Tropical systematic and random error energetics based on NCEP (MRF) analysis-forecast system – A barotropic approach *Part I: in physical domain*

S DE* and D R CHAKRABORTY

Indian Institute of Tropical Meteorology, Dr Homi Bhabha Road, P.O. NCL, Pashan, Pune 411 008, India.

**e-mail: sde@tropmet.res.in*

Deterministic predictability in the perspective of systematic and random error and their growth rates and different components of growth rate budgets like flux, pure generation, mixed generation and conversion in energy/variance form are investigated in physical domain for medium range tropical (30°S–30°N) weather forecast using daily horizontal wind field at 850 hPa up to 5-day forecast for the month of June, 2000 of NCEP (MRF) model.

The study reveals the following:

- The Indian peninsula, the Indonesian region and their adjoining areas over 10°N–20°N latitudinal belt show a large amount of forecast error variance indicating that cumulus parameterization process may play a major role in the generation of tropical systematic error.
- Sparse observational networks over the tropical region are attributed to the uniform spread of random error over the continental as well as oceanic area. The results suggest that generation of random error in some geographical locations is perhaps due to the inefficient description of sensible heating process in the model.
- As far as growth rates are concerned, systematic error growth rate increases at initial forecast time and attains maximum value at 2-day forecast then it remains unchanged for rest of the forecast days. Whereas, the growth rate of random error is nearly invariant at 1 and 2-day forecasts and then it increases slowly at subsequent forecast time.
- Analyzing the flux, pure generation, mixed generation and conversion terms involved with the components of systematic and random error growth rate budget, it is shown that the components have their large variance in those regions where the respective error predominates.

1. Introduction

One of the most far reaching problems to be completely solved for the meteorological community since the time of invention of NWP by Richardson in the year 1922 is atmospheric predictability. Since then, atmospheric predictability is not considered a subject of predicting the future state of the atmosphere in practice but up to what extent it can be possible to forecast weather using complete

theoretical knowledge of physical laws that govern the atmosphere. Thompson (1957) first dealt with the subject through the study of initial state error in the large scale atmospheric flow to give a satisfactory answer to the question if increasing the data density coverage is the only mean to combat the problematic return in the accuracy of forecasts, raised by the then National Meteorological Service. Thompson has made two important conclusions in this regard. Firstly, doubling

Keywords. Systematic error; random error; error energy and predictability.

the cost of maintenance and overall density of the observing stations reduces 50% rms error of vector wind. Secondly, to remove the discrepancy generated between the practical forecastability and the theoretically calculated maximum range of predictability, the error in prediction depends not only on the atmospheric initial state but on other parameters also like period of forecast, size of transient disturbances, vertical wind shear, rms wind anomaly and the static stability of atmosphere.

Lorenz (1963) in his classical paper dealt with the other facet of the subject atmospheric predictability and studied the deterministic behaviour of the flow using three shallow water equations representing the convective process in a dissipative system instead of a conservative system. As a result, he obtained solutions which are nonperiodic and chaotic in nature and while applying these solutions to the large scale atmospheric flow Lorenz clearly mentioned that much greater uncertainty is involved resulting in the failure to observe the state of the atmosphere and in the inadequate model formulation governing the physical processes in comparison to the uncertainty generated from the indeterministic flow. So, it is too early to deal with the error associated with the indeterministic flow in predictability studies. Following this paper, most of the researches are performed assuming the atmospheric system as the deterministic system.

Systematic error, an error with a non-zero mean over a large number of realizations, is identified as a stationary component of total forecast error and is generated mainly from the inadequacy in the model formulation and parameterization of physical processes. Random error, identified as a deviation from non-zero mean at each realization, is a transient part of the total forecast error. It is comprised of errors in initial conditions due to poor observational coverage and error generated from inaccurately describing the forcing functions of the model.

