HADLEY CENTRE
FOR CLIMATE PREDICTION AND RESEARCH

The role of observations in climate prediction and research.

by

D.J.Carson

CRTN 8

January 1991

CLIMATE RESEARCH TECHNICAL NOTE

Hadley Centre
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY

CLIMATE RESEARCH TECHNICAL NOTE NO. 8

THE ROLE OF OBSERVATIONS IN CLIMATE PREDICTION AND RESEARCH

by

D J CARSON

Hadley Centre for Climate Prediction and Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY
U.K.

THE ROLE OF OBSERVATIONS IN CLIMATE PREDICTION AND RESEARCH

by

D J CARSON

Hadley Centre for Climate Prediction and Research
Meteorological Office, London Road, Bracknell, Berkshire RG12 2SY

(Background notes prepared for the WCRP Workshop on the Planning of the Global Climate Observing System, 14-15 January 1991, Winchester, UK.)

1. INTRODUCTION

Many uncertainties surround our understanding of, and ability to model, the climate system. These, and other technical limitations, make prediction of the nature, magnitude and timing of climate change, particularly at scales smaller than global, unreliable at present. Increasing concern over the need to detect, understand and predict global and regional environmental changes has highlighted the need to establish high-quality observational datasets relating to the earth's atmosphere, oceans, cryosphere and land-surface. The basic climate record needs to be better established from geological times up to the present and decisions must be taken now to ensure more comprehensive global data coverage of the climate system in future.

A prerequisite to designing and implementing a Global Climate Observing System (GCOS) is to identify the principal uses and users of the data to be collected in order to justify such a costly, long-term, internationally-collaborative operation.

The major requirements for data in support of climate prediction and research, the latter to be interpreted in its widest sense, are for:

a. **monitoring** the climate and its variability at global and regional scales, thereby enabling quantification of natural climatic fluctuations on a range of temporal and spatial scales and the **detection of climate change**

b. **attribution of climate change** to particular causes, as in the search for a **fingerprint** which will enable detection and attribution of climate change due to an enhanced greenhouse effect resulting from man-made emissions of greenhouse gases

c. **detection (and attribution) of the environmental impacts of climate change**

d. **diagnostic studies** to elucidate the behaviour of the climate system and its component parts, viz., atmosphere, oceans, land surface, cryosphere, etc, including studies of the mechanisms of natural climatic variability

e. **<u>development and testing of hypotheses</u>** relating to local and global climate variations and to the degree of predictability of climatic phenomena

f. **<u>process studies</u>**. Special data are needed in support of detailed research studies of a wide variety of complex dynamical, physical, chemical and biological processes which help govern the state and evolution of the climate system. Such specialised datasets are likely to need to be highly-resolved in time and space and therefore gathered for a limited period over a restricted area of the globe. High-priority process studies include cloud-climate and air-sea interactions

g. **<u>providing boundary conditions for climate models</u>**. These include surface properties such as albedo, roughness length, vegetation index, soil physical parameters, etc. Some such data are also needed to identify changes in relevant surface properties, e.g. due to deforestation (see also c above)

h. **<u>initialisation of climate model integrations</u>**, especially of the oceanic and cryospheric components, and also some of the land-surface hydrological characteristics such as the 'soil moisture content'

i. **<u>validation of climate models</u>**. A very wide range of data, including palaeoclimatic data, are needed to assess the performance of the models being used for climate simulation and prediction. Model behaviour is compared with that of the 'observed' climate, often leading to further development and improvement of the models

j. **<u>data-assimilation techniques for climate model development</u>**. Aspects of the climate models can be improved through the use of climate observations in sophisticated four-dimensional data-assimilation schemes, such as those used with state-of-the-art operational weather forecasting models.

Although model and observational studies are becoming increasingly interdependent, it is convenient for the points I wish to focus on primarily in this note to consider the requirements a-f above as being, in the main, model-independent, whereas g-j are directly related to the development and use of climate-simulation and prediction models. A few comments on the data requirements for climate monitoring and diagnostic studies are given in Section 2. Section 3 concentrates on the observational needs of climate-simulation and climate-prediction models, i.e. items g-j above.

