Seasonal Tropical Cyclone Forecast – Part 1

Lee Sai Ming
Hong Kong Observatory,
Hong Kong, China

Typhoon Committee Roving Seminar
30 Nov – 3 Dec 2010, Thailand

香港天文台
HONG KONG OBSERVATORY

Seasonal TC forecast Part 1 - Contents

- Introduction
- HKO's experience
- Recent attempts to improve TC forecast
- Data & methodology of current method
- Verification results

Seasonal TC forecast Part 2 - Contents

- Application of the forecast method to other coastal cities (Manila & Da Nang)
- Real life examples from HKO

香港天文台
HONG KONG OBSERVATORY

What I do

- My division: Climate Change and Climate Prediction
- My team: Long range forecast

1. Annual outlook for Hong Kong
2. Seasonal forecast (4 times a year) for HK
3. Internal monthly forecast
4. Joint pilot project on forecasting yield collected at reservoirs

- ENSO update

1. Internal and for general public

The need for monthly and seasonal TC forecast

Long-range TC forecast supports:

- Risk assessment
- Disaster prevention / reduction and preparedness planning
- Policy decision
- Product pricing, e.g. insurance

WMO IWTC
(International Workshop on Tropical Cyclones)

- To examine current knowledge, forecast and research trends
- IWTC VII was held in La Reunion, 15-20 Nov 2010
- IWTC VI was held in San Jose, Costa Rica, Nov 2006
- WMO Bulletin 56 (4): Seasonal Tropical Cyclone Forecasts

WMO Bulletin 56 (4) – a very comprehensive overview of seasonal TC forecasts

Group	Basins	Type	Website
City University of Hong Kong, China (CityU)	Western North Pacific	Statistical	http://aposf02.cityu.edu.hk
Colorado State University, USA (CSU)	Atlantic	Statistical	http://hurricane.atmos.colostate.edu
Cuban Meteorological Institute (INSMET)	Atlantic	Statistical	http://www.met.inf.cu
European Centre for Medium-Range Weather Forecasts (ECMWF)	Atlantic Australian Eastern North Pacific North Indian South Indian South Pacific Western North Pacific	Dynamical	http://www.ecmwf.int (collaborating agencies only)
International Research Institute for Climate and Society (IRI)	Atlantic Australia Eastern North Pacific South Pacific Western North Pacific	Dynamical	http://iri.columbia.edu/forecast/tc_fcst/
Macquarie University, Australia	Australia / southwest Pacific	Statistical	http://www.iges.org/ellfb/past.html
Meteorological Office, United Kingdom (MetOffice)	North Atlantic	Dynamical	http://www.metoffice.gov.uk/weather/tropicalcyclone/northatlantic
National Meteorological Service, Mexico (NSM)	Eastern North Pacific	Statistical	http://smn.cna.gob.mx
National Climate Centre, China	Western North Pacific	Statistical	http://bcc.cma.gov.cn
NOAA hurricane outlooks	Atlantic Eastern North Pacific Central North Pacific	Statistical	http://www.cpc.noaa.gov http://www.cpc.noaa.gov http:// www.prh.noaa.gov/hnl/cphc
Tropical Storm Risk (TSR)	Atlantic Western North Pacific Australian region	Statistical	http://tsr.mssl.ucl.ac.uk

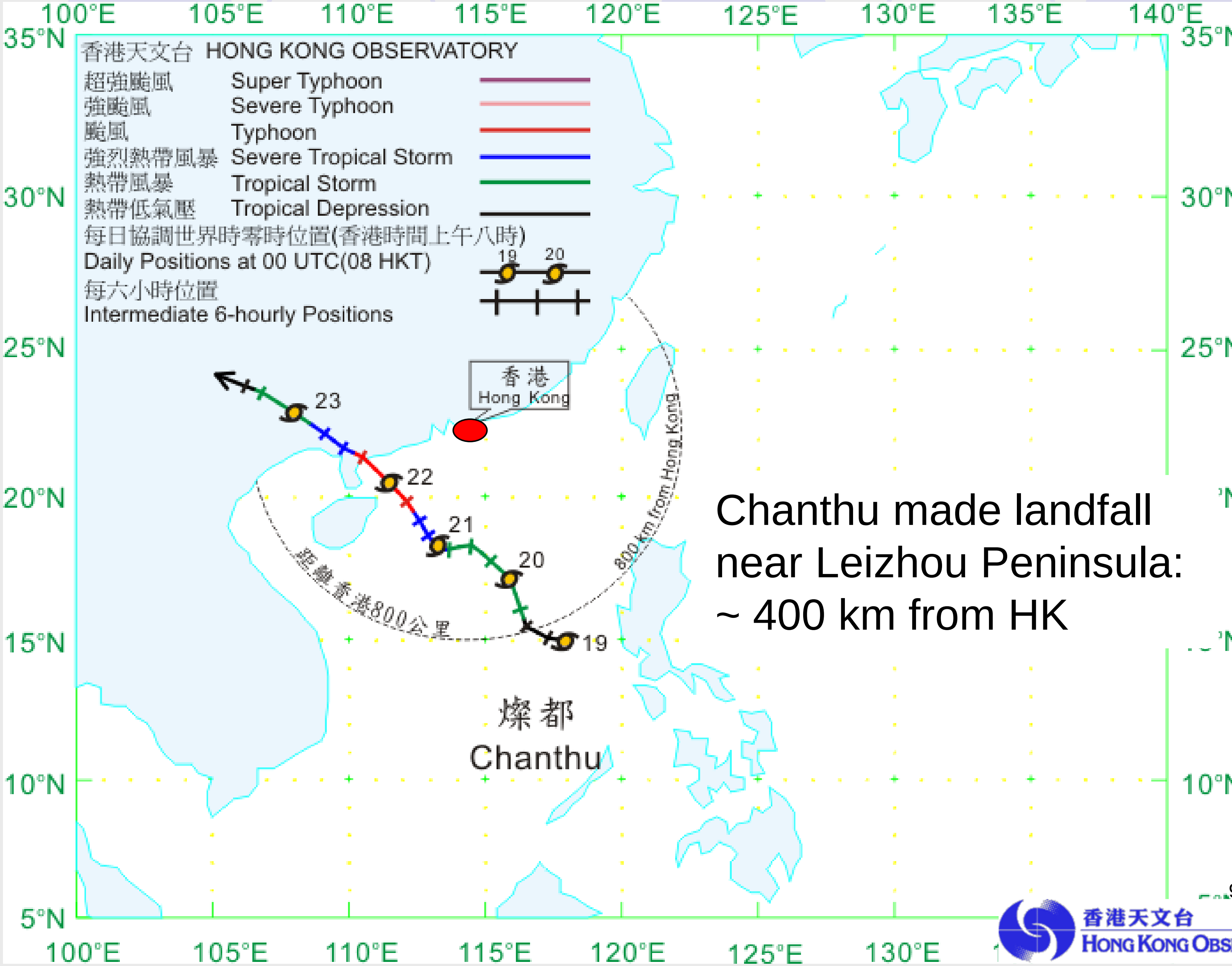
WMO Bulletin 56 (4)

- Websites of forecast producing centres
- Methods used by centres:
  - statistical (in use since the early days)
  - dynamical (getting more important)
- Forecast products:
  - number of TC / named storms
  - ACE index (Accumulated cyclone energy)
  - mean position of TC
  - number / probability of landfalling TC

Seasonal TC forecast

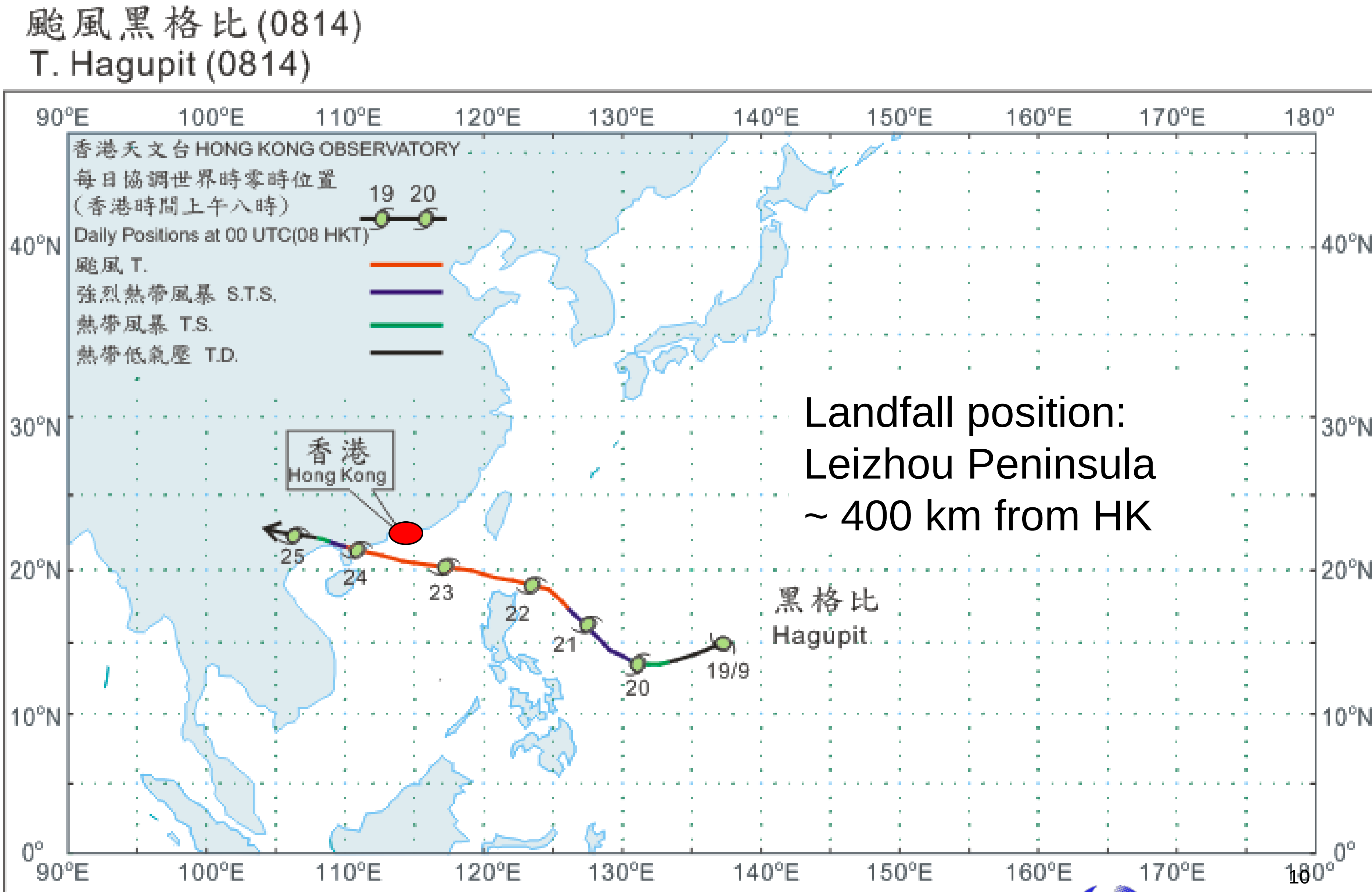
- No. of TC / named storms / ACE index over an ocean basin
  - No region-specific information
  - How to use these forecasts?
- No. of TC landfalls
  - TC affecting a city does not need to make landfall there
  - 2 examples from Hong Kong

Typhoon Chanthu, July 2010



Chanthu made landfall near Leizhou Peninsula: ~ 400 km from HK

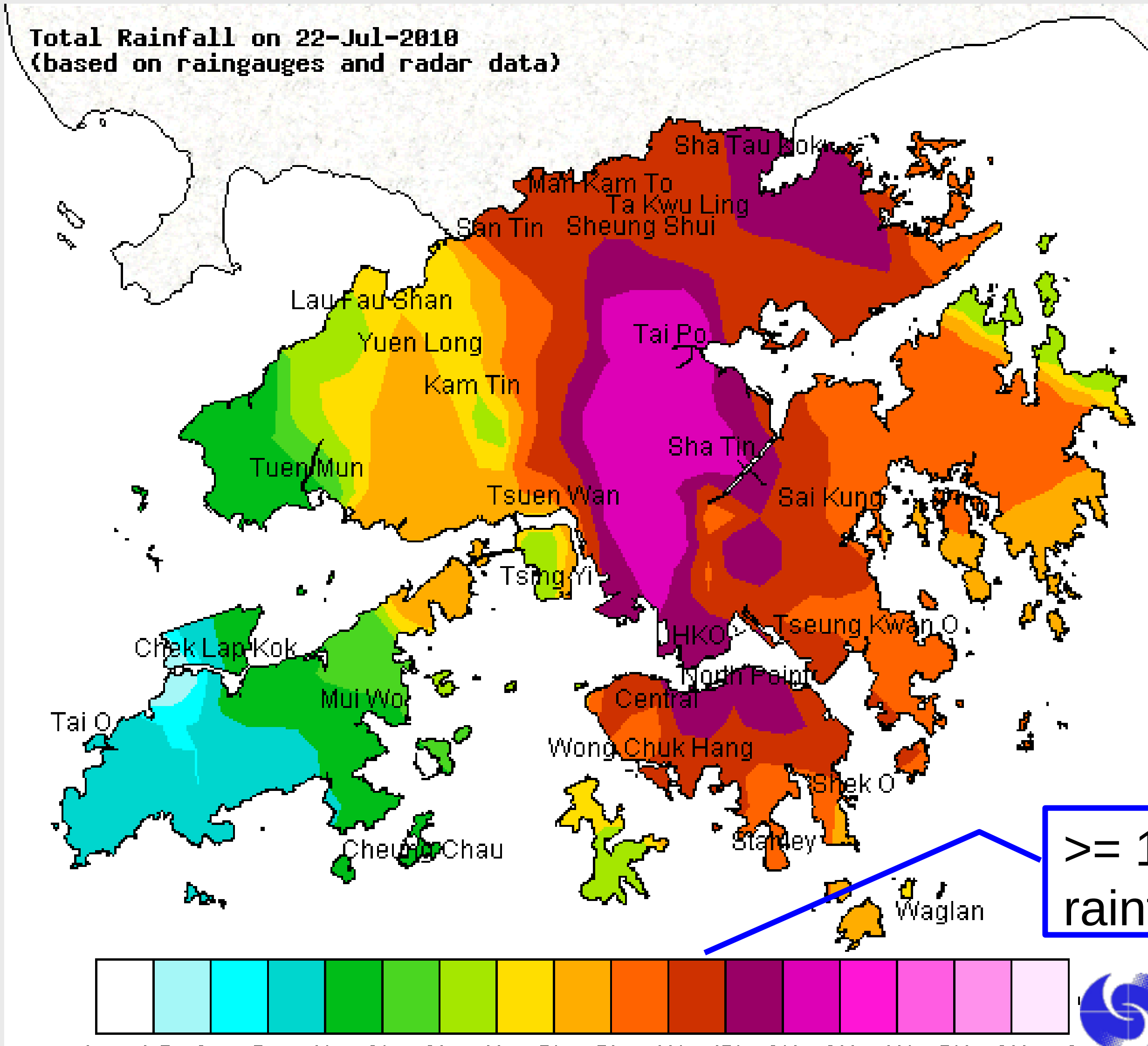
Typhoon Hagupit, September 2008



Landfall position:
Leizhou Peninsula
~ 400 km from HK

Heavy rain brought by Typhoon Chanthu (July 2010)

