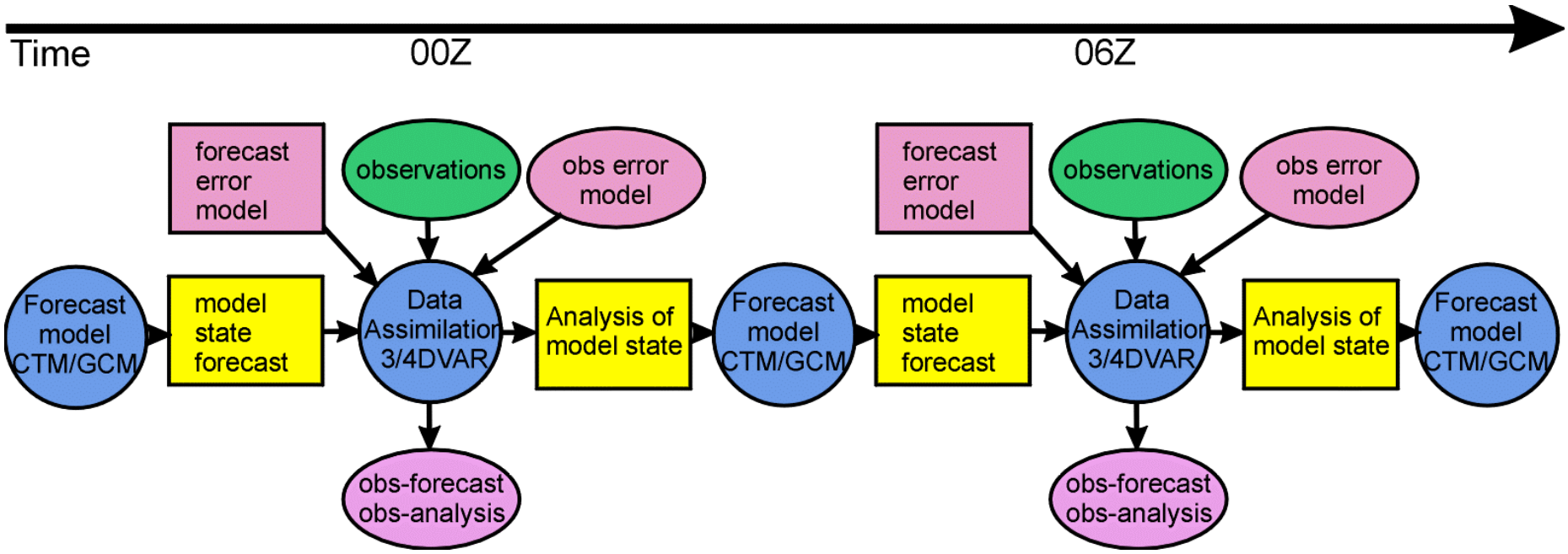
Data Assimilation Cycles



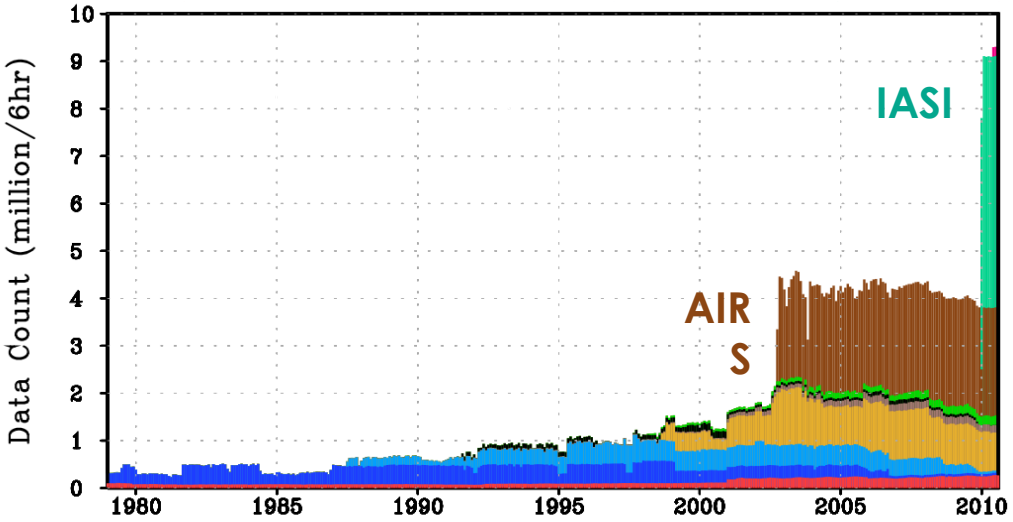
[From Saroja Polavarapu]

3. Upcoming Changes to NRT GEOS-5

Beginning in mid-June, GEOS-5.7.3 will replace GEOS-5.2.0 as our NRT system (MERRA will continue unchanged)

Resolution will be increased to 0.25°×0.3125° with updates to the model and enhanced capabilities in the analysis

Number of observations processed in each six-hour window of GEOS-5



Plot by Ron Gelaro

Hyperspectral sensors (AIRS & IASI) greatly increase the number of observations each analysis step. After thinning and QC, about 1/3 this number are assimilated. GEOS-5.7.3 includes IASI and GPSRO.

[Presented by Steven Pawson (NASA GMAO) at the IGC5 meeting at Harvard University, May 2011]

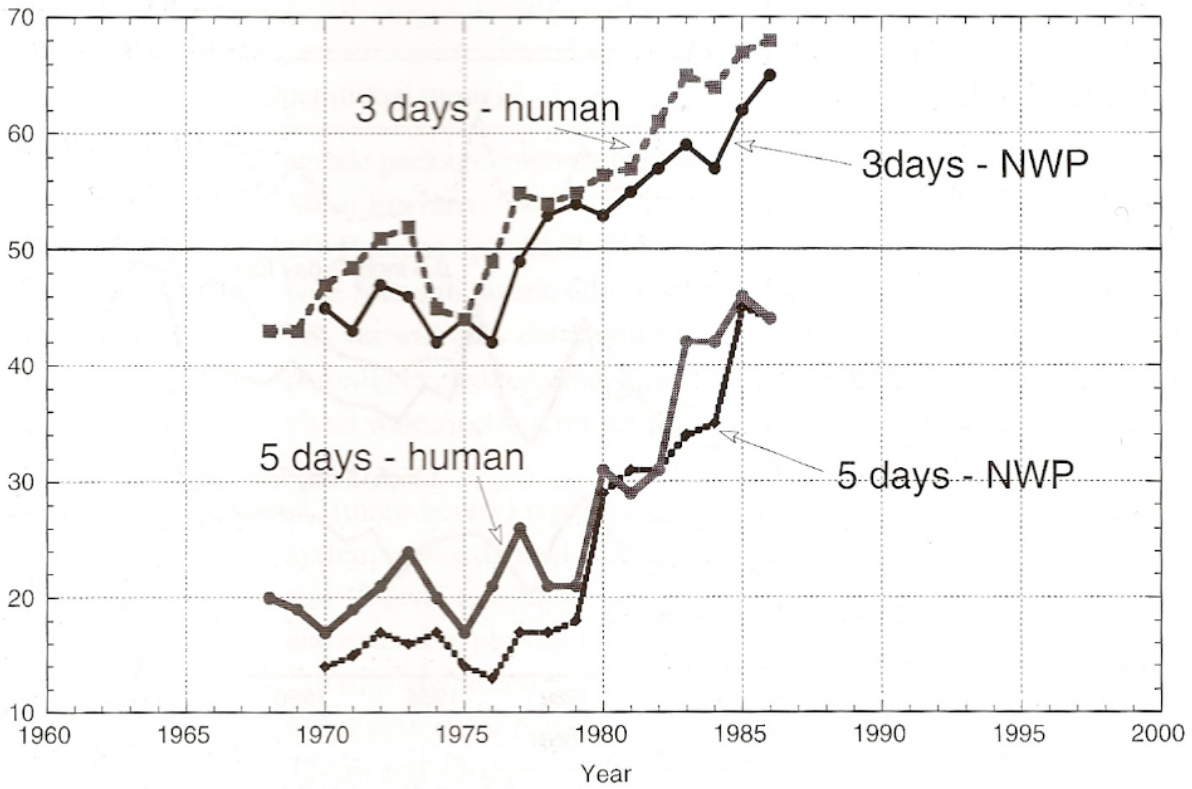


Figure 1.5.2: Hughes data: comparison of the forecast skill in the medium-range from NWP guidance and from human forecasters.