It is worth stressing that the above needs and uses encompass a variety of types of data, for example :

- **'stand-alone' instrumental data**, including data from both in-situ and remote-sensing (in particular satellite-borne) instruments. Examples of the former are surface and upper-air temperatures, rainfall, etc. The latter category also includes surface and upper air temperatures, and cloud cover, etc.

- **'blended' data**, i.e. a dataset produced from more than one data source; e.g. sea surface temperatures derived from both satellite- and ship-borne sensors.

  (Climate studies are by their very nature global in character and this necessitates extensive use of space observations to provide the necessary coverage and continuity. Space observations are, however, in many respects difficult to interpret and special blending techniques are required in order for them to provide reliable information for climate research. In particular, they need to be combined with in-situ observations to resolve ambiguities in their interpretation, to provide higher precision, and to enable estimates of fields of derived quantities, such as surface heat and moisture fluxes.)

- **data assimilated by operational weather forecasting models**. This is a powerful technique for the optimal combination of space-based and in-situ observations. It provides 'complete' global datasets which are internally, dynamically and physically consistent; e.g. 3-dimensional, global wind fields. However, biases may be introduced by changes to the models used.

- **'derived' data** i.e. data calculated from other directly measured or assimilated data; e.g. derived surface fluxes of heat, moisture and momentum. Information on land-surface characteristics, such as albedo and surface roughness, are usually 'derived' from a 'blend' of a variety of data types.

2. DATA REQUIREMENTS FOR MONITORING, DETECTION AND DIAGNOSTIC STUDIES

Studies of climatic processes often require short-lived specialised regional-scale datasets and associated field programmes. Studies of climate variability, change and predictability demand long-time-scale (decades to centuries or more) datasets which are homogeneous in time, i.e. without bias or with a constant bias. Studies of global environmental change (including climate change) clearly require global datasets, and regional variability, change and predictability are being studied increasingly in the context of the corresponding global-scale characteristics. The climate datasets to be assembled must reflect these diverse needs.

The 'operational' and research needs will also often lead to a hierarchy of datasets for a given variable, e.g. a real-time dataset (used for operational weather forecasting, and for monthly and seasonal predictions), a delayed-mode dataset (containing more data than the real-time set, and useful for preliminary monitoring and research studies), and a slow-mode dataset (containing all available data with instrumental corrections, e.g. historical in-situ sea surface temperature datasets).

Beside the routine production of analyses for climate research using data-assimilation techniques, it is essential to continue to process selected in-situ and remotely-sensed observations independently from the integrated data-assimilation systems. This is because we need information on the atmosphere and oceans (for example, ocean and land surface temperature analyses) which are free from the biases that arise from the inclusion of (changing) modelling assumptions. For selected variables, long time series of observations need to continue to be carefully processed to produce analyses which are as free as possible from the influence of changes to the observing or processing techniques introduced over the period. 'Pure-data' archives are necessary for this purpose. A particularly important example for validating climate models is the need for a reliable, homogeneous archive of atmospheric data from radio-sondes.

For monitoring climate, quantifying climate variability and detecting climate change, a long historical perspective is needed, in contrast to the normal data requirements in operational weather forecasting. In this context, it is important to realise that many historic data are not yet in a form for use on computers. Indeed many of the data needed for climate purposes are not expected to be processed through data-assimilation schemes in the medium-term future. Criteria for the collection and assembly of datasets for climate monitoring include :

- data are required for different parts of the climate system, particularly for atmosphere and oceans
- there is a need to establish reliable homogeneous quasi-global historic 'base-line' datasets
- current data need to be homogeneous with historic data; this is a difficult task and specialised techniques need to be developed to achieve the necessary high levels of homogeneity
- satellite data need to be blended with more-conventional data
- near real-time data are required to monitor short-period weather variability on timescales ranging from a day to a month. Few daily data are available on the global scale and this prevents worldwide monitoring of extreme events such as frosts.

In the context of short-period weather variability it is worth noting the additional data needs for long-range (monthly - seasonal) forecasting. This activity requires daily, real-time global datasets with special requirements for particular regions (e.g. high-resolution climate data for the UK are needed for the long-range forecasting activities within the UK Met Office). Also, long-range forecasting based on statistical methods needs long, homogeneous historical datasets.