(more than 100 mm of rainfall recorded in a couple of hours)



>= 100 mm of rainfall

Flooding in Tai O after Typhoon Hagupit (Sep 2008)

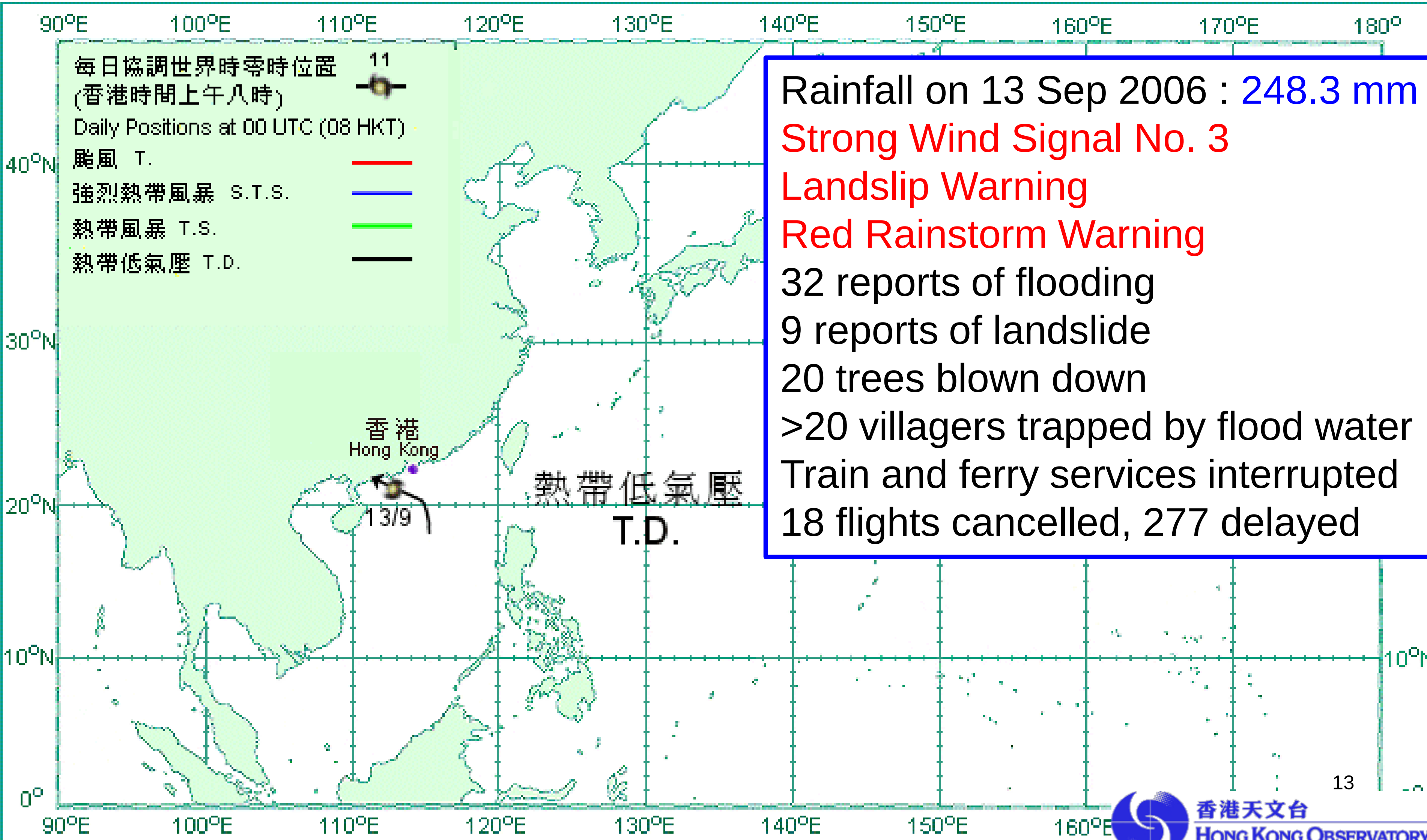
(storm surge caused by Hagupit)

(courtesy of TVB)

香港天文台 HONG KONG OBSERVATORY

Tropical Depression can also be devastating

熱帶低氣壓：九月十二至十三日
Tropical Depression of 12 - 13 September 2006



Rainfall on 13 Sep 2006 : 248.3 mm
Strong Wind Signal No. 3
Landslip Warning
Red Rainstorm Warning
32 reports of flooding
9 reports of landslide
20 trees blown down
>20 villagers trapped by flood water
Train and ferry services interrupted
18 flights cancelled, 277 delayed

HKO's experience

- HKO has been issuing the annual outlook for HK since 2001 (disseminated over the Internet)
  1. Yearly rainfall of HK in tercile category (above normal, near normal, or below normal)
  2. Yearly number of TC "affecting" HK, e.g. 5-6 TC

- TC "affecting" HK = TC necessitating the issuance of local warning signals

HKO's experience

- Old methodology: an ENSO-based conceptual model:
  - El Niño year – fewer TC affecting HK
  - La Niña year – more TC affecting HK
- For each ENSO situation (El Niño, La Niña, neutral), there is an empirical distribution of yearly no. of TC affecting HK
- Prob of no. of TC affecting HK =

  ∑ prob(no. of TC affecting HK | ENSO situation of the year) * prob(ENSO situation)

  [summation over the three possible ENSO situations]

Problems with the old methodology

- Classification of the ENSO status of the year can be difficult at times: e.g. 2010
  - El Niño during the 1st half of the year
  - La Niña during the 2nd half of the year
- High uncertainty in the ENSO forecast (the annual outlook is issued in March)
- ∑ prob(no. of TC affecting HK | ENSO situation of the year) * prob(ENSO situation)
  - A strong tendency towards the climate normal
- TC "affecting" HK = TC necessitating the issuance of local warning signal
  - Subjective judgment involved, not entirely objective
- A recent study shows that El Niño's impact on TC activity affecting HK (another definition) is not significant. La Niña's impact is confined to late season.

Recent attempts

- Can we apply the Poisson regression model to do the TC forecast?
- Attempted a Poisson regression model trained by actual Niño 3.4 SST anomaly
  - A perfect prognosis approach
  - Only one single predictor
- Dynamical model output (digital/numerical data) are available on the web. Can they be utilized to formulate the Poisson regression model? How?

A recent investigation

- A statistical-dynamical method to forecast monthly, seasonal and annual TC activity "affecting" a region/city.
- Correlate regional TC activity with large scale dynamical climate model forecast
- The method is still evolving. Comments and suggestions are most welcome.

Data

- TC data source: HKO TC best track dataset (include all TC categories)
- Dynamical model data source: WMO designated **G**lobal **P**roducing **C**entres for Long Range Forecasts

WMO GPC

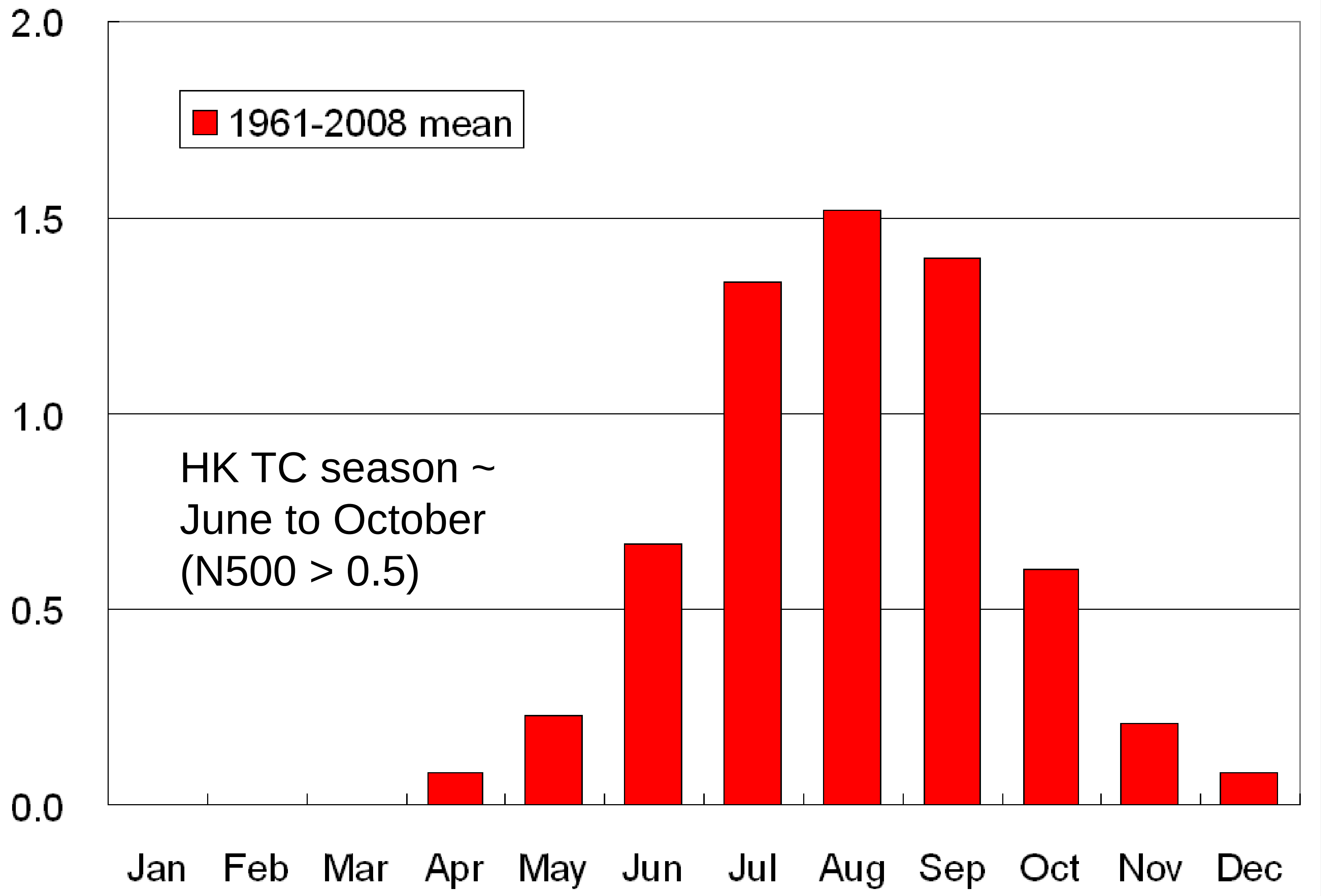
http://www.wmo.int/pages/prog/wcp/wcasp/clips/producers_forecasts.html

- Bureau of Meteorology (BoM), Australia
- China Meteorological Administration (CMA) / Bejing Climate Center (BCC)
- Climate Prediction Center (CPC), NOAA, USA [http://www.cpc.ncep.noaa.gov/]
- European Centre for Medium-Range Weather Forecasts (ECMWF)
- Japan Meteorological Agency (JMA) / Tokyo Climate Center (TCC) [http://ds.data.jma.go.jp/gmd/tcc/tcc/index.html]
- Korea Meteorological Administration (KMA)
- Meteo-France
- Met Office (United Kingdom)
- Meteorological Service of Canada (MSC)
- South African Weather Services (SAWS)
- Hydrometeorological Centre of Russia

TC activity affecting a city

- Definition: No. of TC coming within a certain range and a certain period of time
- Hong Kong: N500 [within 500 km of HK]
- Long term mean of annual N500 ≈ long term mean of annual Nsig [issuance of warning signals]

