectangle pour une fonction du système et une bande rectangulaire pour un dépôt de données).

On a représenté sur la figure 2 le DFD de niveau 0 correspondant à SISCLIMA, tel qu'il a été modélisé par l'équipe mixte d'analystes et de climatologues. A ce niveau, le plus général de la conception, sept fonctions ont été identifiées pour schématiser l'ensemble des opérations requises par les climatologues, et présentées précédemment.

Plusieurs notions sont ainsi introduites ou précisées et constituent autant de solutions proposées par les professionnels du traitement de données aux problèmes posés par les thématiciens :

- la notion de tables du système, fichiers où sont rangées toutes les données de correspondances indépendantes de la station (codification, conversion, valeurs de coefficients, etc.), apporte souplesse et garantie d'intégrité au système. Cette information est accessible en lecture par les climatologues qui administrent les stations : seul le responsable scientifique, spécialiste qui définit et gère le système, est autorisé à la mettre à jour ;

- l'introduction d'un fichier de caractéristiques de station satisfait à la nécessité de conserver à tout moment l'état d'équipement de la station. Les valeurs limites propres à chaque paramètre, utilisées pour la validation et estimées par le climatologue responsable de la station considérée, y sont également stockées, de telle sorte qu'elles deviennent également modifiables ;

- il est apparu nécessaire de dissocier l'opération de validation de celle de saisie, de façon à permettre la saisie des trois horaires journaliers par un opérateur, généralement l'observateur lui-même. Trois niveaux d'accès (opérateur, climatologue et responsable scientifique) sont ainsi mis en évidence dans le système, le contrôle correspondant est ainsi réalisé par un jeu de « mots de passe ». La philosophie de traitement par lots correspondant à la gestion des observations instantanées n'est perçue que sur le DFD de niveau 1 de cette fonction (fig. 3). La phase de critique est une opération automatique qui se solde par l'émission d'un état de données suspectes, analysé par le climatologue, les corrections qui pourront être déduites de cette analyse sont introduites en utilisant l'étape de saisie, fermant de cette façon le cycle bien connu saisie-validation-correction ;

- la distinction entre fichier de données brutes et fichier opérationnel est fréquente dans les systèmes de stockage d'information : elle fait appel à des notions de fiabilité, cohérence, sécurité, restriction d'accès, niveau d'agrégation, etc. Pour SISCLIMA, le fichier brut est le fichier des observations horaires, tel qu'il est produit par l'étape