[Kalnay, 2003]

Variational Continuous Assimilation of TMI and SSM/I Rain Rates: Impact on GEOS-3 Hurricane Analyses and Forecasts

ARTHUR Y. HOU, SARA Q. ZHANG,* AND ORESTE REALE+

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

[Monthly Weather Review, 2004]

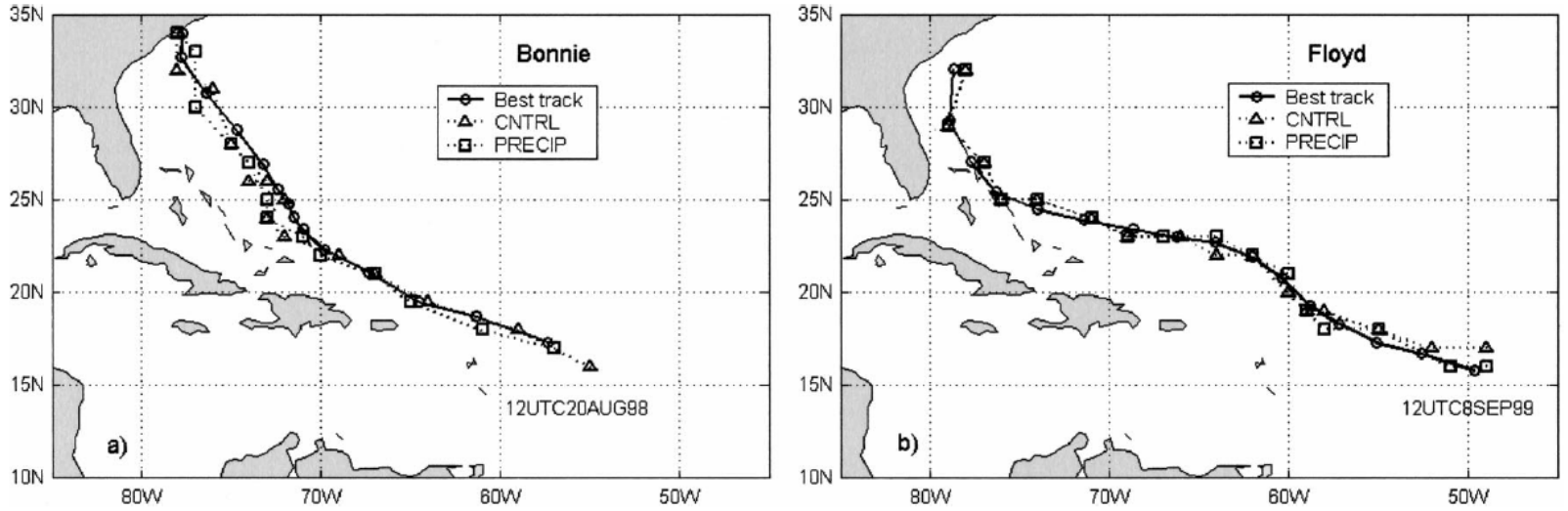


FIG. 4. Analyzed 12-h positions of the minimum surface pressure of (left) Bonnie and (right) Floyd (rendered to the closest integer degree in latitude and longitude) from CNTRL and PRECIP analyses, compared with NOAA best-track locations. The first analysis time of the track is marked for each storm.

Datasets: Rainfall data from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), And the Special Sensor Microwave Imager (SSM/I)

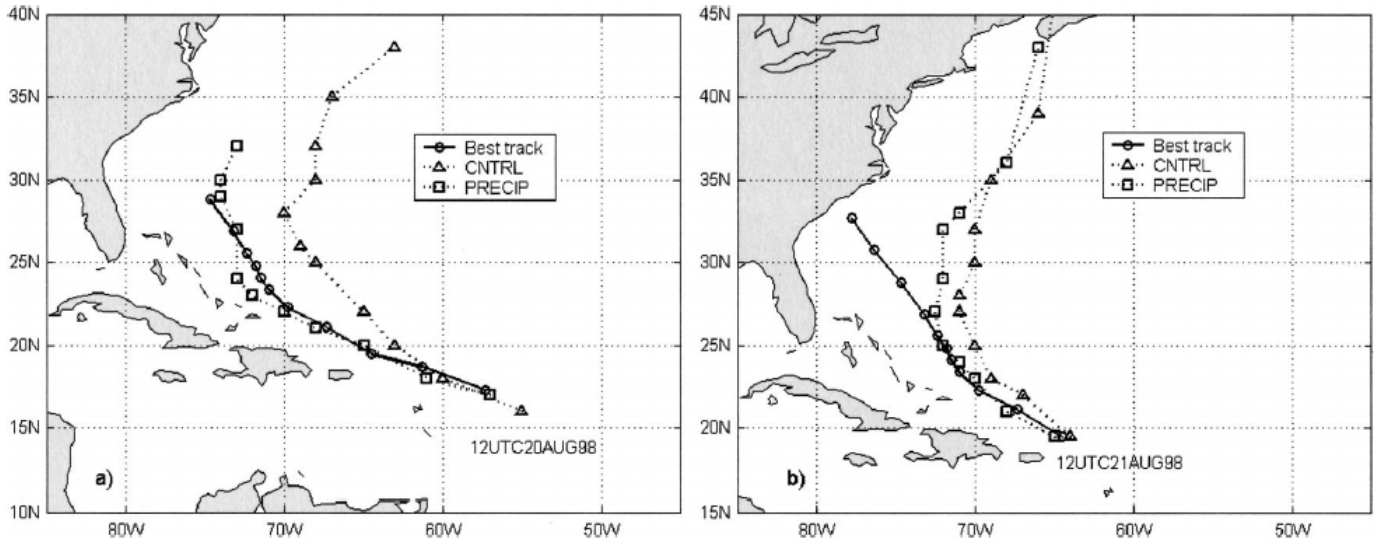


FIG. 10. Five-day Bonnie track forecasts. Triangles and squares mark forecasts (dashes) initialized with CNTRL and PRECIP analyses, respectively. The storm positions rendered to the closest integer degree in latitude and longitude are plotted every 12 h. The circles show positions of the best track (solid). (a) Forecasts issued from 1200 UTC 20 Aug 1998. (b) Forecasts issued from 1200 UTC 21 Aug 1998.