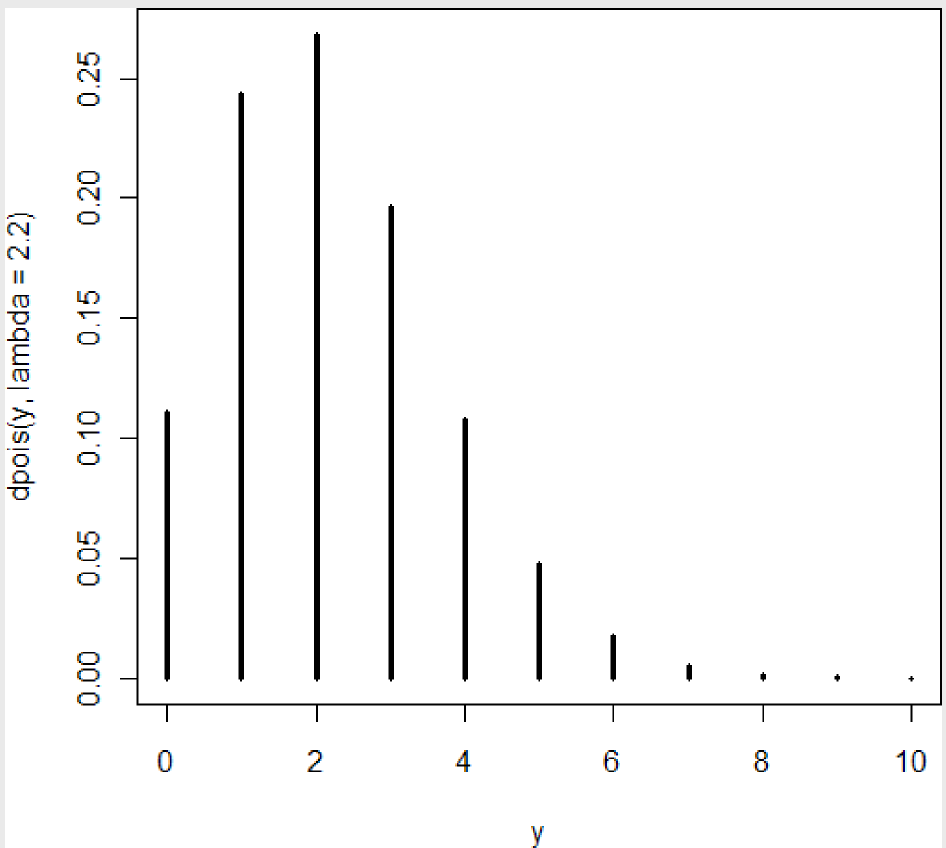
Monthly N500 of Hong Kong



Methodology

- N500 is a count parameter
- Can be modelled by the Poisson distribution

$$p(y,\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad y = 0,1,2,\ldots$$



香港天文台 Hong Kong Observatory

Methodology

- N500 is a count parameter
- Can be modelled by the Poisson distribution

$$p(y,\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad y = 0,1,2,\ldots$$

- The Poisson dist. belongs to the family of exponential dist.

$$f(y;\theta) = s(y)t(\theta)e^{a(y)b(\theta)}$$

where *a, b, s, t* are known functions

Methodology

- The Poisson dist. can be written in the canonical form:

$$f(y;\theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

where $a(y)=y$, $b(\theta)=\log\theta$, $c(\theta)=\theta$, $d(y)=-\log y!$

- We can formulate a generalized linear model to forecast the expected value of y

Ref.: Dobson, A. J, A. G. Barnett (2008):

An Introduction to Generalized Linear Models

Generalized Linear Model (GLM)

$$g(E(Y_i)) = x_i^T \beta \qquad i = 1, 2, \ldots N$$

Y_i = response variable

E = expectation of the dist.

g = link function (monotone, differentiable)

x_i = covariates or explanatory variables or predictors (p x 1 vector)

β = model parameters (p x 1 vector)

N = no. of realizations

This formulation is also known as Poisson regression model when Y_i comes from a Poisson distribution.

Generalized Linear Model (GLM)

$$g(E(Y_i)) = x_i^T \beta \qquad i = 1, 2, \ldots N$$

GLM is an extension of the classical linear regression model

In the classical linear model:

Y_i comes from a normal distribution

g = identity function

Generalized Linear Model (GLM)

$$g(E(Y_i)) = x_i^T \beta \qquad i = 1, 2, \ldots N$$

Y_i : monthly N500 derived from HKO TC best track data

g : natural log (canonical link)

β : maximum likelihood estimators to be found by an iterative weighted least squares (IWLS) procedure

x_i = ??? (what predictors do we need?)

Predictors

- Variability of TC activity is governed by atmospheric and oceanic conditions
- The predictors should be able to describe the atmospheric and oceanic conditions well
- Physical variables predicted by global climate model are good candidates

NCEP CFS

- Climate Prediction Center, NOAA, USA: a WMO designated **G**lobal **P**roducing **C**entre (GPC) of Long Range Forecasts
- CPC provides digital long range forecast and hindcast [generated by the NCEP **C**limate **F**orecast **S**ystem]
- Hindcast data used in this study: 1981-2008
- 12-hourly data, need to calculate monthly means

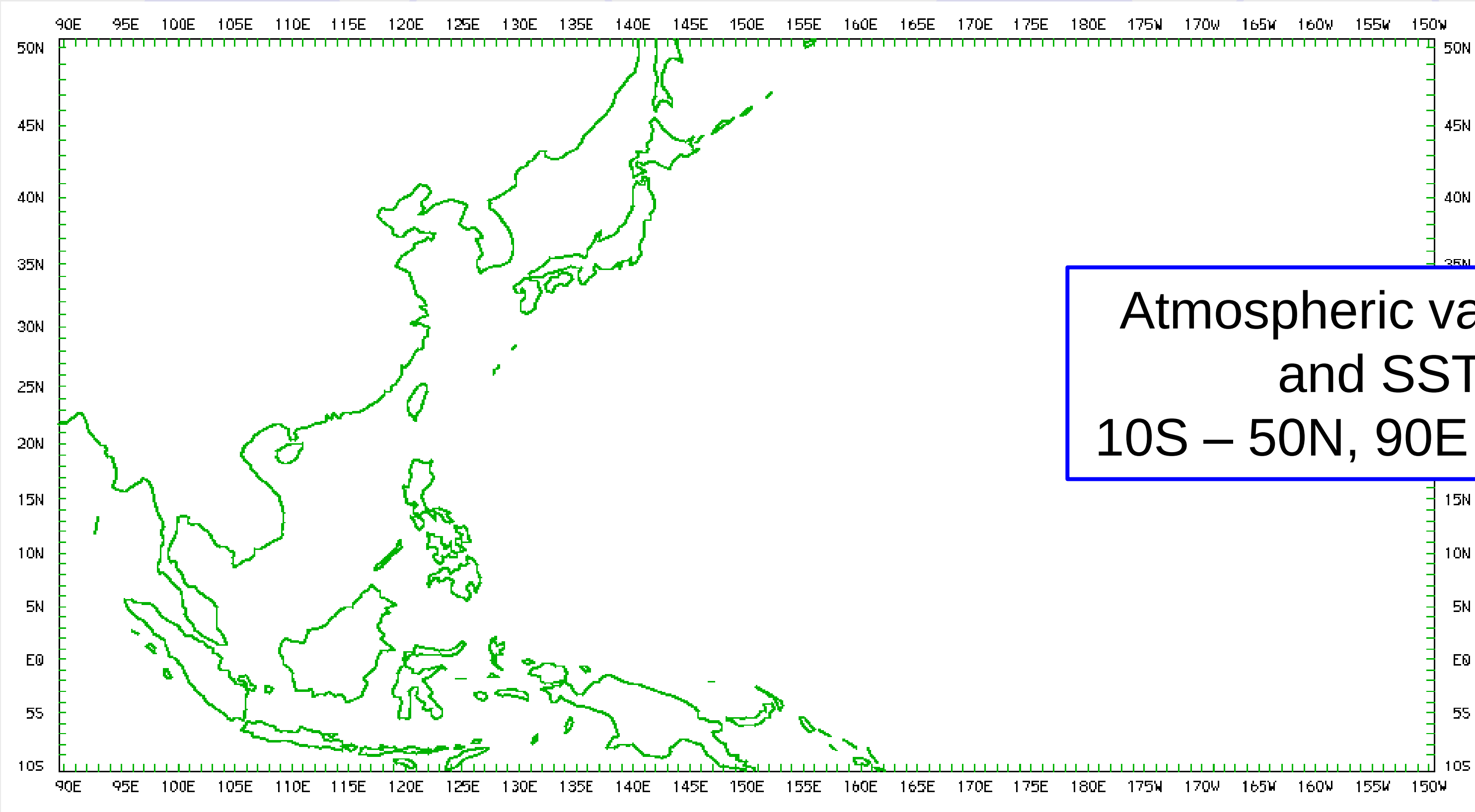
NCEP CFS

- Physical variables: [a total of 26]
  - mslp, 2m temperature, precipitation rate, precipitable water, SST
  - 850 hPa u, v wind, gph, streamfunction, velocity potential
  - 700 hPa gph, 500 hPa gph
  - 200 hPa u, v wind, gph, streamfunction, velocity potential
  - vorticity, divergence, vertical wind shear, thickness [these are derived elements]
- **9-month lead time**

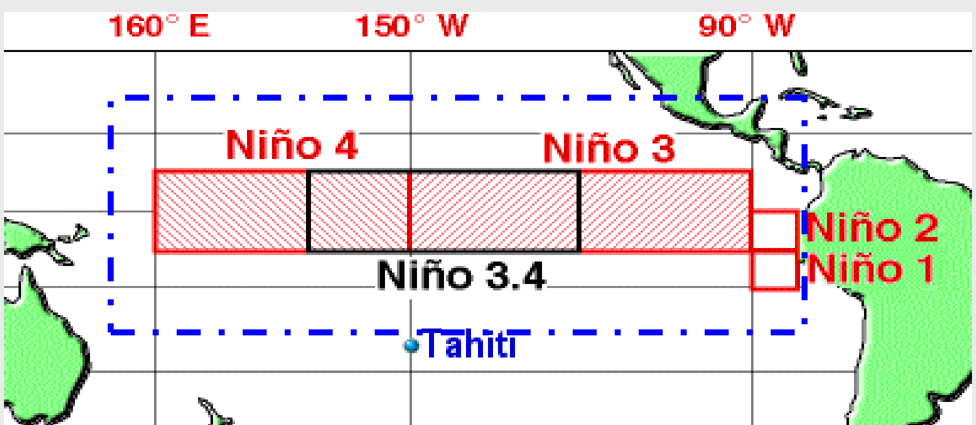
Consideration of the spatial coverage of the physical variables

- We are forecasting N500, but we shouldn't just look at a circle of 500 km in radius.
- We should consider a region where tropical cyclones develop, move and traverse toward HK
- A region where large-scale atmospheric circulations govern TC genesis and movement reside
- Also consider ENSO's effect on TC activity, i.e. SST of equatorial Pacific

Spatial coverage of predictors used in regression



Atmospheric variables and SST
10S – 50N, 90E – 150W



Eq. Pacific SST:
15S – 15N, 150E – 80W

Hugh amount of data

- Horizontal resolution of data:
  - 1 lat. x 1 lon.   for SST
  - 2.5 lat. x 2.5 lon.   for other elements
- No. of data grid points = 1225 [for each element]
- Impossible to regress on 1225 x 26 predictors with just 28 years of observations
- Can the data be condensed or compressed?

Empirical Orthogonal Function Analysis

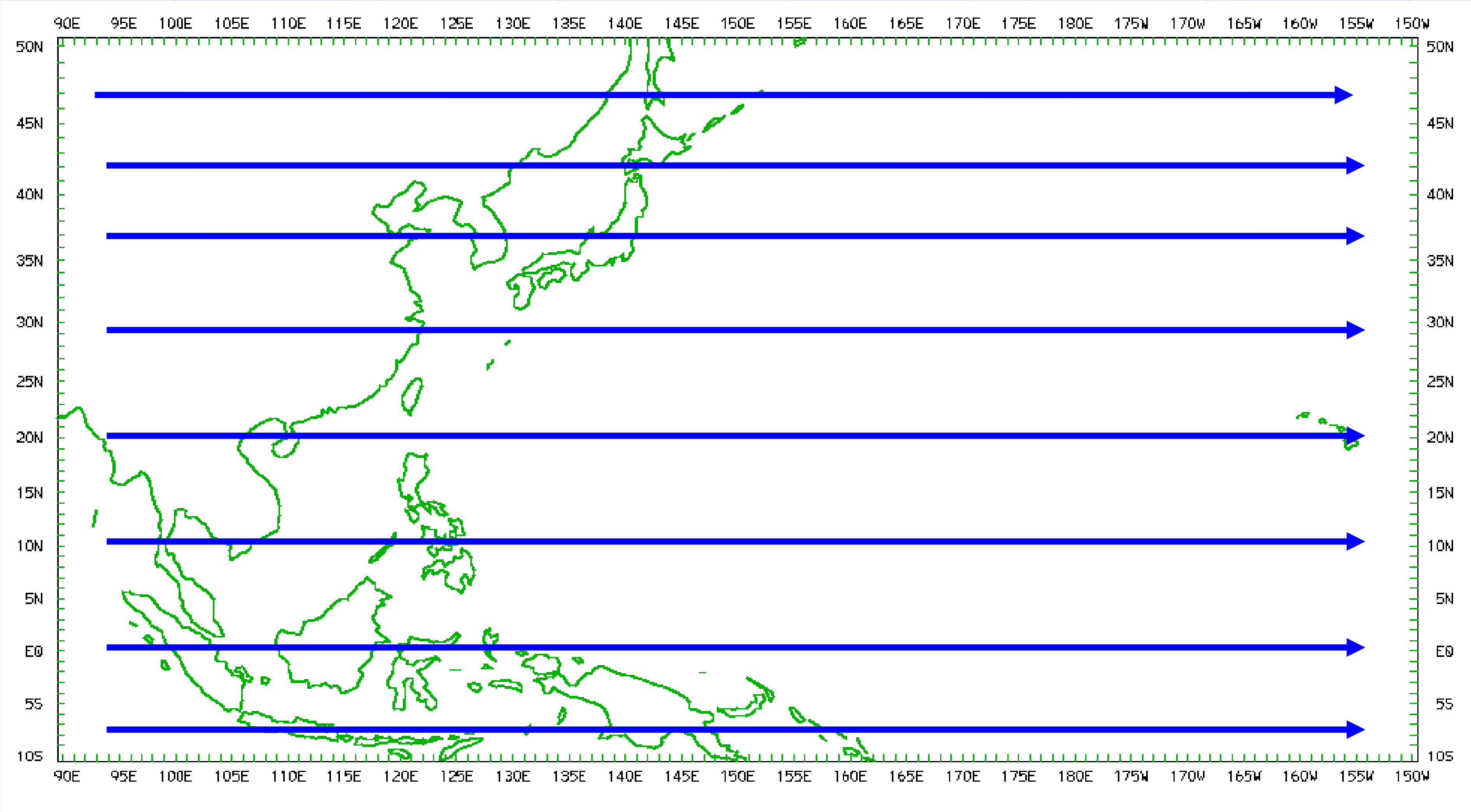
- Same as Principal Component Analysis (PCA)
- A powerful tool for data compression or dimensionality reduction in meteorology and oceanography