Extensive research has already been performed through statistical and empirical methods like objective score, anomaly correlation etc. and by empirically formulated error growth rate budgets over global and extratropical regions. Kanamitsu (1985) and Heckley (1985) both studied the geographical distribution of forecast error in temperature, wind field and moisture etc. over the tropical region but they have not discussed the error growth rate. Lorenz (1982), Dalcher and Kalanay (1987) discussed the error growth by empirically parameterizing the error growth rate budget using 100 days ECMWF data of 10 days forecast. N. Surgi (1989) investigated the nature of systematic error of FSU global spectral model carrying out sensitivity experiment using winter data of 1978/79 and expressed that orography and model physics

are considered to be the possible causes for errors. Kamga *et al* (2000) depicted ECMWF model biases over tropical Africa using summer data of 1995 and conjectured possible suggestions for potential improvement of ECMWF model and data system. Recently, Roy Bhowmik (2004) has estimated the systematic error in IMD operational model over Indian region using wind, temperature, geopotential and relative humidity.

The investigation on error growth, by formulating the dynamical error growth rate equation derived from the primitive equations following the dynamical-empirical approach to be the best method prescribed by Lorenz (1969) was first performed by Boer (1984, 1993, 1994) to study the atmospheric predictability. In dynamical-empirical approach, each dependent variable of error growth rate equation is expressed in terms of the estimated form of error energy / variance.

Keeping in view the above introduction, first time authors in this field, have concentrated on documenting the geographical distribution of systematic and random error and their growth rate budgets which consist of different terms that govern the error growth in a forecast model using the horizontal wind field over the global tropical (30°S–30°N) belt. In this study authors adopt the dynamical error growth rate equations given by Boer (1993) with modifications based on tropical dynamics. Each term of the error growth rate budget equations and the systematic and random error itself are expressed in error energy/variance form. They are evaluated in geographical domain at different forecast time and are analyzed.

2. Choice of parameter

Boer used the systematic and random error growth rate equations at 500 hPa level for extratropical regions. So, naturally the equations are geostrophic and the basic variable is the geopotential height. For suitable application of the same equations over the tropical region, total wind field is used instead of geopotential height as the equations are not geostrophic over the tropics. Furthermore, the forecast of height field is not as appropriate as that of wind field due to the low variability of geopotential height at the subsequent forecast time over the tropical region (Kanamitsu 1985). Other than the above mentioned reasons, there is much less error over the tropical region (30°S–30°N) compared to that over extratropical region in the geographical distribution of systematic error of 10-day forecast of 500 hPa geopotential height (Dalcher and Kalanay 1987). Therefore it is preferred to choose the wind field as a parameter in error analysis and predictability studies over the tropical region.

3. Systematic and random error and their growth rate equations

Error in the forecast wind field may be written as, $Ve = Vf - Vo$; $ue = uf - uo$; $ve = vf - vo$; where V is the total wind field and u, v are the component wind in zonal and meridional direction and “e”, “f”, “o” refer to the error, forecast and observed (analyzed) part in total, zonal and meridional wind respectively. Further, forecast error in the wind field may be partitioned as, $ue = ues + uer$; $ve = ves + ver$; where ues, ves are the time mean error averaged over all days for a fixed forecast time and are termed as systematic error in zonal and meridional wind respectively. uer, ver are the deviation from the time mean at each day for a fixed forecast time and are named as the transient or random error in respective zonal and meridional wind. Now the systematic and random error kinetic energies are a function of forecast time (t) only and are expressed as

$$Ks(t) = \langle ks(t) \rangle = \left\langle \frac{1}{2} Ves \bullet Ves \right\rangle \quad (1)$$

$$\text{and } Kr(t) = \langle kr(t) \rangle = \left\langle \frac{1}{2} Ver \bullet Ver \right\rangle, \quad (2)$$

where $Ves = (ues, ves)$ and $Ver = (uer, ver)$ are the systematic and random error vector of total wind. The over bar represents the spatial average and $\langle \rangle$ represents the ensemble average over all days at a fixed forecast time and the dot represents the dot product. Systematic error (equation 1), which is already time averaged, may be written as

$$Ks(t) = \frac{1}{2} \overline{Ves \bullet Ves}. \quad (3)$$

Spatially averaged systematic and random error energy growth rate equations are used following Boer (1993) and are expressed as,