[Hou et al, 2004]

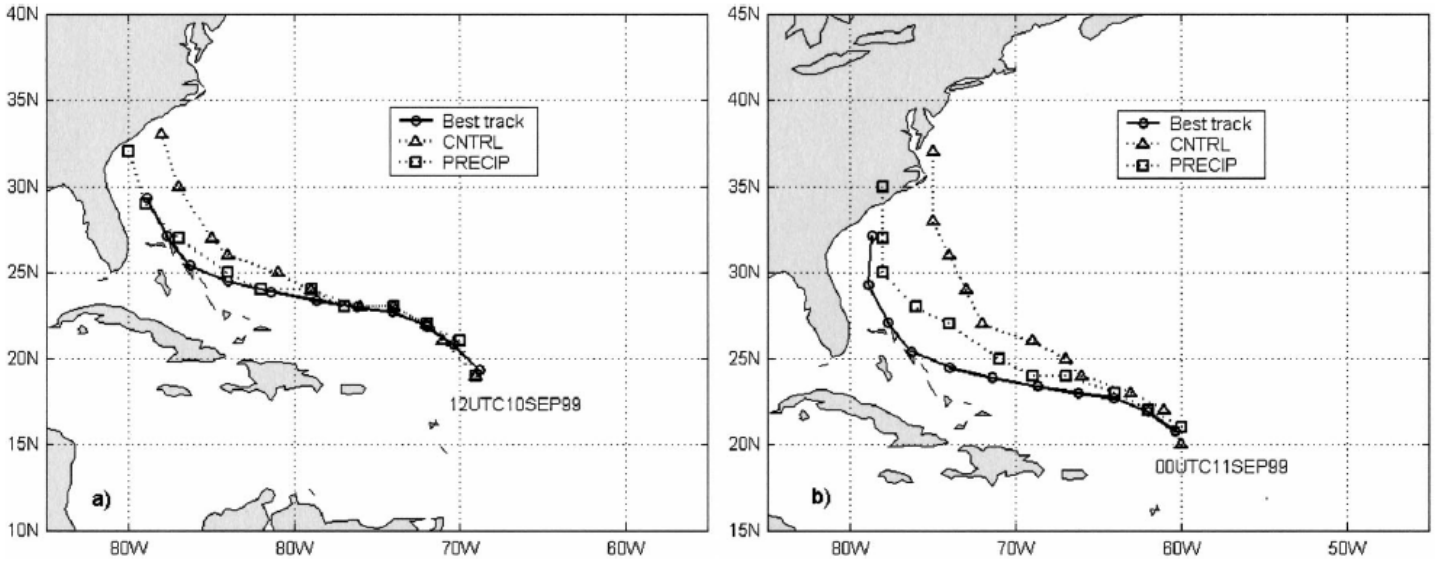


FIG. 11. Five-day Floyd track forecasts. Triangles and squares mark forecasts (dashes) initialized with CNTRL and PRECIP analyses, respectively. The storm positions rendered to the closest integer degree in latitude and longitude are plotted every 12 h. The circles show positions of the best track (solid). (a) Forecasts initialized from 1200 UTC 10 Sep 1999. (b) Forecasts issued from 0000 UTC 11 Sep 1999.

[Hou et al, 2004]

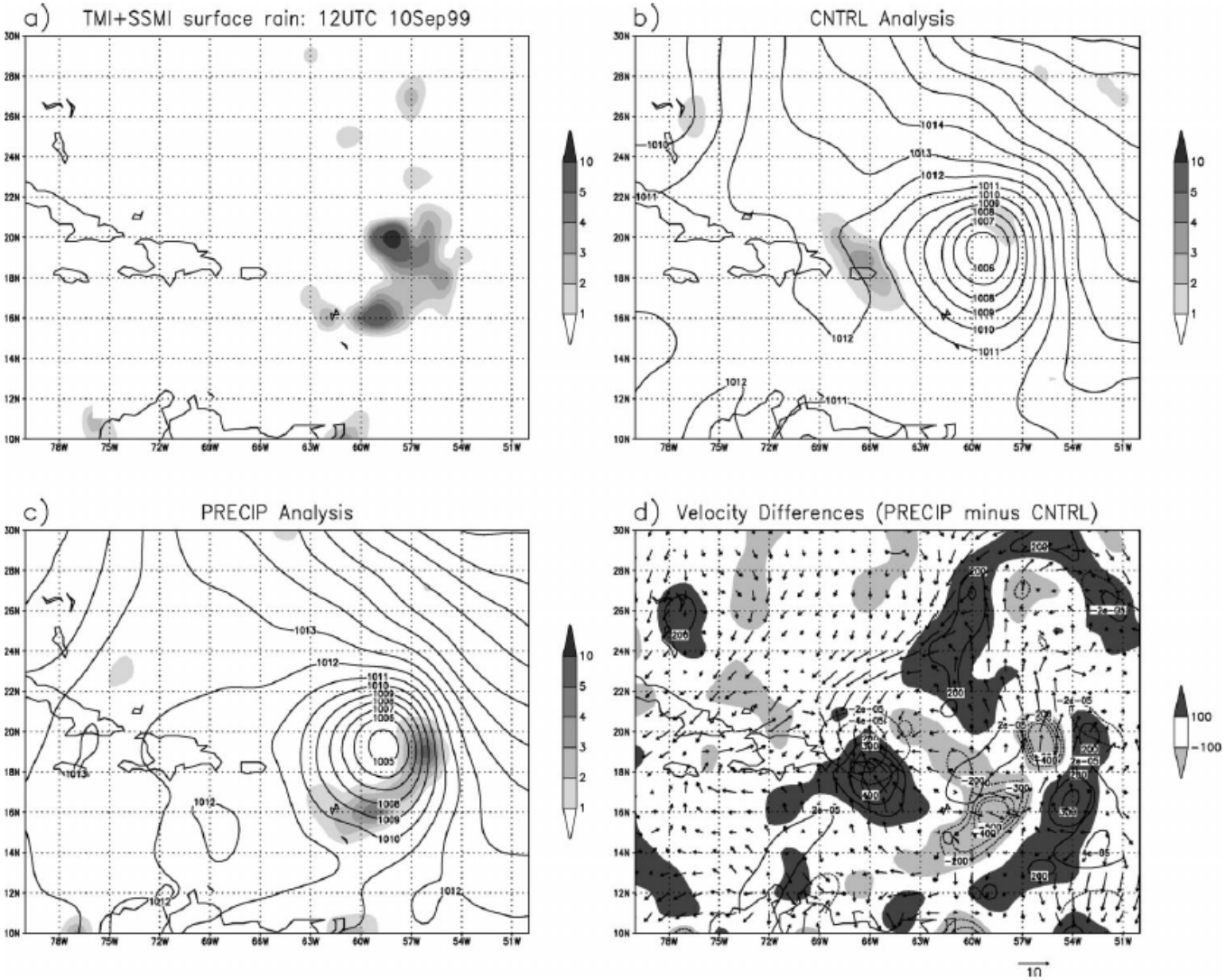


FIG. 7. (a) Combined TMI and SSM/I observations of Floyd surface rain in mm h^{-1} at 1200 UTC 10 Sep 1999. (b) Surface rain in mm h^{-1} (shaded) and SLP in hPa (contours) in CNTRL at the same analysis time. (c) Same as (b), except for PRECIP analysis. (d) Changes between PRECIP and CNTRL analyses in 500-hPa omega velocity in hPa d^{-1} (shaded, with negative values indicating rising motion), divergence in s^{-1} (contour interval of 2×10^{-5} s^{-1} with zero omitted), and horizontal winds in m s^{-1} at 200 hPa. The vector scale for 10 m s^{-1} is given for reference. Note that the heavy observed rainfall maximum in Fig. 7a is flagged by the preanalysis QC check since the background rain is zero at this location (Fig. 7b) and the O–B exceeds 5 mm h^{-1}.