Example

- Suppose $\boldsymbol{x(t)}$ is the time series (28 years) of 500 hPa gph (standardized anomaly) over the EOF analysis domain (i.e. 1225 points, or 1225-dimensional).
- After EOF analysis, $\boldsymbol{x(t)} = \sum \alpha_i (t)\, \boldsymbol{e_i}$
- where i runs from 1 to 28, $\boldsymbol{e_i}$ is the i^{th} EOF and α_i is the i^{th} principal component
- Note that $\boldsymbol{e_i}$ s are constant vectors (eigenvectors)
- Hence, 1225 data points are compressed into 28 principal components.
- The closer $\boldsymbol{x(t)}$ resembles a particular $\boldsymbol{e_i}$, the larger α_i is.

Ref.: Wilks D. S. (2006): Statistical Methods in the Atmospheric Sciences

Example: 500 hPa gph of July 2001



We work on 28 PCs instead of the 1225 data points.

Remark

- The eigenvectors are normally found by solving the eigenvalue problem of the covariance (or correlation) matrix.
- In our case, we have 1225 data grid points but the time series is only 28 years long.
- Eigenvalues starting from 29 are all zero.
- We have to use the singular value decomposition (SVD) method. Outcome: 28 EOFs

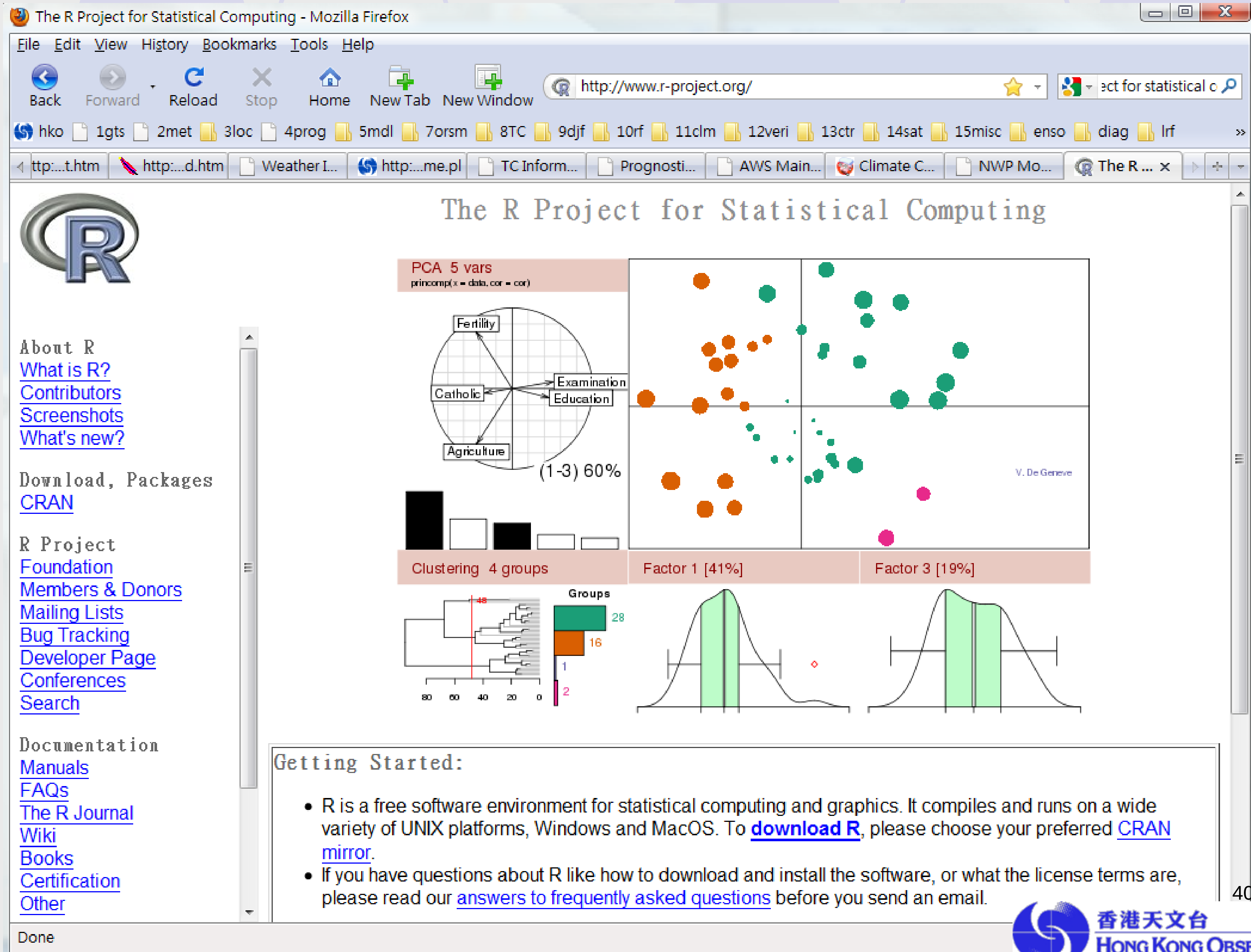
Ref.: Wilks D. S. (2006): Statistical Methods in the Atmospheric Sciences

Selection of predictors and combinations

1. Fit a single predictor GLM, search for skilful single predictor
2. Fit a multiple predictor GLM [predictors from step 1], filter out redundant predictors by stepwise regression

Cross-validate the ‘reduced’ or simplified GLM [from step 2], search for top performers

A no-cost tool



香港天文台 Hong Kong Observatory

The R software

1. To work on GLM:

- model <- glm(y ~ x1 + x2 + x3 + x4, family=poisson)
- model$coef gives the model parameters

2. To work on principal component analysis:

- model <- prcomp(x)
- predict(model) gives the PCs
- model$rotation gives the eigenvectors

Generalized Linear Model (GLM) example

$$g(E(Y_i)) = x_i^T \beta \qquad i = 1, 2, \ldots 28$$

Y_i : July N500 derived from HKO TC best track data

g : natural log (canonical link)

β : maximum likelihood estimators to be found by an iterative weighted least squares procedure

x_i = PCs of CFS July hindcasts (initial conditions dated at the end of June)

香港天文台 HONG KONG OBSERVATORY

Selection of single predictors

1. No. of potential predictors = 26 x 28 = 728
2. Fit a single predictor GLM:
   - glm(y~x, family=poisson)
3. Search for skilful single predictor
   - **summary**(glm(y~x, family=poisson))
   - Check if the p-value of the estimated parameter for x is less than a certain threshold, e.g. 0.05