Some suggested datasets for climate monitoring are listed in Table 1. Many of the qualifying column entries in Table 1 are somewhat tentative.

Diagnostic studies of data to explore climate mechanisms and processes, especially studies of low-frequency weather variability, have some additional and different requirements. The overriding need is for 3-dimensional, high-resolution, physically and dynamically self-consistent global datasets. This is generally only feasible by use of data assimilation allied to a global model. Data are needed at least twice daily (preferably 6-hourly) to avoid biases that can be introduced by diurnal effects. Homogeneity is another important requirement, but cannot always be achieved. Many of the datasets to be analysed are of derived physical and dynamical properties of the atmosphere and, where possible, the oceans, e.g. diabatic heating rates, eddy fluxes, ageostrophic quantities, etc.

3. DATA NEEDS FOR CLIMATE MODELLING AND PREDICTION

Some comments are offered on the following aspects of the data requirements in connection with the development and use of global numerical models for studying and predicting climate and climate change:

- boundary conditions
- initialisation
- validation
- data-assimilation techniques for model development.

3.1 Data for model boundary conditions

This encompasses a wide range of data which are prescribed in order to run climate models in various configurations (viz., atmosphere-only, ocean-only, coupled atmosphere-ocean). They may be held constant or updated prescriptively throughout a model integration, but they are not determined prognostically or diagnostically within the model itself. A range of boundary conditions is needed for application at the land-surface or air-sea interface, in the upper reaches of the atmosphere, throughout the atmosphere, oceans and soils, and at the lower ocean boundary.

The need to prescribe such boundary conditions depends to a very large extent on the state of development of the model being used, the configuration it is being used in, and what it is being used for. As models of the full climate system develop and as they are used increasingly for making predictions of transient global and regional climate change, then correspondingly more degrees of freedom will be given to such models and the requirements for boundary conditions will evolve in response. Some quantities acknowledged at present as boundary conditions will become full prognostic variables in the models, whilst new boundary conditions will be needed for new and more complex representations of the sub grid-scale processes (i.e. new parametrizations will require new boundary conditions).

Examples of important boundary conditions requiring better global datasets are :

Atmosphere-only models :

- sea surface temperatures (SSTs)

- sea ice extent, concentration, thickness

- ozone distribution

- land-surface characteristics (albedo, surface roughness length, vegetation type, etc)

- land ice extent and thickness

- orographic height (and sub grid-scale variance).

Ocean-only models :

Aside from ocean bottom topography, the prime requirement here is for data to provide the surface forcing needed for ocean model integrations, namely:

- surface stress fields

- wind mixing

- heat fluxes (incoming solar and longwave radiation, sensible and latent heat fluxes) and freshwater (precipitation less evaporation) fluxes

- river runoff.

Two basic approaches may be made to provide the appropriate forcing for the model, though in practice a combination of the two is frequently used.

The fluxes may be specified directly, in which case it is necessary to use a 'flux-correction' technique whereby the net heat flux, Q_o, is modified by a feedback term which helps to constrain the predicted sea surface temperature, T_p, in the model to the observed field, T_o, depending on the size of the feedback parameter, k. i.e the applied heat flux, Q, is given by:

$$Q = Q_o + k(T_o - T_p).$$

A similar approach may be used for the freshwater flux, which is then constrained to the observed salinity field.

Alternatively, the basic meteorological parameters necessary to drive the turbulent fluxes can be specified and the fluxes derived via appropriate bulk formulae as the integration proceeds and with reference to the modelled SST (or ice surface temperature) where appropriate. In this case the datasets required are:

- surface wind field
- windmixing
- surface air temperature
- surface air humidity
- surface pressure
- incoming solar and longwave radiation/cloudiness.

Coupled atmosphere-ocean general circulation models (AOGCMs) :

Particular boundary datasets required here are:

- ozone distributions
- land-surface characteristics, including land ice extent and thickness
- topographic data for both land and oceans.