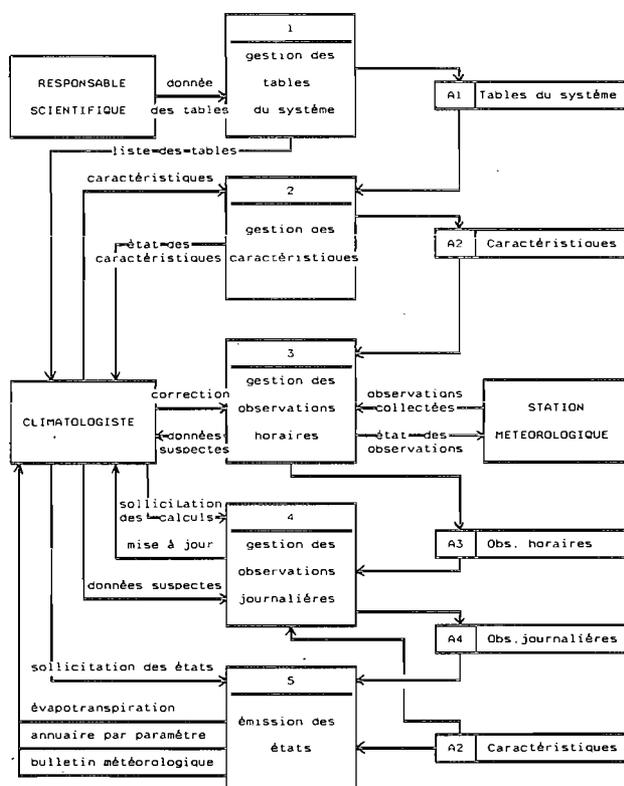


Fig. 2. - Diagramme général de flux de données (DFD) de SISCLIMA

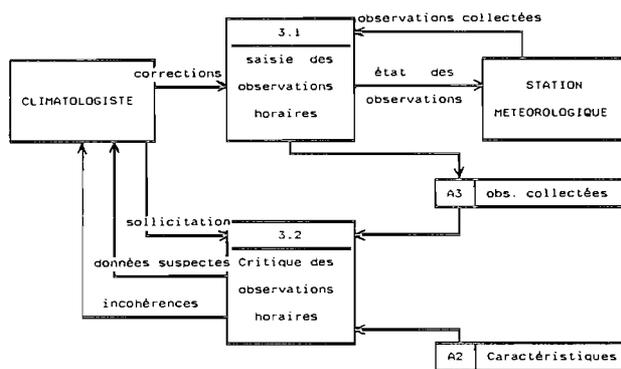


Fig. 3. - DFD de la gestion des observations horaires

saisie-corrrection. L'accès à ce fichier est limité à l'administrateur de la station. Le fichier opérationnel est constitué du fichier des informations journalières, généré par unités mensuelles, et associé à chaque donnée élaborée un code permettant d'en évaluer l'origine et/ou la qualité ;

- la fonction d'émission des états regroupe toutes les applications préalablement définies par le climatologue à partir du fichier opérationnel. Pour satisfaire les besoins les plus divers des utilisateurs de l'information agroclimatique, le développement d'une interface capable de mettre le fichier opérationnel de SISCLIMA à la disposition d'un progiciel de traitement statistique est prévu, et devra permettre la réalisation des applications non prévisibles a priori.

4.2. DICTIONNAIRE DES DONNÉES

Les diagrammes de flux de données guident le reste de l'étape de conception du système qui nécessite encore l'intervention du spécialiste en climatologie. Celle-ci se poursuit par une description précise de tous les éléments mis en évidence dans les DFD (entités externes, fonctions, dépôts et flux d'informations).

La partie de définition des flux comprend une estimation du volume et de la périodicité, en plus de la composition détaillée en terme de champs, mais reste indépendante du support physique (formulaire, grille d'écran, rapport, etc.) qui sera spécifié ultérieurement. Les champs identifiés conduisent à la création d'un dictionnaire des variables, élément central de la mise au point du système.

Pour SISCLIMA ce dictionnaire est géré par un progiciel adéquat et contient par variable un enregistrement formé, entre autres, des attributs suivants :

- un nom symbolique interne, mnémonique de sept caractères, qui sera utilisé par tous les programmes développés ;
- une description précise qui sera élaborée par le responsable scientifique ;
- un nom symbolique externe, usuellement utilisé par les spécialistes en climatologie ;
- les caractéristiques concernant le domaine des valeurs possibles pour cette variable (type, nombre de caractères ou de chiffres significatifs, masque d'édition) ;
- l'origine de la variable, c'est-à-dire comment elle est introduite dans le système (saisie, obtenue par extraction dans une table ou calculée par algorithme). La détermination de cet attribut oblige le climatologue à préciser les algorithmes de calcul qu'il utilise, y compris les situations d'exception (ignorance d'un ou plusieurs des facteurs qui interviennent dans les formules de calcul) ;
- le type de validation à appliquer sur la variable, s'il y a lieu. Dans la plupart des cas il s'agit de comparer la valeur de la variable aux limites maximales et minimales stipulées, qui seront alors indiquées dans cet enregistrement ;
- éventuellement, un algorithme spécial de cohérence avec d'autres variables du système.

4.3. SPÉCIFICATION

La phase suivante consiste à définir, pour chaque flux identifié lors de la modélisation du système, un support physique et, en fonction de ce dernier, la forme suivant laquelle les divers champs constitutifs de ce flux seront rangés.