Summary of fitting glm

```
Call:
glm(formula = y ~ x, family = poisson)

Deviance Residuals:
    Min        1Q       Median       3Q        Max
-2.4386  -1.5282  -0.2943   1.2617   4.0834

Coefficients:
                 Estimate   Std. Error  z value    Pr(>|z|)
(Intercept)   5.23952    0.15424    33.969   <2e-16   ***
x              -0.05273    0.00587    -8.983    <2e-16   ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The small p-value indicates that x is likely a good predictor of y.

Selection of combinations of predictors

1. Fit a single predictor GLM, search for skilful single predictor
2. Fit a multiple predictor GLM [predictors from step 1]
   - glm(y~x1+x2+x3+x4, family=poisson)
   - No. of combinations > 2 x 10^8
   - Hence randomly select a limited no. of combinations
   - Max. no. of predictors = 6
   - Max. no. of combinations = 12000 (say)

The R Book suggests that the max no. of predictors should be no more than 1/3 of the data points (i.e. ~9 in this case).

Selection of combinations of predictors

1. Fit a single predictor GLM, search for skilful single predictor
2. (a) Fit a multiple predictor GLM with 6 predictors at most
   - glm(y~x1+x2+x3+x4+x5+x6, family=poisson)
2. (b) Filter out redundant predictors by stepwise regression (both backward and forward)
   - model<-glm(y~x1+x2+x3+x4+x5+x6, family=poisson)
   - stepAIC(model, direction='both')

Selection of combinations of predictors

1. Fit a single predictor GLM, search for skillful single predictor
2. Fit a multiple predictor GLM, filter out redundant predictors by stepwise regression

Cross-validate the 'reduced GLM' [from step 2], search for top performers

Cross-validation of the regression model

1. Hide the observation of 1 year
2. Estimate the GLM parameter from the rest of the observations and the predictors
3. Verify the GLM forecast against the hidden observation
4. Rotate the process through 28 years

Observation	Predictor 1	Predictor 2	Predictor 3	Predictor 4
		Forecast Y1		
		Forecast Y2		
		Forecast …		
		Forecast …		
		Forecast Y27		
		Forecast Y28		

Verification

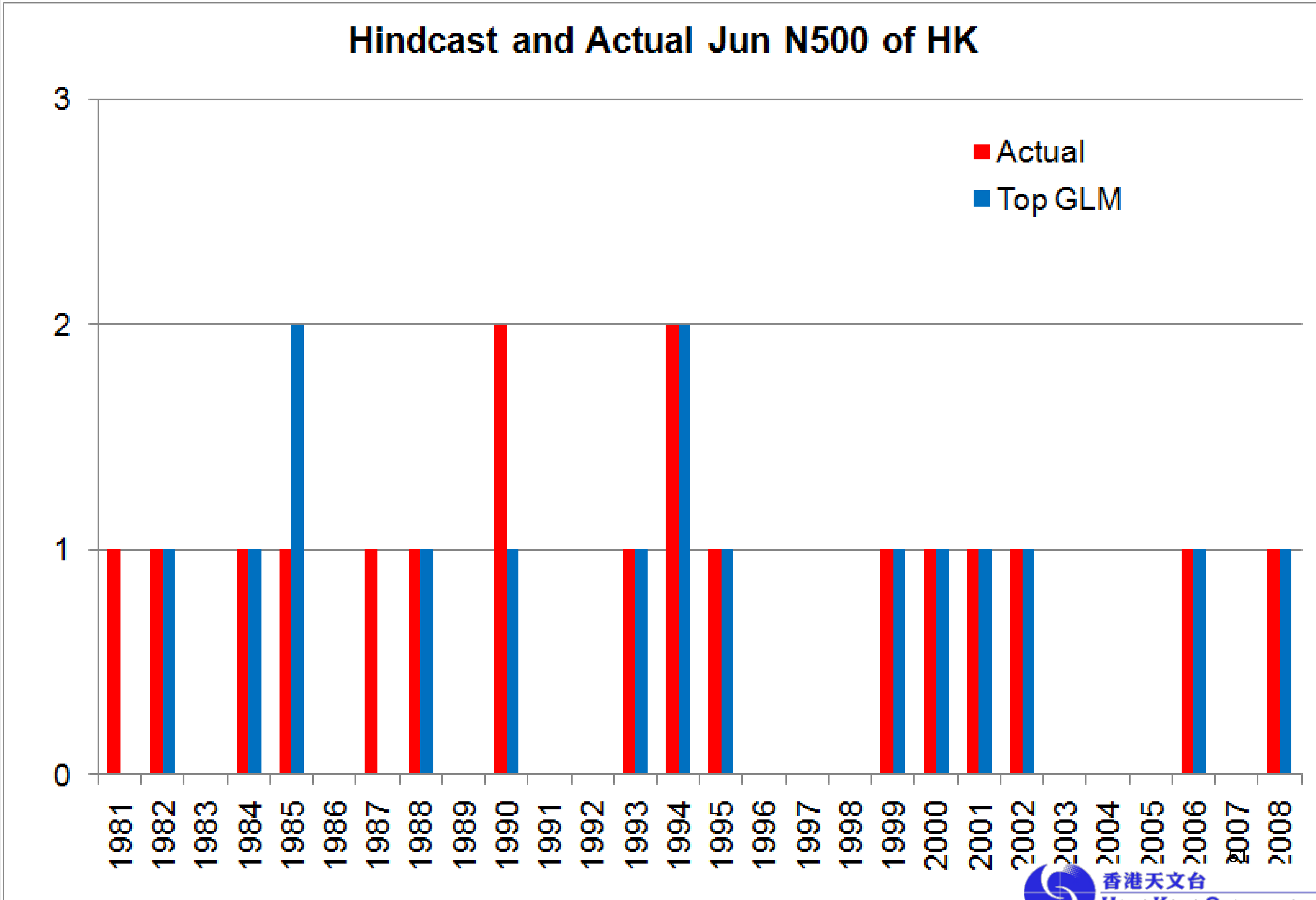