3.2 Data for model initialisation

These are data needed to set the initial values of variables which will then be updated, usually prognostically, within the model itself. It is often not understood that climate modelling and climate prediction are not initial-value problems in the same way that short- to long-range forecasting are initial-value problems - at least as far as initialising the atmosphere is concerned. Initialisation of the atmosphere is not a critical factor in climate modelling or prediction; it is quite acceptable to use model-compatible data from a single time from a previous climate model run or a numerical weather prediction analysis, and quite unnecessary and of no additional value to demand real-time global analyses.

In the current state of development of climate models, it may be more critical to initialise aspects of the land-surface prognostic variables, if not starting from a previously derived model dataset. In particular, there is modelling evidence of long-lasting effects of the initial values used for some of the land-surface hydrological characteristics. Snow cover, soil moisture content and soil temperatures are examples of variables which may need to be initialised by 'spinning up' the model through at least a seasonal cycle.

The main current (and probably future) requirements for initialising climate models concern the oceans, whether in ocean-only models or coupled AOGCMs. Global fields of profiles of temperature, salinity and currents are required to provide a realistic initialisation of ocean models. Currently the Levitus global temperature and salinity dataset is used and the currents generated as the model runs. Repeated insertion of Levitus data may be useful and valid for high-resolution ocean models in this context (such as that used in NERC's FRAM project). Long (centuries or more), costly computer integrations are necessary for the models to come into equilibrium with any specified forcing. Note that the final state may depart significantly from the 'observed' state.

Aspects of the cryosphere also require careful initialisation in climate models. In particular, as a minimum, sea ice extent needs to be initialised. There is also need for sea ice concentrations and thicknesses, although the latter are particularly difficult to achieve.

In coupled AOGCMs a major problem is to bring the component systems into mutual adjustment with one another. The length of model integration necessary to achieve sufficient adjustment (and estimate flux corrections) for a given purpose may vary from decades to millenia. The initial data requirements for AOGCMs are as above. There is no doubt that initialisation of the ocean component of the climate system presents by far the greatest challenge in this area of climate modelling and prediction. Note that careful specification of the 'observed' state is particularly necessary for TOGA-related predictions and hindcasts (e.g. the prediction of ENSO events).

3.3 Data for model validation

The IPCC WG1 Report identifies three categories of data needed for climate model validation :

a. variables important for description of the atmospheric and oceanic circulation

   e.g. **atmosphere** : mean sea-level pressure; wind and temperature profiles; variability as indicated by eddy kinetic energy

   ocean : surface dynamic height (not often calculated in models); temperature, salinity and current structure; distribution of tracers; eddy statistics

b. variables critical in defining climate change

   e.g. **atmosphere** : surface air temperature; precipitation; soil moisture; monthly means plus interannual and daily variability

   ocean : SST; ocean mixed-layer depth

c. variables important for climate feedbacks

   e.g. snow cover; sea ice; clouds and their radiative effects.

Note that the above categories a-c are not mutually exclusive and many types of observations fit equally well into more than one.

Validation of selected regional aspects is also of relevance :

   e.g. occurrence of ENSO events in coupled models; monsoon phenomena

There are again particular considerations when verifying ocean models. On the seasonal timescale data are needed to validate simulations of the upper ocean mixed layer. Data on the spread of transient tracers provide valuable verification of decadal timescale changes in the ocean. On the timescale of a century or more data on deep ocean temperature and salinity structures, as well as of other tracers, will be of value.

Particular requirements for validating results from coupled AOGCMs include data on surface fluxes, SST and sea ice (extent, concentration and thickness).

3.4 Additional comments on the needs for boundary conditions, initialisation and validation of climate models

Table 2 lists some of the recognised data requirements for initialising, validating and providing boundary conditions for climate models.

Data are needed with a range of temporal resolutions : monthly means and variances, both on a climatological basis and for individual months; daily data are explicitly required in some cases as are data on the diurnal cycle of some quantities.

Future model assessments would benefit particularly from improved datasets on :

- precipitation and evaporation rates over the oceans
- evapotranspiration, soil moisture and snow depth over land
- clouds
- ocean properties (temperature, salinity, currents, etc)
- sea ice.