A ce niveau, et à ce niveau seulement, les ressources disponibles pour l'opération concrète du système sont prises en considération, en même temps qu'est établie une tactique de mise au point. Pour SISCLIMA il a été choisi de développer le système sur une configuration minimale de micro-ordinateur de type CP/M 80 avec 64 K de mémoire centrale et deux unités de disquettes de 400 K, disponibles à l'EMBRAPA dans plus de 70 % des cas.

Le dessin des formulaires, des grilles d'écran et des états, le libellé des messages et le contenu des instructions qui constituent le manuel d'utilisation du système sont soumis à l'approbation écrite du responsable scientifique. La structure et l'organisation des fichiers et la décomposition en sous-systèmes complètent la documentation de SISCLIMA, avant que ne soit élaborée une seule ligne de codification.

La sélection des méthodes d'accès aux fichiers, réalisée au cours de cette étape, est étroitement liée à la prévision d'utilisation qui peut en être faite. Dans le cas des tables du système et des caractéristiques de station, l'organisation séquentiel-indexée permet de satisfaire les modifications fréquentes sans allonger sensiblement les temps d'accès. En ce qui concerne les fichiers de données, observations horaires et journalières, l'organisation directe s'avère la mieux adaptée, dans la mesure où le nombre d'enregistrements est fixe pour chaque année.

4.4. CODIFICATION ET IMPLÉMENTATION

Le choix d'un langage de programmation structuré, permettant une relative portabilité de l'application développée sur plusieurs types de micro-ordinateurs, en même temps qu'un niveau acceptable de performances et la disponibilité de l'organisation séquentiel-indexée est plutôt restreint. L'utilisation du Pascal et plus spécialement de la version TURBO Pascal munie de l'option TURBO Access répond de manière très satisfaisante aux exigences exprimées précédemment.

La réalisation de ces deux dernières étapes ne présente pas de caractéristiques originales. On se bornera à mentionner ici quelques éléments significatifs qui méritent d'être brièvement soulignés, dans la mesure où ils pourraient servir d'orientation dans la mise au point d'un système semblable :

- la périodicité du traitement respecte évidemment celle observée dans le traitement traditionnel manuel ; les formulaires obtenus à chaque journée de collecte d'information sont regroupés par lots mensuels sur une station déterminée. Toutes les ressources du système (saisie, critique, émission d'états, etc.), sont ainsi utilisables sur une période minimale d'un mois ;
- les données horaires et journalières, correspondant à une année d'observation sur une station déterminée, sont contenues sur un seul volume (disquette). Les caractéristiques de la station en vigueur pour cette période y sont également consignées, de telle sorte que les données d'une disquette « station-année » sont complètes et aptes à un traitement ultérieur, prévisible pour des applications régionales ;
- SISCLIMA contient des ressources complètes de « back-up » et prend totalement en charge la sécurité des informations. Il s'agit là d'une préoccupation fondamentale : l'utilisateur se trouve déchargé de toute préoccupation concernant la maintenance de copies de ses informations. Pratiquement chaque information apparaît sur deux volumes physiques, un original et une copie réciproquement identifiables. Toute opération de modification de l'état de l'un des fichiers du système (tables, caractéristiques, observations horaires et journalières), s'accompagne automatiquement d'une opération de copie de sécurité. Les contrôles correspondants sont réalisés par le biais de fichiers dits de paramètres, manipulés lors de l'ouverture des fichiers principaux de SISCLIMA. Seul les originaux sont utilisés, lorsqu'un dommage intervient, l'original est régénéré à partir de la copie.

CONCLUSION

SISCLIMA a été conçu pour mettre à la disposition de l'utilisateur des données sûres sans pertes excessives d'information. Pour aboutir à ce résultat l'intervention humaine a été réduite au minimum et les données vérifiées par plusieurs niveaux de tests. Afin de rendre ces tests plus puissants on les a adaptés aux conditions climatiques particulières de chaque station. Pour cela il a été nécessaire de fournir des informations complémentaires contenues dans un système de tables dont le nombre et le contenu sont modifiables à volonté.