$$\begin{aligned} \frac{\partial ks}{\partial t} = & -\nabla \bullet \left[\left(\frac{Ves \bullet Ves}{2} + Ves \bullet Ver \right) \cdot Vf \right] \\ & + \left[\overline{(uer \cdot Vo + uer \cdot Ver) \bullet \nabla ues} \right. \\ & \quad \left. + \overline{(ver \cdot Vo + ver \cdot Ver) \bullet \nabla ves} \right] \\ & - \left[\overline{(ues \cdot Ves \bullet \nabla uo + ves \cdot Ves \bullet \nabla vo)} \right. \\ & \quad \left. + \overline{(ues \cdot Ver \bullet \nabla uo + ves \cdot Ver \bullet \nabla vo)} \right] \\ & + \overline{Ves \bullet Rs} \end{aligned} \quad (4)$$

and

$$\begin{aligned} \frac{\partial kr}{\partial t} = & -\nabla \bullet \left[\frac{Ver \bullet Ver}{2} \cdot Vf \right] \\ & - \left[\overline{(uer \cdot Vo + uer \cdot Ver) \bullet \nabla ues} \right. \\ & \quad \left. + \overline{(ver \cdot Vo + ver \cdot Ver) \bullet \nabla ves} \right] \\ & - \left[\overline{(uer \cdot Ver \bullet \nabla uo + ver \cdot Ver \bullet \nabla vo)} \right. \\ & \quad \left. + \overline{(uer \cdot ves \bullet \nabla uo + ver \cdot ves \bullet \nabla vo)} \right] \\ & + \overline{Ver \bullet Rr}. \end{aligned} \quad (5)$$

Ensemble average of the equations (4) and (5) are taken for all days at each forecast time. All the velocity fields in the above equations are actual winds instead of geostrophic winds as used by Boer, for more meaningful application of the error growth rate equations over the tropical region. The above equations deal with the barotropic processes. The systematic and random error energy growth rates in the l.h.s. of the equations (4) and (5) are mainly governed by the convergence or divergence of flux of error represented by the first term within the third bracket on the r.h.s. of each of the equations, conversion from random to systematic error associated with the second term in the third bracket of both equations, generation of systematic and random error due to the non-linear barotropic processes associated with the third term of the respective equation. The generation term may be partitioned into the pure and mixed components as some portion of error after generation is converted into the other form of error. The first part within the first bracket, which is purely generated contains either systematic or random component associated with the observed flow and is termed as pure generation term. The other part, which is converted contains mixed form of errors and is termed as the mixed generation term. The last term of each equation indicates the source-sink term, which represents all processes other than the barotropic process such as baroclinic process, dissipation, computational error, etc. Each term of equations (4) and (5) is computed from data except the source-sink term which is the residual term, estimated by taking the difference between the l.h.s and r.h.s. for both equations (Boer 1993).

4. Data and model

In this experiment, daily analyzed and 1–5 day forecast horizontal wind fields of NCEP medium range forecast (MRF) model at 850 hPa available at $2.5^\circ \times 2.5^\circ$ grid size are used over the global tropical (30°S – 30°N) belt. The whole data set contains

35 days data of 00GMT from 1st June, 2000 to 5th July, 2000.

In NCEP (MRF) model, primitive equation dynamics are expressed in terms of vorticity, divergence, the logarithmic of surface pressure, specific humidity and virtual temperature. The horizontal resolution of the model is T126 with 28 unevenly spaced sigma levels. The main time integration scheme is leapfrog semi implicit (gravity and zonal advection of vorticity) scheme. Arakawa-Schubert scheme is used for penetrative convection whereas shallow convection is parameterized by Tiedtke. Mean orography is taken into account from silhouette. Boundary layer effect is typically felt at the first five levels above the surface (at sigma = 0.995, 0.981, 0.960, 0.920 and .856). Surface solar observation is determined from surface albedo and long wave emission from Planck equations. Soil moisture is represented by single bucket model (Manabe).