1. Round the GLM forecast to the nearest integer and take it as the count forecast [count forecast instead of a floating point number forecast will be issued in reality]
2. Sort the GLMs according to performance, i.e. no. of correct count forecast
3. Look for top performers

Climatology of HK N500, 1971-2000

The mode is used as the climatology forecast, a benchmark for performance comparison.

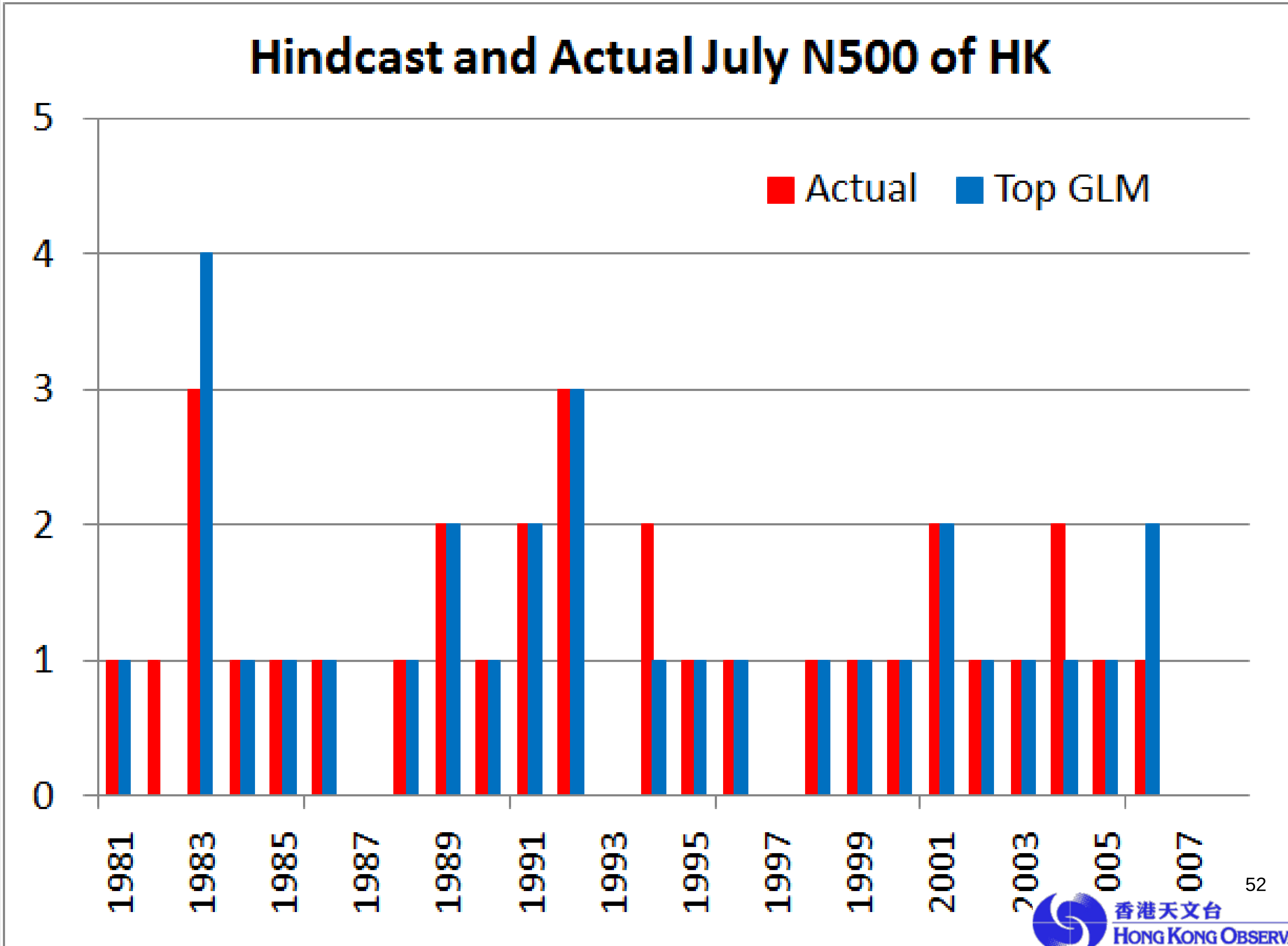
	Mode	Mean
Jun	1	0.77
Jul	1	1.37
Aug	1	1.43
Sep	1	1.47
Oct	0	0.70

Hindcast Vs Actual June N500, 1981-2008



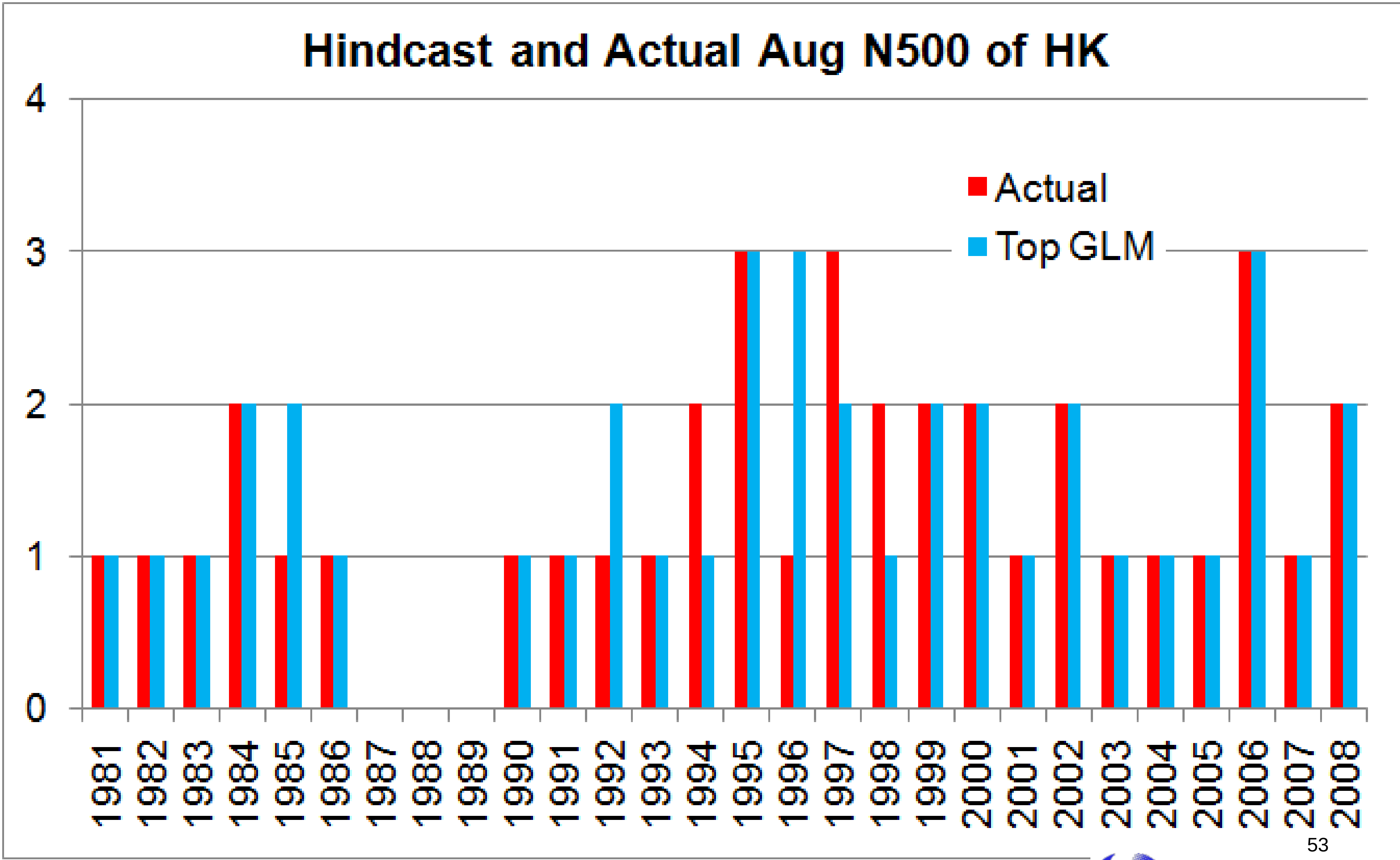
香港天文台
HONG KONG OBSERVATORY

Hindcast Vs Actual July N500, 1981-2008

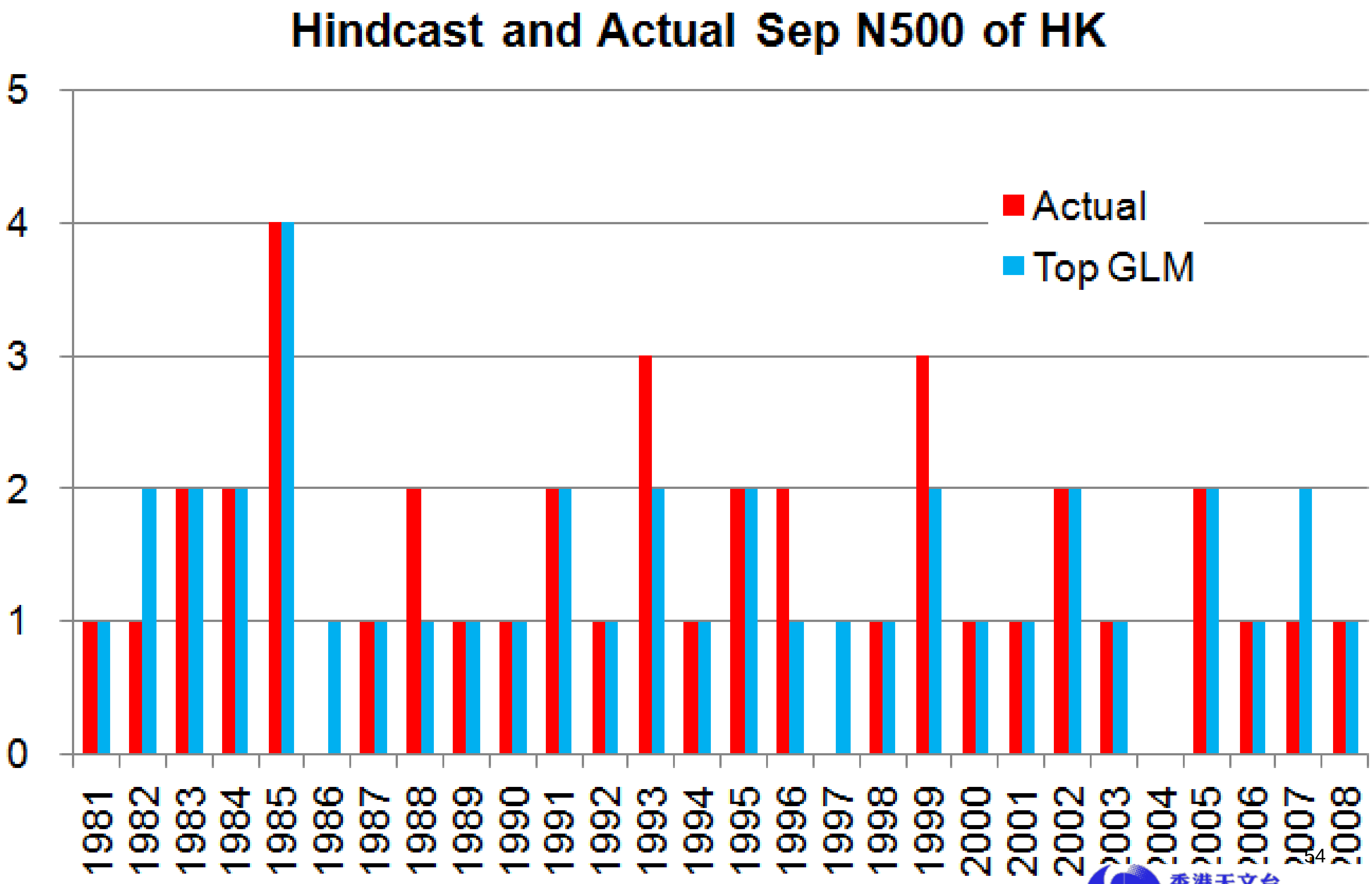


香港天文台
HONG KONG OBSERVATORY

Hindcast Vs Actual August N500, 1981-2008

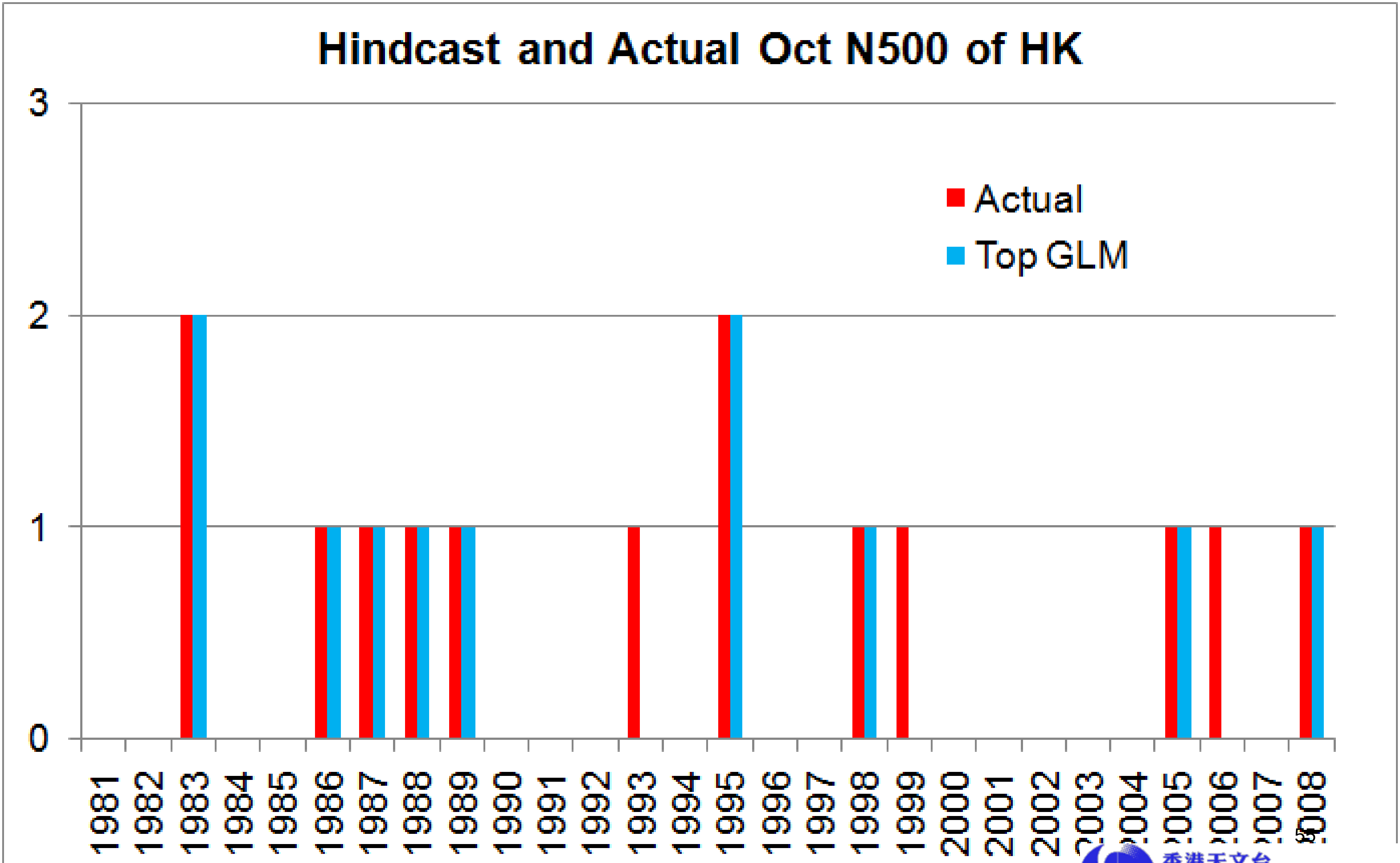


Hindcast Vs Actual September N500, 1981-2008



香港天文台 HONG KONG OBSERVATORY

Hindcast Vs Actual October N500, 1981-2008



Performance comparison
No. of correct count forecast in 1981-2008

	Climatology (mode) 1971-2000	Top GLM	Gain (%)
Jun	14	24	**71**
Jul	16	23	**44**
Aug	15	22	**47**
Sep	14	20	**43**
Oct	16	25	**56**

Test for significance (permutation test)

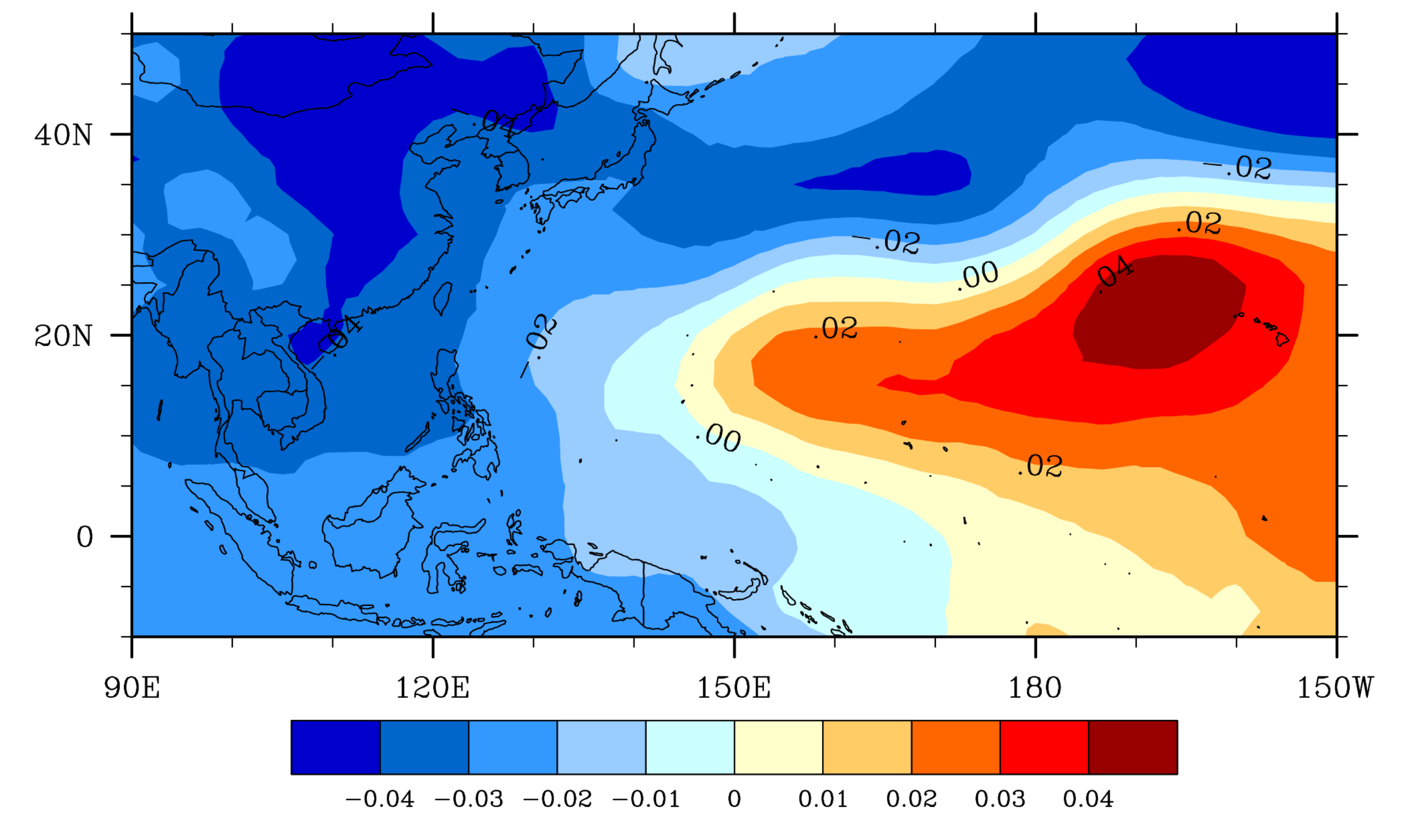
1. Define T = No. of correct forecast by GLM – no. of correct forecast by climatology
2. Null hypothesis H_0 : T is zero.
3. Alt. hypothesis H_a: T is greater than zero.
4. Use perm.test()

Jun	Reject H_0 at 5% significance level
Jul	Reject H_0 at 5% significance level
Aug	Reject H_0 at 5% significance level
Sep	Reject H_0 at 10% significance level
Oct	Reject H_0 at 5% significance level

Ref.: Wilks D. S. (2006): Statistical Methods in the Atmospheric Sciences

Physical Interpretation

Negative mslp anomalies over the southern part of the South China Sea and seas near the Philippines favour TC formation.

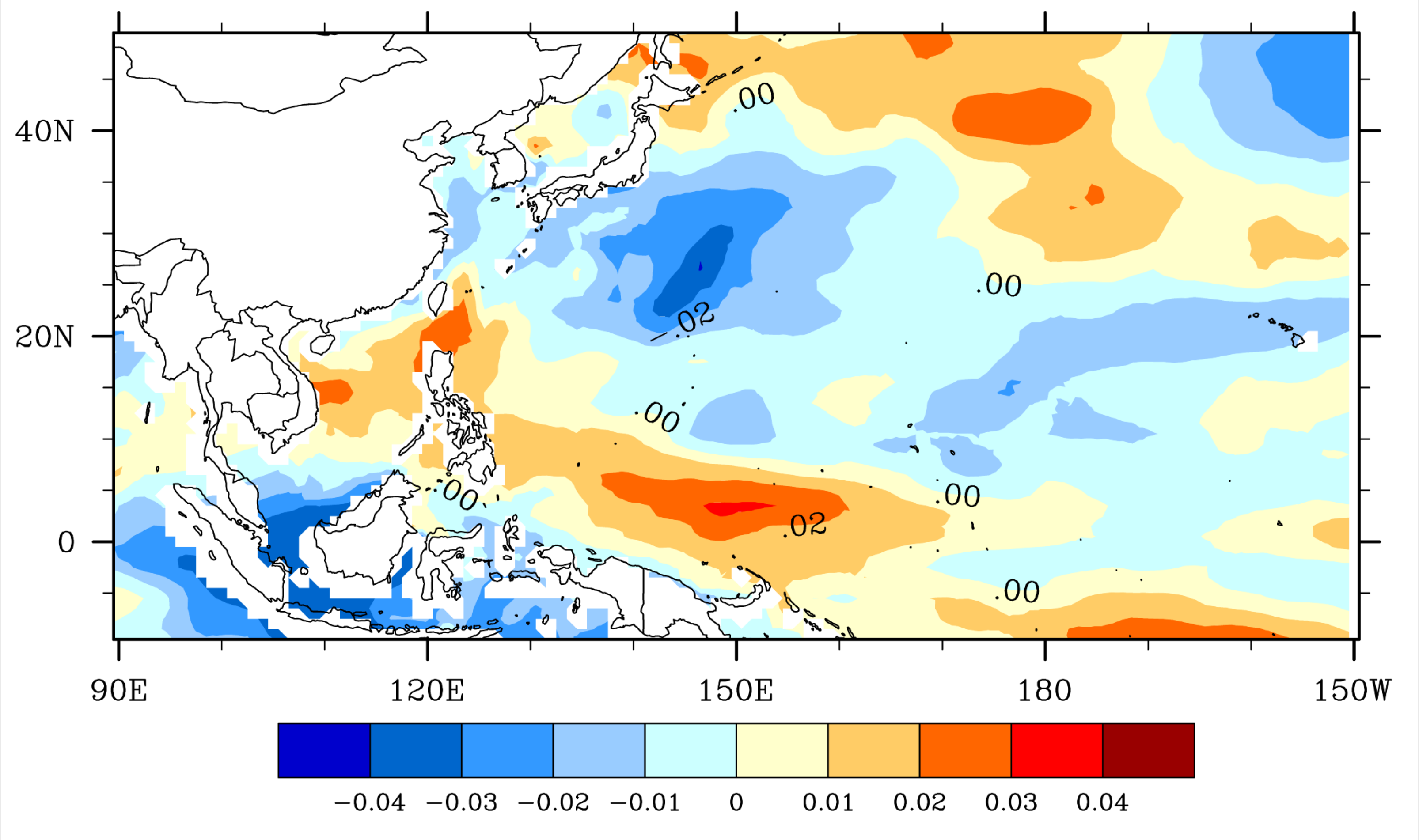


The 2nd EOF of mean sea level pressure of June

香港天文台 HONG KONG OBSERVATORY

Physical Interpretation

Positive SST anomalies favour TC genesis.

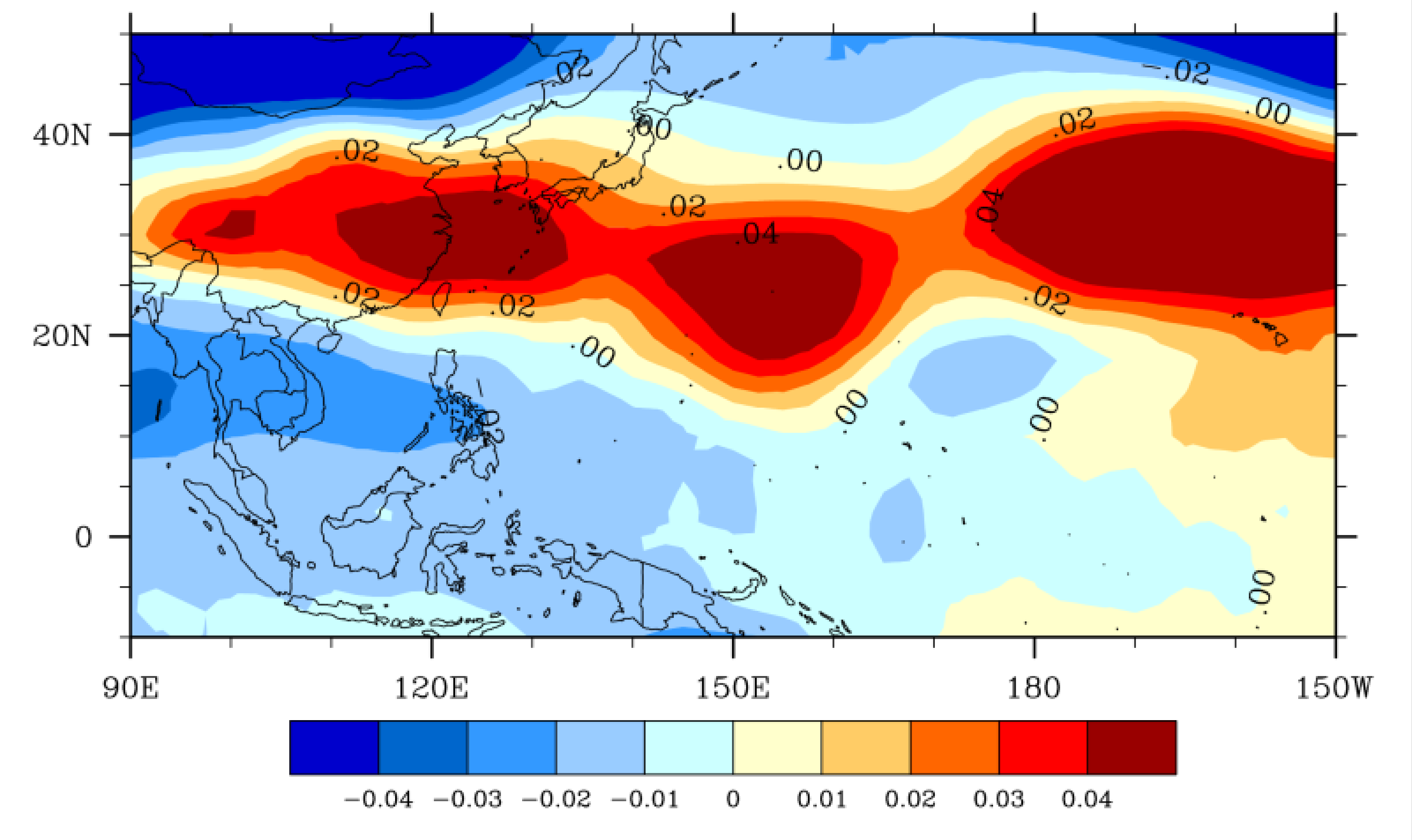


The 7th EOF of sea surface temperature of June

香港天文台 HONG KONG OBSERVATORY

Physical Interpretation

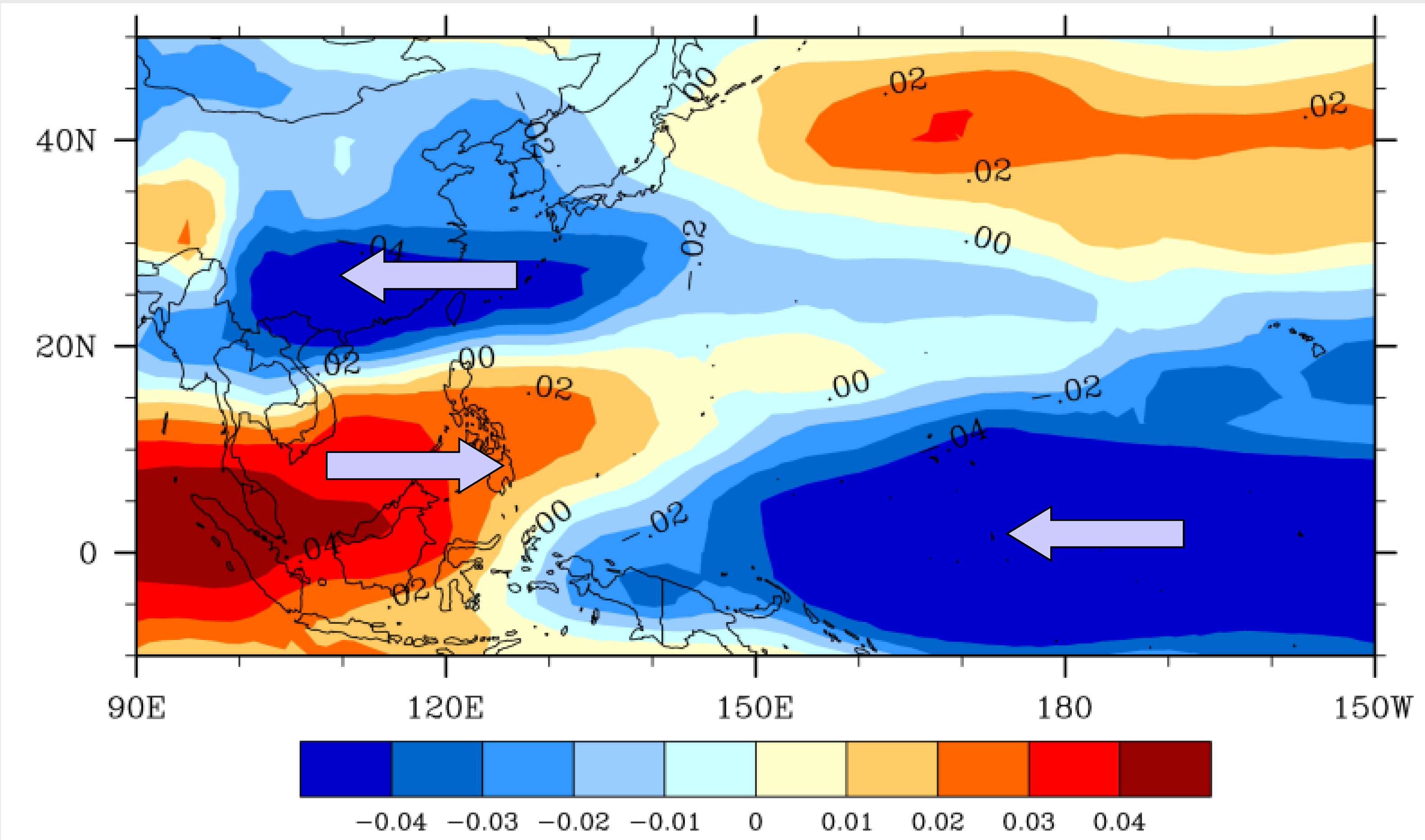