[Hou et al, 2004]

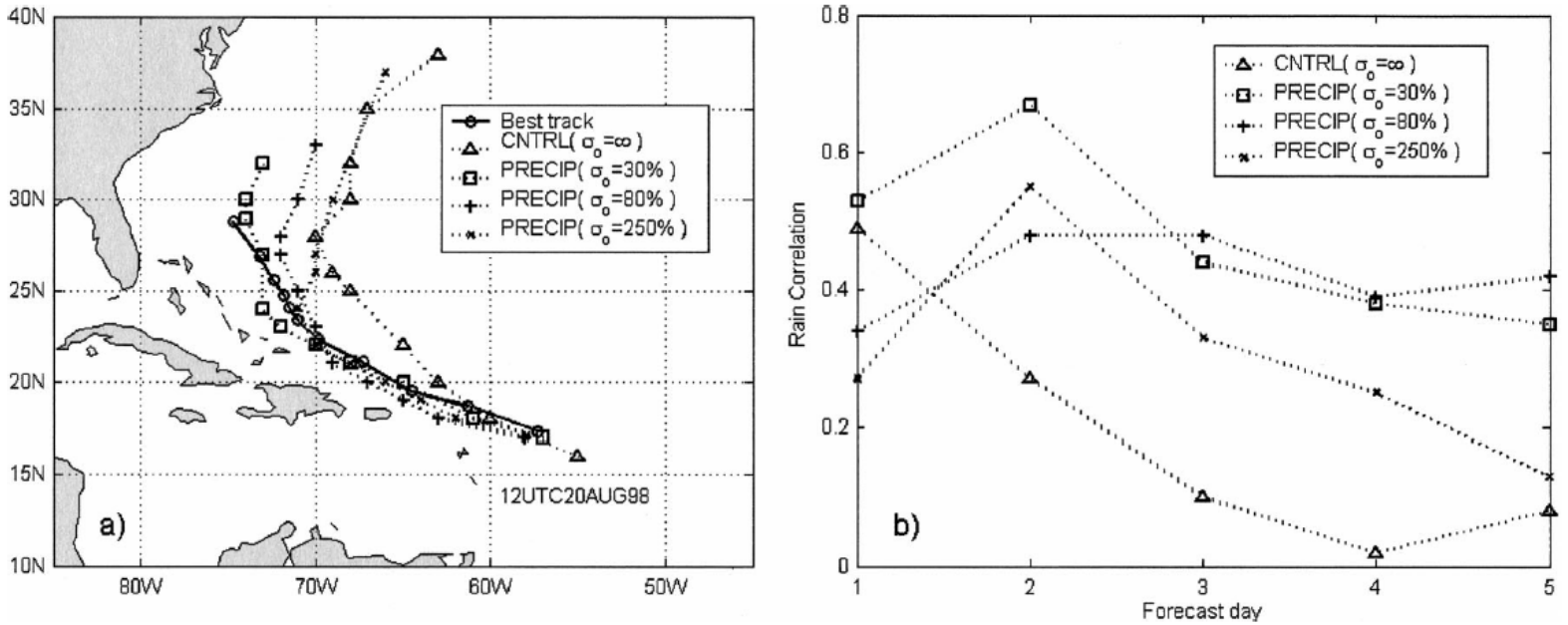


FIG. 12. Sensitivity of a 5-day Bonnie forecast from 1200 UTC 20 Aug 1998 to TMI and SSM/I rainfall information retained in the initial condition. (a) Comparison of 5-day track forecast with the NOAA best-track analysis. (b) Spatial correlations between 5-day precipitation forecasts and combined TMI and SSM/I daily rain rates. The legend identifies the forecasts by the error standard deviation for rain (σ_o) in the analysis used for initial condition.

[Hou et al., 2004]

The MACC reanalysis: an 8 yr data set of atmospheric composition

A. Inness[1], F. Baier[2], A. Benedetti[1], I. Bouarar[3], S. Chabrillat[4], H. Clark[15,16], C. Clerbaux[3], P. Coheur[14], R. J. Engelen[1], Q. Errera[4], J. Flemming[1], M. George[3], C. Granier[3,6,7], J. Hadji-Lazaro[3], V. Huijnen[8], D. Hurtmans[14], L. Jones[1], J. W. Kaiser[1,5,9], J. Kapsomenakis[12], K. Lefever[4], J. Leitão[10], M. Razinger[1], A. Richter[10], M. G. Schultz[11], A. J. Simmons[1], M. Suttie[1], O. Stein[11], J.-N. Thépaut[1], V. Thouret[15,16], M. Vrekoussis[12,13], C. Zerefos[12], and the MACC team

[1]ECMWF, Reading, UK
[2]DLR, Oberpfaffenhofen, Germany
[3]UPMC Univ. Paris 06, Université Versailles St-Quentin, CNRS/INSU, LATMOS-IPSL, Paris, France
[4]BIRA-IASB, Brussels, Belgium
[5]King's College London, London, UK
[6]NOAA/ESRL and CIRES, University of Colorado, Boulder, CO, USA
[7]Max Planck Institute for Meteorology, Hamburg, Germany
[8]KNMI, De Bilt, the Netherlands
[9]Max-Planck-Institute for Chemistry, Mainz, Germany
[10]IUP-Bremen, Bremen, Germany
[11]Forschungszentrum Jülich, Jülich, Germany
[12]Research Center for Atmospheric Physics and Climatology, Academy of Athens, Athens, Greece
[13]EEWRC, The Cyprus Institute, Nicosia, Cyprus
[14]Spectroscopie de l'Atmosphère, Service de Chimie Quantique et Photophysique, Université Libre de Bruxelles (ULB), Brussels, Belgium
[15]Université de Toulouse, UPS, LA (Laboratoire d'Aérologie), Toulouse, France
[16]CNRS, LA (Laboratoire d'Aérologie), UMR5560, Toulouse, France

Correspondence to: A. Inness (a.inness@ecmwf.int)

Received: 8 October 2012 – Published in Atmos. Chem. Phys. Discuss.: 5 December 2012
Revised: 20 March 2013 – Accepted: 21 March 2013 – Published: 18 April 2013

Abstract. An eight-year long reanalysis of atmospheric composition data covering the period 2003–2010 was constructed as part of the FP7-funded Monitoring Atmospheric Composition and Climate project by assimilating satellite data into a global model and data assimilation system. This reanalysis provides fields of chemically reactive gases, namely carbon monoxide, ozone, nitrogen oxides, and formaldehyde, as well as aerosols and greenhouse gases globally at a horizontal resolution of about 80 km for both the troposphere and the stratosphere. This paper describes the assimilation system for the reactive gases and presents validation results for the reactive gas analysis fields to document the data set and to give a first indication of its quality.

Tropospheric CO values from the MACC reanalysis are on average 10–20 % lower than routine observations from commercial aircrafts over airports through most of the troposphere, and have larger negative biases in the boundary layer at urban sites affected by air pollution, possibly due to an underestimation of CO or precursor emissions.

Stratospheric ozone fields from the MACC reanalysis agree with ozonesondes and ACE-FTS data to within ±10 % in most seasons and regions. In the troposphere the reanalysis shows biases of −5 % to +10 % with respect to ozonesondes and aircraft data in the extratropics, but has larger negative biases in the tropics. Area-averaged total column ozone agrees with ozone fields from a multi-sensor reanalysis data set to within a few percent.

NO_2 fields from the reanalysis show the right seasonality over polluted urban areas of the NH and over tropical biomass burning areas, but underestimate wintertime NO_2

Table 1. Satellite retrievals of reactive gases that were actively assimilated in the MACC reanalysis. PROF denotes profile data, TC total columns, TRC tropospheric columns, PC partial columns, and SOE solar elevation. PC SBUV/2 data consist of 6 layers between the surface and 0.1 hPa. NRT (near-real-time) data are available within a few hours after the observation was made, and are being used in operational forecast systems. For periods towards the end of the MACC reanalysis period, NRT data were used for some of the species when no offline products were available.