5. Results and discussions

Kanamitsu (1985) reported that the main reasons for tropical weather forecast error are:

- inadequate parameterization of cumulus convection,
- error in initial conditions due to the insufficient data density coverage, and
- inaccuracy in describing sensible heating process in the model.

Recent study (Kamga *et al* 2000) shows that boundary layer process may play an appreciable role in model bias over the tropical region. Now, from the definition of systematic and random errors it is seen that both types of error are strongly influenced by the above mentioned causes. So, it is necessary to compute and analyze the systematic and random errors and their growth rates, as well as the different components like flux, conversion, pure generation and mixed generation governing the error growth rate of both types of error in physical (geographical) domain. At each grid point, both types of error result from the overall effects of the causes mentioned above. It is very difficult to separate the error due to an individual cause. The dominance of one cause may be explained by the distribution of systematic and random error over certain geographical regions.

Computations of systematic and random error variances, their growth rates, and the different components governing the error growths are performed in the physical domain using equations (2), (3), (4) and (5), utilizing the above mentioned data. Figures 1 and 2 show the geographical distribution of systematic and random error variance in m^2/sec^2 for 1–5 days forecast of 850 hPa horizontal

wind field over 30°S – 30°N global belt. The following are the important findings from these diagrams.

- The magnitude of systematic and random error variance and their area of coverage increases with forecast time.
- In the systematic error variance (figure 1), the Indian peninsula and adjoining oceanic areas, the Indonesian region and some parts of north Australia (10°S – 20°N , 60°E – 130°E) and the African region (10°N – 20°N) show dense and large amounts of error. There is no systematic error in the south Indian and Pacific Oceans. Small pockets of systematic error are found in the Mexican region and in the northern part of South America.
- The distribution of random error variance (figure 2) in the 4 and 5-day forecasts shows its large amplitude and almost uniform spread over the Indian peninsula and adjoining regions, the African region located between 10°N and 20°N , the Saudi Arabian region, the Mexican region, northern south American region, Indian Ocean, south Pacific and some portion of the north Pacific Ocean. It is found to be comparatively more intense over the three areas comprised of the Indian region (10°N – 20°N , 80°E – 100°E), the Mexican region (10°N – 20°N , 95°W – 105°W) and the Saudi Arabian region (25°N , 45°E) in the 5-day forecast.

The large magnitude and the dense areas of systematic error variance shown in figure 1 are the areas of land and ocean contrast. In the month of June during the Northern Hemispheric summer monsoon period, intense differential heating and the passage of ITCZ over these regions may lead to transform the regions into the convective areas where the cumulus parameterization process is critical to the forecast. In view of the boundary layer process, frictional effect is less felt due to the smooth terrain over the oceanic area. Model configuration shows that boundary layer process is less effective above the vertical level of sigma 0.856. So, the boundary layer process may not take the dominant role in the generation of systematic error at 850 hPa over the region of study. Hence, it may be inferred that the tropical systematic error is generated mainly from the inadequacy in formulation of the cumulus parameterization process.

Appearance of random error variance in both the continental and oceanic areas (figure 2) implies that the error is generated mainly due to the poor data density coverage for sparse observational networks over the tropical region. The three regions of large random error energy (Indian region, Mexican region and Saudi Arabian region) are also approximately the large recipients of sensible heat flux from the earth's surface to the atmosphere

Figure 1. Spatial distribution of systematic error variance in m^2/sec^2 at different forecast time.

Figure 2. Spatial distribution of random error variance in m^2/sec^2 at different forecast time.

according to Budyko's (1974) global sensible heat flux climatology for the month of June (figure 3). Though the authors are not dealing explicitly with sensible heat flux for this error study, the horizontal wind field plays a major role in the formulation of the sensible heat flux. Therefore, from the definition of random error it may be conjectured that the above mentioned three large regions of error may be due to inaccurately describing the sensible heating process in the model.