Positive 500 hPa gph anomalies help to prevent TCs from re-curving to the northeast too early.



The 4th EOF of 500 hPa geopotential height of June

Physical Interpretation

Strong low level lateral shear over the northern part of the South China Sea favours TC development.

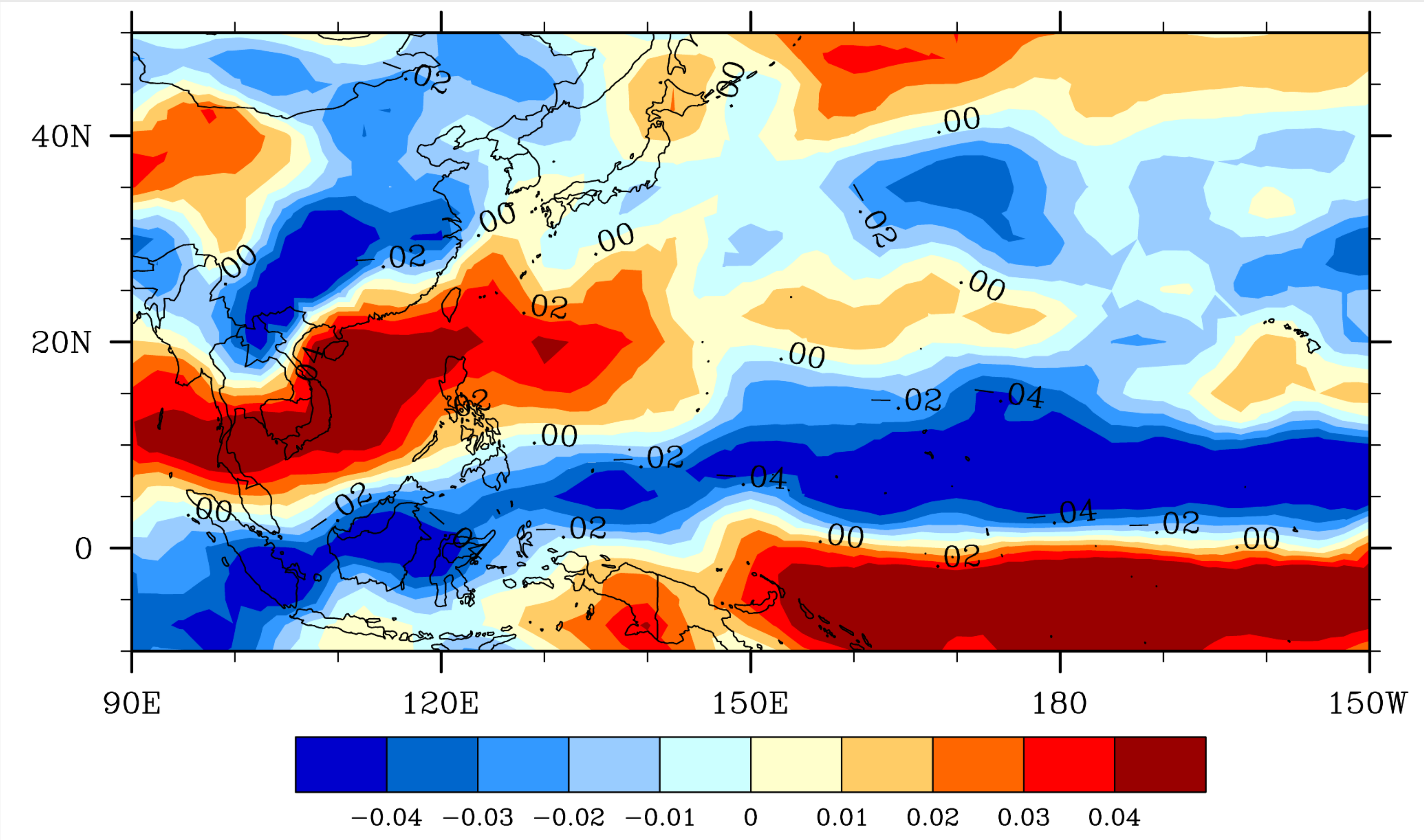


The 1st EOF of 850 hPa zonal wind of October

香港天文台 Hong Kong Observatory

Physical Interpretation

Positive low level vorticity over the northern part of the South China Sea favour TC development.



The 1st EOF of 850 hPa vorticity of October

Multi-GLM combination – skill enhancement

- Weigel et al., 2008: Can Multi-model Combination Really Enhance the Prediction Skill of Probabilistic Ensemble Forecasts? *Quarterly Journal of the Royal Meteorological Society*
- A message with respect to deterministic forecasts: combination of similarly skilful models can enhance prediction skill

Multi-GLM combination – skill enhancement

- We have already sorted the GLMs according to performance
- The top performers are models of similar skill
- We obtain the multi-model combination by taking the mode of the GLM forecasts, i.e. a voting process

	GLM1	GLM2	GLM3	GLM4	GLM5	MMC
1981	1	0	1	1	0	1
1982	1	2	2	0	2	2
..	..	..	..	..	..	..
..	..	..	..	..	..	..
2007	0	2	0	0	0	0
2008	2	1	1	1	1	1

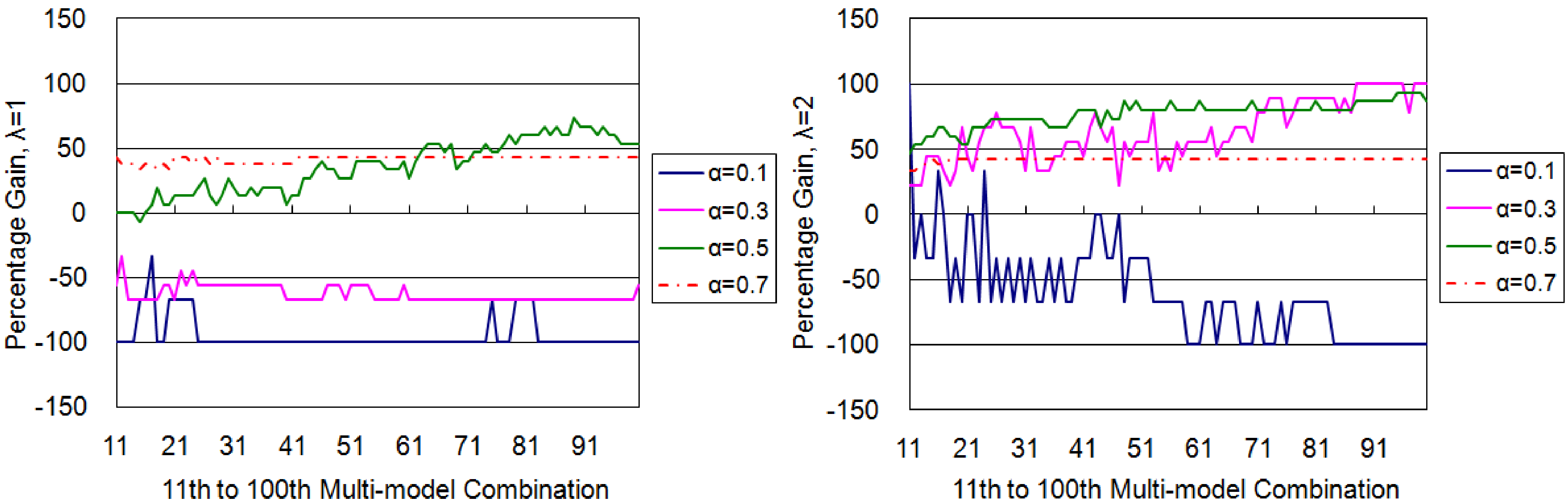
Performance comparison
No. of correction forecast in 1981-2008

	Climatology (mode) 1971-2000	Top GLM	Multi-GLM (mode of top 20 GLM)
Jun	14	24	27
Jul	16	23	25
Aug	15	22	26
Sep	14	20	23
Oct	16	25	28

Simulation to illustrate the MMC method

- Generate 30 years of artificial observations (random draws from a Poisson distribution with parameter λ)
- Generate 1000 artificial models of the same skill α
  - α = 0 : 0% of 30 forecasts are correct
  - α = 0.5 : 50% of 30 forecasts are correct
  - α = 1 : 100% of 30 forecasts are correct
- Simulate the multi-model combination by taking the mode, i.e. a voting process among the artificial models

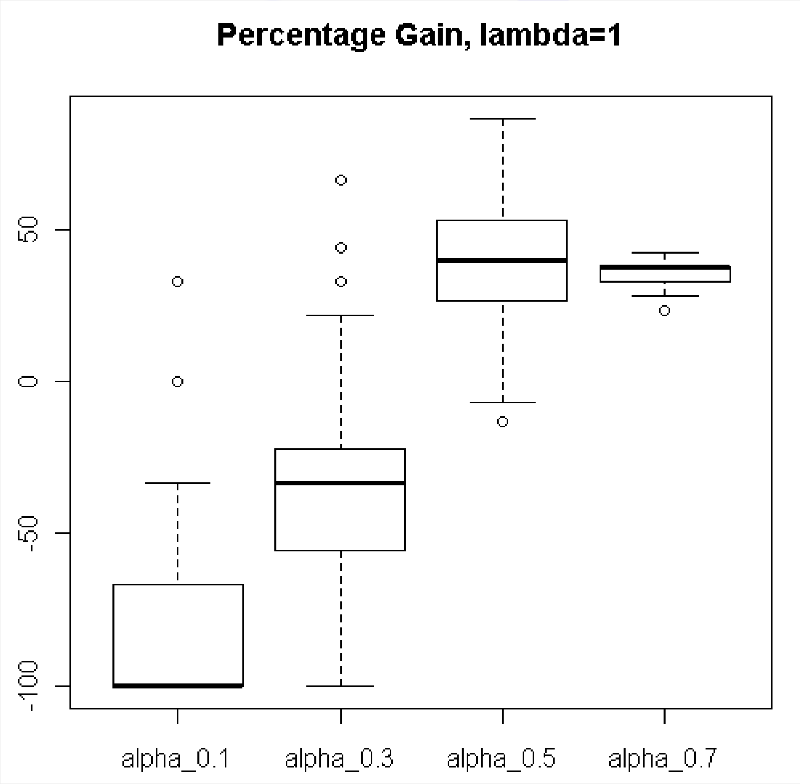
One simulation of 1000 artificial models (try λ = 1 and λ = 2)

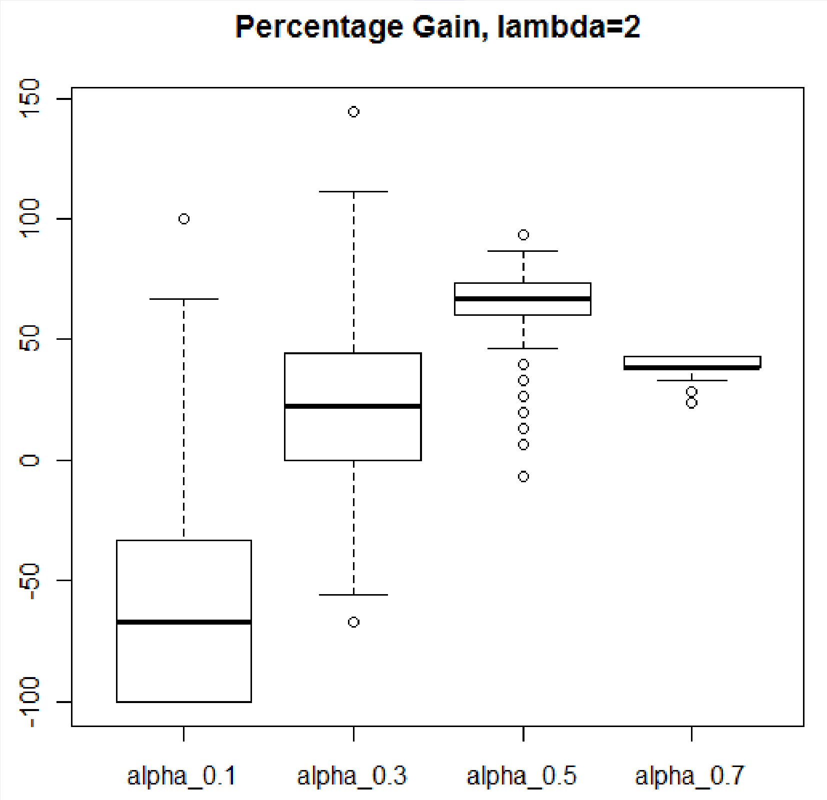


1. Negative gain for combinations of unskilful models