Sensor	Satellite	Provider	Version	Period	Type	Data usage criteria	Reference
GOME	ERS-2	RAL		20030101–20030531	O_3 PROF	Used if SOE > 15° and 80° S < lat < 80° N	Siddans et al. (2007)
MIPAS	ENVISAT	ESA		20030127–20040326	O_3 PROF	All data used	Carli et al. (2004)
MLS	AURA	NASA	V02	20040808–20090315, NRT data from 20090316	O_3 PROF	All data used	Waters et al. (2006)
OMI	AURA	NASA	V003	From 20041001, NRT data 20070321–20071231	O_3 TC	Used if SOE > 10°	Bhartia and Wellemeyer (2002); Levelt et al. (2006)
SBUV/2	NOAA-16	NOAA	V8	From 20040101	O_3 PC	Used if SOE > 6°	Bhartia et al. (1996)
SBUV/2	NOAA-17	NOAA	V8	From 20030101	O_3 PC	Used if SOE > 6°	Bhartia et al. (1996)
SBUV/2	NOAA-18	NOAA	V8	From 20050604	O_3 PC	Used if SOE > 6°	Bhartia et al. (1996)
SCIAMACHY	ENVISAT	KNMI		From 20030101	O_3 TC	Used if SOE > 6°	Eskes et al. (2005)
IASI	METOP-A	LATMOS/ULB		From 20080401	CO TC	Used if 70° S < lat < 70° N	George et al. (2009); Clerbaux et al. (2009)
MOPITT	TERRA	NCAR	V4	From 20030101, NRT data after 20100323	CO TC	Used if 65° S < lat < 65° N	Deeter et al. (2010)
SCIAMACHY	ENVISAT	KNMI	V1.04	20030101–20070630	NO_2 TRC	Used if SOE > 6° and 60° S < lat < 60° N	Boersma et al. (2004)
SCIAMACHY	ENVISAT	KNMI	V1.1	From 20070911	NO_2 TRC	Used if SOE > 6° and 60° S < lat < 60° N	http://www.temis.nl; Wang et al. (2008)

[Inness et al., 2013]

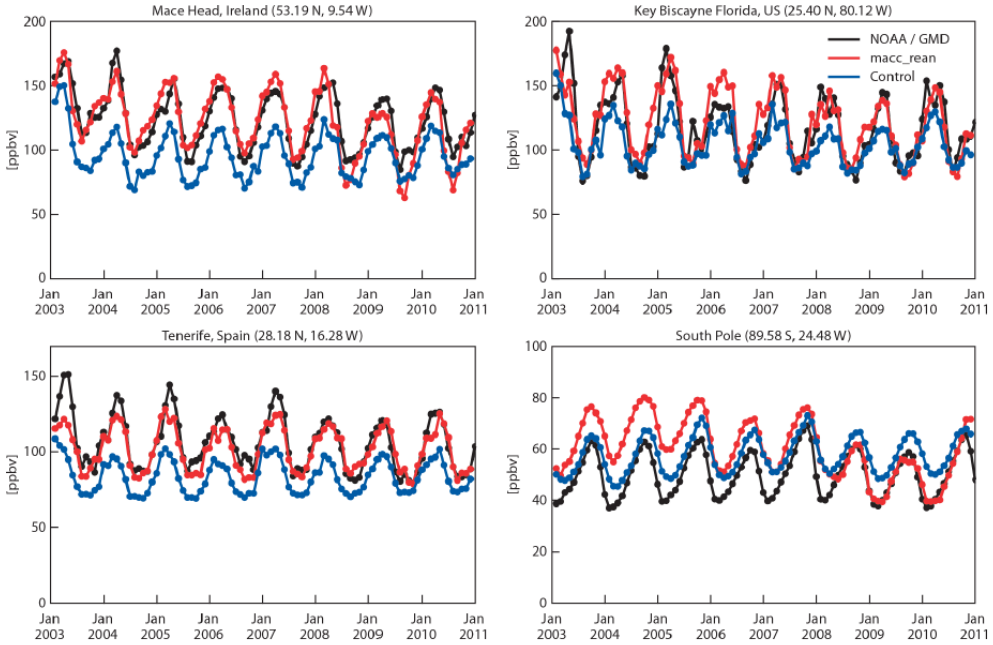


Fig. 5. Time series (2003–2010) of monthly mean CO concentrations (ppbv) from the MACC reanalysis (red), the control run (blue), and from NOAA/GMD ground-based measurements (black) over Mace-Head (top left), Key Biscayne (top right), Tenerife (bottom left), and South Pole (bottom right) stations.

[Inness et al., 2013]

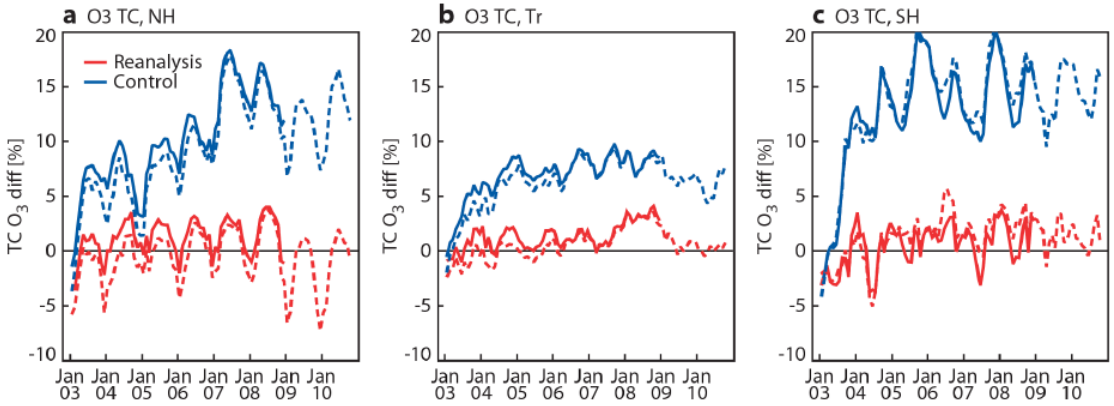


Fig. 10. Time series of the mean difference of the MACC reanalysis minus total O_3 columns from the multi-sensor reanalysis (solid) and SCIAMACHY fields (dashed) in % averaged over (**a**) the NH extratropics (30° N–90° N), (**b**) the tropics (30° S–30° N), and (**c**) the SH extratropics (90° S–30° S) for the years 2003 to 2010. Red lines show the MACC reanalysis and blue lines the control run.