With regard to the potential of climate datasets produced by means of sophisticated data-assimilation techniques there is a need for climate modellers to explore more thoroughly how 'operational' archives can be utilised more comprehensively for model validation. There is still considerable capital to be extracted from existing datasets and scope for carefully-considered reanalyses of data from the recent past, using a fixed model.

There is increasing need to ensure that data are made available in machinable form, compatible with what models require. In that context, more-uniform practices need to be adopted in the retention of the model data which are needed increasingly for model intercomparison studies e.g. snow-cover frequency and depth, extremes and means of daily near-surface temperature.

3.5 Data for assimilation techniques for development of climate models

There are insufficient observations at any one time to determine the state of the atmosphere (even less for the oceans). Therefore we need to invoke any additional information we have, and this is available indirectly as the knowledge of the behaviour and structure of the atmosphere and, to a lesser extent, of the oceans, which we have encapsulated in the formulations of our operational weather forecasting and climate models. In particular, knowledge of evolution with time is embodied in operational forecasting systems and this enables the use of data distributed in time. Such models also provide a dynamically and physically consistent means of representing the atmosphere and oceans, and of deriving fluxes and other diagnostic quantities.

Assimilation is the process of finding the model representation which is most consistent with the available observations. Given observations distributed in time and space, and a forecast model, we can perform a 4-dimensional data assimilation. This is normally done by adding observations as the model is integrated forward in time. The current model state summarizes in an organized way the information from earlier observations and, given new observations, the model state is modified to be consistent with these and the earlier information.

It should be emphasised that at any given time the model state usually contains more information than can be extracted solely from the currently available observations. However, only variables which are well represented in the weather forecasting model can be assimilated sensibly in this way. There is considerable need and scope therefore for developing data assimilation techniques to process new types of observations within the framework of model assimilation in order to calculate the misfit between observed (or subsequently derived) properties and those produced (or diagnosed) from climate model integrations. Such observational and model comparisons can be used to validate and improve the representation of the physical processes in models and, conceivably, when such parametrizations have been demonstrated to represent an observed property adequately, then the model can be developed into a full assimilation using inverse methodology.

ACKNOWLEDGMENTS

The above notes on the role of observations in climate prediction and research were compiled following discussions with several colleagues in the Meteorological Office. I am indebted in particular to Messrs H Cattle, C K Folland, A C Lorenc, J F B Mitchell, D E Parker and P R Rowntree.

TABLE 1: Data for climate monitoring

PARAMETER	DATA SOURCES		DATA TYPES		PRIORITY	ACCURACY OR PRECISION	DESIRABLE ACCURACY OF LONG TERM CHANGES	SPATIAL RESOLUTION		TIME RESOLUTION
	IN SITU	REMOTE	PURE	ASSIMIL.				HORIZONTAL	VERTICAL	
SST	✓	✓	✓	✓	XXX	0.2^oC local or region 0.1^oglobal	$\leq 0.1^oC$	100km		2 weeks
subsurface ocean temp.	✓		✓	?	XX	0.1^oC	0.05^oC	500km?	20m above thermo cline 0.2km below	monthly
marine surface air temp.	✓		✓	✓	XX	0.2^oC local and region 0.1^oC global	$\leq 0.1^oC$	500km?		monthly
land surface air temp.	✓		✓	✓	XXX	0.2^oC local and region 0.1^oC global	$\leq 0.1^o$	500km		monthly and daily
atmospheric temperature	✓	✓	✓	✓	XXX	0.5^oC local 0.1^oC global	$\leq 0.2^oC$	500km	1 to 2km	monthly
humidity	✓	✓	✓	✓	XX	5-7% specific humidity	1% specific humidity	500km	2 to 3km	monthly
surface wind	✓	✓	✓	✓	XXX	1m/s	$\leq$ 1m/s	500km		monthly
winds aloft	✓	✓	✓	✓	XX	1m/s	$\leq$ 1m/s	500km	1 to 2km	monthly
surface pressure	✓		✓	✓	XXX	1hPa	$\leq$ 1hPa	500km		daily
atmospheric geopotential heights	✓	✓	✓	✓	XXX	10m	$\leq$ 10m	500km	1km	daily