The growth rates of both components of error and the terms associated with their growth rate budget are expressed in m^2/sec^3 in the physical domain. The values are shown after multiplying the original values with 10^5 for their conversion from day to second. For both components of error, the large growth rate values (figures 4 and 5) appear in those areas where the respective error energy has their large magnitudes. Figure 4 reveals that the systematic error growth rate attains maximum value at 2-day forecast and then it remains almost invariant for the rest of the forecast days. There is no significant change in random error growth rate at 1-day and 2-day forecast except one or two pockets then the growth rate increases slowly at subsequent forecasts (figure 5). Figures 6 and 7 show the geographical distribution of flux in systematic and random error variance in the growth rate budget. Positive and negative values indicate respectively the convergence and divergence of flux of error by the forecast flow. In the systematic error flux distribution (figure 6), there is a strong convergence shown at 20°E and 40°E in the African region located between 10°N and 20°N on 3–5 day forecasts and in the Indian region (20°N , 75°E) on 5-day forecast. In the random error flux (figure 7), there are pockets of strong convergence of error in the southwest Pacific Ocean (20°S , 160°E) on 3–5 day forecasts, southeast Pacific Ocean (25°S , 140°W) and the Mexican region (20°N , 110°W) on 3-day forecast and in the Indian region (15°N , 70°E) on 4-day forecast.

By comparing the geographical distribution of pure and mixed generations for both types of error it is seen that the values are larger in pure generation than in the mixed generation. Figures 8 and 9 represent the geographical distribution of pure generation of systematic and random errors in the growth rate budget, respectively. Positive and negative values in pure generation imply the generation and dissipation of error respectively. Systematic error is generated mostly in the 10°N – 20°N latitudinal belt of the African region at 2–5 days forecast. The maximum amount of random error is generated over the Indian Ocean (5°S , 80°E) in 2 to 4-day forecasts and in the central Pacific Ocean (25°S – 30°S , 150°W) in 2 to 5-day forecasts. Mixed generation terms are not shown

in figures due to their insignificant values. As the mathematical expression of conversion in equations (4) and (5) is the same, figure 10 represents the geographical distribution of conversion for both types of errors. The positive value represents the non-linear conversion of random to systematic error and the negative value represents the opposite. There are large pockets of conversion from random to systematic error in the Indian region (20°N , 80°E) at 4-day and 5-day forecast, whereas conversion of systematic to random error is biased on oceanic region particularly on north Indian Ocean and some areas of west and central Pacific in the 4 and 5 day forecasts. In addition to these, the Mexican region also shows some amount of systematic to random error conversion.

In the above analysis it is understood that different components of the growth rate budget for both types of errors show their large magnitude in those areas where respective error predominates. In the geographical distribution of the residual term for systematic and random errors, values are generally higher since the residual term is comprised of the errors generated due to all other processes other than barotropic activity (figures are not shown).

6. Conclusions

Detailed computation, estimation and analysis of tropical systematic and random error and the different terms related to the growth rate budgets in energy/variance form based on the above mentioned data may lead to the following inferences.

- In the geographical distribution, tropical systematic error is confined largely to the Indian peninsula and adjoining areas, the Indonesian region and some parts of north Australia and the African region which are the convergence zones with large land–ocean heat contrast indicating that cumulus parameterization process plays a major role in the generation of tropical systematic error.
- Unlike the systematic error, random error distribution is quite different, with strong errors over oceanic regions, particularly the Indian Ocean, south Pacific and some portion of the north Pacific Ocean. Random error appears over the continents of the Indian region, 10°N – 20°N of the African region, the Saudi Arabian region, some portion of Mexico and south America, showing that it is basically generated due to the poor data density coverage over the tropics. In addition, there are three large pockets of random error in the Indian, Mexican and Saudi Arabian regions which are approximately large recipients of sensible heat flux also, implying the error due

Figure 3. Sensible heat flux from the earth's surface into the atmosphere ($\text{kcal cm}^{-2} \text{ month}^{-1}$) for the month of June (Budyko 1974).

Figure 4. Spatial distribution of the rate of change of systematic error ($\times 10^5$) in m^2/sec^3 at different forecast time.