[Inness et al., 2013]

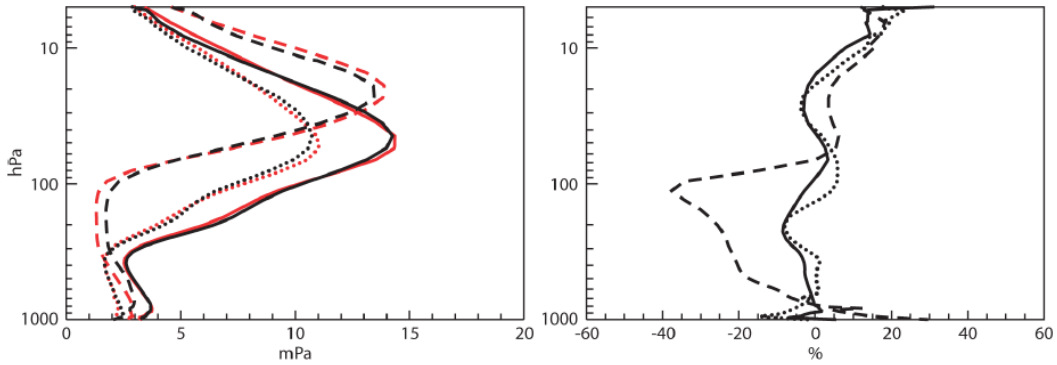


Fig. 16. Time mean ozone profiles (left) in mPa from the MACC reanalysis (red) and ozonesondes (black), and averaged ozone bias in % of MACC reanalysis minus ozonesondes (right) averaged over the period January 2003 to December 2010. Solid lines show means for the NH extratropical stations (i.e. north of 30° N), dashed lines for tropical stations (30° S–30° N), and dotted lines for SH extratropical stations (south of 30° S).

[Inness et al., 2013]

Data Assimilation for a Coupled Ocean–Atmosphere Model. Part II: Parameter Estimation

DMITRI KONDRASHOV

University of California, Los Angeles, Los Angeles, California

CHAOJIAO SUN*

Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, and Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland

MICHAEL GHIL+

University of California, Los Angeles, Los Angeles, California

(Manuscript received 29 January 2008, in final form 9 April 2008)

ABSTRACT

The parameter estimation problem for the coupled ocean–atmosphere system in the tropical Pacific Ocean is investigated using an advanced sequential estimator [i.e., the extended Kalman filter (EKF)]. The intermediate coupled model (ICM) used in this paper consists of a prognostic upper-ocean model and a diagnostic atmospheric model. Model errors arise from the uncertainty in atmospheric wind stress. First, the state and parameters are estimated in an identical-twin framework, based on incomplete and inaccurate observations of the model state. Two parameters are estimated by including them into an augmented state vector. Model-generated oceanic datasets are assimilated to produce a time-continuous, dynamically consistent description of the model's El Niño–Southern Oscillation (ENSO). State estimation without correcting erroneous parameter values still permits recovering the true state to a certain extent, depending on the quality and accuracy of the observations and the size of the discrepancy in the parameters. Estimating both state and parameter values simultaneously, though, produces much better results. Next, real sea surface temperatures observations from the tropical Pacific are assimilated for a 30-yr period (1975–2004). Estimating both the state and parameters by the EKF method helps to track the observations better, even when the ICM is not capable of simulating all the details of the observed state. Furthermore, unobserved ocean variables, such as zonal currents, are improved when model parameters are estimated. A key advantage of using this augmented-state approach is that the incremental cost of applying the EKF to joint state and parameter estimation is small relative to the cost of state estimation alone. A similar approach generalizes various reduced-state approximations of the EKF and could improve simulations and forecasts using large, realistic models.

[Monthly Weather Review, 2008]

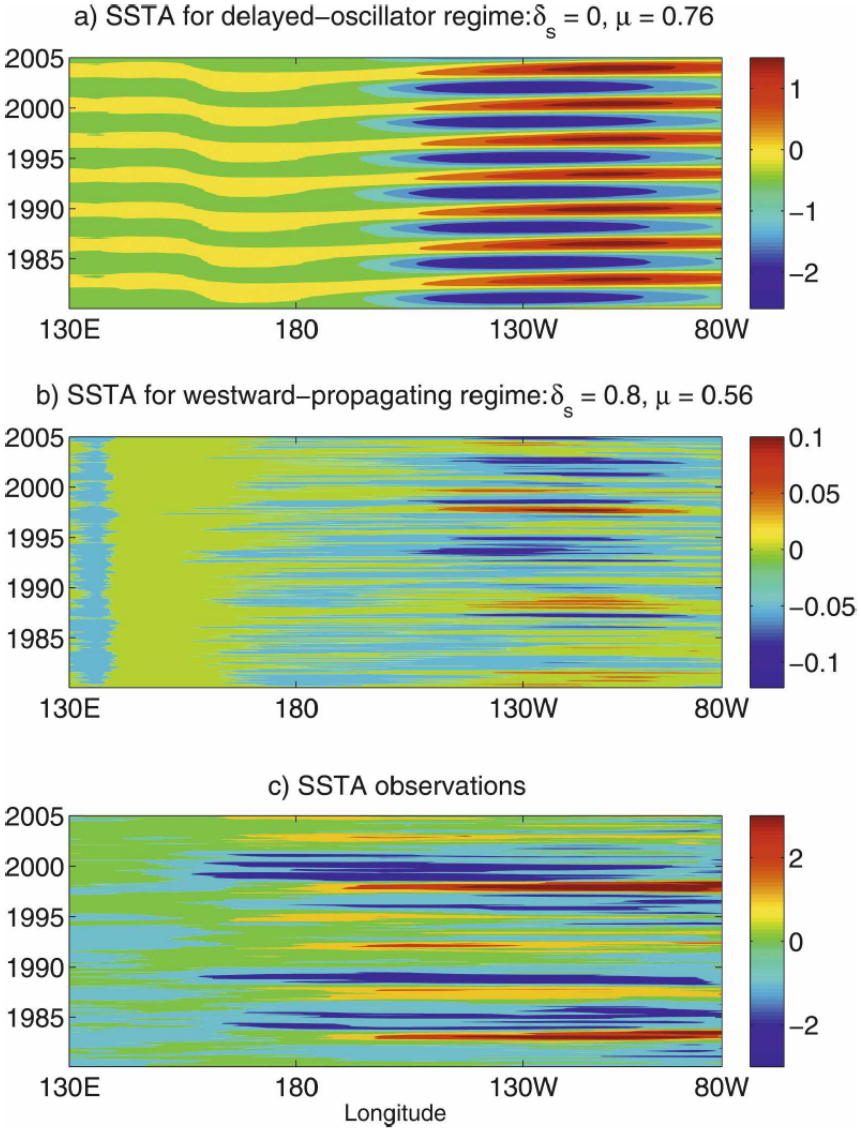


FIG. 1. Time evolution of equatorial Pacific SST anomalies (SSTA; °C) for the (a) delayed-oscillator and (b) westward-propagating modes of the ICM; and (c) observations of monthly SST data from the Climate Data Library at the IRI/LDEO (see online at http://ingrid.ldeo.columbia.edu/).

[Kondrashov et al., 2008]