2. Positive gain is possible for α >= 0.5

香港天文台 HONG KONG OBSERVATORY

1000 simulations of 1000 artificial models consider the 20th MMC





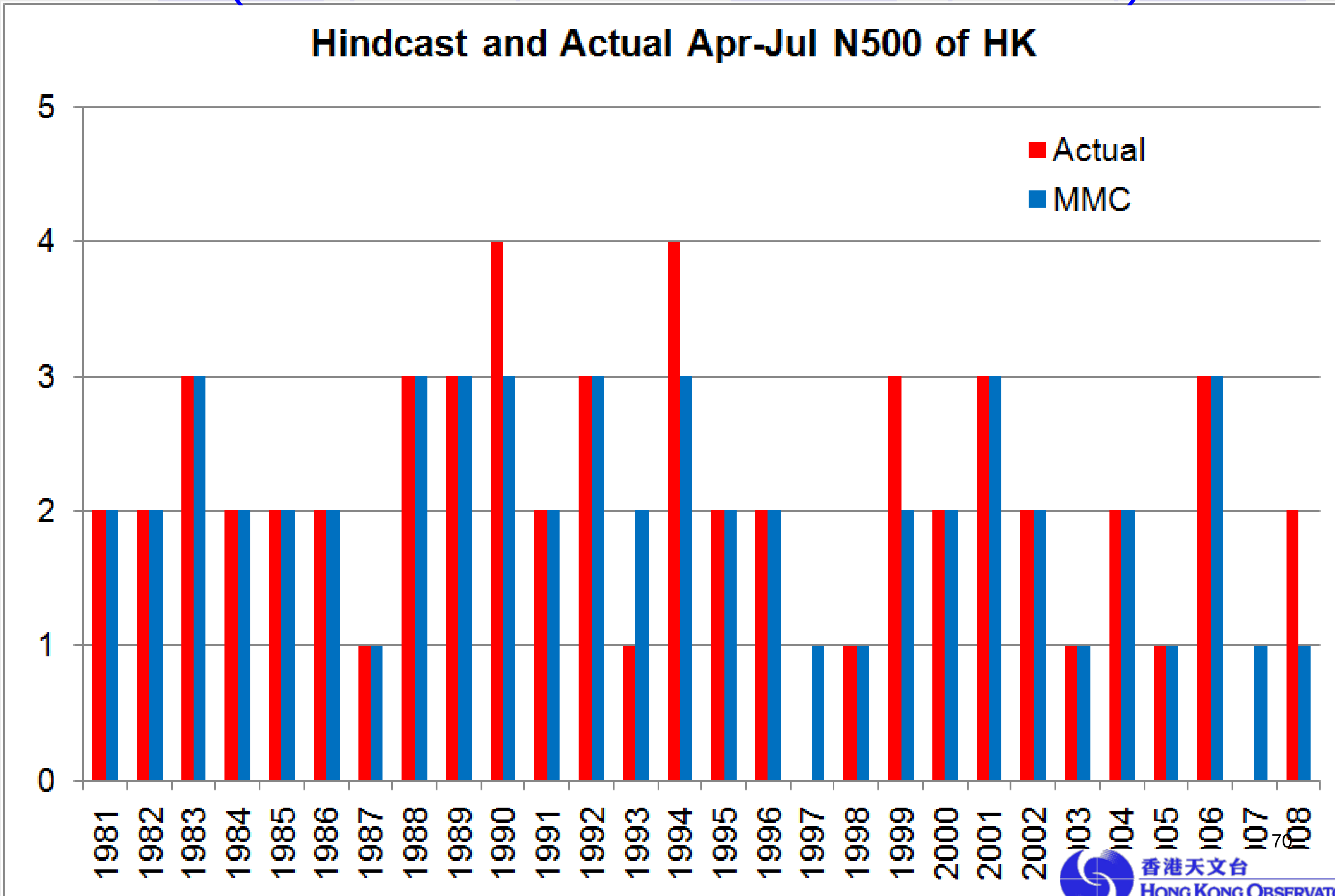
1. Positive gain should generally be expected for α >= 0.5

Seasonal and annual forecast

- Consider CFS model runs initialized at the end of Feb
- Can produce forecast for the whole TC season of HK
- Two forecast periods: Apr-Jul, Aug-Nov

Hindcast and Actual Apr-Jul N500, 1981-2008

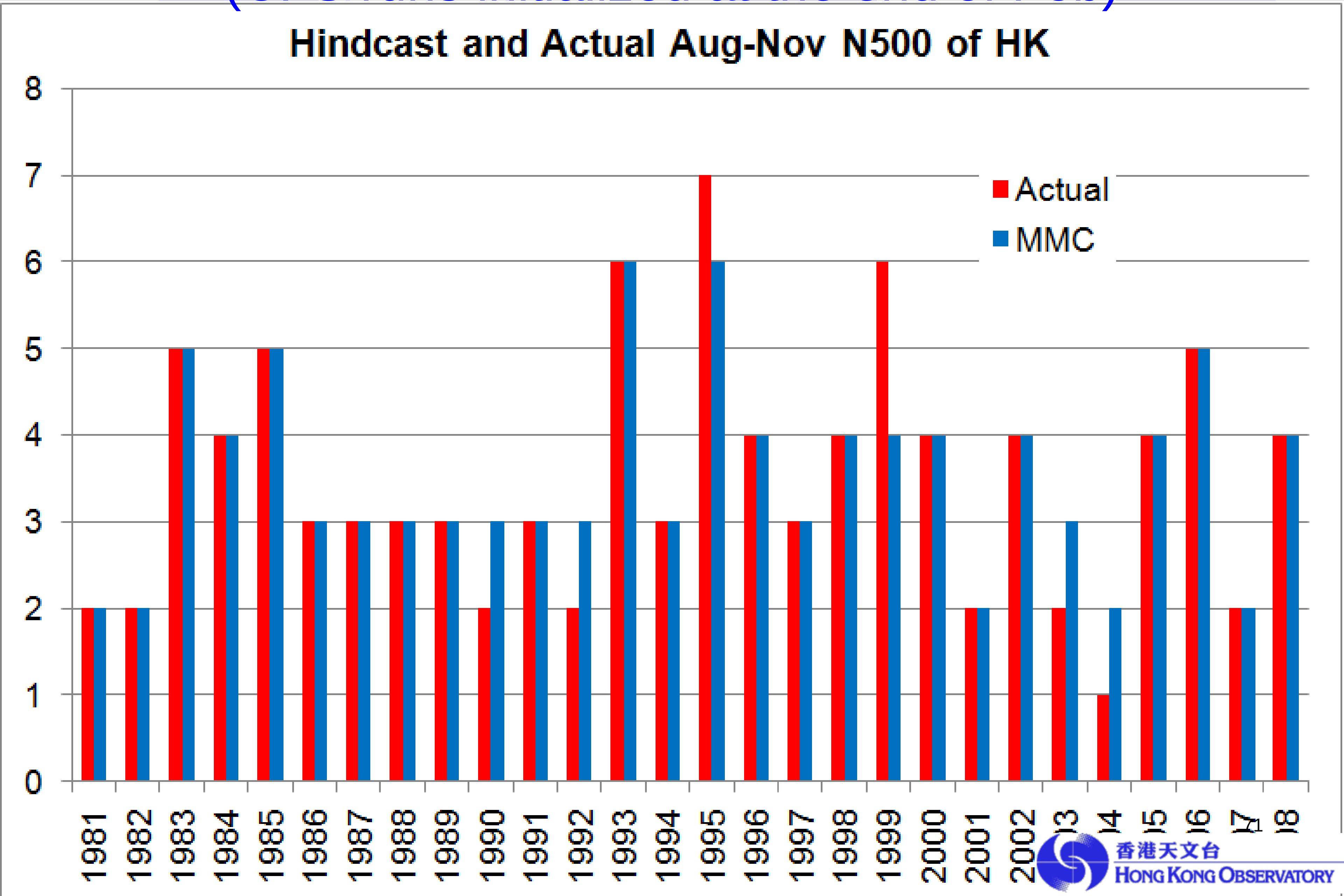
(CFS runs initialized at the end of Feb)



香港天文台
HONG KONG OBSERVATORY

Hindcast and Actual Aug-Nov N500, 1981-2008

(CFS runs initialized at the end of Feb)



香港天文台
HONG KONG OBSERVATORY

Performance comparison
No. of correction forecast during 1981-2008

	Climatology (mode) 1971-2000	Multi-GLM	Gain (%)
Apr-Jun	12	21	75
Aug-Nov	7	23	229

Conclusion

- Monthly/seasonal/annual TC forecast can be formulated in terms of Poisson GLM
- Dynamical climate model (e.g. NCEP CFS) forecast data contain a lot of predictive information
- Further skill enhancement is made possible by multi-GLM combination

Remarks

- Too many predictors: the possibility of irrelevant predictors getting high scores by chance exists
- The single predictors are found based on the whole dataset, hence the verification skill may have positive bias. The whole process of finding the predictors and regression equations should be cross-validated.
- Not all EOF can be easily interpreted
- Better to verify the floating point forecast instead of the count forecast. Can use the floating point forecast and the associated probability distribution to deal with uncertainty.

A Major upgrade of CFS

- CFS will be upgraded on 18 Jan 2011
- A new set of hindcast will be produced
- Both spatial and temporal resolution of forecast and hindcast will increase
- More pressure levels in the vertical direction
- More forecast cycles per day
- Big jump in data volume

Thank you

Acknowledgement:

The Hong Kong Observatory gratefully acknowledges NOAA/CPC for providing CFS forecast and hindcast data on the web to support research and seasonal forecasting operation conducted by the Observatory.