Figure 5. Same as figure 3 but for the rate of random error growth ($\times 10^5$).

Figure 6. Same as figure 3 but for flux ($\times 10^5$) of systematic error energy growth rate budget.

Figure 7. Same as figure 3 but for flux ($\times 10^5$) of random error energy growth rate budget.

Figure 8. Same as figure 5 but for systematic error pure generation ($\times 10^5$) in error energy growth rate budget.

Figure 9. Same as figure 6 but for random error pure generation ($\times 10^5$) in error energy growth rate budget.

Figure 10. Same as figure 3 but for conversion ($\times 10^5$) between systematic and random errors.

to inaccurately describing sensible heating processes in the model over the tropical region.

- Growth rate of systematic error is maximum at 2-day forecast, then it remains almost the same for rest of the forecast days but that of random error is nearly invariant during initial forecast time and then increases slowly at subsequent forecasts.
- Different components of growth rate budget like flux, pure generation and conversion terms in the systematic and random error show their bias in those areas where the respective error is also large.

To study the behaviour of the systematic and random error at the two consecutive seasons, extended study can be performed in a similar way using daily data of summer and winter season with more number of forecast days. Partitioning the wind field into rotational and divergent parts and by suitably modifying the error growth rate equations, contribution of systematic and random error and their growth rate budget for rotational and divergent motions may be investigated separately at different vertical levels. The impact of various sensitivity experiments (convective process, boundary layer process) in the systematic and random error may be investigated to explore the possible mechanism for the generation of error. Comparative study during El-Nino and non-El-Nino years may also be performed.

Acknowledgements

The authors wish to express their gratitude to Dr. G B Pant, Director, IITM for his keen interest and for providing facilities to carry out the work. Thanks are due to Dr. (Mrs) P S Salvekar for her constant help and support. The authors are especially indebted to Prof. Robert S Ross, Dept.

of Meteorology, Florida State University, USA for his valuable suggestions and overall improvement of the paper. Authors are also thankful to Brian Doty, COLA for using GrADS software.

References

- Budyko M I 1974 *Climate and Life*; New York: Academy Press, pp. 497
- Boer G J 1984 A spectral analysis of predictability and error in an operational forecast system; *Mon. Weather Rev.* **112** 1183–1197
- Boer G J 1993 Systematic and random error in an extended range forecasting experiment; *Mon. Weather Rev.* **121** 173–188
- Boer G J 1994 Predictability regimes in atmospheric flow; *Mon. Weather Rev.* **122** 2285–2295
- Dalcher A and Kalanay E 1987 Error growth and predictability in operational ECMWF forecasts; *Tellus* **39A** 474–491
- Heckley W A 1985 Systematic errors of the ECMWF operational forecast model in tropical regions; *Q. J. R. Meteorol. Soc.* **111** 709–738
- Kanamitsu M 1985 A study of predictability of ECMWF Operational Forecast Model in the tropics; *J. Meteor. Soc. Japan* **63** 779–804
- Kanga A F, Fongang S and Viltard A 2000 Systematic errors of the ECMWF operational model over tropical Africa; *Mon. Weather Rev.* **128** 1949–1959
- Lorenz E N 1963 Deterministic nonperiodic flow; *J. Atmos. Sci.* **20** 130–141
- Lorenz E N 1969 Three approaches to atmospheric predictability; *Bull. Amer. Meteorol. Soc.* **50** 345–349
- Lorenz E N 1982 Atmospheric predictability experiments with large numerical model; *Tellus* **34** 505–513
- Richardson L F 1922 *Weather prediction by numerical process*; London: Cambridge Univ. Press, (Reprinted Dover, 1965), pp 236
- Roy Bhowmik S K 2004 Systematic errors of IMD operational NWP model; *Mausam* **55(1)**, (to appear)
- Surgi N 1989 Systematic errors of the FSU global spectral model; *Mon. Weather Rev.* **117** 1751–1766
- Thompson P D 1957 Uncertainty of the initial state as a factor in the predictability of large scale atmospheric flow pattern; *Tellus* **9** 275–295