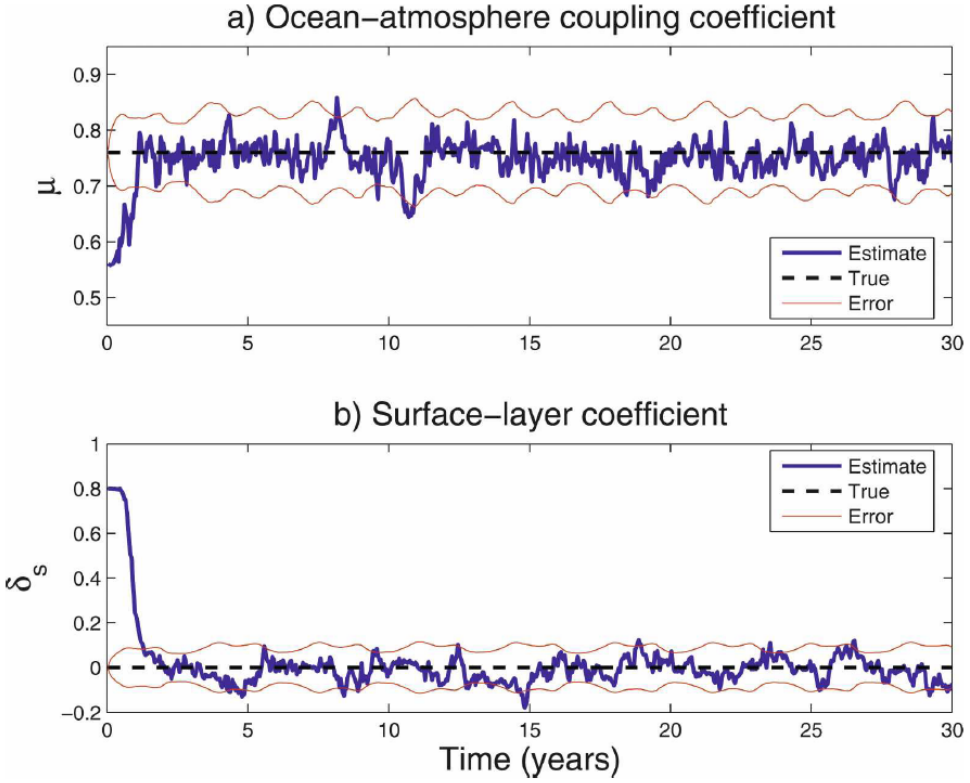


FIG. 2. Time evolution of estimated parameters (a) μ and (b) δ_s for the identical-twin assimilation run with the observations for SST data taken from 15 equatorial locations in the east-central Pacific. The correct (true) values of μ and δ_s are equal to 0.76 and 0, respectively.

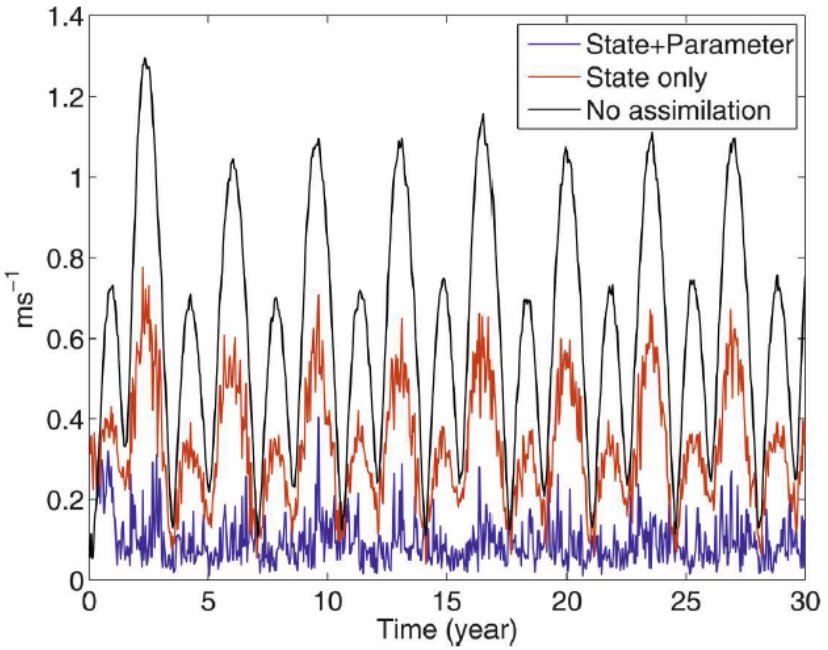


FIG. 3. RMS errors in the zonal surface current for the identical-twin experiment of Fig. 2. Blue and red lines are for actual errors with and without parameter estimation, respectively; and the black line indicates the pure forecast error.

[Kondrashov et al., 2008]

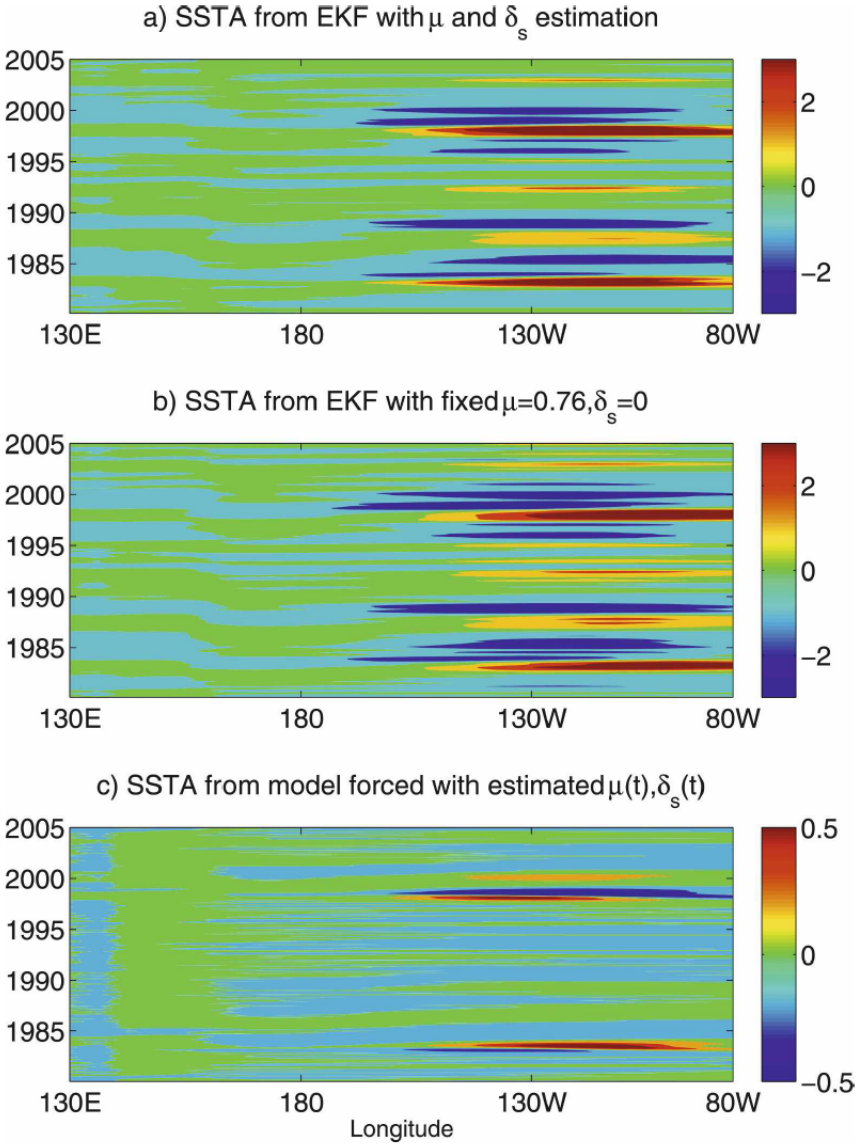


FIG. 5. Time–longitude plots (Hovmöller diagrams) of SST anomaly results (°C) for three EKF experiments: (a) combined state and parameter estimation, (b) state estimation only with constant parameters from the delayed-oscillator mode, and (c) model simulation with a time history of estimated $\mu(t)$ and $\delta_s(t)$, but with no further data assimilation.