TABLE 1

PARAMETER	DATA SOURCES		DATA TYPES		PRIORITY	ACCURACY OR PRECISION	DESIRABLE ACCURACY OF LONG-TERM CHANGES	SPATIAL RESOLUTION		TIME RESOLUTION
	IN-SITU	REMOTE	PURE	ASSIMIL.				HORIZONTAL	VERTICAL	
precipitation over land	✓	✓	✓	?	XXX	10% daily 5% monthly	≤ 2%	variable		monthly and daily
precipitation over ocean	if possible	✓	✓	?	XX	20%	≤ 5%	500km		5 days?
cloudiness variables	✓	✓	✓		XXX	complex (10% of total amount)	complex (≤ 5% of total amount)	250km	1km	monthly or less
sea ice extent	✓	✓	✓	?	XXX	5% open water	1% open water	100km ?		monthly
sea ice thickness	✓	?	✓	?	XXX	10% or 0.5m?	2% or 0.1m?	100km		monthly
snow extent	✓	✓	✓	?	XXX	3%	1%	100km?		3 days
snow depth	✓	?	✓	?	X	20% ?	< 5% ?	100km?		3 days
salinity	✓	?	✓	?	XX	2%	1%	250km	0.2km	monthly
near surface ocean currents	✓	✓	✓	✓	XXX	2cm/s	5%	20km in major currents	0.2km	monthly
sea surface height	✓	✓	✓	✓	XXX	5cm	2cm	200km?		monthly or less
soil moisture	few	?	?	?	XXX	10% of field-capacity	2% of field capacity	100km ?	10cm	monthly and 5-daily
vegetation		✓	✓		XX	10% of total biomass	5%? of total biomass	100km		monthly
planetary radiation budget components		✓	✓	?	XXX	$10 W/m^2$	$\leq 0.5 W/m^2$	250km ?		monthly
solar " constant"		✓	✓		XXX	0.05%	0.01% ($\sim 0.1 W/m^2$)			monthly
atmospheric CO_2 concent.	✓		✓		XXX	0.5ppm	0.5ppm	global		1 year
atmospheric diagnostics	✓	✓		✓	XXX	5%	1%	500km	1 to 2km	daily and 6-hourly

TABLE 2

DATA REQUIRED TO PROVIDE INITIALISATION, VALIDATION AND BOUNDARY CONDITIONS FOR CLIMATE MODELS (TENTATIVE)

	BOUNDARY	INITIALISATION	VALIDATION
ATMOSPHERE			
T, q, V (surf & u/a)	* (surf)		●
Cloudiness	*		●
Radiative fluxes (surface, vertical profiles & TOA)	*		●
Precipitation	*		●
Surface fluxes	*		●
Ozone	●		●
LAND SURFACE			
albedo	●		●
snow depth		●	●
vegetation	●	(●)	●
runoff	●		●
soil moisture		●	●
surface temperature		●	●
topography	●		
roughness	●		
SEA ICE			
extent/concentration	●	●	●
thickness	●	●	●
roughness	●		
OCEAN			
dynamic height		(●)	●
surface T, S, V	●(SST)	●	●
T, S, V structure		●	●
mixed-layer depth		●	●
wave characteristics- (height etc)			●
tracers		●	●
water type (plankton, CO_2 etc)	●		

* OCEAN

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 1	Oct 1990	Estimates of the sensitivity of climate to vegetation changes using the Penman-Monteith equation. P R Rowntree
CRTN 2	Oct 1990	An ocean general circulation model of the Indian Ocean for hindcasting studies. D J Carrington
CRTN 3	Oct 1990	Simulation of the tropical diurnal cycle in a climate model. D P Rowell
CRTN 4	Oct 1990	Low frequency variability of the oceans. C K Folland, A Colman, D E Parker and A Bevan
CRTN 5	Dec 1990	A comparison of 11-level General Circulation Model Simulations with observations in the East Sahel. K Maskell
CRTN 6	Dec 1990	Climate Change Prediction. J F B Mitchell and Qing-cun Zeng
CRTN 7	Jan 1991	Deforestation of Amazonia - modelling the effects of albedo change. M F Mylne and P R Rowntree
CRTN 8	Jan 1991	The role of observations in climate prediction and research. D J Carson