Cette dernière particularité donne une très grande souplesse à la phase saisie-vérification des données. C'est la seconde caractéristique de SISCLIMA : son « adaptabilité ». Conçu pour nécessiter une configuration minimale et écrit dans un langage reconnu pour sa portabilité, il peut être adapté à la majorité des stations climatiques de l'EMBRAPA moyennant un minimum de transformations. Enfin, si le nombre des applications développées à partir du fichier opérationnel reste faible, la conception du système permet d'envisager leur multiplication sans problème majeur : le fichier opérationnel comporte déjà pré-calculées les grandeurs intervenant le plus fréquemment en agroclimatologie (durée du jour, tension moyenne de vapeur d'eau, etc.), quant à l'information complémentaire elle peut être introduite à l'aide de nouvelles tables ou directement en utilisant des programmes interactifs.

Ces nouvelles fonctions sont à définir par l'utilisateur lui-même à l'occasion du fonctionnement de SISCLIMA, l'une d'entre elle paraît cependant évidente : c'est le calcul du bilan hydrique d'une culture pendant une période déterminée. Enfin il est envisagé la création d'une interface mettant le fichier opérationnel à la disposition d'un progiciel statistique, ce qui permettra de multiplier les applications.

Manuscrit accepté par le Comité de Rédaction le 18 mai 1987

BIBLIOGRAPHIE

- ANGSTRÖM (A.), 1924. - Solar and terrestrial radiation. *Quart. J. R. Met. Soc.*, 50 : 121-5.
- BAILLE (A.) et al., 1982. - Memento rayonnement. *Institut National de la Recherche Agronomique, Département de Bioclimatologie*, Avignon, 60 p.
- BLANEY (H. F.) et CRIDDLE (W. D.), 1962. - Determining consumptive use and irrigation water requirements. *USDA (ARS) Tech. Bull.*, 1275, 59 p.
- BENAVIDES (G. J.) et LOPEZ (D.), (J.), 1970 - Formula para al calculo de la evapotranspiracion potencial adaptada al tropico (15° N, 15° S). *Agron. Journ.*, 20(5) : 335-45.

- DOORENBOS (J.) et PRUITT (W. O.), 1975. – Les besoins en eau des cultures. *Organisation des nations unies pour l'alimentation et l'agriculture (FAO), Rome, Bulletin d'irrigation*, 24, 198 p.
- GANE (C.) et SARSON (T.), 1979. – Structured systems analysis : tools and technics. Prentice Hall, Englewood Cliffs, N.J., 253 p.
- HARGREAVES (G. H.) et CHRISTIANSEN (J. E.), 1973. – Water use, ERTS readout and climate. Department of agriculture and irrigation engineering. Utah State University. Logan, Utah.
- MENEGUELLI (N. do A.), ASSIS (D. S.) et SECHET (P.), 1983. – SISOLOS : Manual de uso. *EMBRAPA-SNLCS. Documentos*, 4, Rio de Janeiro -RJ-245 p.
- MENEGUELLI (N. do A.) et SECHET (P.), 1984. – SISOLOS : Guia de entrada. *EMBRAPA-SNLCS. Documentos*, 9, Rio de Janeiro -RJ- 91 p.
- MILDE (L. C. E.), 1985. – Manual de utilização do sistema CADE. Parte 1 : usuario. *CEPLAC, rapport mulgr*. Itabuna -Ba-, 24 p.
- PENMAN (H. L.), 1948. – Natural evaporation from open water, bare soil and grass. *Royal Soc. London Proc. Serie : A*, 193 : 120-146.
- TURC (L.), 1961. – Évaluation des besoins en eau d'irrigation, évapotranspiration potentielle. *Ann. Agron.*, 12 (1) : 13-49.
- WORLD METEOROLOGICAL ORGANISATION, 1981. – Meteorological aspects of the utilization of solar radiation as an energy source. *W. M. O. technical note*. 172, Genève, 198 p.
- WORLD METEOROLOGICAL ORGANISATION, 1986. – Clicom project (climate data management system). *W. M. O. World Climate Data Program*, 131, Genève, 32 p.