[Kondrashov et al., 2008]

2.2.1 Probability density functions

A random variable is completely described by its **probability distribution function**. For a continuous r.v., the probability distribution function, $F_{\mathrm{x}}(x)$, is given by:

$$F_{\mathrm{x}}(x) = P(\mathrm{x} \leq x) = \int_{-\infty}^{x} p_{\mathrm{x}}(s)ds. \tag{2.1}$$

(It is only continuous r.v.s that we will encounter in this course.) The function $p_{\mathrm{x}}(x)$ is the **probability density function** or p.d.f. From this definition and the fact that $0 \leq p_{\mathrm{x}}(x) \leq 1$, it is clear that $F_{\mathrm{x}}(x)$ is a non-decreasing function and that $F_{\mathrm{x}}(-\infty) = 0$ and $F_{\mathrm{x}}(+\infty) = 1$. In addition,

$$\int_{-\infty}^{+\infty} p_{\mathrm{x}}(s)ds = 1.$$

It turns out that we almost always deal with p.d.f.'s rather than with probability distribution functions. In fact, often when we refer to an underlying distribution of a process, we are actually speaking of the the p.d.f. of the r.v., since one may often be deduced from the other.

2.2.3 Moments of a distribution

It is very difficult to determine the complete p.d.f. of a random variable, in practice. Often, it may be sufficient to determine the properties of the p.d.f. Other times, these properties may simply be the only information which can be practically obtained.

The **mean** or expected value of x is

$$\mu = \mathcal{E}\{x\} = \int_{-\infty}^{+\infty} xp(x)dx.$$

In general, for any function, $f(x)$:

$$\mathcal{E}\{f(x)\} = \int_{-\infty}^{+\infty} f(x)p(x)dx.$$

If $f(x) = x^n$, then

$$\mathcal{E}\{x^n\} = \int_{-\infty}^{+\infty} x^n p(x)dx$$

defines the nth moment of **x**. The first moment is the mean. The nth central moment is defined by

$$\mathcal{E}\{(x - \mathcal{E}(x))^n\} = \int_{-\infty}^{+\infty} (x - \mathcal{E}(x))^n p(x)dx.$$

The second central moment is the **variance**:

$$\text{var}(x) = \mathcal{E}\{(x - \mathcal{E}(x))^2\} = \mathcal{E}\{x^2\} - (\mathcal{E}\{x\})^2.$$

2.2.2 Normal distribution

An r.v. has a normal or Gaussian distribution if its p.d.f. is given by

$$p_{\mathrm{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{2.4}$$

where

$$\mu = \int_{-\infty}^{+\infty} x p_{\mathrm{x}}(x) dx$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 p_{\mathrm{x}}(x) dx.$$

The normal density is important because it has some very nice mathematical properties. One important property is that the normal density is completely defined by its mean and variance and is written

$$\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2).$$

Variance is a measure of dispersion about the peak so that a small σ^2 corresponds to a sharp peak while a large σ^2 corresponds to a flat peak. The normal distribution is also important because it appears quite often in practice. One possible reason for this is given by the "Central Limit Theorem" which says that a super-position of independent random variables always tends toward normality regardless of the individual distributions involved. Therefore, if noise is due to a superposition of many small contributions, it may be reasonable to assume normality.

2.3.1 Joint density and distribution functions

The r.v.s x and y are jointly distributed if they are defined in the same probability space. The joint p.d.f. is then $p_{\mathrm{xy}}(x, y)$. The marginal densities of x and y are

$$p_{\mathrm{x}}(x) = \int_{-\infty}^{\infty} p_{\mathrm{xy}}(x, y)dy$$

$$p_{\mathrm{y}}(y) = \int_{-\infty}^{\infty} p_{\mathrm{xy}}(x, y)dx.$$

Let us define an n-vector, $\mathbf{x}$ as

$$\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n)^{\mathrm{T}}.$$

A realization of the random vector, $\mathbf{x}$ is then

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}.$$

As noted in Todling (1999), all of the previous definitions can be written in vector notation. For example, the p.d.f. is

$$p_{\mathbf{x}}(\boldsymbol{x}) = p_{\mathrm{x}_1\mathrm{x}_2\ldots\mathrm{x}_n}(x_1, x_2, \ldots, x_n).$$

The probability distribution is

$$\begin{aligned} F_{\mathbf{x}}(\boldsymbol{x}) &= \int_{-\infty}^{\boldsymbol{x}} p_{\mathbf{x}}(\boldsymbol{x}')d\boldsymbol{x}' \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_{\mathrm{x}_1\ldots\mathrm{x}_n}(x_1', \ldots, x_n')dx_1' \ldots dx_n'. \end{aligned} \tag{2.23}$$

The probability density function is defined as the derivative of the distribution function:

$$p_{\mathbf{x}}(\boldsymbol{x}) = \frac{\partial^n F_{\mathbf{x}}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial^n F_{\mathbf{x}}(\boldsymbol{x})}{\partial x_1 \ldots \partial x_n}. \tag{2.24}$$

x, y are statistically independent if $p_{\mathrm{xy}}(x_i, y_j) = p_{\mathrm{x}}(x_i)p_{\mathrm{y}}(y_j)$ for all x_i, y_j.

Example 2.8 *Darts (Brown, p35)*

Let the position of a hit be given by its (x,y) coordinates. After sufficient practice, assume that the scatter in position is unbiased in both directions. The joint p.d.f. is then

$$p_{\mathrm{xy}}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{2.10}$$

The marginal densities are:

$$\begin{aligned} p_{\mathrm{x}}(x) &= \int_{-\infty}^{\infty} p_{\mathrm{xy}}(x, y)dy = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \\ p_{\mathrm{y}}(y) &= \int_{-\infty}^{\infty} p_{\mathrm{xy}}(x, y)dx = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}. \end{aligned}$$

The expectation of the product of two r.v.s is of special interest:

$$\mathcal{E}(\mathrm{xy}) = \int_{-\infty}^{\infty} xyp_{\mathrm{xy}}(x, y)dxdy. \tag{2.11}$$

If x and y are independent, then

$$\mathcal{E}(\mathrm{xy}) = \int_{-\infty}^{\infty} xp_{\mathrm{x}}(x)dx \int_{-\infty}^{\infty} yp_{\mathrm{y}}(y)dy = \mathcal{E}(x)\mathcal{E}(y). \tag{2.12}$$

If $\mathcal{E}(\mathrm{xy}) = \mathcal{E}(\mathrm{x})\mathcal{E}(\mathrm{y})$, then x, y are uncorrelated.

If x, y are independent, then they are uncorrelated.

If x, y are uncorrelated, **this does not mean that** they are independent.

If $\mathcal{E}(\mathrm{xy}) = 0$, x and y are said to be orthogonal.

The **covariance** of x and y is

$$cov(\mathrm{x}, \mathrm{y}) = \mathcal{E}\{(\mathrm{x} - \mu_x)(\mathrm{y} - \mu_y)\}.$$

The correlation coefficient is defined as

$$\rho = \frac{\mathcal{E}\{(\mathrm{x} - \mu_x)(\mathrm{y} - \mu_y)\}}{\sqrt{\mathcal{E}\{(\mathrm{x} - \mu_x)^2\}}\sqrt{\mathcal{E}\{(\mathrm{y} - \mu_y)^2\}}}.$$

Note that if x=y, then ρ=1, and if x=-y, ρ=-1. If x and y are uncorrelated, then

$$\begin{aligned}
\mathcal{E}\{(\mathrm{x} - \mu_x)(\mathrm{y} - \mu_y)\} &= \mathcal{E}\{\mathrm{xy} - \mu_x\mathrm{y} - \mu_y\mathrm{x} + \mu_x\mu_y\} \\
&= \mathcal{E}\{\mathrm{xy}\} - \mu_x\mu_y - \mu_y\mu_x + \mu_x\mu_y \\
&= \mathcal{E}\{\mathrm{x}\}\mathcal{E}\{\mathrm{y}\} - \mu_x\mu_y \\
&= \mu_x\mu_y - \mu_x\mu_y \\
&= 0. \qquad (2.13)
\end{aligned}$$

Thus if x and y are uncorrelated, then ρ=0.